

2018

What Airbnb Reviews can Tell us? An Advanced Latent Aspect Rating Analysis Approach

Yi Luo

Iowa State University

Follow this and additional works at: <https://lib.dr.iastate.edu/etd>



Part of the [Advertising and Promotion Management Commons](#), [Business Administration, Management, and Operations Commons](#), [Management Sciences and Quantitative Methods Commons](#), and the [Marketing Commons](#)

Recommended Citation

Luo, Yi, "What Airbnb Reviews can Tell us? An Advanced Latent Aspect Rating Analysis Approach" (2018). *Graduate Theses and Dissertations*. 16403.
<https://lib.dr.iastate.edu/etd/16403>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

**What Airbnb reviews can tell us?
An advanced latent aspect rating analysis approach**

by

Yi Luo

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Hospitality Management

Program of Study Committee:
Liang (Rebecca) Tang, Major Professor
Tianshu Zheng
Eric D. Olson
Gang Han
Linda S. Hagedorn

The student author and the program of study committee are solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2018

Copyright © Yi Luo, 2018. All rights reserved.

TABLE OF CONTENTS

	Page
TABLE OF CONTENTS.....	ii
LIST OF FIGURES.....	iv
LIST OF TABLES.....	v
NOMENCLATURE.....	vi
ACKNOWLEDGMENTS.....	vii
ABSTRACT.....	ix
CHAPTER 1. INTRODUCTION.....	1
CHAPTER 2. LITERATURE REVIEW.....	12
The Rise of the Sharing Economy.....	12
Economic Factors.....	15
Technical Factors.....	15
Social Factors.....	16
Political Factors.....	17
Airbnb.....	20
Customer Experience.....	22
Lodging Experience.....	22
From WOM to eWOM.....	24
Word-of-mouth.....	24
eWOM.....	26
User-generated Content.....	28
Motivation for Creating UGC.....	29
User-generated Content and Travel.....	30
Online Customer Reviews.....	33
eWOM and Reviews.....	36
Big Data and Text Mining.....	41
Sentiment Analysis.....	43
Other Text Mining Techniques.....	45
Textual Analysis.....	46
Topic Modeling.....	47
Machine Learning.....	47
Semantic Analysis.....	48
Aspect-based Sentiment Analysis.....	49
Latent Aspect Rating Analysis.....	50
Plutchik's Wheel of Emotions.....	52
The NRC Emotion Lexicon.....	53

CHAPTER 3. METHODOLOGY.....	56
Data Collection and Analysis.....	56
Sampling.....	56
Data Analysis.....	57
Data Analysis Procedures.....	58
Data Pre-processing.....	59
Information Extraction.....	60
Information categorization.....	61
Aspect Segmentation.....	62
Sentiment Detection.....	62
Latent Rating Regression.....	64
LRR Model Estimation.....	66
CHAPTER 4. ANALYSIS AND RESULTS	69
Sample Descriptive Analysis.....	69
Aspect-based Sentiment Analysis.....	76
Latent Rating Regression.....	81
CHAPTER 5. DISCUSION.....	91
Discussion of the Results.....	91
Implications.....	93
Theoretical Implications.....	94
Practical Implications.....	96
Limitations and Future Research Directions.....	102
REFERENCES.....	105

LIST OF FIGURES

	Page
Figure 3.1. LRR graphical representation.....	65
Figure 4.1. Worldcloud of the tokened reviews.....	73
Figure 4.2. Airbnb listings locations summary.....	73
Figure 4.3. General review sentiment analysis	74
Figure 4.4. Co-occurrence networks of reviews	75
Figure 4.5. Emotional level of the lodging aspects.....	79
Figure 4.6. Airbnb cheap listings locations summary.....	87
Figure 4.7. Airbnb expensive listings locations summary.....	88

LIST OF TABLES

	Page
Table 4.1. Review information extracted from the Airbnb website	70
Table 4.2. Summary of Airbnb listings' overall ratings	70
Table 4.3. Descriptive summary of Airbnb reviews with different overall ratings	71
Table 4.4. Maximum and minimum review length.....	71
Table 4.5. Summary of average review length by star rating	72
Table 4.6. Summary of LDA seed words of each aspect.....	76
Table 4.7. Summary of boot-strapping key words for each aspect.....	77
Table 4.8. Summary of key word basic emotion levels for each aspect	80
Table 4.9. Summary of predicted overall ratings.....	81
Table 4.10. Summary of predicted aspect ratings.....	82
Table 4.11 Listings with different aspect ratings.....	84
Table 4.12. Summary of predicted aspect weights	84
Table 4.13. Summary of average aspect weights.....	86
Table 4.14. Aspect weights comparison between expensive and cheap Airbnb listings	87
Table 4.15. Aspect rating comparison between expensive and cheap Airbnb listings	89
Table 4.16. The top five listings with their highest product/service aspect weights	90

NOMENCLATURE

ABSA	Aspect Based Sentiment Analysis
EM	Expectation–maximization Algorithm
LARA	Latent Aspect Rating Analysis
LDA	Latent Dirichlet Allocation
LRR	Latent Rating Regression Model
MAP	Maximum a Posteriori Probability
ML	Maximum Likelihood Estimator
P2P	Peer to Peer Accommodation
TM	Topic Modeling
UGC	User-generated Content

ACKNOWLEDGMENTS

I would like to thank my committee chair Dr. Liang (Rebecca) Tang, with whom I have worked for the past five years. I have been tremendously fortunate to have a major professor who has provided me with innumerable opportunities for academic and professional growth. Dr. Tang's prudence and perseverance toward research and teaching will continue to inspire me throughout my life. I would also like to thank my department committee members Dr. Tianshu Zheng and Dr. Eric Olson, whose insightful support and guidance, and enthusiastic input, have motivated me to persist in pursuing my Ph.D. degree and to continuously strive to improve my doctoral dissertation. I would like to express deep appreciation to my outside department committee members, Dr. Linda Hagedorn and Dr. Gang Han. Dr. Hagedorn (from the School of Education) graciously agreed to be on my committee even though she will retire this August, and she has provided me tirelessly with advice for the completion of my dissertation and for my academic development. I am very grateful to Dr. Han, who has provided me with many details regarding text mining techniques.

In addition, I would like to dedicate this dissertation to the text mining experts in the field of computer science. I want to express my gratitude to Dr. Hongning Wang, Dr. Yue Lu, and Dr. Chengxiang Zhai for introducing me to the LARA model. It would not have been possible for me to complete my dissertation without their support in sharing the original Java code. I would also like to give special thanks to the computer science specialists Marina Kraeva, Yau Chan, Yau Mo, Zhicheng Zhao, and Lakshay Ahuja, who have spent countless hours of their personal time helping me to survive Python coding.

I wish to thank to my parents. I greatly appreciate and value what you taught me throughout my childhood. Thank you for all of the encouragement and support you've given me during my study abroad period. Your underlying unconditional love inspires me to be a better person. Finally, I thank my AESHM colleagues, the department faculty, and staff for making my time at Iowa State University a wonderful experience.

ABSTRACT

There is no doubt that the rapid growth of Airbnb has changed the lodging industry and tourists' behaviors dramatically since the advent of the sharing economy. Airbnb welcomes customers and engages them by creating and providing unique travel experiences to "live like a local" through the delivery of lodging services. With the special experiences that Airbnb customers pursue, more investigation is needed to systematically examine the Airbnb customer lodging experience. Online reviews offer a representative look at individual customers' personal and unique lodging experiences. Moreover, the overall ratings given by customers are reflections of their experiences with a product or service. Since customers take overall ratings into account in their purchase decisions, a study that bridges the customer lodging experience and the overall rating is needed. In contrast to traditional research methods, mining customer reviews has become a useful method to study customers' opinions about products and services. User-generated reviews are a form of evaluation generated by peers that users post on business or other (e.g., third-party) websites (Mudambi & Schuff, 2010).

The main purpose of this study is to identify the weights of latent lodging experience aspects that customers consider in order to form their overall ratings based on the eight basic emotions. This study applied both aspect-based sentiment analysis and the latent aspect rating analysis (LARA) model to predict the aspect ratings and determine the latent aspect weights. Specifically, this study extracted the innovative lodging experience aspects that Airbnb customers care about most by mining a total of 248,693 customer reviews from 6,946 Airbnb accommodations. Then, the *NRC Emotion Lexicon* with eight emotions was employed to assess the sentiments associated with each lodging aspect. By applying latent

rating regression, the predicted aspect ratings were generated. With the aspect ratings, the aspect weights, and the predicted overall ratings were calculated.

It was suggested that the overall rating be assessed based on the sentiment words of five lodging aspects: communication, experience, location, product/service, and value. It was found that, compared with the aspects of location, product/service, and value, customers expressed less joy and more surprise than they did over the aspects of communication and experience. The LRR results demonstrate that Airbnb customers care most about a listing location, followed by experience, value, communication, and product/service. The results also revealed that even listings with the same overall rating may have different predicted aspect ratings based on the different aspect weights. Finally, the LARA model demonstrated the different preferences between customers seeking expensive versus cheap accommodations.

Understanding customer experience and its role in forming customer rating behavior is important. This study empirically confirms and expands the usefulness of LARA as the prediction model in deconstructing overall ratings into aspect ratings, and then further predicting aspect level weights. This study makes meaningful academic contributions to the evolving customer behavior and customer experience research. It also benefits the shared-lodging industry through its development of pragmatic methods to establish effective marketing strategies for improving customer perceptions and create personalized review filter systems.

CHAPTER 1. INTRODUCTION

Information technologies that have emerged with the development of Web 2.0 have resulted in the rapid growth of websites that encourage sharing, collaboration, and user-generated content online (Kaplan & Haenlein, 2010). Sharing is the most basic economic behavior in human societies and is a form of exchange that has existed for thousands of years (Hellwig, Morhart, Girardin, & Hauser, 2015). In recent years, open-source software, file sharing programs, online forms of collaboration, and P2P financing tools are aspects of the new sharing economy phenomenon (Hamari, Sjöklint, & Ukkonen, 2016). Well-known examples of sharing economy innovations such as open-source software repositories (e.g., LinShare and Ares Galaxy), collaborative online encyclopedias (e.g., Baidu Baike), social content sharing platforms (e.g., Facebook, Instagram, etc.), and peer-to-peer document sharing websites (e.g., GitHub). In 2011, *TIME Magazine* nominated the sharing economy as one of “10 ideas that will change the world” (Walsh, 2011). In the early 1990s, the sharing economy demonstrated a shift in how people gain access to goods and circulate them in the marketplace. This innovative economic form is recognized to be a divergence from conventional models, because it concentrates not on ownership, but on access to assets or resources (Stephany, 2015). Hamari, Sjöklint, and Ukkonen (2016) identified the phenomenon of the emergence of the sharing economy from an array of developments in technology that have made sharing physical and non-physical products (e.g. goods and services) easier and simpler through a variety of IT sources available on the Internet. The sharing economy has extended to the sharing of such diverse products and services as rides, spare rooms, tool sharing (e.g., screwdrivers), relationship advice, and even legal expertise.

As consumers express greater concern about the effects of climate change and yearn for increased social embeddedness as expressed through local and communal consumption (Botsman & Rogers, 2010; Schor, 2016), the sharing economy has emerged as an appealing alternative for many consumers. It is both a technological and economic phenomenon comprised of P2P activities to obtain, share, or give access to goods/services that are coordinated through various online communities and bolstered by IT and communications developments; increased consumer awareness; the rapid growth of collaborative Internet communities; and new activities in sharing and social commerce (Cohen & Kietzmann, 2014; Kaplan & Haenlein, 2010; Stephany, 2015; Hamari, Sjöklint, & Ukkonen, 2016). The sharing economy is particularly salient because today's consumers are generally motivated to participate in "collaborative consumption" communities by a desire to promote the social good and environmentally friendly activities (Belk, 2014; Möhlmann, 2015; Schor & Fitzmaurice, 2015).

A number of significant social, economic, and technological shifts in the past five years have radically enlarged the sharing economy as a new aspect of the travel industry. A significant amount of studies primarily studied the different services offered by sharing economy platforms, with the emerging understanding that customers typically pursue unique, non-traditional experiences when making sharing economy purchases and using sharing economy services (Schor, Walker, Lee, Parigi, & Cook, 2015). In the context of hospitality and tourism, sharing economy travelers tend to pursue the feeling of "living like a local" for a short time during their travels (Zervas, Proserpio, & Byers, 2014). The popularity of the sharing economy in tourism reveals the consumer desire to visit authentic local and join in activities with local communities. With a rise in trust in strangers and the desire to experience

the local lifestyle, even well-to-do tourists partake in the sharing economy. Sharing the home of a local stranger differs significantly from the outdated stereotype of the “ignorant tourist” that quickly snaps a few photos at a destination and departs without interest in the local lifestyle (Yannopoulou, Moufahim, & Bian, 2013). Sharing economy platforms enable tourists to enjoy this experience because they offer various rental options/services outside of high-traffic, typically touristy areas in the so-called tourist ghetto (Trivett & Staff, 2013).

Airbnb is an innovative platform for accommodation-sharing services that connects hosts and guests (Lu & Kandampully, 2016). Airbnb accommodations typically offer greater space and amenities, such as free Wi-Fi, kitchens, and onsite laundry facilities, thus creating a more home-like atmosphere. In 2015, it was noted that a daily average of more than 50,000 people rent rooms from approximately 2,000,000 rentals offered on Airbnb in 65,000 cities in 191 distinct countries (Airbnb, 2015). A 2013 market report (Lee & Kim, 2018) stated that Airbnb rental offerings typically are priced 21.2% lower for houses and 49.5% lower for single rooms as compared with hotel rates. Travelers tend to stay longer in Airbnb properties on average compared to traditional hotel room stays and rate their satisfaction higher with Airbnb than with hotels (respectively, 3.8 nights for Airbnb as opposed to 2.1 nights for hotels; 4.72 stars with Airbnb in contrast to 4.04 stars with hotels) (Certify, 2015).

As a result of the rapid worldwide growth of Airbnb, hospitality scholars have taken notice and begun to study Airbnb’s unique operational system and its role in the sharing economy (Lee & Kim, 2018). Airbnb is one of the most well-known examples of a thriving sharing-economy business (Le & Kim, 2018) and research on the company generally follows the trends of the sharing economy. Previous literature on Airbnb can be categorized into three streams: (a) discussions of the uniqueness of Airbnb accommodations and travel

experiences compared to those offered by traditional hotels (e.g., Lehr, 2015; Zervas, Proserpio, & Byers, 2014); (b) examinations of the characteristics of P2P sharing transactions (e.g., Botsman & Rogers, 2010; Rothkopf, 2014); and (c) studies regarding the unique legal and financial issues surrounding Airbnb (e.g., Ert, Fleischer, & Magen, 2016; Fraiberger & Sundararajan, 2015). However, only a few sporadic studies have investigated customers' written reviews evaluating their experiences of sharing economy-based accommodations. Similar to traditional travel e-commerce websites, Airbnb allows guests to share their assessments on cleanliness, accuracy, check-in processes, communication with hosts, location, and value (Lehr, 2015).

Previous researchers have helped to improve practices in the area of tourism and hospitality by identifying the determinants of customer experience and the effects of different marketing strategies on customers' perceptions and satisfaction (Kim & Kim, 2004; Jeong & Mindy Jeon, 2008; Ye, Law, & Gu, 2009). Many indicators of experience quality used in the conventional hospitality industry are not well suited to accommodation offers in the sharing economy (of which the majority are personal assets used for residential purposes) because of the different type of experience that sharing economy customers seek (Guttentag, 2015). Therefore, it is necessary to identify a new set of experience quality indicators for the sharing economy. In addition, due to the distinctive characteristics of sharing-economy accommodation services, particularly the availability of idle assets and non-professional business owners (Botsman & Rogers, 2010), it is useful to reexamine the influence of determinants relevant to the conventional hospitality industry. However, based on a review of previous literature, it was found that very few researchers have investigated the factors determining customer experience through the lens of sharing economy-based

accommodations. Moreover, as with traditional e-travel commerce, Airbnb tourists are able to evaluate their experiences by leaving reviews commenting on cleanliness, accuracy, check-in processes, communication with hosts, location, and value (Lehr, 2015), providing significant data that is critical to analyze.

To investigate the customer evaluation of experience, previous studies have primarily relied upon traditional qualitative, quantitative, or mixed methods such as questionnaire surveys or focus groups in order to find the experience dimensions and create measurement items to provide empirical evidence. Guo, Barnes, and Jia (2017) noted that the existing research methods require a trade-off to be struck between the performance of estimation and the costs associated with sample collection. The newly emerging technologies of Web 2.0 spurred the creation of a variety of types of user-generated content (UGC) forms on different types of websites such as online social networks, online communities, and booking platforms, where guests are able to discuss their experiences with products/services with other users (Plank, 2016). Statistic Brain (2017) revealed that 81 percent of travelers find user reviews important. In the market domain, UGC has sparked researchers' interest (e.g., Marine-Roig & Clavé, 2015; O'Connor, 2008; Tang, Fang, & Wang, 2014). Such content broadly includes ratings, reviews, photos, videos, social posts, and Q&A participation (Akehurst, 2009), with the most common form of UGC being reviews (Dhar & Chang, 2009). Reviews are a new, e-commerce-based type of WOM about not only products and services, but their providers (Chen & Xie, 2008). Consumers increasingly use others' reviews to retrieve information on destinations, accommodations, attractions, activities, and experiences (Gretzel & Yoo, 2008; Park & Gretzel, 2007; Zhou, Ye, Pearce, & Wu, 2014). Prior research has also shown that reviews influence consumer decision-making because

consumers deem reviews as more informative and trustworthy than company-generated information (Filiéri, 2016; Senecal & Nantel, 2004). Previous studies indicated that reviews can influence sales in a positive manner (Öğüt & Onur Taş, 2012; Mudambi & Schuff, 2010). Specifically, some researchers noted a correlation between very positive ratings and product sales growth (Clemons, Gao, & Hitt, 2006). Tourism practitioners are paying increasing attention to the fact that customers are highly interested in their fellow consumers' opinions about products/services because of the strong influence that such interest exerts (Bjørkelund, Burnett, & Nørvåg, 2012). With the rapid expansion of e-commerce, people show a greater likelihood to share their thoughts about and hand-on experiences with products or services they have purchased. These reviews are important for both business organizations and customers (Zhuang, Jing, & Zhu, 2006). Hospitality and tourism is a particularly information- and service-oriented industry. Moreover, the field involves customer-based service in a context where multiple factors, such as noise, nearby construction, weather, and even customer expectations, may impact customer experience and the evaluation of services (Bjørkelund, Burnett, & Nørvåg, 2012). Such events can influence a consumer's overall judgment at any given time, resulting in a dynamic customer evaluation process, which in turn can influence his/her review. Identifying the ways in which the online evaluation process is in flux can offer opportunities for customers and practitioners to understand vast quantities of opinion data. However, the number of reviews on a popular product can number in the hundreds or thousands, which poses a challenge for customers reading them in an effort to make informed purchasing decisions. Furthermore, it also makes it difficult for companies to track and manage customer opinions (Shi & Li, 2011).

Given the advantages and benefits that can be obtained from mining customer reviews, many researchers recently have transferred their interests from traditional research methods to team up computational linguistics knowledge, network crawling, statistical methods, and machine-learning techniques to aggregate, analyze, and interpret so-called text mining analytics for marketing purposes, such as extracting trending topics and sentiments and recognizing opinions about products (Liu, 2012). Specifically, SA and opinion mining are excellent options for various market intelligence applications (Pang & Lee, 2008) in order to extract opinions from unstructured documents, and thus inform strategies for public relations, reputation management, trend prediction, and public viewpoint tracking (Rambocas & Gama, 2013). In these years, social media studies usually employed opinion mining techniques (e.g. Asur & Huberman, 2010; Hutto & Gilbert, 2014). For example, Twitter data, movie reviews, and blog content have been used to predict box-office revenues for movies (Annett & Kondrak, 2008; Go, Bhayani, & Huang, 2009; Thet, Na, & Khoo, 2010). The analysis of mood states expressed on Twitter was used to predict the stock market in a study by Bollen, Mao, and Zeng (2011), while other studies identified expert investors that microblogged and carried out SA of opinions on stock prices (Oh & Sheng, 2011; Ranco et al., 2015).

In the context of hospitality, sentiment analysis can be employed to summarize reviews and extract opinions from textual data in order to provide overall perspectives on customers' lodging experience evaluations. This technique enables consumers to save time and can facilitate the decision-making process. It also permits hotel management staff to discern customer opinions about their lodging offerings and services, making use of constructive feedback to improve their services (Bucur, 2015). However, Chaovalit and Zhou

(2005) indicated that mining opinions and sentiments from online product reviews is a complicated undertaking that necessitates more than text mining. Thus, research in this area is not only important for natural language processing, but potentially for the fields of management sciences, social sciences, political science, and economics, given that they are all areas affected by the opinions of individuals.

There have been several recent studies on joint sentiment/topic extraction (Mukherjee & Liu, 2012; Titov & McDonald, 2008; Zhang, Mei, & Zhai, 2010). These methods were used to analyze customers sentiment of several topics considered as clusters of object features through mining their reviews. This approach is known as aspect-based sentiment analysis (ABSA) and it can perform SA of review texts in greater depth (Thet et al., 2010). Specifically, it is founded in the concept that opinions are comprised of positive or negative sentiments and targets (that is, opinions, topics, or aspects). In Liu of examining language elements such as sentences, clauses, phrases, paragraphs, or documents, ABSA directly analyzes the opinion itself (Liu, 2012). Nonetheless, contemporary methods pose challenges when digesting and exploiting large quantities of reviews in light of the lack of support to understand individual's opinions in term of topics, which is a more fine-tuned level of analysis. As such, it has been suggested that examining overall ratings to learn about reviewers' feelings about the various aspects matters, since hotel reviewers may give an identical overall rating but for different motives (Wang, Lu, & Zhai, 2010). To achieve the envisioned in-depth, highly detailed comprehension of customer reviews, the latent aspect rating analysis (LARA) was developed by Wang, Lu, and Zhai (2010) to provide a new approach to text mining that analyzes review opinions by topical aspect in order to unearth latent aspect ratings and the importance of their weights in making an overall judgment as

expressed by the overall rating. The LARA method operates under the premise that reviewers give aspect ratings based on the weighted combination of the sentiment polarities derived from the review that discuss their corresponding aspect, and an overall rating comprised of the weighted combination of all aspect ratings in a review (Wang et al., 2010).

In the last approximately ten years, researchers have performed a significant amount of study in the area of sentiment analysis. Of particular research interest is determining if words, phrases, or documents have positive or negative polarities in terms of favorable/unfavorable sentiments as expressed toward the targets under review (Thet et al., 2010). Nonetheless, there is inadequate research regarding emotions as expressed in text and automatic analysis of such text. Analysis is complicated by the fact that emotions are frequently expressed via facial gestures (Fortenbaugh, 1970; Russell, 1994) and through a wide variety of words. The *NRC Emotion Lexicon* is an annotation dictionary with a large lexicon of terms and their emotional associations. The lexicon includes terms, their associated emotions, and measures of correlation strength between the terms and the emotions. This dictionary employs the emotions of anger, anticipation, disgust, fear, joy, trust, sadness and surprise, which are generally viewed as basic, prototypical emotions (Plutchik, 1980). By applying the *NRC Emotion Lexicon* in text mining studies, the relative aspect weights are more precise, since customers choose different words to express their feelings toward a product or service and this in turn affects their overall ratings.

The present study was undertaken with the intent to use text-mining techniques to determine the hidden relationships between salient customers' sentiments toward different lodging experience aspects from their Airbnb reviews and the overall ratings of Airbnb accommodations. Specifically, the author first applied aspect-based sentiment analysis to

more effectively discern Airbnb customers' experiences while staying with hosts. By mining a vast number of Airbnb customer reviews, it was possible to extract the lodging information that pertain to the factors of the lodging experience that are most valued by Airbnb customers and categorized the extracted information into different aspects such as service, location, communication, etc. Using the *NRC Emotion Lexicon* and the sentiment words in the reviews which described the aspects, each aspect was assigned a sentiment value.

Next, the relationship between the aspects and the reviews' overall ratings was investigated by applying LARA in order to investigate Airbnb customer ratings pertaining to each lodging experience aspect sentiment that was extracted in the previous step. This method first identified the aspect ratings based on the sentiment values, with the aspect ratings, the aspect weights were generated. In this study the different lodging experience aspects are the independent variables, and the Airbnb accommodations' overall ratings are the dependent variable. The latent rating regression (LRR) model, which is the key algorithm of LARA, was used to discern the relationship between endogenous and exogenous variables. In the process of LRR, it is assumed that an overall rating is created via the weighted amalgamation of the latent ratings across the aspects. The rating of aspects and weights were estimated, wherein the weights modeled the relative emphases placed on individual aspects in the process of developing an overall rating. Through this method, it was possible to determine the levels of emphasis placed on each aspect in the formation of overall ratings of Airbnb accommodations.

The study offers a theoretical foundation for future hospitality studies on the sharing economy. The new proposed text analysis method could be a useful tool to enable scholars to better understand customer rating behavior and experience evaluation. The present work also

contributes to the literature by mining customer reviews from the text mining and big data perspective; in particular, it investigates customer lodging experience evaluation by using extracted related aspects and examines their impact on customers' rating judgments.

Compared with the exiting sentiment mining studies, this study investigated customer sentiment in greater depth using eight emotions, as opposed to merely positive/negative analysis. This study advances the LARA by deconstruct the model into two stage including ABSA and LRR and provides supplements to the exiting LRR by adding eight basic human emotions into computing. Thus, a more precise customer sentiment toward services or products can be derived, which provides a new angle on investigating customer behavior. In the field of hospitality and tourism, this research opens up a brand-new direction in review mining by converting quantitative review data into qualitative sentiment polarities, and further analyzing the relationship between qualitative sentiment polarities and ratings.

This study also provides practical advice to Airbnb investors and hosts regarding methods to improve the star rating system, emphasizing the importance of addressing the sub-ratings of the various lodging dimensions that more accurately reflect what customers really value and care about. A new review filter system is also recommended to help customers find what they want quickly and efficiently. Furthermore, the results of the study provide industry practitioners with valuable insights regarding the types of unconventional accommodations offered by sharing economy-based rentals. Finally, the study is expected to provide stakeholders with the means to increase profits and further develop their businesses.

CHAPTER 2. LITERATURE REVIEW

Chapter 2 delivers both a general background and a theoretical foundation for the study. Specifically, this chapter offers a current literature review on the sharing economy and the theoretical foundations of customers experience and customer WOM. Then, review and examine the different textual data analysis methods that employed in the previous investigations. To deeper investigate the different emotions that reflect customer experience, Plutchik's Wheel of Emotions and *NRC Emotion Lexicon* were discussed. Based on the discussion, the advanced LARA model is proposed that deconstructed customer rating behavior into ABAS and LRR stages.

The Rise of the Sharing Economy

After Botsman and Rogers (2010) first introduced the concept of shared social and economic activities in recent years, a number of academicians have sought to develop definitions for the sharing economy. For example, Bardhi and Eckhardt (2012) indicated that the sharing economy is based not on owning products, but on paying for temporary access to assets. Frenken, Meelen, Arets, and Van de Glind (2015) defined the sharing economy as consumer-to-consumer, temporary-access transactions involving underutilized physical assets ("idle capacity"). Schaefer, Lawson, and Kukar-Kinney (2016) further defined this economic model as "market-mediated transactions that provide customers with temporarily limited access to goods in return for an access fee, while the legal ownership remains with the service provider" (p. 571).

Based on these attempts to define the sharing economy, Frenken et al. (2015) proposed three characteristics of the concept. First, the sharing economy involves renting/leasing products on consumer-to-consumer platforms, in contrast to the business-to-

consumer product/service economy. This characteristic indicates the major difference between the sharing economy market paradigm and the traditional two-sided paradigm. Sharing economy suppliers include individual, nonprofessional decision-makers that are not companies or professional agents (Li, Moreno, & Zhang, 2015). Second, the sharing economy involves temporary access to, as opposed to the transfer of ownership of, a good. Botsman (2015) stated that “sharing” is a varied concept that involves transactions and contact between individual consumers and individual providers and that it does not include the second-hand economy that involves consumers selling or giving away preowned items on websites such as Facebook and eBay. Third, the sharing economy involves the efficient use of physical assets, as opposed to the delivery of services, between private individuals via online platforms that bring consumers together in an on-demand economy. Added value for both parties is achieved by uniting buyers and sellers “on board” (Rochet & Tirole, 2004). Rifkin (2014) employed the term “sharing economy” to refer to a “hyperconnected” economic model. Using digital platforms, sharing economy members connect customers to whatever may be needed at the current moment, in contrast to the conventional supply-and-demand model between customers and companies.

Both the business and academic worlds have focused increasing attention on the new sharing economy (Hamari et al., 2016; Malhotra & Van Alstyne, 2014; Zervas et al, 2014). The extant literature regarding the sharing economy falls under several subdomains, one of which is concerned with the psychological basis of sharing, access, and ownership per se (Belk, 2014; Hamari et al., 2016; Heinrichs, 2013; Zervas et al, 2014). A primary driver of the sharing economy is the opportunities for social interaction that it offers (Belk, 2014). Rothkopf (2014) stated that sharing economy websites such as Airbnb provide creative,

progressive ways to connect with others. Anyone with Internet access can become a supplier, which represents the most profound innovation of digitalization. By connecting together sharing economy platforms and social media, a supplier can create a trustworthy and credible image, which is essential in order to attract potential customers (Matzler, Veider, & Kathan, 2015). Such platforms serve as generic “ecosystems” that link potential customers to a vast multiplicity of offerings from providers ranging from private individuals to international corporations.

Other researchers have focused on the legal characteristics of the sharing economy (Kassan & Orsi, 2012; Koopman, Mitchell, & Thierer, 2014; Schor, 2016) or on topics related to the nature of peer-to-peer markets (Cohen & Sundararajan, 2015; Fraiberger & Sundararajan, 2015; Sundararajan, 2014). With the rising sharing economy, the problem of unfair competition has appeared because some sharing economy suppliers evade paying taxes or following regulations (Heo, 2016). Guttentag (2015) discussed the legal and taxation issues surrounding sharing economy platforms, providing an overview of the existing regulatory fluidity and offering potential solutions. Heo (2016) also pointed out that sharing economy platforms still face certain legal roadblocks and stakeholder issues.

The sharing economy and collaborative consumption have been main and rising phenomena in numerous industries involving millions of users and businesses (Möhlmann, 2015). A number of critical social, economic, technological, and other changes in the late 2010s have fueled the growth of the sharing economy, now a significant player in the e-commerce industries. These influential factors are discussed in detail below.

Economic Factors

The primary extrinsic motives of the new economy's platforms include financial benefits, practical needs, and approval from others. The extant literature regarding the motivations for and attitudes about participation in the sharing economy ("collaborative consumption") cites economic benefits as the most significant driver (Schor, 2016). Being a sharing economy provider enables individuals to make use of idle assets (such as vehicles) and earn income, thus reducing ownership expenses and providing typically more economical options (e.g., compared to traditional car rentals) to consumers. As such, the sharing economy offers expanded and less costly choices. In particular, within the context of tourism and hospitality, travelers are more aware of idle or excess assets (Zervas et al, 2014). From the perspective of sharing development, a significant advantage of the sharing economy's technological platforms is that they are not costly to build (John, 2013), which contributes to the diversity of sharing products or services and makes almost everyone capable of sharing their own underutilized resources.

Technical Factors

Finally, technological advances that have facilitated and widely expanded peer-to-peer transactions by taking them online have extended the sharing economy beyond traditional transactions with family, friends, and neighbors. The radical growth of social media and mobile technology has been integral in matching sharing economy supply and demand along a vast network and with a reasonable level of trust (Hsu, Ju, Yen, & Chang, 2007). Trusting unknown providers is a prerequisite for consumers of car- and home-sharing. To this end, the majority of sharing economy websites incorporate social networking features to reduce the anonymity of transactions and build consumer trust in providers (Teubner,

Adam, Camacho, & Hassanein, 2014).

The implementation of online payment systems in the sharing economy has reduced opportunities for fraud, with companies such as Airbnb serving as middlemen between consumers and providers (Ranchordás, 2015). Prior to the emergence of the new sharing economy, peer-to-peer models for accommodation rentals and other goods and services required that customers transfer payments directly to providers, a process that many consumers perceive as riskier than dealing with a sharing company with a solid online reputation.

Social Factors

As consumers' ideas about the concept of value change (Brand & Rocchi, 2011), the opinions that have dominated people's mindsets in the past few decades have shifted from the traditional, industrial economy that focuses on the ownership of a product to an experience-based economy. Furthermore, Oskam and Boswijk (2016) stated that the information-based economy that focuses upon self-actualization has shifted toward a so-called transformational economy that involves consumers seeking meaningful life experiences. The paradigm has changed from one of mass production (with its attendant focus on branding and marketing) to knowledge-based platforms and value networks, generating higher levels of social engagement and awareness (Green, 2007). As a result, new communication technologies have emerged that allow consumers to become co-creators of value (Oskam & Boswijk, 2016).

The primary intrinsic motives fueling the sharing economy are social and environmental factors (Schor et al., 2015). The availability of sharing networks, social media, and review sites is also critical in explaining consumers' and providers' willingness to

participate in sharing business activities (Cohen & Kietzmann, 2014). Many older platforms (such as NeighborGoods, GrubWithUs, and LooseCubes) that offered social value, but no obvious revenue models have folded as new social media websites have emerged to bring about attitudinal and social changes, including a willingness to share opinions, recommendations, and information with strangers via review- and peer-based systems that were previously inconceivable (Schor & Fitzmaurice, 2015). The transparency inherent in the new platforms and service providers is a strong influence on creating trust, a central tenant and social validation of sharing economy services (Guttentag, 2015). Changes in societal norms and consumer tastes have also strongly driven sharing economy growth, a phenomenon that traditional travel companies would be wise to track (Schor et al., 2015). This area of the economy has been growing rapidly, and social norms have not yet been fully utilized in light of the changing realities (Teubner, 2014).

Political Factors

The sharing economy has also engendered risks. As mentioned in the previous section, attendant legal issues have been debated since the emergence of the sharing economy. With the rapid growth of sharing economy websites, the related legislation (e.g., laws or guidelines) has not been particularly well-established in most areas (Bergen & Guggenheim, 2016). In the area of sharing spaces, numerous rentals are illegal in some regions, particularly in residential neighborhoods (Streitfeld, 2014). Moreover, governmental authorities are aware of the proliferation of unlicensed rentals, and tax and registration revenue issues (Kaplan & Nadler, 2015) may also be a barrier to further development of the sharing economy. Although some sharing economy platforms have made adjustments with government officials in various areas in order to collect local taxes on behalf of individual

suppliers, the influence of legal concerns surrounding sharing-economy platforms on consumers' perceptions negatively impacts consumers' purchase intentions.

Since economic benefits are the main motivation for using sharing economy platforms, the unregulated pricing systems decrease the chance for clients to reap benefits from promotional bonuses (e.g., discounts), especially for rental businesses (Wang & Nicolau, 2017). Because hosts on such platforms determine pricing by themselves, it may be difficult to set promotions on a large scale. Jung et al. (2016) pointed out that because the sharing rental business hasn't yet been highly developed, it could take advantage of benefits such as lower rates (that do not include paying tax) in order to better compete with traditional lodging options.

In conclusion, as an increasing number of customers join the sharing economy, the private and commercial fields will continue to merge and interact. It has been shown that the sharing economy will continue to experience long-term growth; proponents of sharing have opened a door through which many users have followed (Teubner, 2014). Thus, managing sharing economy resources, offerings, and hosts efficiently is a matter of considerable importance for both users and suppliers, as well as society as a whole.

Kaplan and Nadler (2015) stated that companies in the sharing economy may vary widely in terms of the services they offer, but share three common traits: a reliance on recent technological advances to satisfy longstanding consumer demands in ways not previously possible; a position parallel to well-established industries that they have disrupted by the sharing economy's provision of innovative alternatives; and operation in interstitial areas of the law; presenting new and very different issues that could not be foreseen when the existing governing statutes/regulations were enacted.

Guttentag (2015) posited that tourists rent accommodations on Airbnb for not only the experiential value but also the economic benefits offered. Pricing is broadly recognized to be one of the principle critical determinants of customer purchase intentions when booking accommodations (Chiang & Jang, 2007). In Wang and Nicolau's (2016) study, Airbnb and hotel rates were compared for 11 cities in the United States and in Toronto (the largest Canadian city). It was found that Airbnb's prices were more competitive than those of hotels in seven out of the 12 cities studied. Wang stated that Airbnb offered better pricing particularly in the Northeastern and Pacific Northwestern U.S. cities examined. Moreover, Wang and Nicolau (2017) indicated that tourist-reported cost reductions are the main reason for consumer preference of sharing economy-based accommodation rentals.

The modern sharing economy exists for the purpose of achieving profits for individuals (Schor et al., 2015). The related online platforms for peer-to-peer sharing of goods/services offer innovative methods for users to earn income by "sharing" their possessions for a fee. The willingness to share goods and services with others in exchange for money is one of the most important motivators for why so many people want to share their unused resources (Hamari et al., 2016).

Thus, the sharing economy has received increasing attention in the past few years by scholars in different fields because of its significant impact on lowering costs and increasing business performance and profitability. Teubner (2014) summarized the literature related to the sharing economy and indicated that there remains a need for further investigation of the sharing economy, particularly with regard to the legal, economical, and behavioral aspects involved. What immediately strikes the eye is that very few previous studies have used qualitative approaches in investigating these aspects and challenges of the sharing economy.

Airbnb

Previous literature on Airbnb can be grouped into three categories. First, the uniqueness of Airbnb accommodations is discussed mainly from the perspective of its difference with regard to the lodging and travel experiences it offers. Outside the academic world, there is a wealth of salient studies regarding the comparison of traditional hotels and Airbnb accommodations, and the subsequent influence of Airbnb on the traditional lodging industry. Liu and Mattila (2017) suggested that Airbnb provides an especially unique customer experience that is radically different from that offered by hotels, and that this uniqueness should be exploited when promoting the site. The growth of Airbnb has not necessarily come at the expense of the traditional hotel industry. However, research has affirmed that the sharing economy is not a threat to the traditional hotel industry.

Nonetheless, Airbnb does enable overall market growth for a wider range of businesses (Kaplan & Nadler, 2015). Liu and Mattila (2017) explored Airbnb's online advertising strategies, which differ from traditional lodging companies by their emphases on "feeling at home (e.g., belongingness)" and an "atypical place to stay (e.g., uniqueness)" (p.33). A large number of studies exist that have primarily studied the social interaction benefits of using Airbnb. For example, Sundararajan (2014) observed that Airbnb has made significant investments in creating community and a sense of partnership, as well as in the dissemination of best practices. The author noted that Airbnb recently held a host convention featuring a number of sessions on being an optimal provider, and also has held regular host groups for knowledge sharing and integrated a host application with embedded hospitality standards and guidelines. Airbnb guests are able to get to know their hosts to various extents, and sometimes owe them some of their fondest travel memories (Dao & Vu, 2016).

Second, characteristics of P2P sharing transactions have been explored by Airbnb researchers. For example, Ert, Fleischer, and Magen (2016) showed that the likelihood of Airbnb providers attaining bookings is affected by the perceived trustworthiness of hosts' photos by conducting controlled experiments that assessed the effect of hosts' photos on guest choice to book. Edelman and Geradin (2015) identified the multiple methods of efficiencies that Airbnb platforms present, including reduced transaction costs; enhanced allocation of resources; and superior information and pricing efficiencies.

Third, other researchers have focused on the legal and financial issues pertaining to Airbnb. For instance, McNamara (2015) discussed the numerous responses and reactions that Airbnb has elicited, including state and local attempts to monitor peer-to-peer rental businesses; struggles to change federal laws; and potential outcomes of leaving Airbnb unregulated. Beyond the trends of the sharing economy, an increasing number of studies have examined Airbnb consumer behavior. Lee and Kim (2018) investigated brand personality with the aim of understanding travelers' perceptions of the Airbnb brand in the lodging industry. The results showed notable perceptual differences amongst high- and low-involvement travelers in the dimensions of excitement, sincerity, competence, and ruggedness. Wang and Nicolau (2017) identified sharing accommodation price determinants in the digital marketplace by mining Airbnb customer reviews. The results of ordinary least squares analysis and quantile regression analysis revealed that Airbnb consumers value the three determinants of super host status, greater numbers of listings, and verified identities as signs of quality.

Customer Experience

Previous scholars stated that “the customer experience originates from a set of interactions between a customer and a product, a company, or part of its organization, which provoke a reaction” (Gentile et al., 2007, p. 397). Meyer and Schwager (2007) indicated that customer experience has previously been defined as an internal, subjective response on the part of the customer to various direct/indirect contacts with a company along many “touch points”. The direct forms of contact typically are customer-initiated and take place during the purchasing and use of a product or service (Richardson, 2010). In contrast, indirect contact usually occurs during unplanned contact with the products, brands, or services of a company, often via positive or negative word-of-mouth, advertisements, media stories, customer reviews, etc. (Richardson, 2010). The concept of customer experience entails a “total experience” that includes searching for a product/service, purchasing, consumption, and various after-purchase stages, all potentially involving a variety of retail-sector channels (Verhoef et al., 2009).

Lodging Experience

Academicians have long sought to define customer experience. Prior to the emergence of customer experience as a concept, literature in the tourism field typically employed the concept of service quality and applied it in investigations of service attribute quality under controller supply, founded in Parasuraman, Zeithaml, and Berry’s expectancy disconfirmation theory (1988). Fick and Ritchie (1991) countered this approach, asserting that it fails to properly address holistic and affective factors that make contributions to service experience quality overall. Otto and Ritchie (1996) subsequently made a distinction between the attribute-based approach and their holistic/gestalt approach towards service

quality, stating that customers make internally referenced evaluations based on the self, as opposed to externally based evaluations of the service environment. The latter approach takes into account both supplier-provided attributes as well as the attributes inherent in customers themselves (Fick, Brent, & Ritchie, 1991).

Various hospitality/tourism studies have tried identified the dimensions of customer experience, with a variety of definitions of experience proposed. Simultaneously, researchers have attempted to take dimensionality approaches to customer experience. Hemmington (2007) classified hospitality experience into five dimensions: “generosity, the host–guest relationship, theater/performance, “numerous small surprises”, and safety/security” (p. 16). Knutson, Beck, Kim, and Cha (2009) examined the hospitality setting and proposed four customer experience dimensions: environment, driving benefit, accessibility, and incentive; while Walls (2013) developed the two broadly defined dimensions of human interaction and physical environment. Other researchers have undertaken micro-level examinations of customer experience by considering sensory aspects such as sound, sight, smell, touch, and taste (e.g., Pine & Gilmore, 2011; Schifferstein & Desmet, 2007). Specifically, Ren, Qiu, Wang, and Lin (2016) proposed seven items to represent basic aspects of the accommodation customer experience, suggesting that hotels should ensure room cleanliness, comfortable showers, quietness, comfortable room temperatures, and an odor-free environment. They also cited the importance for budget-conscious travelers of positive staff attitudes, hotel aesthetics, and location. The majority of the dimensions examined, however, were situated in the luxury hotel context or in other service settings and are thus limited in their applicability to the realm of shared accommodations.

Accommodation customer experience encompasses the types of memorable encounters experienced by guests during their travels. For hotels, the customer experience includes a continuum from making reservations through the selected channel to the actual stay (Knutson & Beck, 2004). Hosts with Airbnb specifically seek to manage this journey for their guests with their unique strategies for positioning, moving their guests along the continuum from expectations about their trips to their reviews once they return home (Brochado, Troilo, & Shah, 2017; Walls, 2013). Both luxury hotels and Airbnb hosts must design the customer experience to deliver their hospitality products/services with a focus on the customer perspective (Bharwani & Jauhari, 2013). Providers cannot view the customer experience as a static or passively, provider-designed entity. Rather, customer experience is a personal, dynamic, and proactive co-creation built by hosts and guests alike in unique, real-time service encounters (Neuhofer, Buhalis, & Ladkin, 2014).

From WOM to eWOM

Word-of-mouth

The widespread effects of word of mouth (WOM) have driven the critical need for businesses to gain an understanding of customer opinions and utilize them strategically to benefit their organizations. Among the first studies to examine the influential effects of WOM, Arndt (1967) described WOM as “oral, person-to-person communication between a perceived non-commercial communicator and a receiver concerning a brand, a product, or a service offered for sale” (p.190). Westbrook (1987) described WOM twenty years later as non-formal customer communication aimed at fellow customers regarding the ownership, characteristics, and sellers of goods/services. Subsequently, numerous studies have sought to elucidate the interplay of informal consumer communications, rumors, and purchase

decisions. The extant research on traditionally defined WOM primarily has focused upon its socio-psychological, managerial, and economic aspects. For instance, Babin, Lee, Kim, and Griffin (2005) found that WOM is one of the strongest influences on consumer purchase behaviors for food products and household goods. The results of studies by numerous other scholars since then have also affirmed the effects of WOM on purchase behaviors (e.g., Bansal & Voyer, 2000; Cheung, Lee, & Rabjohn, 2008; Ye, Law, Gu, & Chen, 2011). Brown and Reingen (1987) extended the concept of strong versus weak ties and stated the importance of combining WOM analysis at both the network and the individual levels. Laczniak, DeCarlo, and Ramaswami (2001) used attribution theory to look at the potential negative impacts of WOM on the variety of possible customer responses to negative WOM.

It has been empirically established that WOM has a significant role in influencing the decisions of tourists (Jalilvand & Samiei, 2012). Specifically, extant studies show that a visitor's overall satisfaction results in the likelihood to revisit a particular destination; share E-WOM; post reviews in online tourism forums; and share personal experiences and recommendations regarding travel products and destinations (Liu & Park, 2015). As such, industry practitioners emphasize the importance of encouraging travel consumers to engage online in "service stories". The widespread prevalence of Web 2.0 is a major factor in facilitating such eWOM via a wide range of channels for communicating online (Chan & Guillet, 2011). Of note is the use of social media, in which Web 2.0 has had a radical impact by allowing consumers to share information freely and efficiently; post the details of their travel experiences; and share their opinions with their online travel peers across the globe at any time (Hennig-Thurau, Gwinner, Walsh, & Gremler, 2004; Sigala, Christou, & Gretzel,

2012). Conventional WOM has now been overtaken by digital or “e” WOM in online venues. It is imperative given the current environment to gain a thorough understanding of the online behaviors amongst consumer groups sharing online. More specifically, it is crucial for businesses to develop extensive insights into the behavioral factors influencing users during online decision-making and purchasing (Robinson, Goh, & Zhang, 2012). Over time, consumer sophistication has reached new levels in searching for objective and candid WOM in order to make informed decisions. The travel sector is particularly sensitive to this increased consumer savvy, with Web 2.0 and Travel 2.0 playing important roles in the development of tourists as “knowledge consumers” who have applied the tenets and behaviors of social websites to the industry of hospitality (Buhalis & Law, 2008). Digital technology played an enormously influential role in new travel experiences because of its role in the emergence of worldwide distribution systems; computer and online reservation systems; dynamic packaging; mobile, multimedia, wearable, and virtual reality technologies; and so-called “smart tourism”. The unprecedented technology advances taken these years created entirely new functionalities and radically expanded the numerous potential venues.

eWOM

eWOM was defined as “any positive or negative statement made by potential, actual or former customers about a product or company, which is made available to a multitude of people and institutions via the Internet” (Hennig-Thurau et al., 2004, p. 39). One integral aspect of Web 2.0 is the facilitation of disseminating rumors and opinions in digital forums (“eWOM”). Litvin, Goldsmith, and Pan (2008) described eWOM as “all informal communications directed at consumers through Internet-based technology related to the usage or characteristics of particular good and services, or their sellers” (p. 461), including

producer-consumer communication and consumer-consumer communication, both of which are integral in the flow of WOM. In light of the perception of online reviews as a trusted information source regarding products/services, Nonnecke, Andrews, and Preece (2006) studied how members in online communities interact and create value through posting reviews, recommendations, and other content. In particular, they examined the important decision of community members to either participate in discussions or to simply “lurk” without contributing content. One salient concern of eWOM scholars, such as Schindler and Bickart (2005); Reza Jalilvand and Samiei (2012); and See-To and Ho (2014); is the motivation to share eWOM and the impact of such sharing on purchase decisions. As such, Schindler and Bickart (2005) identified three motives of consumers searching for eWOM: (a) the use of information to make purchase decisions; (b) the desire for a sense of community and support; and (c) entertainment value. In the study, it was found that the subjects utilized eWOM in order to make numerous decisions about large and small purchases that had utilitarian or hedonic value. It was found that subjects who were driven primarily by information-seeking motives tended to prefer reading direct comparisons of products/brands, and often focused on negative information regarding alternatives, which is often found in online customer reviews. It was also found that subjects sought out eWOM in order to bolster support for decisions that they had already made and to find communities pertaining to products and services. The study subjects also sought out positive information to bolster support/confirm previous decisions, and many stated their primary motivation for reading eWOM was entertainment, with discussion forums as the most popular source.

Initially, eWOM was perceived as a straightforward conversion of traditional WOM onto the Internet (Baka, 2016). Dellarocas (2003), however, noted that e-reputation systems

placed the traditional networks for WOM on a vastly larger scale. Furthermore, the move online has had a radical impact on the nature and power of WOM due to the greatly increased scope and magnitude, as well as the entirely new stylistic and cultural implications of the digital platform. Litvin et al. (2008) compared in-person WOM to eWOM and stated that “far different from physical WOM, eWOM can create virtual relationships and communities with influence far beyond the readers and producers of WOM; it actually creates a new type of reality by influencing readers during their online information searches” (p. 10). Not only can it be posited that eWOM has transformed our ways of communicating, living, and judging our lived experiences on both the personal and the organizational levels, it can also be stated that eWOM seems virtually inescapable in the lives of anyone who uses the Internet (Baka, 2016).

User-generated Content

Moens, Li and Chua (2014) defined user-generated content (UGC) as any content type generated in online platforms by users, while Burgess, Sellitto, Cox, and Buultjens, (2009) proposed UGC as the opposite of traditional forms of media/marketing because it originates from the lay consumer, not the marketing professional. User-generated content exists most frequently as supplementary to digital platforms such as social media sites, and can consist of content such as blog posts, Wikipedia entries, comments, e-commerce reviews, and customer uploaded videos (Kurtz, Mak, & Werndorfer, 2006). User-generated content websites are equivalent to eWOM marketing, in which a consumer shares his/her opinions, beliefs, views, and experiences regarding products and services with other consumers on the Internet (Burgess, et al., 2009). The content of blogs in particular can be extensively mined for information to facilitate customer profiling, acquisition, and engagement, as well as

brand awareness and reinforcement; customer service; and reputation management (Laboy & Torchio, 2007). Dhar and Chang (2009) cited the rich new opportunities for customer storytelling afforded by UGC, which can help build trust-based relationships for potential customers. Such opportunities are important aspects of the rapidly increasing power of UGC in consumer purchase decisions (Jiménez & Mendoza, 2013).

Motivation for Creating UGC

The rise of the development of the internet and its extensive use has provided consumers with ever-expanding access to peer opinions beyond individuals' real-life friends, family members, and acquaintances. Many studies have shown that consumers are motivated to create UGC for their own benefits (e.g., Burgess et al., 2009; Daugherty, Eastin, & Bright, 2008; Holsing & Olbrich, 2012), especially for the advantages in knowledge sharing and seeking that they obtain. Thus, the knowledge function states that individuals are motivated to interact with others in order to gather information and thus better organize and comprehend their environments. In other words, humans are instinctually motivated to make sense of their experiences. Researchers can employ this tendency in understanding how UGC is created to help individuals be aware of their environments, the topics at hand, and even their selves through the intrinsic knowledge gained during the content-sharing process. The process of creating UGC is value-expressive and enables individuals to share their values and self-concepts and experience an enhanced self-image in the eyes of peers with similar values. As a result, UGC creators can experience inherent gratification and increased self-esteem by sharing content in online communities with other members with whom they share important beliefs and principles. Such validation is gratifying to the self and to one's worldview. Furthermore, creating UGC can serve an ego-defensive function and can buffer participants

from personal insecurities or external threats, thus providing defenses to self-image. User-generated content creators with ego-defensive motives can share online as a way of assuaging self-doubt, experiencing a feeling of belonging, and reducing guilt when not making contributions.

The four functions above are considered by researchers to comprise the main constructs in understanding attitudinal motivations; however, contemporary scholars have further explored and clarified additional aspects. Knutson (1973) extended the value-expressive function by focusing on social adjustment motivation, which involves the expression of behaviors and/or attitudes that others find agreeable. Social adjustment motivation has been expanded to explain motivations grounded in relationships with others and distinguishes between the individual's internal beliefs and his/her desire for external relationships, regardless of moral values. Notably, the social function is a driving force that explains why individuals seek out chances to interact with their friends or share in those activities that are favorably perceived by esteemed others (Clary et al., 1998). Thus, the social function is a significant motivator in creating UGC since it entails social interaction and information sharing on a widespread level. The creators and consumers of UGC may also be motivated by the social function due to the ways in which esteemed reference groups might view membership in a given online community.

User-generated Content and Travel

Internet has overtaken a critical new role in tourism and travel planning as compared to the traditional travel agency (Burgess, et al., 2009). With the assistance of Internet, “a new type of user is emerging, one who acts as his or her own travel agent and builds a personalized travel package” (Werthner & Ricci, 2004, p. 101). Tourists become more and

more gathering travel information independently and decision making regarding travel destinations, products, and services, with diminished or no assistance from intermediaries. Travelers today may make decisions based not on curated images in advertisements or brochures, but on informal images posted online by customers on social media and networking websites. User-generated scores and reviews are expected components of travel agency websites such as Expedia.co. and Priceline.co. Users themselves have a direct impact on the display order of information and generate a significant impact on potential customers' decision making.

In the past, tourists and travelers searching on the Internet were forced to rely on the information provided on the travel intermediaries websites, travel service providers, and tourism destination marketers, with a degree of assurance of the validity of the sources. Later, Chung and Buhalis (2008) indicated that “type and relative importance of information sources have changed over time” (p. 72). Specifically, the huge volume of information obtainable from the Internet seemingly endless content providers has resulted in concerns about the trustworthiness and validity of information. In the early 2000s, Senecal and Nantel (2004) showed that a considerable number of consumers may feel distrustful of information that appears to promote the creator's self-interests, and thus prefer to rely on the product and service evaluations of their fellow consumers. In the travel e-commerce context, this has resulted in an increasing number of tourists preferring to seek out UGC to learn about products because they trust these sources more than information provided by sellers (e.g., travel agents, hotel companies) (Gretzel, Yoo, & Purifoy, 2007).

Many scholars have also proposed that noncommercial information is generally considered more credible and objective; thus, peer-generated content is often regarded as

more trustworthy (Amaral, Tiago, & Tiago, 2014; Kwak, Kim, & Zimmerman, 2010). Gretzel et al. (2007) indicated that commercial website user reviews are deemed less trustworthy and objective by consumers than reviews posted on dedicated, non-commercial review sites that offer no benefits to the commercial entities being reviewed. TripAdvisor is currently the most well-known dedicated UGC content review website in the travel and tourism industry, and the upsurge in UGC on such sites has had an indisputable impact on tourists' decisions.

In a 2007 study by Gretzel, Yoo, and Purifoy, statistical data was mined from Complete, Inc. and analyzed. The results suggested that nearly 50 percent of consumers buying travel products had used UGC planning for their trip and almost one out of three stated that the UGC they viewed had proven useful in making their decisions. Harwood (2007) cited research findings from Nielsen/Netratings that claimed that UGC websites were deemed the most credible/reliable source of information by more than one-fifth of respondents, a figure that is nearly double that of their closest rival, travel agency websites. The majority of online users view travel consumer reviews as prefer to provide reliable, latest, and pleasant content, in contrast to information from professional travel service providers (Ye et al., 2011). Consumers who travel frequently perceived peer traveler reviews as superior, thus showed the greater likelihood to be strongly influenced by this type of content (Gretzel & Yoo, 2008). The authors also found that more than 50 percent of users relied on online reviews every time when planning pleasure trips, primarily at the early trip planning phase in order to gather ideas and narrow their selections. A lesser number of users reported that they consulted peer traveler reviews in the later stages of trip planning in order to confirm their choices (Gretzel, 2006). A more recent study revealed that online consumer

reviews are especially important for consumers choosing accommodation products, with significantly less relevance when choosing other types of travel products (Park, Xiang, Josiam, & Kim, 2014).

Online Customer Reviews

Reviews generated by customers that post online are described as “peer-generated product evaluations posted on company or third-party websites” (Mudambi & Schuff, 2010, p. 186). Currently, online reviews left by customer could be accessed for an extensive variety of product/service categories, including experience goods (e.g., videos) and search goods (e.g., cellular phones) (Ye, Li, Wang, & Law, 2014). These reviews offer users information not only about products and services, but also about the post-consumption experience (e.g., value for money, quality, and overall judgement). Online customer review is a critical information source that helps consumers and marketers learn about product quality (Chen & Xie, 2008). Customers now seek out and read reviews online before making purchases more than at any other time in history.

Mudambi and Schuff (2010) conducted a study to determine how helpful customer reviews are perceived using as background. They assumed that “helpfulness as a measure of perceived value in the decision-making process reflects information (i.e., online review) diagnosticity” (Mudambi & Schuff, 2010, p. 186). They found out that customer review elaborateness has a positive effect on perceived helpfulness. Liu (2012) studied the credibility of consumer reviews using sentiment analysis to mine review text. Using the dual-process theory, he discovered that customers have a tendency to focus on a variety of information in previous reviews. For examples, peripheral cues (e.g., customer ratings and rankings) are deemed helpful during the information-searching process; in contrast, central

information processing (e.g. looking at the total amount of words or negative words in a review) matters to consumers when making evaluations of similar products/services. Hu, Liu, and Zhang (2008) reached the conclusion that customer website reviews are perceived as inferring product/service quality and reducing product/service uncertainty, which aids consumers in making final purchase decisions.

Many studies have looked at the relationship between product sales and online reviews. A 2006 study by Chevalier and Mayzlin discovered that a positive relationship exists between customer reviews on book retailer platforms (e.g. Amazon.com) and product sales. It was found in a subsequent study that the average numerical rating (or valence) and the number of reviews on such sites (Duan, Gu, & Whinston, 2008) were critical in predicting movie ticket sales at box offices. Additionally, in 2006, Clemons, Gao, and Hitt discovered that craft beer sales were significantly impacted by online reviews; specifically, by the rating variances and the strong influence of the 25 percent most positive reviews. Researchers have also examined the effects of the extent of review writers' identity disclosures in evaluations of product quality on product sales (e.g., Hu et al., 2008; Hu, Pavlou, & Zhang, 2006). While some researchers believe that high consumer scores on products mean increased online sales, Moe and Trusov (2011) showed that this is not necessarily the case. Specifically, they indicated that heterogeneous consumer tastes can be accounted for by the different ways that consumers employ the opinions of fellow consumers in deciding whether or not to purchase products.

In the field of hospitality and tourism, numerous academicians have looked at the effects of online reviews on decision making when purchasing an entire travel package (Pan & Fesenmaier, 2006), choosing hotels (Ye, Law, & Gu, 2009), and choosing restaurants (Lu,

Ba, Huang, & Feng, 2013). Researchers in the field have also estimated the market shares of travel products based on online reviews (e.g., Sparks, Perkins, & Buckley, 2013; Ye, Zhang, & Law, 2009), while Vermeulen and Seegers (2009) looked at how online hotel reviews affected consumer considerations and observed that hotel awareness is increased by online reviews and that travelers are aided by online reviews when developing consideration sets. Ye et al. (2011) assumed that the reviews quantity includes the linear sales function and then estimated the effects of customer reviews topic on hotel sales. It was discovered that the ratings of reviews and room pricing are critical in order to predict online hotel room sales. Zhang, Ye, Law, and Li (2010) examined online information that reflected context-specific restaurant variables (e.g., food quality and service), along with customer overall ratings and the amount of reviews that reflected the online popularity of restaurants.

Web reviews are helpful in the service-oriented industry to investigate customer perceptions (Zhang et al., 2010). In comparison with customer interviews and questionnaires, customer reviews have proven to be an effective, innovative approach for investigating customer perceptions. A number of studies have employed online reviews in order to explore customer perceptions and customer satisfaction. Pantelidis (2010) performed content analysis on online comments about different types of restaurants and revealed that food, service, price, ambience, decor, and menus all affect guest satisfaction. Zhang et al. (2010) showed that the higher quality of customer reviews indicated, the more popular of restaurants are, while Ryu, Han and Jang (2010) analyzed online review data and revealed that the physical environment and service quality in quick-casual restaurants are important determinants of customer satisfaction. The results of experimentation undertaken in Vermeulen and Seegers' (2009) study indicated valence of customer reviews, reviewer expertise, and hotel familiarity

are independent factors capable of enhancing consumer consideration of hotels. In the present study, the author utilized online review data to examine customers' judgement of quality and value.

eWOM and Reviews

The negativity effect, which state that negative customer reviews have a larger effect on customers compared with positive reviews, is more pronounced for female consumers, according to a study by Mo, Malik, and Coulson (2009). Zhu and Zhang (2010) bolstered the idea that reviews could serve as an overall customer satisfaction reflection, which consequently impact the decision-making processes of other possible customers seeking for product information online strongly (Chevalier & Mayzlin, 2006). Hsieh, Hsieh and Tang (2012) examined how eWOM is disseminated from the perspective of message and investigated the elements that contribute to making engaging online videos and influencing viewers' intentions to forward videos to others. They further studied online video persuasiveness and employed the dimensions of content, source, and channel to look at three factors with potential influence: awareness of persuasive intent, multimedia effects, and perceived humor.

From the opinion mining perspective, Robinson et al. (2012) looked at the ways in which customers interpret and are influenced by consumer reviews, as well as examined online review textual factors in order to develop an effective approach to employ raw data from opinion text in online reviews to extract reliable, accurate, useful, and impactful information. The factors that enhance the persuasiveness of online opinion text were identified and employed to develop approaches for opinion mining. Few studies have been

undertaken in hospitality area in spite the growing importance mining social media contents and its impact on hospitality marketing. Prior studies primarily focused on the issues of:

1. Tourists' narratives as digital WOM and their marketing potential (Confente, 2015);
2. The impact of user reviews and blogs (Rong, Li, & Law, 2009);
3. The role and profiles of particularly helpful reviewers ("opinion leaders") in online social travel networks (Lee, Law, & Murphy, 2011);
4. The performance of hotel companies in viral and social media marketing (Chan, & Guillet, 2011).

The present author proposes that the foci of the above-described studies on certain issues pose specific limitations. These limitations include the issues under exploration, the perspectives utilized by the researchers, and the types of media that were investigated. As a result, the findings of these studies are not generalizable to all types of social media, and the literature cited above exemplifies how there is insufficient research into social media's use by tourists as a vehicle of WOM. This is particularly noteworthy because social media has increasing influence over interpersonal communication and thus online reviews and eWOM have ever-greater importance.

The present study next examines the issues critical to developing a theoretical framework for understanding the impacts of eWOM. For example, online communication is bipolar, consisting of sender/narrator and recipient/receiver messages. In the former, influential individuals (friends, relatives, opinion leaders) or less influential individuals (strangers, acquaintances) give eWOM in the form of recommendations or reviews. Opinion leaders are individuals who (in real life or online) offer informed, trusted opinions and are

important and influential (Litvin et al., 2008). Such leaders influence those seeking opinions; interpret meanings for others (Lee et al., 2011); and influence the decisions of community members online and in real life (Yoo & Gretzel, 2008). According to this communication perspective, sources/media, motivation, and mediating variables are the most important issues involved.

According to research, tourists who post eWOM and customer reviews on tourism products/services are motivated by various emotions (Söderlund & Rosengren, 2007), specifically pleasure, satisfaction, and sadness (e.g. Nyer, 1997); by altruism (Gretzel et al., 2007); by the need to exhibit reciprocity (Dellarocas, Fan, & Wood, 2004); and by the simple pleasure derived from sharing travel expertise and experiences, particularly in the post-trip process of sharing (Litvin et al., 2008). These motivations have been deemed the primary ones that drive WOM contributions.

For the eWOM recipient, the mediating variables and the outcomes are critical issues. There are four mediating variables that exercise influence over recipients:

Source evaluation. Communication researchers define source evaluation in terms of the credibility of the message, source, medium, and media; the trustworthiness of the information and source; and the reliability of the source (Moutinho, Ballantyne, & Rate, 2011; Savolainen, 2007). An important mediator in source evaluation is trust in peer community members (Yeh & Choi, 2011).

Brand familiarity. In consumer behavior psychology, brand familiarity is defined as a heuristic that uses prior actions or schemas as the basis for future actions in new (yet familiar) situations (Ashcraft, 2008). This heuristic results in an increased likelihood of repeat purchases of same-brand products. The familiarity heuristic posits a customer rule of

thumb that past behavior (such as buying the product of a certain brand) is likely correct and therefore it is desirable to repeat the behavior (Sundaram & Webster, 1999).

Socio-metric integration. This concept denotes the degree to which a consumer is integrated into his/her community (e.g., the academic community) (Litvin et al., 2008) or a brand community (Yeh & Choi, 2011).

Memory. In the field of psychology, memory is defined as the processes that encode, store, and retrieve information. Encoding involves the physical and chemical stimulation of the senses in order to convey information from the world outside of the individual. Recall and recognition memory are distinguished in the literature (Carlson, 2010). The former encompasses memory tasks requiring the individual to discern if she or he has previously encountered given stimuli (e.g., words, images). The latter involves tasks requiring the individual to retrieve information that was learned previously.

The digital context provides numerous new opportunities for travelers and tourists, as well as for businesses and destinations. Specifically, some social media sites offer opportunities for both businesses and customers interactively to post numerous opinions on the website (Edelman, 2007). As such, this variable was included in the framework for both perspectives. The dissemination of eWOM has four anticipated outcomes: (a) purchase decision influence (specifically, positive WOM will increase purchase likelihood; negative WOM will decrease purchase likelihood) (Hsieh et al., 2012); (b) product evaluations (Dellarocas, 2003); (c) consumer loyalty intentions (Gruen, Osmonbekov, & Czaplewski, 2005), and (d) consumer empowerment (Bae & Lee, 2011; Litvin et al., 2008). The concepts “source credibility” and “degree of involvement” require more detailed clarification, which follows (Sotiriadis & Van Zyl, 2013, p.112).

Source credibility is related to reputation and can be defined as the level of trust in an individual. The degree of trust is influenced by the receiver's belief in a communicator's honesty, concern for others, and short-term unalterable opinions (Helm & Mark, 2007). Online reputation systems aid users in reducing uncertainty about tourism product quality/performance by helping tourists identify trustworthy individuals that can help in decision making (Helm & Mark, 2007; Zhu & Zhang, 2010). The previous research suggested that both trustworthiness and expertise affect the credibility of a source, thus resulting in more positive evaluations of reviews (Hertzum, Andersen, Andersen, & Hansen, 2002). Notably, individuals have a tendency to rely on experts when lacking expertise in a particular area, a fact that was underscored by Lee, Park, and Han (2011).

Keng, Liao and Yang (2012) identified the relationship between product involvement in consumer experience and perceived risk. The intangible nature of tourism services (which cannot be assessed before consuming) causes consumers to perceive them as quite risky (Havitz & Dimanche, 1990; Middleton, Fyall, Morgan, & Ranchhod, 2009). Havitz and Dimanche (1990) suggested that tourism practitioners should pay close attention to how the involvement construct interrelates to search behavior and promotion-based stimuli in light of the fact that tourism service purchases require high levels of involvement. Purchases are therefore deemed high-risk, particularly with regard to emotional risks. The evaluations of a reference group are instrumental in consumers' decision-making processes. For this reason, recommendations and suggestions have greater influence and place greater demands on marketers of tourism products. The above-discussed literature has allowed scholars to discern the most salient aspects and issues of eWOM and elucidate the concepts and terms involved.

Big Data and Text Mining

Xiang, Schwartz, Gerdes and Uysal (2015) defined big data as very large data sets obtained through a variety of sources such as Internet traffic; mobile transactions; social media; clickstreams; intentionally captured content on sensor networks and in retail transactions; and from a variety of additional domains including healthcare, bioinformatics, and finance (Chen, Chiang, & Storey, 2012; Raghupathi & Raghupathi, 2014; Savage, 2014).

The goal of big data analytics is the generation of new insights (often in real time) to supplement and complement traditional, generally static sources of data such as statistics, surveys, and archival sources (Davenport, Barth, & Bean, 2012). The growing application of big data analytics to business intelligence is employed to better understand market characteristics, customers, products, competitors, business environments, technology impacts, and strategic stakeholders (e.g. alliances and suppliers) (Wixom, et al., 2014). There are numerous case studies and examples to show how big data analytics could be applied in order to discover/resolve business problems (Chen et al., 2012). Goh, Heng, and Lin (2013) observed that consumer-generated content of social media, in particular, are desirable sources of public and community data.

Text mining (also referred to a knowledge discovery from textual databases or text data mining) (Hung & Zhang, 2012; Tan, 1999) involves the extraction of non-trivial, meaningful knowledge or patterns in unstructured text sets. Aggarwal and Zhai (2012) defined text mining as the analysis of data in natural-language texts, while other researchers have viewed text mining as an extension of data mining (Chau, Shiu, Chan, & Chen, 2007). Text mining serves to process unstructured information and extract meaningful numeric indices from such information, a process that generally involves ‘turning text into numbers’

(Krallinger, Valencia, & Hirschman, 2008). Text mining seeks to derive high-quality information from text and is usually performed via the recognition of data patterns (Aggarwal & Zhai, 2012). Numeric indices make information accessible for further analysis or data mining (statistical and machine learning) algorithms (Meyer et al., 2008; Sebastiani, 2002).

User-generated content streams can result in the creation of vast amounts of data over time, which can be utilized for pragmatic assistance and can supplement traditional research methods in order to identify salient issues (Gopalkrishnan, Steier, Lewis, & Guszczka, 2012). Greater numbers of studies in the past decade that have examined online consumer reviews from the perspective of big data (e.g., Chen & Xie, 2008; Gretzel & Yoo, 2008; Sparks & Browning, 2011) and proposed that user-generated content is a rich data source for extracting the determinants of satisfaction in the field of hospitality and tourism. Specifically, researchers have noted that the huge amount of the online communities' users can involve in and participate in the content generation, which resulting "wisdom of the crowds" (Surowiecki, 2005, p. 2), and have posited the use of user-generated content for enterprises concerned with better understanding consumer demands, particularly for hotels and restaurants.

The majority of previous marketing and tourism studies applied the UGC analysis to assess the impact of ratings on hotel websites (e.g., Schuckert, Liu, & Law, 2016; Ye et al., 2009; Zhu & Zhang, 2010). Ye et al. (2009) indicated that online ratings are numerical and easy to comprehend; in contrast, online reviews are text-based and typically contain much more information than can be analyzed by traditional statistical and econometric methods. Sparks and Browning (2011) called for an expansion of the sparse empirical findings from

huge amount of online review data sets to help understand both consumer satisfaction and the determinants of consumer satisfaction. The traditional evaluations of the lodging experience cannot comprehensively measure the consumer experience in the context of the rising sharing economy accommodations platforms such as Airbnb. However, few attempts have been made in investigating Airbnb customer lodging experience. Moreover, to the knowledge of the present author, no previous research has investigated the customer lodging experience by mining the sentiments in reviews under the umbrella of big data. This study aimed to fill these research gaps.

Sentiment Analysis

The term “sentiment analysis” emerged in 2003 (Nasukawa & Yi, 2003), opinion mining was surfaced in 2003 as well (Dave, Lawrence & Pennock, 2003). However, investigation about human sentiments and opinions had been conducted earlier (Boshoff & Van Eeden, 2001; Godbole, Srinivasaiah, & Skiena, 2007; Pang & Lee, 2008; Yi, Nasukawa, Bunesco, & Niblack, 2003). Sentiment analysis, also referred to as opinion mining, is defined as the analysis of opinions, attitudes, sentiments, appraisals, evaluations, and emotions towards entities (and their attributes) including products, services, individuals, organizations, events, issues, and topics (Liu, 2012). Nasukawa and Yi (2003) extended the definition of sentiment analysis as a method to “identify text fragments that denote a sentiment about a subject within documents rather than classifying each document as positive or negative towards the subject” (p. 71). The term “sentiment analysis” is more widely used in industries, but the terms “opinion mining” and “sentiment analyses are often employed in academia (Dey & Haque, 2009; Ortigosa, Martín, & Carro, 2014). They essentially are all the identical field of study (Pak & Paroubek, 2010; Pang & Lee, 2008).

Sentiment analysis and opinion mining are derived from artificial intelligence; natural language processing; and information retrieval and extraction (Nasukawa & Yi, 2003).

Sentiment analysis includes the following steps: (a) the location of relevant documents about a specific topic or for a specific purpose; (b) the pre-processing of the documents (e.g., single-word tokenization and extraction of salient information); and (c) the classification of sentiments that characterize the product/organization (Schmunk, Höpken, Fuchs, & Lexhagen, 2013). Liu (2012) stated that text mining has a broader scope; in contrast, sentiment analysis is a form of text mining focused on the identification of contained opinions, sentiments, and subjective statements, especially in online UGC.

Sentiment analysis can represent a large problem space with various names and various tasks, including opinion mining/extraction; sentiment mining/analysis; review mining; and subjectivity, affect, and emotion analysis (Liu, 2010), which now all fall under the headings of sentiment analysis or opinion mining (Liu, 2010). There are three primary levels of sentiment analysis classification: sentence-level, document-level, and aspect-level. The first serves to classify opinion documents according to whether they express positive or negative opinions or sentiments and assesses the entire document as a basic unit of information that covers a single topic. Sentence-level sentiment analysis classifies the sentiment expressed in an individual sentence (Medhat, Hassan, & Korashy, 2014), while aspect-level sentiment analysis finds the overall sentiment of an entity as well as the sentiment of the entity aspects under discussion (Schouten & Frasincar, 2016).

Two main approaches exist in sentiment analysis, namely the supervised learning approach and the unsupervised learning approach (Aggarwal & Zhai, 2012). In the former, parts of the data, such as observations or measurements, are labeled according to pre-defined

classes such as “like” or “dislike”. Then, the rest of the data is treated as the test data, which uses the model derived from training dataset to classify data into these classes. Unlike supervised learning, the labels of the data of unsupervised learning are unknown.

Other Text Mining Techniques

There is a long history of research on both linguistics and natural language processing (e.g., Clark, Fox, & Lappin, 2013; Manning & Schütze, 1999; Jurafsky, 2000) and a dearth of research regarding consumer opinions and sentiments prior to 2000. Since then, it has become a highly active area of research, due to its wide range of applications in a plethora of domains (Hu, Bose, Koh, & Liu, 2012), and the sentiment analysis industry has flourished in light of the proliferation of commercial applications, thus fueling much research (Liu, 2012).

Nonetheless, sentiment analysis is limited in its effectiveness, as are all data mining techniques. For example, opinion mining seeks to classify an opinion along a polar spectrum (Lak & Turetken, 2017), with the ends of the spectrum corresponding to positive/negative feelings about brands, products, or people. Furthermore, all feelings are subjective according to the individual and can be irrational (Guess, 2015). While the use of large amounts of data is critical when measuring sentiment, an individual’s feelings toward a product/brand are subject to indirect influence by one or several factors. In addition, sentiments often change over time based on the individual’s mood, world events, and other factors. It is therefore critical to examine data from a time-related standpoint (Guess, 2015). Finally, sarcasm and irony are difficult if not impossible for machine identification when considered in isolation (Wang, 2013).

Textual Analysis

Textual analysis is a method employed by communication researchers with the purpose of interpreting and describing characteristics of visual or recorded messages (Fairclough, 2003). Its purpose is the description of the structure, content, and function of messages contained within texts (McKee, 2003). Texts fall in two general categories: transcripts of communication and outputs of communication, the latter of which are generally more readily available (Fairclough, 2003).

There exist four main approaches to text analysis: content analysis, rhetorical criticism, interaction analysis, and performance studies. In content analysis, occurrences of specific messages and embedded message characteristics are identified, enumerated, and analyzed (Holsti, 1969). This approach shows greater concern for which meanings are associated with messages than with frequency at which message variable occur (Graneheim & Lundman, 2004). Rhetorical criticism involves a systematic method to describe, analyze, interpret, and evaluate messages embedded within texts in terms of their persuasive force (Bormann, 1972) and is a type of criticism that uses rhetorical principles to study interactions between text, author, and audience. Interaction analysis defines interaction as a multifaceted act that requires considerable knowledge from communicators, along with the capability to coordinate behavior with other communicators. Performance studies is defined as “the process of dialogic engagement with one’s own and others’ aesthetic communication through the means of performance” (Denzin & Lincoln, 2011, p. 411). For the purpose of this study, none of these four methods of textual analysis are suitable for detecting the sentiment of reviews.

Topic Modeling

Topic modeling (TM) is a technique that is useful to identify the overriding (dominant) themes in a huge set of documents and to work with a large text corpus (Wallach, 2006). The primary idea of TM is to extract topics from a body of text and discern what the topics are based on the words that are used in the text (Lu & Zhai, 2008). Topic analysis involves classifying documents from a corpus based on topic and is similar to text classification. Topic model is a statistical model for revealing the main topics in a set of textual data (Hong & Davison, 2010), and TM is frequently used as a text-mining tool to learn about the hidden semantic structures in a body of text (Aggarwal & Zhai, 2012).

Topic relations can be extracted from a variety of co-occurrence relations; thus, documents from discussion forums on the internet can theoretically produce meaningful relationships among topics (Steyvers & Griffiths, 2007). For example, synonyms are words in a topic with similar meaning, and are also similar to words in similar topics. Another example is polysemy, defined as words with different meanings which can appear in other topics simultaneously. The use of TM allows for the disambiguation of the meanings of words from those of other, similar topics (Williamson, Wang, Heller, & Blei, 2010). Moreover, for large document sets, the scalability issue of topic modeling may arise. Another considerable drawback of TM is that the number of topics is set and therefore must be determined in advance.

Machine Learning

Machine learning is the branch of computer science defined as “the computer’s ability to learn without being explicitly programmed” (Samuel, 1959, p. 411). Machine learning investigates the architecture of algorithms to learn from data and make data-based

predictions. In other words, machine learning algorithms enable researchers to make predictions and decisions driven by data, not just by static program instructions (Witten, Frank, Hall, & Pal, 2016). Machine learning is strongly related to mathematical optimization, which provides theories, methods, and application domains to machine learning. Machine learning is at times confused with data mining (Blum & Langley, 1997). To clarify, text mining is a critical sector of data mining that uses machine-learning techniques (classification, clustering, predictive modeling, and association rules) to uncover relationships and meaning in content (Wunnava, 2015). As mentioned above, sentiment analysis is a typical text mining technique to analyze unstructured data, and employed some machine learning methods, such as tokenization (identifying distinct elements, such as n-grams or words), parsing, term reduction (grouping words with similar meaning via similarity measures and synonyms), stemming (reducing word variants to bases), and parts-of-speech tagging (using POS tags) which help discern facts and relationships. In conclusion, machine learning contains several important techniques that are fundamental to some sentiment analysis processes, such as data preprocessing and classification.

Semantic Analysis

Semantic analysis is a method to obtain and build structure for unstructured data, such as social media posts and social network chatter. However, this method does not preconceive whether a content is related or how it is related to others (Baroni, Dinu, & Kruszewski, 2014). Semantic analysis distills important, useful information from large bodies of unstructured data, explores the meaning of natural speech in online posts, and reveals particular foreign language meanings. Semantic analysis enables scholars to cluster various data components based on similarity, rather than on preset classifications such as positive,

negative, and neutral (Guess, 2015). Because the aim of this study is to study Airbnb users' evaluations of their lodging experiences via mining their online reviews, detecting user sentiments toward different aspects is a key step. Based on the discussion above, sentiment analysis is more suitable for this research design.

Aspect-based Sentiment Analysis

The majority of early sentiment analysis research primarily consisted of classifying negative and positive sentiments in a binary system in order to offer predictions about the overall sentiments expressed in review documents. In contrast, a number of more recent studies have employed aspect-based sentiment analysis (ASBA), which offers the advantage of greater depth of analysis (Thet et al., 2010). The ABSA systems usually employ textual data such as messages or reviews about particular products (for example, cell phones) on social media as input and seek to detect the aspects discussed most frequently (the main features) of an entity (for example, a cell phone screen or battery) and determine the overall positivity or negativity of reviews according to each aspect (estimating the average sentiment). Although various ABSA systems have been developed based on prototypes (Liu, 2012), no ABSA task decomposition has been established. Moreover, there is a lack of measures to evaluate ABSA system subtasks.

The aforementioned earlier studies primarily conducted document-level analysis in order to assign sentiment orientations to documents (Moraes, Valiati, & Neto, 2013; Yessenalina, Yue, & Cardie, 2010; Zhang, Zeng, Li, Wang, & Zuo, 2009). Since then, researchers have gone into greater depth and performed sentiment analysis at the sentence level to examine discrete aspects of objects under review (Boiy & Moens, 2009; Liu, 2012; Wilson, Wiebe, & Hoffmann, 2005), a generally more sophisticated approach. In 2004, Hu

and Liu mined data from consumer reviews of electronics (including cell phones, digital cameras, music players) and identified the product feature aspects (such as screen size and product quality) of the products in order to make predictions about the positivity or negativity of opinion sentences. The prevalence of positive or negative opinion-oriented words was interpreted as whether or not positive/negative were expressed at the sentence level. Blair-Goldensohn et al. (2008) developed a system to summarize review sentiments regarding local services (hotels, restaurants, department stores, etc.). The resulting system was employed to extract relevant aspects (such as food, experience, and value) of these service-oriented products and aggregate sentiment per aspect as revealed through aspect-relevant text according to positive-negative polarity values. The study employed a dual (lexicon-based and maximum entropy) approach in order to classify each individual sentence in a review as neutral, positive, or negative. Ding, Liu and Yu (2008) also identified a sentence-level from the lexicon-based approach; however, they analyzed short passages consisting of few sentences as a single sentence. They also examined the challenges of using binary-valued sentiment orientations but refrained from attempting to assign sentiment scores.

Latent Aspect Rating Analysis

The fast-growing accumulation of textual opinion data on the Internet has raised compelling new challenges in text mining and has resulted in numerous studies in an effort to extract increasing amounts of useful information from reviews and better understand the customer evaluation experience. In order for large numbers of online product reviews to be quickly and accurately interpreted by users, detailed opinion information must be provided for multiple topical aspects (such as battery life) of a product entity (Wang et al., 2010). This need has prompted recent research attempts to utilize opinion mining to conduct “fine-

grained” sentiment analysis. In the majority of these studies (e.g., Liu & Seneff, 2009), the algorithms used are capable of identifying ratings or sentiment orientations for specific (topical) aspects and providing detailed, useful opinion summaries. Nonetheless, users demand greater information than that which can be decomposed from overall ratings into specific aspect ratings (Wang et al., 2011). Specifically, a hotel that receives five stars out of five for the aspect of “value” may nonetheless be considered very costly by the general standards of reviewers (such as business travelers) that place much more emphasis on “service”. Vacationing reviewers who are more concerned with price, however, may deem the same hotel reasonable in this aspect.

Wang et al. (2010) sought to discern these types of variations in reviews and infer not only aspect ratings but also the relative emphases made by different reviewers on various product aspects, resulting in a new method for opinion mining named LARA. The LARA process involves using review document text sets with overall ratings about entities as input and generating as output: (a) ratings on sets of predefined aspects of the entities, (b) the relative weights that reviewers placed on individual aspects when making reviews. As such, LARA both decomposes overall ratings into discrete ratings on different topical aspects (for example, a three-star rating for “value” versus a two-star rating for “room”) and infers when reviewers place higher weights on particular aspects (such as value, from which it could be concluded that a hotel truly is expensive or inexpensive). It is highly useful to elucidate reviewers’ inferred aspect weights, since they can be employed to analyze rating behaviors via business intelligence applications (Wang et al., 2010).

The LARA process thus serves to analyze review sets that have been assigned overall ratings and determine individual reviewers’ latent ratings on each topical aspect and their

relative weights in determining the overall judgment (Wang et al., 2010; 2011). This analysis method provides a significant range of application tasks. Specifically, latent aspect ratings lend themselves to immediate, aspect-based opinion summaries, while aspect weights can be directly used to analyze reviewer rating behaviors. Used together, these two forms of analysis can enable entities to then be ranked according to aspect-level ratings by employing text from reviewers with similar aspect weightings.

Wang et al. (2010) conceived of a two-step process based upon a unique regression model for latent rating. First, they used several “seed words” that described different aspects and a bootstrapping algorithm was employed to identify the words that belong to each aspect. Then, they used a generative LRR to gain the ratings of each aspect and their weights by using the customer review and overall rating. Latent regression rating specifically operates under the assumption that an overall rating is formed through the weighted sum of ratings across aspects. Wang et al. (2010) further proposed that each aspect rating is produced by the weighted combination of word features in which the weights are indicative of corresponding sentiment polarities. Because the ratings of different aspects are not observable, the aspect rating (that is, the response variable of the LRR model) is considered latent.

Plutchik’s Wheel of Emotions

There are benefits to investigating the types of emotions experienced and expressed by customers during and after purchasing, as well as the ways in which customer emotions change over time (Munero, Montero, Mozgovoy, & Sutinen, 2013). The basic, primary emotions experienced by individuals have been prominently proposed via fundamental classifications in the field of psychology by researchers such as Frijda (1986), Ekman (1992), and Plutchik (1980). Plutchik’s wheel of emotions is comprised of eight primary human

emotions in four sets of two contrasting emotions each: sadness and joy; fear and anger; trust and disgust; and anticipation and surprise. Plutchik operated under the assumption that the eight basic emotions were complete in so far as any expressed emotion can be related or subsumed by one of the eight options. He stated that the eight emotions were culturally independent (Chafale & Pimpalkar, 2014) and illustrated the connectivity of his ideas on emotion via a color wheel. Plutchik posited that, like actual colors, emotions that are primary can possess differing levels of intensity; specifically, that each emotion can be expressed via three degrees of intensity, and that primary emotions can be merged to form emotions of greater complexity. Irie, Satou, Kojima, Yamasaki, and Aizawa (2010) exemplified this concept by showing that the anger group of emotions comprises anger, rage, and annoyance.

These works have been adapted widely in subsequent research in the area of sentiment analysis (SA), a field of research with ongoing salience in text mining (e.g., Medhat et al., 2014; Nasukawa & Yi, 2003), which is concerned with analyzing the emotional attitudes of individuals. The majority of such studies has primarily conveyed research findings on the overall sentiment orientations of documents but failed to perform deeper mining of emotion information in textual data. The use of lexicon-based approaches in SA can enable researchers to discern the various levels of emotions that individuals have toward topics or events.

The NRC Emotion Lexicon

The need to develop lexical resources for use in SA has been a subject of concern in the field of computational linguistics. Wilson et al (2005) compiled a list of English-language words in categories of “positive” and “negative” and titled it *The Opinion Finder Lexicon*. Bradley and Lang (1999) developed a lexicon of affective norms for words in English know

as ANEW. Nielsen (2011) applied ANEW to analyze Twitter data and created the *AFINN lexicon*. Esuli and Sebastiani (2007) and later Baccianella, Esuli, and Sebastiani (2010) extended the well-established lexical database Wordnet via the introduction of sentiment polarities to synsets and developed SentiWordnet. Thelwall, Buckley, and Paltoglou (2012) addressed the use of lexicon resources for strength estimation and created SentiStrength. Finally, Mohammad and Turney (2013) released NRC, an emotion estimation lexicon resource in which words in English were assigned emotion ratings based on Plutchik's (1980) wheel of emotions.

The *NRC Emotion Lexicon* is comprised of a considerable set of words provided by individuals and tagged according to emotion. Mohammad and Turney (2013) employed the use of tagging with the crowdsourcing Amazon Mechanical Turk platform and developed a lexicon comprised of more than ten thousand distinct words in English that they annotated by adapting Plutchik's wheel of emotions and that can be tagged into more than one category each. The four-opposing emotion sets of sadness-anger, joy-trust, anticipation-disgust, and surprise-fear were utilized in the development of the new lexicon. In addition, words in the *NRC Emotion Lexicon* were tagged by positive-negative polarity classes, which are not considered in this work. The *NRC Emotion Lexicon* system focuses on providing sentiment analysis for any type of product review and efficiently identifying sentiments toward products in order to convey information about why a given product is the best of its type among many other products.

In summary, the sharing economy as a critical, emerging sector of the hospitality industry has unparalleled marketing potential (Zervas et al., 2014) because of the unique travel experience it offers. Thus, customer expectations for travel experiences are

continuously switching from traditional sightseeing activities to living like a local (Sundararajan, 2014). Airbnb, as a leading sharing accommodation platform that is now used worldwide, has gained enormous attention from academicians in the field of hospitality.

Previous studies have shown that the role of the customer review is an important part of UGC and eWOM within the contexts of hospitality and tourism. Both customer reviews and their corresponding ratings are a vital consideration in customers' booking behaviors (e.g., Casalo, Flavian, Guinaliu, & Ekinci, 2015; Ögüt & Onur Taş, 2012; Tsao, Hsieh, Shih, & Lin, 2015). However, very little research to date has explored the specific relationship between customer experience with Airbnb accommodations and the ratings of the accommodations in a hospitality context by mining their reviews. Given the numerous advantages of using large-scale textual data, the necessity for more highly integrated attempts to mine customer lodging experience reviews and understanding their rating behaviors has recently risen to prominence in the hospitality industry. Therefore, the LARA approach emphasizing the role of different latent aspect ratings of lodging experience in customer perception of overall rating formation is a critical focus (Wang et al., 2010).

Based on the discussion above, the primary purpose of this study is to examine the relationship between different lodging aspects and their roles in the formation of an overall rating for an Airbnb listing. To achieve this goal, this study applies ABSA to mine customer reviews and extract the lodging aspects that customers value. In addition, LARA was employed to further investigate the hidden aspect ratings and weights of different aspects and their roles in forming overall ratings.

CHAPTER 3. METHODOLOGY

The previous chapter emphasizes the role of customers experience towards sharing accommodations in the formation of their overall ratings. LARA model has been proven and improved to be a useful, compatible method to mining the relation. This chapter further explains the review collection, and analysis procedures used to identify the latent relationship between customers' reviews and their overall ratings in detail. This chapter begins with a discussion of review collection and analysis, followed by the step by step data analysis procedures.

Data Collection and Analysis

Sampling

Consumers usually express their public online sentiments on forums such as blogs, discussion boards, and product review sites, as well as their personal pages on social networks like Facebook, Instagram and Twitter. Because it is nearly impossible to manually analyze data, specialized programs such as R, Python, and Java frequently are employed in the process. Pak and Paroubek's (2010) used Tweepy, a Python library for accessing the Twitter API, in order to obtain Twitter data programmatically. Some other studies applied the automatic online website crawler by using some predefined key words or selecting criteria (e.g. Ghani, Probst, Liu, Krema, & Fano, 2006; Maedche & Staab, 2001; Thelwall, 2001).

In the present study, the automatic Python crawler procedure was used to collect Airbnb customer reviews and ratings. Data were collected from December 18–29, 2017 with the use of an automated Web crawler (see Stringam and Gerdes, 2010 for details). Airbnb.com was crawled and customer reviews were extracted for all Los Angeles listings.

Los Angeles is one of the 10 largest U.S. cities according to the most up-to-date statistics by the U.S. Census Bureau population (2007). Los Angeles is the top U.S. destination for international tourists, and travelers thus have a multiplicity of different needs when traveling there, in comparison to other tourism destination. Los Angeles is one of the main cultural centers in the U.S., and tourists therefore seek to get a taste of the inimitable local culture when staying there. Because greater Los Angeles covers a large geographic area, there are many attractions that can be far-flung from one another, allowing hosts to have unique and greater opportunities to market their offerings. All Los Angeles listings and their reviews and ratings were crawled in adherence with the website's Robot Exclusion Standard (Koster, 1994). Data was collected for 7,537 listings and 250,439 reviews. From the pages Airbnb website, Excel files were extracted that contained listing IDs, listing overall ratings, listing prices, listing locations, and reviews given for Airbnb accommodations in the city.

Data Analysis

The present study first describes the main text analytics process employed (specifically, aspect-based sentiment analysis), including its multiple steps: data pre-processing, aspect extraction (including: information extraction and categorization), aspect segmentation (key word boot-strapping), sentiment detection, and statistical association analysis. Textual data pre-processing typically first requires processes such as lowercasing; stringing; removing stop words and punctuation; and tokenizing and tagging words. The system extracted explicitly expressed Airbnb experience aspects in each review. The next step was aspect extraction. By using latent dirichlet allocation (LDA), the actual words and phrases indicating lodging experience were extracted into aspects as homogeneous aspects (typically also known as topics), with each aspect representing a unique set of aspects

included in the hotel service category. Different from the previous studies, key word bootstrapping was applied to get more sentiment related words. Then, the sentiment words that described the aspect were identified for each aspect and a sentiment value was assigned given the context of the key words. The sentiment words were classified into eight basic sentiment categories (anger, anticipation, fear, disgust, joy, sadness, surprise, and trust), and a numeric value was assigned to each emotion dimension with the assistance of the *NRC Sentiment and Emotion Lexicon*.

Since LARA assumes that the overall rating of customer experience depends on the sum of aspect ratings based on different weights, which are reflective of the relative customer preference of each aspect, the first step of LRR was to calculate the aspect ratings by the sum of the sentiment word values of each aspect using the *NRC Emotion Lexicon* dictionary with eight different sentiments (anger, anticipation, fear, disgust, joy, sadness, surprise, and trust) based on Plutchik's (1980, 1994) psycho-evolutionary theory. By utilizing the expectation-maximization (EM) algorithm to examine the maximum *a posteriori* estimation, the most probable settings of aspect weights in the reviews were generated. Maximum likelihood estimation was then used to find the most probable set of model parameters. The following section describes the data analysis procedures in detail.

Data Analysis Procedures

Several groundbreaking studies that applied text mining techniques to online reviews date to the early 2000s. These investigations used sentiment analysis on reviews posted on social media websites to determine overall opinions on particular trending topics (e.g., Jo & Oh, 2011; Lin & He, 2009; Pak & Pariybek, 2010). This study follows most of these studies by applying the traditional sentiment analysis consisted of the following main steps:

Data Pre-processing

Text preparation is the simple act of filtering extracted data prior to analysis and involves the identification and elimination of non-textual and irrelevant content. The first step in the present study was to pre-process the textual data to classify various phrases, parts of speech, and named entities. Pak and Pariybek (2010) primarily employed four functions in text preparation. First, since they “crawled” tweets online, they obtained some useless information (such as URL links), and the Repackage program was used to remove this unnecessary information. Next, they converted all of the posts into “bags” of words, by dividing sentences into separate words using the tokenization function. As mentioned in the first section, not all of the words (such as “I”, “am”, “is”, and “are”) in a sentence are useful for sentiment analysis. These words are referred to as stop words. Pak and Pariybek (2010) removed stop words to reduce the overall number of words. To enhance the accuracy of the results, the negation problem was solved by using n-grams. This technique eliminates, for example, the two bigrams such as “do not” and “like not”.

A coding schema was devised to ensure validity and reliability and to supervise the process of extracting aspects related to Airbnb customer reviews. Xiang et al (2015) noted that this schema differs from the more typical sentiment analysis that is mainly employed to identify subjective evaluations of products. Important nouns from the extant literature about guest experiences with Airbnb services, such as “reservation”, “arrival”, “on-site experience”, and “departure” (Kotler, Bowen, Makens, & Baloglu, 2006) were added into the code schema.

Words reflecting guest experience aspects were retained in the dictionary, except for: (a) stop words like “about”, “can”, “does”, and “a/an”; (b) abbreviation like “LA”, “CA”,

“I’m”, and “aren’t”; (c) highly ambiguous words such as “go”, “do”; e) words related to Airbnb location such as “Los Angeles”, “California”, and “United States”.

Information Extraction

Aspect extraction can be described as pulling conceptually relevant linguistic entities from the corpus (Hearst, 2003), and is also known as information extraction. In other words, information extraction involves the automatic extraction of specific, structured data from unstructured/semi-structured text-based natural language (Soderland, 1999).

In Pak and Pariybek’s (2010) study, they did not mention in their study about the techniques that employed in terms of extraction features stage. However, feature extraction is an important aspect of traditional sentiment analysis (Pang & Lee, 2008; Pannala, et al., 2016; Wilson, Wiebe, & Hoffmann, 2005). As far as feature extraction is concerned, techniques such as unigrams, bigrams, part of speech tags, and combinations of these features are commonly used by linguistics and computer science scholars (e.g., Hollander et al., 2016; Türkmen, Ekinici, & Omurca, 2016).

Information extraction most often involves extracting frequent nouns and noun phrases (Liu, 2010) and looking for explicit expressions about aspects in the form of nouns/noun phrases from a large set of reviews in a particular domain. Hu and Liu (2004) employed a data mining algorithm to identify nouns/noun phrases using a part-of-speech tagger. The resulting occurrence frequencies were then tabulated and only the most frequent nouns/noun phrases were retained. Experimentation was conducted to determine the frequency threshold. However, this approach was found to be less effective due to that fact that, when consumers comment using a variety of verbs to state the truth or their experience of a given entity. These phases are typically convergence in their vocabulary and the reviews.

As a result, beside the frequently employed nouns/noun phrases, verbs can also tend to be genuine and salient information that hidden in customers' reviews.

Topic modeling is a valid method to extract meaningful information from a large amount of textual data. This study applied the most common topic modeling method LDA to extract valuable information, which used as the seed words in the later boot-strapping procedure.

Information categorization

Aspect categorization can be defined as a process in which a given set of words is assigned into groups of entities with members in the same category. As in any review or comment, people often evaluate a service from different aspects (Liu, 2012). For example, security/safety, guestroom cleanliness, and check-in speed are all aspects considered when evaluating the perception of hotel performance (Oh, 1999). Recognizing the cross-entity resemblance by aggregating similar entities into aspects aids in the identification of an order in the multifaceted environment, without which the individual experience of an entity would be entirely unique and non-extendable to ensuing encounters with related entities. The process is a supervised classification technique of with a set of pre-classified documents as the training set, and a supervised classification (Liu, 2010).

Topic modeling methods actually perform both information expression discovery and categorization at the same time in an unsupervised manner, because topic modeling is used for classification in a document collection (Andrzejewski, Zhu, & Craven, 2009). Since the present study is innovative in applying LDA for information extraction, the algorithm will automatically classify the information into different aspects.

Aspect Segmentation

Since the previous step only provides limited number of words in each aspect, a bootstrapping algorithm designed by Wang et al. (2010) was applied to provide a greater number of words related to each aspect. By using the aspect seed words that they determined manually and customer review contents, Wang et al. (2010) assigned every phrase that has the most term overlapping into aspect using a bootstrapping algorithm. The process for the aspect segmentation algorithm follows:

Established on their proposed annotation of aspect, the dependencies between each aspect and word were calculated using the Chi-Square (χ^2) statistic proposed by Wang et al. (2010), and the phrase which as high dependencies were included in the corresponding aspect keyword documents obtained from the previous step. This calculation was repeated unless the key word list of each aspect exceeded the iteration limitation or was unchanged.

In study, both verbs and nouns were generated from the information extraction stage by using LDA, and these words were used as the seed words in to the aspect segmentation to get the keywords. Thus, the total key words after this stage include nouns, verbs, adjective etc.

Sentiment Detection

There are two broad classifications of sentiments: positive and negative. At this stage, each subjective sentence was classified into the following groups: positive, negative, good, bad, like, and dislike. In the supervised learning approach, the process of algorithm learning was applied in both the training dataset and the testing set. Pak and Pariybek (2010) employed this approach in their aforementioned Twitter study. Since no technique existed to

aid scholars in selecting and dividing posts into different sentiment categories, Pak and Pariybek (2010) mutually annotated the training set of 216 posts into positive, negative, and neutral categories. Furthermore, they built Naive Bayes classifiers into the classification of their training set. The trained classifier was used to test the model by predicting the target class of unseen test data to assess the model accuracy. The formula for accuracy was defined “as the number of correct classifications divided by the total number” of test cases (Passmore et al., 2003, p. 9).

Sentiments/opinions and their targets/aspect relationships allow sentiment words to be determined via their identified aspects; aspects can be determined via known sentiment words. In order to do so, sentiment words and aspects are propagated, resulting in the term “double propagation” for this process. Specific dependency relationships between sentiment words and aspects are used to develop extraction rules. The results of a study by Tesniere (1965) revealed that adjectives could be considered sentiment words.

A rule-based method was employed for the extraction process. For example, “A noun on which a sentiment word directly depends through mod is taken as an aspect” (Liu, 2010, p. 80) was one of the rules employed. Likewise, in “The host offered a clean and tidy apartment for use,” the adjectives “clean and tidy” were parsed as depending on the noun “apartment”. Since “cleanliness” is an aspect, “clean and tidy” were extracted as the sentiment words.

The *NRC Sentiment and Emotion Lexicons*, which is an adjective dictionary, is used as the sentiment polarity in the later aspect rating prediction calculation. Since from the previous stage, the key words are mixed nouns, verbs, adjective, etc., the nearest adjective’s sentiment value of the key word is counted into further calculation. If the key word is

adjective, it will directly use as the sentiment polarity from the *NRC Sentiment and Emotion Lexicons*.

Latent Rating Regression

The latent rating regression (LRR) model was designed to formally capture the above-described generation process. After aspect segmentation, a word frequency matrix W_d was generated for each listing d , which provided the normalized frequency of the words belong to each aspect. In this model, W_d was treated as independent variables (for example, the features of the listing d), while the listing rating was treated as the dependent variable (which is predicted the variable). The LRR assumed that overall listing ratings were not directly determined by features of word frequency. Instead, the model used the latency aspect ratings that were more pointedly determined by aspect frequency with combination of their corresponding weights.

The work of Wang et al. (2010) was adapted to show that the review-level-dimensional aspects weight vector s_d was a linear sum of W_{di} and β . $\beta \in \square$ indicated the polarities of sentiment on aspect A_i obtained from the sentiment detection. The weighted sum of aspect rating s_d and aspect weight α_d determined the overall rating. Specifically, it was assumed that the overall rating was a sample obtained from a Gaussian distribution indicating that the overall rating predictions were uncertain.

Wang et al. (2010) further discovered that reviewer emphasis on various aspects can be a complex issue due to the various factors involved. For instance, reviewers may show different preferences for the aspects at hand (that is, business travelers may place an emphasis on Internet service, while honeymooning couples may be more concerned with the listings). Furthermore, aspects may not be independent, particularly when certain aspects

overlap (that is, reviewers interested primarily in cleanliness most likely are more interested in the listings themselves as well). Wang et al. (2010) accommodated reviewer preference diversity by treating the aspect weight α_d of each listing d as a free variable obtained from an underlying prior distribution for the listings as a whole. Multivariate Gaussian distribution was also used as the prior distribution for aspect weights to capture the different dependencies among each aspect.

In reviews using the LRR model in the study, the observed overall rating probability was given as such: where r_d and W_d were the observed in previous analysis listing d , μ , Σ , δ^2 , β was the listing level parameters, and α_d was the latent aspect weight of listing d ; μ , Σ and δ^2 was not dependent upon individual reviewers and were deemed as aspect-level parameters. The LRR model is graphically represented in Figure 3.

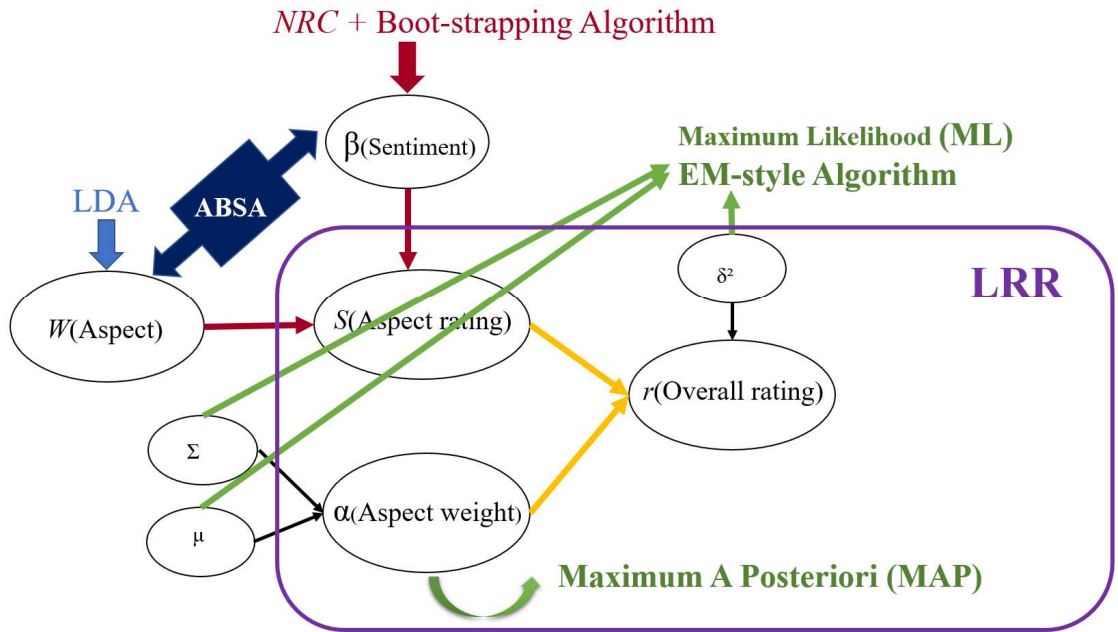


Figure 3.1. LRR graphical representation

Next, the estimation method of maximum a posteriori (MAP) was applied to gain the value of α_d of a listing with the maximum probability. The MAP estimation of listing d was defined as:

$$\mathcal{L}(d) = \log p(\alpha_d | \mu, \Sigma) p(r_d | \sum_{i=1}^{\kappa} \alpha_{di} \sum_{j=1}^n \beta_{dij} W_{dij}, \delta^2)$$

Wang et al. (2010) expanded this method by associating all parameters with regard to α_d in each listing [denoted as $\square(\alpha_d)$] below:

$$\begin{aligned} \hat{\alpha}_d &= \arg \max \mathcal{L}(\alpha_d) \\ &= \arg \max \left[-\frac{(\gamma - \alpha_d^T S_d)^2}{2\delta^2} - \frac{1}{2} (\alpha_d - \mu)^T \Sigma^{-1} (\alpha_d - \mu) \right] \end{aligned}$$

To address the above problem of constraint non-linear optimization, the conjugate-gradient-interior-point method was used with a derivatives formula adapted from Wang et al. (2010) with respect to α_d in the present study:

$$\frac{\partial \mathcal{L}(\alpha_d)}{\partial \alpha_d} = -\frac{(\alpha_d^T S_d - r_d) S_d}{\delta^2} - \Sigma^{-1} (\alpha_d - \mu)$$

LRR Model Estimation

This section contains a discussion of how the model parameters were estimated in the present study applying the maximum likelihood (ML) algorithm. In other words, the ML estimator was employed to obtain the optimal $\Theta = (\mu, \Sigma, \delta^2)$ to maximize the likelihood of overall ratings. Adapted from Wang et al. (2010), the log-likelihood estimator was applied to all the customer reviews:

$$\mathcal{L}(D) = \sum_{d \in D} \log p(r_d | \mu, \Sigma, \delta^2, W_d)$$

Thus, the ML estimate was:

$$\hat{\Theta} = \arg \max_{\Theta} \sum_{d \in D} \log p(r_d | \mu, \Sigma, \delta^2, W_d)$$

For ML estimation, all of the parameter values were first randomly initialized to find the probable $\Theta(0)$. The EM-style algorithm was used to update and increase the parameters iteratively by alternately performing the E-step and the M-step in each iteration as follows:

1. E-Step: For each listing d , the author inferred the aspect rating s_d and aspect weight α_d with $\Theta(t)$ current parameter, which t represents the iteration) based on the discussion above.
2. M-Step: Based on aspect rating s_d and aspect weight α_d obtained from the existing parameters $\Theta(t)$, the updated parameters adapted from Wang et al. (2010) were employed and $\Theta(t+1)$ was obtained through the maximization value of the “complete likelihood” including overall ratings r_d , the aspect ratings s_d , and the aspect weights α_d of listing d . Here, the goal was the maximization of probability of observing all α_d gained the current step. Thus, Wang et al. (2010) updated and developed Gaussian distribution formula based on the ML estimation:

$$\Sigma_{(t+1)} = \frac{1}{|D|} \sum_{d \in D} (\alpha_d - \mu_{(t+1)})(\alpha_d - \mu_{(t+1)})^T$$

Second, a method was determined to update δ^2 . Since α_d was assumed to be known, the update δ^2 to could be generated maximized. To solve this optimization problem, Wang et al. (2010) updated formulae were adapted as following:

$$\begin{aligned}
\delta_{(t+1)}^2 &= \underset{\delta^2}{\arg \max} \left[-|D| \log \delta^2 - \frac{\sum_{d \in D} (r_d - \alpha_d^T S_d)^2}{\delta^2} \right] \\
&= \frac{1}{|D|} \sum_{d \in D} - \frac{(r_d - \sum_{i=1}^k \alpha_{di} W_{di})^2}{2\delta_{(t+1)}^2}
\end{aligned}$$

CHAPTER 4. ANALYSIS AND RESULTS

The purpose of this chapter is to present the results from this study. First, a description of review characteristics is given, including sample descriptive analysis, general sentiment analysis, and word relation analysis. Second, the results of the aspect-based sentiment analysis for the entire sample are presented, followed by description statistics results for each extracted aspect of lodging experience, a summary of the key words of each aspect, and their corresponding sentiment polarities with eight emotional levels. The final section includes the results of the latent rating regression analysis on the data and the perceived applications samples.

Sample Descriptive Analysis

The data for the present study was obtained via an online robot constructed specifically for the research. A total of 250,439 reviews were compiled in an Excel file containing categories presenting the typical attributes of Airbnb consumer reviews (see Table 4.1). Listings that had less than three reviews were deleted because they lacked an overall rating. A total of 248,693 reviews from 6,946 Airbnb listings were retained for further analysis. A summary of all 6,946 listings that were included in this study (with corresponding star ratings) and the number of reviews is presented in Table 4.2 below. Nearly 60% of the Airbnb listings had an overall rating of 5 stars, followed by 4.5 and 4 stars. Very few accommodations had overall ratings that were lower than 4.

Table 4.1. Review information extracted from the Airbnb website

Listing ID	Star rating	Geo Longitude	location Latitude	Price	Comments
13781312	5	118.2704814	33.87541428	125	I finally tried Antico! Absolutely loved it!

As shown in Table 4.2, the most prevalent ratings of the Airbnb listings in Los Angeles were 5 stars (57.745%), 4.5 stars (34.797%), and 4 stars (6.335%). Listings with lower than 4- star ratings comprised only approximately 2% of the total accommodations in this area.

Table 4.2. Summary of Airbnb listings' overall ratings

Overall rating	Total numbers of listing ID (N = 6,946)	%
1.5	1	.014%
2	1	.014%
2.5	5	.072%
3	10	.144%
3.5	61	.878%
4	440	6.335%
4.5	2417	34.797%
5	4011	57.745%

A summary of the review descriptions is shown in Table 4.3. Nearly three in five reviews (58.616%, n = 145,774) rated listings at 5 stars, while 35.98% (n = 89,325) of the reviews rated listings at 4.5 stars, 5.107% (n = 12,701) rated listings at four stars, and less than 1% rated listings at a 3.5 star or lower rating.

Table 4.3. Descriptive summary of Airbnb reviews with different overall ratings

Overall rating	Total numbers of reviews (N = 248,693)	%
1.5	3	.001%
2	3	.001%
2.5	25	.010%
3	52	.021%
3.5	810	.326%
4	12,701	5.107%
4.5	89,325	35.918%
5	145,774	58.616%

The length of a review directly influences customers' sentiments and opinions. Thus, the reviews with the most and least words were extracted, as shown in Table 4.4. The longest review had 680 words and gave an overall rating of 4.5 stars, while the shortest review had only seven words and gave an overall rating of 5 stars.

Table 4.4. Maximum and minimum review length

Listing ID	Overall rating	Length
15631788	5	7
8992330	4.5	680

The average review length (determined by the quantity of tokens before data pre-processing) of each star rating level is summarized in Table 4.5. The table demonstrates that Airbnb listings with 2- or 2.5-star ratings had much longer reviews than listings with other

ratings. For listings with higher ratings (3 stars and higher), the length of the reviews ranged from 67 words to 75 words.

Table 4.5. Summary of average review length by star rating

Overall rating	Avg. length
1.5	69
2	216
2.5	103
3	67
3.5	75
4	74
4.5	67
5	73

Simple preprocessing was then performed on the review data to: (a) convert words into lower case; (b) remove punctuation; (c) remove stop words; and 4) stem each word. Figure 4.2 shows the word cloud for terms extracted from the 248,693 reviews, providing a visual interpretation of the results. The cloud is drawn on the frequency of terms occurring in the textual contents of the reviews; a term occurring more frequently will be depicted in a larger font. Aside from the stop words, all other terms were adopted. The cloud shows a total of 50 terms and indicates that the words “stay” and “place” are clearly the main lodging attributes mentioned by customers, followed by “location” and “clean”. Reviewers also frequently showed an interest in the host, the room, and comfort.

A GIS plot showing the location of the 6,946 listings was generated to show the distribution of the Airbnb accommodations in the Los Angeles area. In Figure 4.1, the listings have the greatest density in the northern part of the city. There is a moderate density of Airbnb accommodations in the area between Santa Monica and the Los Angeles airport.

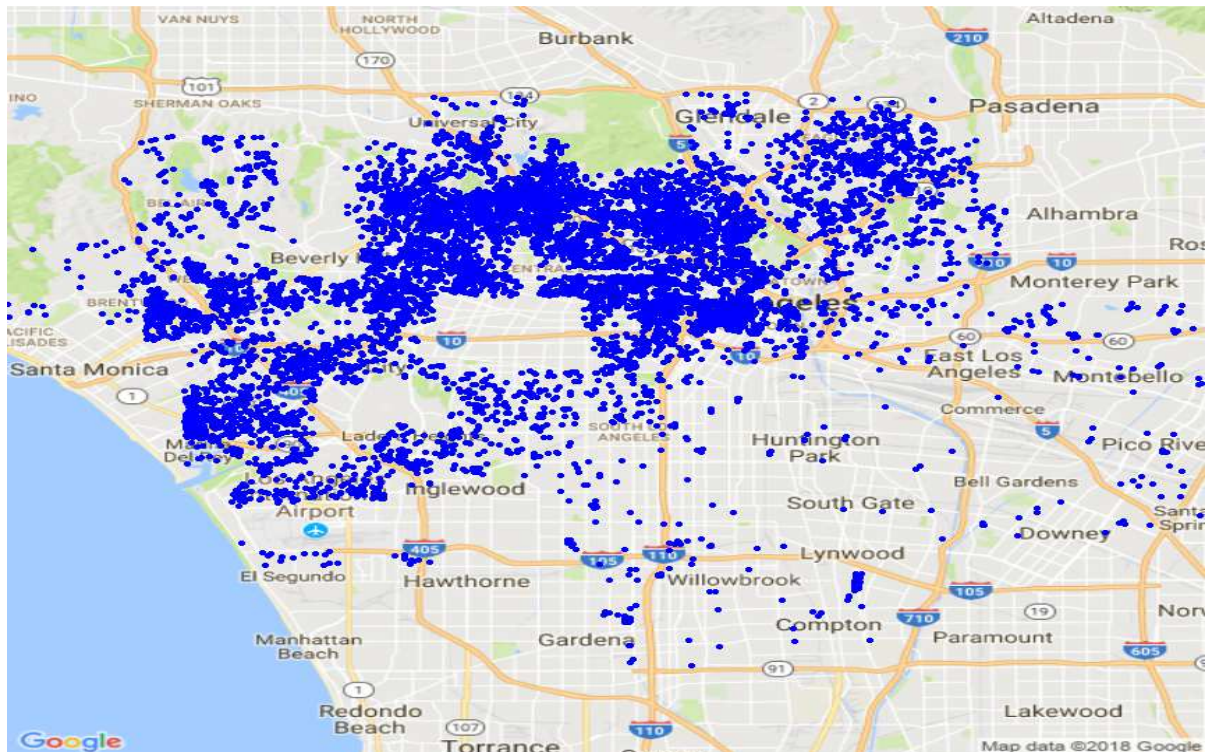


Figure 4.2. Airbnb listings locations summary

A general review sentiment based upon the bag of words was employed to obtain a general sense of the review sentiments by classifying all the reviews into positive, negative, and neutral. Figure 4.3 shows the distribution of reviews with their corresponding sentiments. Almost 95% of the reviews were positive, with the majority of listings rated at 4 or more stars. However, given the limited sentiment results, more in-depth sentiment analysis was needed.

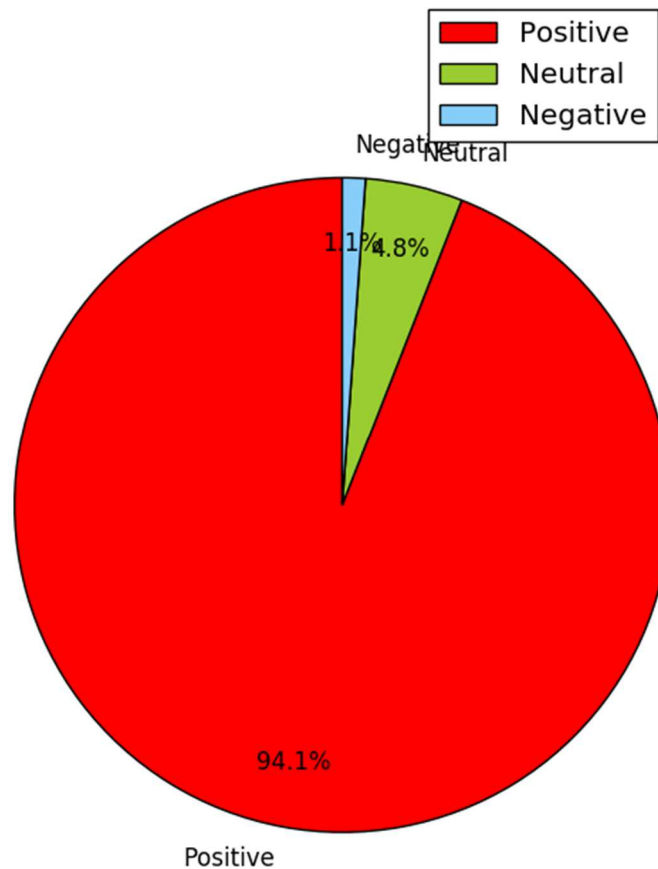


Figure 4.3. General review sentiment analysis

Graphic visualizations of possible relationships between individuals, concepts, organizations, and other entities in textual material are often shown in co-occurrence

networks. This method is a useful method with the uptick in electronic text that is suitable for text mining. Co-occurrence networks can be defined as a “collective interconnection of terms founded in their paired presence in a specific unit of text” (Deokar, Gupta, Iyer, & Jones, 2017, p. 274), with networks generated through the connection of term pairs via criteria that defines co-occurrence.

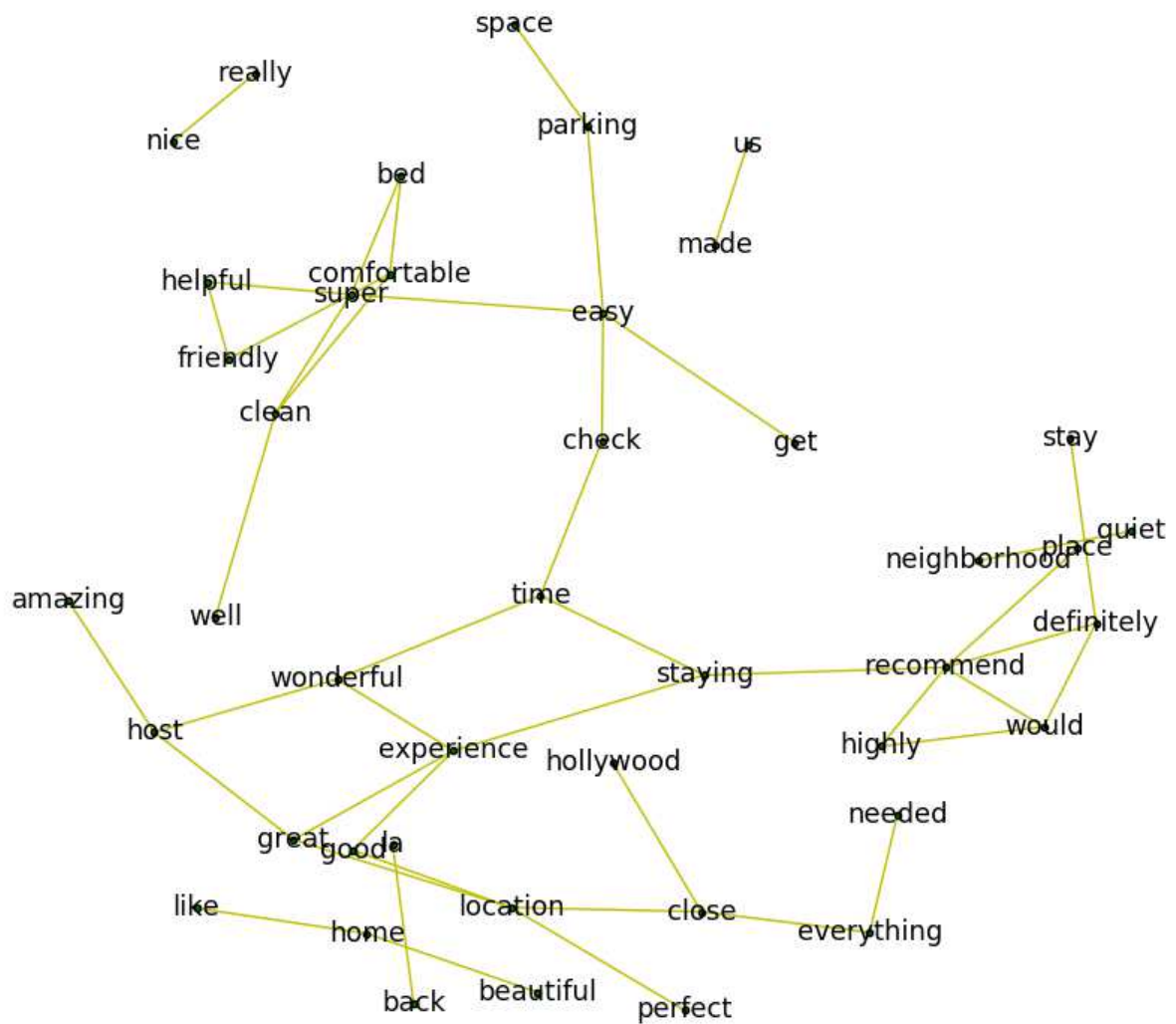


Figure 4.4. Co-occurrence networks of reviews

Figure 4.4 shows that there were some pairs of words with relatively high co-occurrence, such as “help” and “friendly”; “staying” and “recommend”; and so on. This

result that it is not sufficient to identify whether a word belongs to an aspect or not in a large group of textual data. Specifically, some words possess strong hidden relationships with a topic/aspect and thus should be counted as key words when weighting each topic/aspect.

Aspect-based Sentiment Analysis

Table 4.6 shows the major aspects that the LDA aspect modeling method identified from the aggregated review corpus. The perplexity scores assessed the topic modeling goodness-of-fit with five lodging experience aspects extracted, with ten meaningful key words for each aspect (as shown in Table 4.6).

Table 4.6. Summary of LDA seed words of each aspect

Aspect	Seed words obtained from LDA
1	host, communicate, talk, idea, guide, friends, problem, neighbor, recommendations, suggestions
2	stay, experience, neighbor, service, view, hospitality, future, overall, environment, care
3	location, downtown, distance, place, area, Hollywood, west, station, convenience, walking
4	bathroom, bedroom, kitchen, room, space, parking, towel, internet, breakfast, pool
5	money, price, worth, value, quality, truth, recommend, reason, accepted, charge

After the seed words were extracted from the LDA, the five lodging aspects were named as “communication”, “experience”, “location”, “product/service”, and “value”. Specifically, communication indicates the interactions between the host and the guest throughout the pre-purchasing process, the actual stay, and the post-purchase stages. Since the definition of experience is customers’ internal responses to any direct/indirect contact with a firm along a multiplicity of touchpoints (Meyer & Schwager, 2007), the experience

aspect describes the information search, purchase, consumption, and after-sale processes involved in guest experience, often over more than one retail channel (Verhoef et al, 2009). The aspect of location mainly describes the geographical convenience of Airbnb accommodations. Product/service refers to tangible products (e.g., room facilities, kitchen appliances) and intangible services (e.g., meeting customer special needs, offering greetings, or giving directions for using room facilities). The value aspect refers to the objective worth of the economic outcome, and the payoff between cost paid and benefit received (Curhan, Elfenbein, & Eisenkraft, 2010).

After the lodging aspects and the related seed words were identified, the aspect segmentation algorithm was employed to boot-strap additional words that related to the five lodging aspects. Specifically, the 50 seed words were employed as the words that described each aspect. Through aspect segmentation, more key words that were related to the seed words were obtained. After the segmentation, 101,457 total key words for the five lodging aspects were boot-strapped from the reviews. Table 4.7 lists some of the boot-strapped words for each lodging aspect (due to space limitations, only 30 words for each aspect are listed in the table).

Table 4.7. Summary of boot-strapping key words for each aspect

Aspect	Boot-strapping key words
Communication	guidance, guy, guests, questions, people, guest, friendship, opinion, quick, interaction, helpful, suggestions, messages, neighbors, advice, fun, contact, response, email, problems, phone, details, help, text, trouble, attention, meet, complaints, info, concern
Experience	pleasure, nice, thoughtful, kind, check, anything, time, trip, plenty, welcoming, fun, cozy, safe, stayed, perfect, amazing, fantastic, enjoyed, day, feel, kids, quality, cool, style, travel, joy, hope
Location	lake, noise, downtown, convenient, walk, street, spot, sights, city,

Table 4.7. (continued)

Aspect	Boot-strapping key words
	highway, beach, building, blocks, center, areas, shopping, metrolink, Universal, market, drive, bus, store, corner, road, minutes, Beverly, uptown, restaurant, easy, find
Product/Service	bed, coffee, view, pictures, tv, privacy, book, shower, instructions, wifi, Uber, garage, snacks, loft, fridge, arrival, check-in, dryer, drinks, fruit, bikes luggage, facilities, tidy, balcony, sheets, shampoo, laundry
Value	promote, reservation, described, chance, hotel, couple, sale, living, comfy, times, honeymoon, good, super, family, expectation, absolutely, right, waste, economy, excellent, reasonable, better, posting, coupon, accommodations, choice, deal, option, save, rate

The *NRC Emotion lexicon* with eight basic emotion levels (anger, anticipation, fear, disgust, sadness, surprise, joy, and trust) was adopted to assign values to calculate the sentiment polarities of the boot-strapped words for each lodging aspect. Among boot-strapping words, there were nouns, verbs, adjectives, and adverbs. Since the NRC only provides sentiment levels for adjectives, the adjectives near the verbs, nouns, and adverbs of the key words were extracted and accounted for in the calculation of the sentiment polarities using the *NRC Emotion lexicon*. If the boot-strapped words were adjectives, they were directly counted into the sentiment polarities. The top 1,000 words that related to each aspect were selected. Based on the sentiment polarities of the words, a general lodging aspect emotional classification was made, as shown in Figure 4.5. This figure lists the aspects with the emotions most expressed by customers, obtained from the ABSA. For example, it was found that customers often experienced the mixed sentiment of joy and surprise (nearly 50/50) with the communication aspect, whereas the experience aspect was highly associated with the emotion of surprise. A closer analysis showed that the location aspect was mostly

related to the emotion of joy, while the aspects of product/service and value generally evoked the feelings of both joy and surprise. Although both of these two aspects shared similar sentiments, the emotion of joy was nonetheless the dominant one identified by the ABSA.

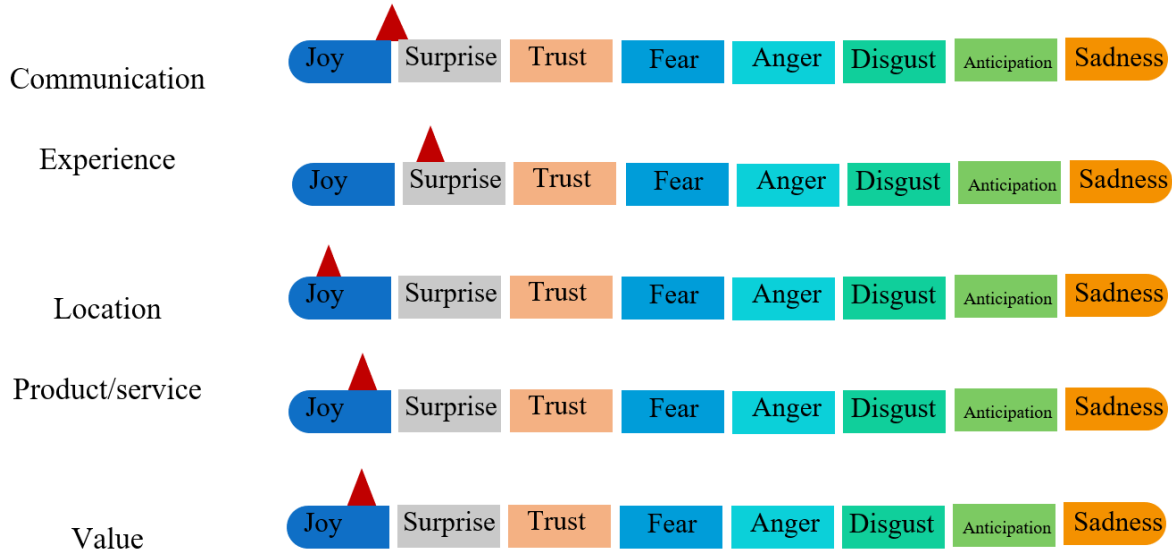


Figure 4.5. Emotional level of the lodging aspects

Next, as part of an in-depth, aspect-based sentiment analysis of the Airbnb review data, some key words with different emotions for each lodging aspect were compiled, as shown in Table 4.8. For example, for the aspect of communication, sometimes anger about potential fees and a sense of disconnection between the host and the guest was expressed in the reviews and was therefore extracted. Furthermore, customers often expressed disgust if their hosts abandoned them after arrival or payment, as well as when guests had to cancel their reservations. It was also noted that customers typically expressed sadness in their reviews when they felt that their complaints weren't resolved by hosts, and that customers reviewing product/service often expressed anger when their listing facilities were inconvenient. Finally, insufficient cleanliness of listings was frequently acknowledged by customers in their reviews with the expression of fear.

Table 4.8. Summary of key word basic emotion levels for each aspect

	Communication	Experience	Location	Product/service	Value
Anger	Fee	Behavior	Guidebook	Response	Description/Photo
Fear	Disconnect	Theft	Far away	Cleanliness	Risk
Anticipation	Neighbor	Explore	Truth	Arrival	Worth/Reward
Trust	Recommend/Suggestions	Help	Convenient	Safe/Food/Profession	Price
Surprise	Child/Difficulty/Urgent	Event Explanation	Restaurant/Shop	Passion/Quick	Discount
Sadness	Complain	Helpless	Rob/Black	Toilet/Bed	Expected
Joy	Entertainment/Explain Enthusiasm	Holiday	Walking-distance	Be Friendly/Smile Beauty	Budget
Disgust	Abandon/Cancel	Dishonesty	Crowded/Noisy	Internet	Advertised

Latent Rating Regression

Since the lodging aspects extracted in this study differed from the aspects provided by the Airbnb website, the overall ratings ranging from 1 to 5 stars served as ground-truth for quantitative evaluations of the latent rating regression. To test the capability of the proposed LARA model, 20 listings with different overall ratings were randomly selected to show the usefulness of this model to predict the overall ratings with given aspects. Among the randomly selected listings with different ratings, five listings for each rating level (i.e., 3.5, 4, 4.5, and 5 stars) were listed in Table 4.9, which provided reliable qualitative evaluation.

Table 4.9. Summary of predicted overall ratings

Listing ID	Real overall rating	Predict overall rating
395413	3.5	3.95672
1124252	3.5	3.389956
4182830	3.5	3.648891
2784245	3.5	3.799363
15395910	3.5	3.748211
8704701	4	4.072723
7845614	4	3.98114
13418551	4	4.013817
1280955	4	3.928527
9774993	4	3.981038
15510747	4.5	4.103087
7623280	4.5	4.078884
13592173	4.5	4.248171

Table 4.9. (continued)

Listing ID	Real overall rating	Predict overall rating
15524053	4.5	4.216781
2256926	4.5	4.338944
15381099	5	4.659496
3505734	5	4.890919
13754009	5	4.785157
751086	5	4.886108
13466172	5	4.850683

For the predicted aspect ratings, 20 listings were randomly selected to present the usefulness of the LRR estimation model, as show in Table 4.10. In this table, while some listings have the same overall ratings, the aspect rating may differ based on the distinct preferences of customers. However, generally it was found that customers preferred to rate experience higher than other lodging aspects, because of the unique experiences that customers obtain from their stays with Airbnb.

Table 4.10. Summary of predicted aspect ratings

Listing ID	Real Overall Rating	Predict Overall Rating	Communication	Experience	Location	Product /service	Value
8925563	4.0	4.215	3.011	4.10	3.000	3.287	2.630
8205212	5.0	4.931	4.000	2.985	4.589	3.495	3.000
15097514	4.5	4.554	3.066	4.252	3.250	3.277	3.348
11660575	4.5	4.434	3.176	4.440	2.920	3.143	3.355
8110495	4.5	4.523	3.10	4.089	3.266	3.281	3.304

Table 4.10. (continued)

Listing ID	Real Overall Rating	Predict Overall Rating	Communication	Experience	Location	Product /service	Value
5061104	4.5	4.499	2.975	4.500	3.295	3.358	2.747
7517023	4.5	4.399	2.911	4.299	3.167	3.252	2.929
1822343	4.5	4.364	3.00	3.041	3.488	3.560	3.0
14203792	4.5	4.451	3.364	4.247	3.212	3.010	2.938
15633296	4.5	4.423	3.059	4.054	3.236	3.280	2.983
8595070	4.5	4.432	3.0	3.0	4.065	4.000	3.0
14774856	5.0	4.571	2.844	4.168	3.246	3.295	3.642
6063911	4.5	4.461	2.864	4.500	3.043	3.378	3.214
1300859	4.5	4.309	3.0	4.439	3.20	2.755	3.047
13814087	4.5	4.387	3.200	4.131	3.220	3.00	2.907
8571568	4.5	4.381	2.934	4.333	3.092	3.166	3.031
14377540	5.0	4.524	3.160	4.440	3.198	3.685	2.700
14419183	4.5	4.542	2.766	4.995	3.331	3.363	2.547
5613279	4.5	4.326	3.0	3.0	3.0	4.821	2.970
898068	4.5	4.304	3.0	3.0	3.125	4.389	2.938

A facile method for judging listing quality is to assess the overall rating; however, this does not offer detailed assessments regarding the aspects' quality and fails to show the differences between the accommodations in terms of aspect level. As mentioned above, for listings with the same overall rating, reviewers can hold unique, distinct opinions on the various aspects. The LRR model can provide detailed information by estimating aspect ratings based on the different aspect weights. In Table 4.11, it is shown that two listings both

have 4.5 overall ratings, but their aspect ratings are different. For the first listing, customers evaluated “experience” more positively than other aspects, while customers felt that the best qualities of the second listing were “product/service” and “location”. Identifying such disparity and providing aspect ratings can aid consumers in making more informed, review-based decisions.

Table 4.11 Listings with different aspect ratings

Listing ID	Real Overall Rating	Predict Overall Rating	Communication	Experience	Location	Product /service	Value
15633296	4.5	4.423	3.059	4.054	3.236	3.280	2.983
8595070	4.5	4.432	3.0	3.0	4.065	4.000	3.0

For Table 4.12, 20 listings were randomly selected to report the predicted aspect weights. As shown in the table, customers valued location more than the other four lodging aspects. One of the main motivations of Airbnb customers is to experience how the locals live in a given destination. If a listing is located in a traditional neighborhood for its area that represents the local culture, customers will value this aspect more than the other aspects.

Table 4.12. Summary of predicted aspect weights

Listing ID	Real Overall Rating	Predict Overall Rating	Communication	Experience	Location	Product /service	Value
31939	4	4.24	.143	.164	.143	.389	.162
15221415	4.5	4.57	.141	.167	.141	.384	.162
7861299	4.5	4.252	.142	.146	.260	.176	.277
6749145	4.5	4.22	.140	.143	.280	.164	.273

Table 4.12. (continued)

Listing ID	Real Overall Rating	Predict Overall Rating	Communication	Experience	Location	Product /service	Value
15217021	4.5	4.893	.149	.149	.405	.149	.149
15400738	5	4.961	.149	.150	.404	.149	.149
15273760	4	4.448	.149	.151	.404	.149	.149
8000680	4.5	4.567	.147	.161	.399	.147	.147
14731931	4.5	4.672	.147	.161	.399	.147	.147
14608919	4.5	4.688	.146	.162	.398	.146	.147
13031874	3.5	3.823	.146	.162	.396	.146	.148
13941545	4	4.378	.146	.162	.397	.146	.149
11601321	4	4.403	.146	.161	.397	.146	.150
987760	4.5	4.75	.146	.160	.397	.146	.151
13021970	4	4.413	.146	.163	.397	.147	.148
247027	4.5	4.754	.155	.161	.365	.146	.174
13193989	3.5	3.759	.146	.161	.396	.146	.151
13100468	4.5	4.797	.146	.163	.396	.146	.150
8205212	5	4.93	.146	.164	.396	.146	.150
5480885	3.5	3.861	.145	.162	.395	.146	.151

In order to further identify the different weights of each lodging aspect, Table 4.13 was created to display the average predicted weight of each aspect. The results show that Airbnb customers value location most, followed by experience, value, communication, and product/service.

Table 4.13. Summary of average aspect weights.

	Communication	Experience	Location	Product/service	Value
Overall average	.138	.183	.370	.137	.172

By inferring the latent aspect weights for individual reviews, the reviewer's relative emphasis on an aspect can be discerned; this can be interpreted as evidence of how the reviewer arrives at a rating. Understanding reviewing behavior can be employed to learn the most influential factors on a consumer's evaluation judgment. For instance, listing prices over \$500 would likely be described as expensive, while listing prices below \$100 would be called cheap. As shown in Figure 4.6, most of the listings with high prices (\$500 and above) are located in West Hollywood and Beverly Hills, while the cheap listings are located in the downtown and airport areas. It was found that Airbnb customers of the expensive listings had higher requirements for the aspects of location, product/service, and lodging (as shown in Table 4.14), in contrast to customers of the cheaper listings. Specifically, high-paying customers sought high service levels, including kitchen facilities in the house, house design, and equipment (e.g., swimming pools, family cinemas, and other forms of home entertainment). These customers also had higher requirements for location, and most of the expensive accommodations provided private views of the hills or the city or were specifically close to locales for events (such as the LA Convention Center, the Staples Center, and LA Live).

Table 4.14. Aspect weights comparison between expensive and cheap Airbnb listings

	Communication	Experience	Location	Product/service	Value
Cheap	.144	.195	.333	.131	.196
Expensive	.139	.178	.377	.139	.167

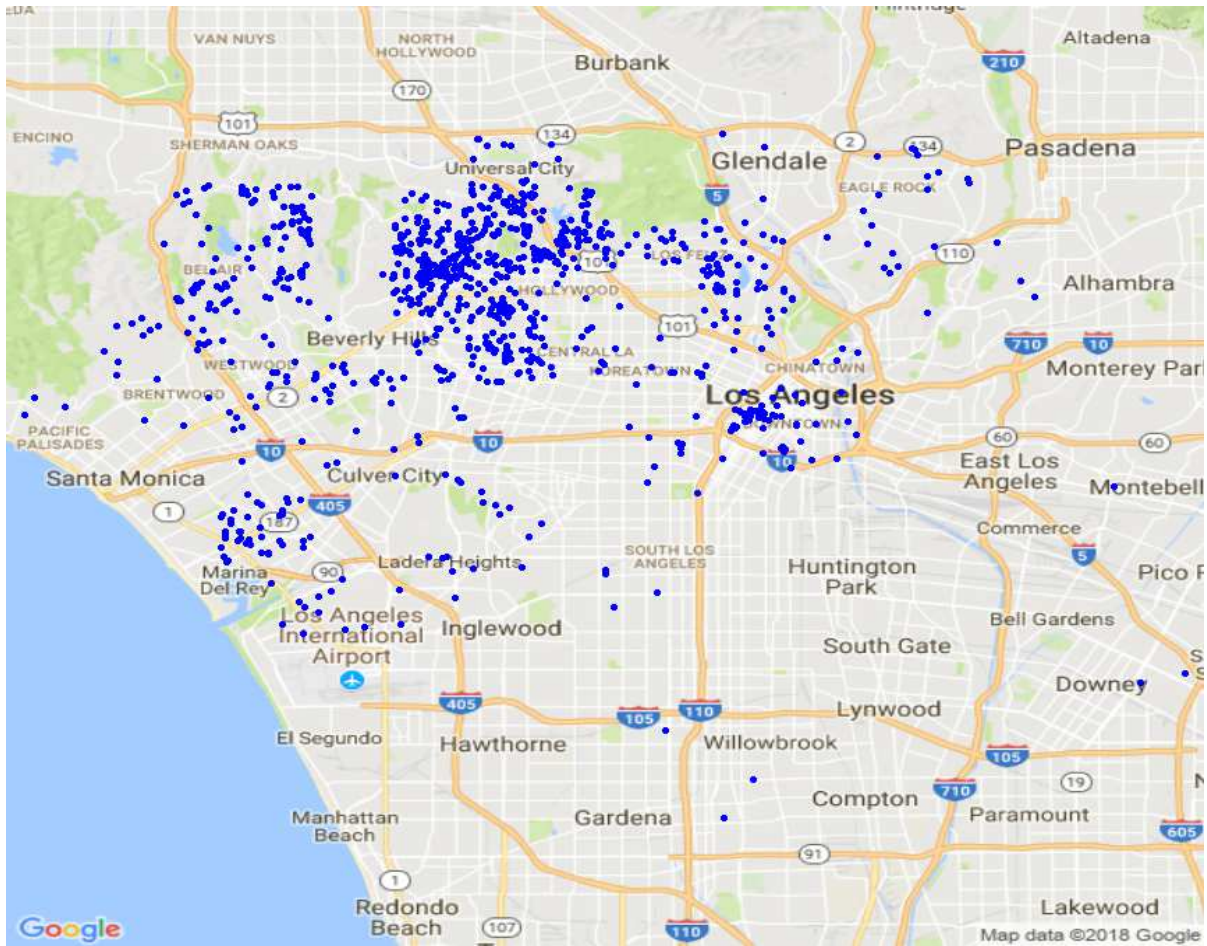


Figure 4.6. Airbnb cheap listings locations summary

Figure 4.7 shows the locations of cheap Airbnb listings in Los Angeles. Besides the location aspect, it was found that the customers of these rentals cared most about value, followed by experience and communication. Overall, most reviewers chose their Airbnb accommodations based on price, but also considered experience and communication as important aspects, due to the “living like a local” feeling that they offer.

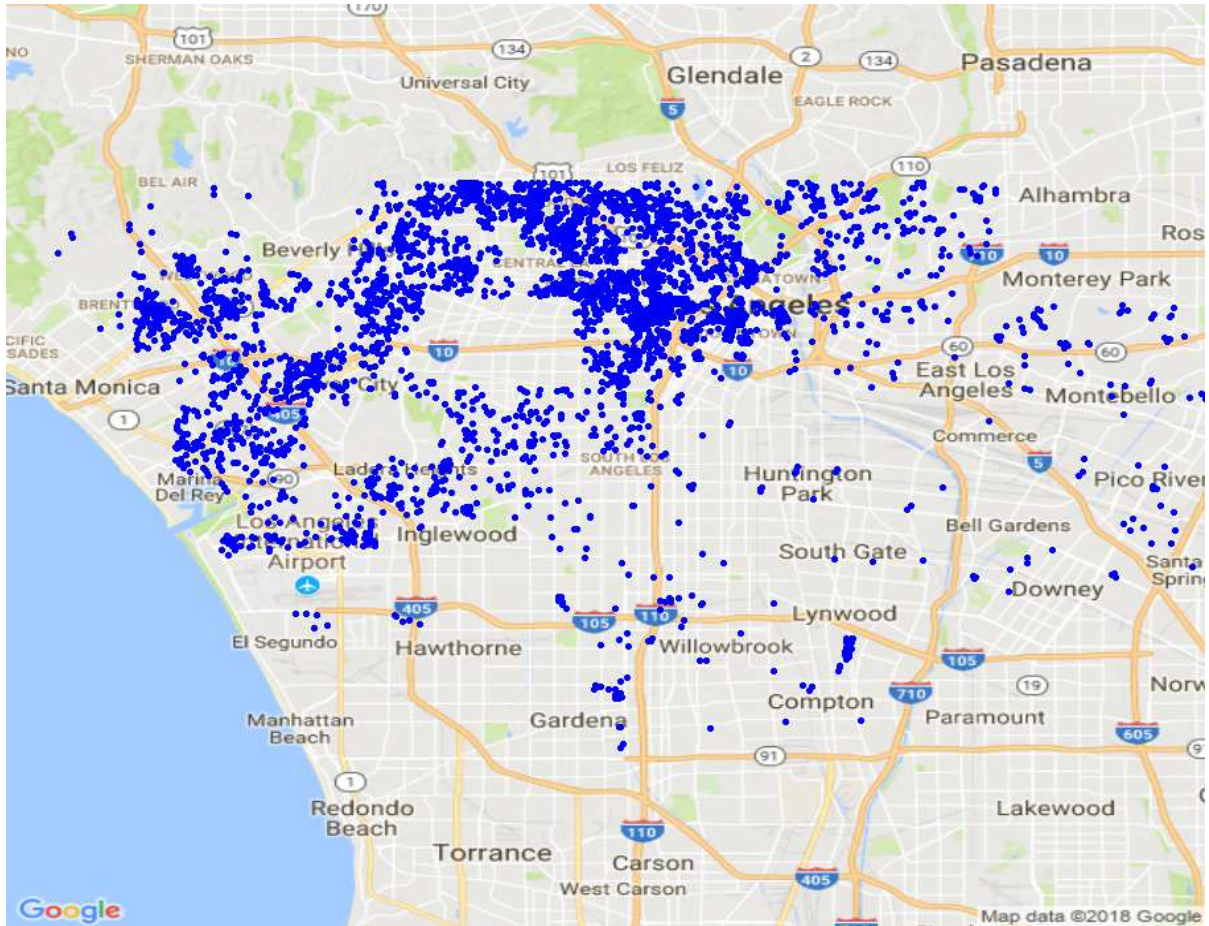


Figure 4.7. Airbnb expensive listings locations summary

To obtain a deeper understanding of Airbnb customers' rating behaviors listings belonging to the same price group (expensive or cheap) but with different overall ratings were selected to reveal the average aspect ratings of these different subgroups of listings, as shown in Table 4.15. Of interest is that reviewers gave expensive rentals high ratings primarily because of the appeal of their products, services, and locations, and gave low ratings due to undesirable listing conditions and overpricing. Reviewers gave the cheap accommodations high ratings primarily, since they offered good pricing/value and fine locations and gave such rentals low ratings for their poor service.

Table 4.15. Aspect rating comparison between expensive and cheap Airbnb listings

	Communication	Experience	Location	Product/service	Value
Cheap	2.924	3.131	3.216	2.391	3.393
Expensive	2.985	3.037	3.413	3.179	2.879

The ability to rank listings according to inferred ratings of aspect is especially useful to consumers. Weighting the different aspects can enable customers to further personalize their searches by only selecting the listings with aspect rating weights similar to those of a specific user. For example, using the weight preference of a listing as a query term, consumers are able to choose listings with weighting preferences that closely match their own and can rank listings based on shared preferences alone. The top five listings with their highest product/service aspect weights, overall ratings, and prices are listed in Table 4.16 as an example.

Table 4.16. The top five listings with their highest product/service aspect weights

Listing ID	Nightly Price	Real Overall Rating	Predicted Overall Rating	Communication	Experience	Location	Product /service	Value
31939	548	4	4.240	.143	.164	.143	.388	.162
14628818	650	5	4.741	.142	.166	.142	.386	.164
15221415	700	5	4.570	.141	.167	.141	.384	.166
298259	285	4.5	4.190	.169	.178	.215	.205	.233
12570324	189	4.5	4.058	.166	.175	.218	.204	.238

CHAPTER 5. DISCUSSION

The prior chapter offers empirical results that support the value of using the advanced LARA framework for predicting customer aspect ratings and the related weights of the different aspects that customers value when writing reviews in the sharing economy context. In summary, the results suggest that there are five lodging aspects that customers most care about: communication, experience, location, product/service, and value. Customers have different sentiments toward different lodging aspects, which lead to different aspect evaluation weights and aspect ratings when they judge their overall stays with the accommodations offered on sharing accommodation platforms. Specifically, customers value the aspects of location and experience more than any other lodging aspects. Chapter 5 specifically discusses how the findings have been interpreted; the implications, both theoretical and managerial, of the findings; the limitations of the study; and directions for further research.

Discussion of the Results

This study proposed a novel approach to (a) examine the underpinnings of rating according to various lodging topics/dimensions considered by customers as expressed in online reviews, and (b) predict an overall rating via the latent weights of each aspect. A text-mining technique with both supervised and unsupervised approaches was employed to identify the most prevalent aspects or dimensions perceived by Airbnb guests when expressing their thoughts and sharing rental experiences online. For topic/aspect extraction, the ABSA of customer reviews revealed meaningful aspects and the eight levels of emotions according to the aspects, which cannot be identified via traditional means.

The five lodging aspects (communication, experience, location, product/service, and value) were extracted in this study by using topic modeling and boot-strapping algorithms to increase both the amount of words belonging to the aspects and the accuracy of the second step of ABSA, which was the method utilized to combine the sentiment analysis of the reviews and the *NRC Emotion Lexicon*. Based on the results, review text expressing joy was less common for the aspects of communication and experience as compared to the other three aspects. Customers were more surprised by the former two aspects due to the unique Airbnb operation process, which differs from that of traditional hotels. Within the aspect of communication, “fee” and “cancel” were found to be the key words that were most salient to customers. For the experience aspect, it was noted that customers expected to explore their destinations. The careful explanation of destination attractions and thoughtful service by hosts can contribute to customers’ positive feelings in this realm. Location was found to be the aspect most valued by customers, who expressed concern in the reviews specifically about convenient transportation and location-related safety issues. Although product/service was found to be the least important aspect to customers, based on the reviews, hosts should nonetheless pay close attention to issues such as cleanliness and Internet reliability. The aspect of value was identified as the main driver of customer booking of sharing accommodations; thus, hosts should ensure that their descriptions match customer expectations upon arrival.

Certain positive key words were mentioned in relation to each aspect with issues such as hosts’ responses to customer questions and host resolution of customer problems, both of which were found to be the points that elicited positive emotions in customers. Thoughtful suggestions and special services provided by hosts are important considerations in

developing customer experience improvements. Prompt action that fulfills customers' needs and resolves problems during the stay; friendliness and smiling; and being readily available to help are all ways for hosts to increase the positive feelings of customers. Reasonable pricing within various customer segments can also contribute to customers' positive attitudes about accommodations.

With the high accuracy performance of the LARA model in predicting overall ratings in customer reviews, the author has shown the underlying relationships between the various emotional expressions in the latent lodging aspects, aspect ratings and the relative weights in the formation of an overall rating. The LRR inferred aspect ratings and weights for the individual listings use the content of the reviews and the overall ratings gleaned from rich online customer reviews. Listing location and experience were identified as the most weighted dimensions for customers' rating Airbnb accommodations. In addition, the study uncovered noteworthy differences in aspect ratings and user rating behaviors with reviews that possessed the same overall ratings. Finally, the study offers evidence that the proposed model can be employed to conduct detailed opinion analysis at the topical aspect level in order to carry out a multiplicity of application tasks, such as opinion summary, listing ranking by aspect rating, and customer rating behavior analysis.

Implications

Since mining UGC has become a pervasive phenomenon in the areas of hospitality and tourism management, a deeper understanding and investigation of customer experience in the overall rating process is extremely important. The findings from the present study provide important theoretical and managerial implications for academicians and practitioners. Specifically, the implications may be beneficial to scholars who seek to evolve research in

the areas of customer experience and rating behavior, and to practitioners (particularly in the lodging industry) when developing useful strategies to create and offer the ideal experiences that customers seek.

Theoretical Implications

The findings of the present study provide theoretical contributions to the literature in several ways. First this study fills the research gap between sharing accommodation platforms and their customer experience. This study adds to the literature of UGC by creating a supplement approach for topic/aspect extraction in UGC with a focus on the analysis of latent ratings in opinion texts. This study also contributes to the research body by elucidating the role of different emotional levels in customer experience expression that in turn influence the overall rating. This study further develops and validates the LARA model as a repertoire of analytical and computational procedures to achieve detailed understandings of opinions and examines the applicability of customer rating behavior in the context of a sharing accommodation platform through a qualitative, empirical, machine learning-based study. Overall, the present study provides new theoretical insights into customer rating prediction in the hospitality and tourism industry.

First, from the theoretical perspective of the sharing economy, this paper contributes to the body of customer experience from the angle of the sharing economy. Because sharing accommodation platforms are a critical aspect of the lodging industry, it is important to evaluate customer experience. Moreover, compared with the traditional hotel, sharing accommodations provide a unique experience to customers. With more and more tourists prefer to express their experience their opinions, experience online. By extracting data from customer reviews, five dimensions and their weights in customers' minds have been

determined. In contrast to research on the traditional hotel customer experience, this study expands the concept of customer experience into five aspects.

Second, this study filled a critical gap by providing a more accurate and efficient method for sentiment analysis and opinion mining that offers a new application of artificial intelligence to online review analysis. With the radical increase in UGC on social media sites, tourism and hospitality researchers have realized the importance of analyzing such content. Prior literature on UGC primarily examined aspects such as customer experience, customer behavior, tourist mobility patterns, service quality, and eWOM, with UGC text as the main type of data used. However, the new trend is the use of specialized methods (e.g., LDA) to handle the Big Data aspects of UGC. In this vein, the present study identified the aspects to examine and applied the boot-strapping method in order to acquire a greater number of related key words per aspects. The results of the study confirm the usefulness of understanding a variety of aspects pertaining to customer experience.

Third, this study revealed the benefits of using text mining and Big Data analytics to identify unique patterns of Airbnb customer rating behaviors in conveniently available Internet UGC. The study utilized data posted specifically on the Airbnb website during a certain period of time, yet the findings reflect in more general ways how consumers discuss personal experiences online in reviews. The hybrid topic or dimension of guest experience shows how the analysis can reveal the semantic differentiations in different lodging aspects in relation to the overall ratings expressed in online reviews. This type of differentiation could not be observed in traditional guest survey studies. Furthermore, very few studies in recent years have mined customers' or travelers' reviews via different level of emotions; instead, the majority has focused on the positive and negative system of classification.

However, these studies have proven the critical role of complex emotions in customer rating behavior. Thus, the findings from this study empirically underscore the influence of different emotion levels according to various lodging aspects on customers' rating behavior formation and extend an empirical understanding of customer experience.

Forth, while text mining analytics from the Big Data perspective is recommended by many researchers as a useful, cross-disciplinary paradigm, it has been sparsely applied in hospitality to the full extent of its capabilities. In particular, many studies have extracted the different lodging topics/aspects from numerous reviews, but the inherent connection between the extracted topics/aspects and overall ratings has received little attention. This study employed and further improved LARA to identify the inherent relationships between different lodging aspects and their various weights in overall ratings. The uniqueness of this study lies in the use of text mining analytics from the Big Data point of view and the delineation of different guest opinions or experiences from stays at Airbnb accommodations. Thus, this study contributes to the theoretical foundation of the determinants and consequences of customer rating behavior.

Practical Implications

An advanced textual data processing method based on both sentiment mining and regression was tested using widely available sources of UGC in this study, a method which is newly emerging. The approach can be applied by hospitality industry practitioners to Big Data gleaned from customers' online reviews. Pragmatically speaking, the study underscored the fact that getting feedback through text mining is of critical value to businesses, and that customer reviews of products/services provide valuable managerial information. Second, while some practitioners have adopted the sentiment mining techniques in the hospitality

field, the results generated by the LARA model suggest the need for a more comprehensive understanding of lodging experience with key words and their corresponding emotions. More importantly, LARA was identified as a helpful review prediction model that can serve as a guideline to develop savvy review recommendation selection systems for travel websites. This work has generated a set of interesting partitioned implications using the LARA model.

Online reviews are typically a form of UGC that is publicly available (Barreda & Bilgihan, 2013), but hotel management can also collect additional data about customer experience via private methods, including Internet-based quality management systems, in order to contact guests via email to solicit evaluations and opinions (Prasad, Wirtz, & Yu, 2014). Hotel management can also get timely guest feedback and identify strengths and weaknesses by mining guest reviews and monitoring, tracking, and managing customer perceptions about their experiences (Prasad et al., 2014). Mining online review text can aid hoteliers in learning about strategies used by competing hotels and developing new products and services. Furthermore, the text mining-based approach employed in the present study can enable hospitality researchers and managers to “hear” every guest’s individual voice as well as perceive the overall picture formed collectively by all customers. In the same vein, these techniques can be employed at the destination level to aid management in evaluating their strengths and weaknesses and developing new marketing approaches.

By comparing aspects between those extracted from customer reviews and those provided by the Airbnb website, this study highlighted the importance of identifying the correct lodging aspects in order to provide valuable information that customers pursue. Based on the findings, it was determined that location and experience are the two lodging aspects about which customers are most concerned. Since customers seek the experience of “living

like a local” in certain desired neighborhoods, the results reveal that location is particularly important in terms of the five lodging aspects and how they shape customer ratings of listings. Based on such information, hosts offering rentals in areas that reflect local customs and traditions can create more effective marketing and advertising strategies by showcasing appealing local features such as beautiful outdoor scenes typical of the locale and of scenes depicting the local customs or characteristics). For listings located in less desirable areas, it is important for hosts to market their offerings by describing convenient area transportation. Features such as parks, bodies of water, airports, and trains are also highly important to travelers evaluating sharing accommodations.

Providing access to the “local experience”, such as local lifestyles, local foods, traditional handicrafts, and local marketplaces, now attracts significantly more customer attention. Emphasizing these unique aspects of the destination can help improve the customer experience. For instance, listing designs can be modified to fit local architectural styles or the city image. One successful example of this approach is the room design at the Hollywood Regency Hotel, which is known for its glamour, sense of drama, and new twist on the old classics. Special outdoor experience factors like water-or landscape views also require highlighting when advertising accommodations, particularly for couples seeking romantic getaways. Offering guidance or education about local foods, markets, or travel sights are also important so that hosts can meet the expectations of their guests, especially for travelers from other countries.

Value is the third most important aspect that matters to customers. From the in-depth ABSA, it was found that, while customers care about price, they are more concerned with whether or not a listing meets their expectations or is worth the money spent on the

accommodation. Thus, hosts should pay close attention to providing cost-effective listing settings and services. With regard to the aspect of product/service, since the majority of customers rent an entire room or apartment, as long as the required facilities are provided by the host (e.g., bathroom, towels, kitchen, and cooking equipment) customer are less concerned about this aspect. In the same vein, renting an individual room or apartment limits opportunities for communication between host and guest during a stay. However, customers are considerably concerned about pre-sale communication.

Even though customers deem negative reviews as more helpful (as compared to positive reviews) in purchase decision-making (Filieri, 2016), extreme negative reviews are less helpful in future decision-making. These types of reviews can influence customer perceptions of a review site and can prompt customers to seek out alternative sites in order to obtain consumer feedback. As a result, it is imperative for websites to provide concise review posting and etiquette guidelines to sustain positive online customer relationships in the digital community. In addition, sharing accommodation hosts should take immediate action to address reviewers' distress as conveyed in reviews before such customers become increasingly disillusioned. One benefit of the LARA model is that sharing accommodation platforms can employ it in developing review methods for management strategies through the categorization and prioritization of customer reviews according to different levels of emotion. According to the service recovery paradox, customers give higher rating to a hotel if it successfully addresses its failures than they would if perfect service had been provided in the first place (Kau & Wan-Yiun, 2006). Sharing accommodation platforms managers and hosts should therefore be aware of reviews conveying excessive negative emotions.

The majority of existing travel booking websites permits viewers to sort reviews according to review rating, date, or author reputation. Homeaway.com reviews, for instance, can be sorted solely by rating; however, this method does not convey whether or not specific reviews are helpful. In contrast, the Airbnb website provides a more comprehensive search method that enables viewers to search according to key word. Nonetheless, the site does not offer a method to quickly filter out those reviews that offer less information and are less helpful to viewers. Since viewers typically have a limited amount of time to search the often-vast amounts of hotel review data available online, a system that allows viewers to rapidly access the most important information about hotels and make online bookings quickly is of the utmost importance in the hospitality industry. The LARA model can enable travel websites to create additional query functions for filtering and ranking reviews according to viewer needs. Some examples of these functions include sorting reviews by the various lodging aspects and by the emotions attached to these aspects, and filtering reviews showing similar lodging preferences. Using this type of algorithm, review and hospitality websites can highlight particularly helpful reviews to better the overall user experience and boost the usefulness of the reviews on the sites.

Based on the findings of this study, it was determined that reviewers chose lodging aspects differently than does the Airbnb website in terms of aspect rating. Specifically, on Airbnb, fewer reviewers care about service as opposed to experience. Thus, the inappropriate choice of aspect ratings by Airbnb means that less valuable information is available to customers. The LARA model allows hotel management staff, investors, and hosts to employ UGC to determine the heterogeneity and importance of the latent dimensions of consumer satisfaction. Moreover, even though Airbnb offers a word search function in its reviews,

some key words under the aspects are absent. For example, the site's search engine only presents reviews containing the term "clean" when providing reviews about accommodations that are deemed clean, while several other key words that denote cleanliness are not used. The LARA model provides a bootstrapping system that can include information from these key words as well.

The present study can help hospitality management to detect which online reviews are most influential to viewers, which is critical in light of the fact that helpful reviews are shown more favor and are therefore more influential in the decision-making process. As such, the reviews deemed most helpful by users should be carefully determined and analyzed. Managers require a system to rapidly identify the most helpful reviews according to different levels of emotion instead of a binary positive-negative criterion. Some reviews contain mixed (positive and negative) emotions about products and thus cannot be categorized in a binary fashion. Therefore, it is critical for hospitality management to analyze a variety of emotions and levels of emotions related to each lodging aspect in order to attract guests and develop adequate remedial methods for complaints according to levels of negative emotions. The findings of this study indicate that managers should primarily emphasize precise or easily comprehensible reviews because they tend to have greater influence on viewers than do other types of reviews.

This study also reveals that reviewers give overall ratings based on the different weights of the various lodging aspects. Using the LARA approach, it is now possible to infer the weights that customers place on such features. Based on the findings, it is proposed that it is possible to develop a smart recommendation system that suggests useful reviews customized to individual travelers. In the envisioned system, customers would be able to

enter their most important lodging aspects according to personal preferences in order to view reviews with similar rating weights. This “smart” recommendation system would provide reviews with the high-quality information that customers seek, according to review quality, review polarity, and reviewer characteristics, with efficiency and ease. Such a system would save time that customers spend on websites searching for helpful reviews for their trips and would increase the usability of booking websites by providing valuable, customized information on the selected accommodations.

In addition to the unique implications for the hospitality industry, this study also provides meaningful managerial implications to sharing economy organizations. There are points of overlap between the customer experience determinants of traditional lodging options and sharing accommodations, since the main purpose of customer sharing accommodations is the same as with traditional lodging: a place to stay during travel. Thus, the aspects such as location, value, and product/service remain the same. However, based on the aspect weights, the differences become significant. The importance of experience and communication play an important role when customers evaluate their experiences staying in sharing accommodations. From this logic, sharing accommodation platforms should add more information about other customers’ experiences, and describe more of the local customs and activities. For sharing hosts, it is essential to have more and exceptionally positive interactions with guests.

Limitations and Future Research Directions

In spite of the valuable theoretical and managerial implications of the study, there are several limitations inherent in the design and methodology of the research that necessitate additional research and investigation. This study is unique in identifying latent lodging aspect

ratings, weights and predicting overall ratings by mining customer review text on the Airbnb website, and represents a first step in comprehending and predicting the relationships between customers' emotional expressions in online reviews and their rating behaviors. However, future studies are necessary in order to conduct additional investigation in greater depth into the issue of mining review data and predicting overall ratings.

First, the model used to extract the latent dimensions of lodging experience was computationally intensive. Nonetheless, although the bootstrapping system proved excellent in providing more key words under each aspect as compared with traditional aspect or topic extraction methods such as LDA, the semantics of each review were not considered in the present study. Some complex expressions such as double negative sentences were not detected in the present study. Various techniques (such as the n-gram) should be tested in order to detect the complex semantic meanings in customer reviews. Moreover, the NRC applied in this study only contains the adjectives. The adjectives sentiment polarities that near the key word that extracted were used into the LRR calculation. It also adds bias into the result of this study.

Second, this study is limited because it focused solely on Airbnb customer reviews for one well-known destination city in the United States. Furthermore, the overall ratings given by customers weren't provided on the Airbnb website, which causes prediction bias. Future studies should extract more aspects from different perspectives by applying different mining techniques. Moreover, since Airbnb customers pursue unique travel experiences, some minor changes are required before extending this model to fit other tourism service review analysis (e.g., restaurants and travel service providers) with the detailed rating information offered on websites other than Airbnb.

Third, this study did not employ the aspect ratings that the Airbnb website provided. Instead, the aspects were based on the results of the wordcloud analysis, which indicated very few aspects in common with the aspects proposed by the Airbnb website. The theoretical framework used in this study should be modified so that it includes a greater number of meaningful variables. In addition, future research should also explore the six aspects that the Airbnb website provides. By the same logic, this study only extracted five lodging aspects from numerous reviews, and there may be some as yet unidentified aspects that are particularly important to customers.

Fourth, the LARA model involves both supervised and unsupervised learning techniques, which limited the generalizability of the study. Building upon this study, future studies should attempt to apply all available supervised learning techniques to build a more reliable prediction model. Future attempts that make use of advances in large-scale computing techniques and related technology could eventually eradicate this limitation. For example, based on this study, future research should focus on mining reviews in order to explore the relationship between review length and product price in an effort to establish the research validity of social media analytics in more economical ways.

Lastly, this study is a pioneering work in classifying reviews according to eight basic emotion levels in order to gain a better understanding of the sentiments expressed in online reviews. However, in sets of customer reviews, more than eight emotions are often expressed. Thus, more detailed emotional classifications are recommended to be employed in future review sentiment mining studies.

REFERENCES

- Aggarwal, C. C., & Zhai, C. (Eds.). (2012). *Mining text data*. Springer Science & Business Media.
- Ahn, T., Ryu, S., & Han, I. (2007). The impact of Web quality and playfulness on user acceptance of online retailing. *Information & management*, 44(3), 263-275.
- Akehurst, G. (2009). User generated content: the use of blogs for tourism organizations and tourism consumers. *Service business*, 3(1), 51.
- Amaral, F., Tiago, T., & Tiago, F. (2014). User-generated content: tourists' profiles on Tripadvisor. *International Journal of Strategic Innovative Marketing*, 1(3), 137-145.
- Andrzejewski, D., Zhu, X., & Craven, M. (2009, June). Incorporating domain knowledge into topic modeling via Dirichlet forest priors. In *Proceedings of the 26th Annual International Conference on Machine Learning* (pp. 25-32). ACM.
- Annett, M., & Kondrak, G. (2008, May). A comparison of sentiment analysis techniques: Polarizing movie blogs. In *Conference of the Canadian Society for Computational Studies of Intelligence* (pp. 25-35). Springer, Berlin, Heidelberg.
- Aristotle, Loveday, T., & Forster, E. S. (1913). *Physiognomonica*. United Kingdom: Clarendon Press.
- Arndt, J. (1967). *Word of mouth advertising: A review of the literature*. Advertising Research Foundation.
- Ashcraft, A. B. (2008). Does the market discipline banks? New evidence from regulatory capital mix. *Journal of Financial Intermediation*, 17(4), 543-561.
- Asur, S., & Huberman, B. A. (2010, August). Predicting the future with social media. In *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 01* (pp. 492-499). IEEE Computer Society.
- Babin, B. J., Lee, Y. K., Kim, E. J., & Griffin, M. (2005). Modeling consumer satisfaction and word-of-mouth: restaurant patronage in Korea. *Journal of Services Marketing*, 19(3), 133-139.
- Baccianella, S., Esuli, A., & Sebastiani, F. (2010, May). Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In *LREC* (Vol. 10, No. 2010, pp. 2200-2204).
- Bae, S., & Lee, T. (2011). Gender differences in consumers' perception of online consumer reviews. *Electronic Commerce Research*, 11(2), 201-214.

- Baka, V. (2016). The becoming of user-generated reviews: Looking at the past to understand the future of managing reputation in the travel sector. *Tourism Management*, 53, 148-162.
- Bansal, H. S., & Voyer, P. A. (2000). Word-of-mouth processes within a services purchase decision context. *Journal of service research*, 3(2), 166-177.
- Bardhi, F., & Eckhardt, G. M. (2012). Access-based consumption: The case of car sharing. *Journal of consumer research*, 39(4), 881-898.
- Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Vol. 1, pp. 238-247).
- Barreda, A., & Bilgihan, A. (2013). An analysis of user-generated content for hotel experiences. *Journal of Hospitality and Tourism Technology*, 4(3), 263-280.
- Belk, R. (2014). You are what you can access: Sharing and collaborative consumption online. *Journal of Business Research*, 67(8), 1595-1600.
- Bergen, S., & Guggenheim, A. (2016). Of Ubers, Tech Users and Elder Abusers: State Legislation to Watch. *ABA Banking Journal*, 108(2), 19.
- Bharwani, S., & Jauhari, V. (2013). An exploratory study of competencies required to co-create memorable customer experiences in the hospitality industry. *International Journal of Contemporary Hospitality Management*, 25(6), 823-843.
- Björkelund, E., Burnett, T. H., & Nørvåg, K. (2012, December). A study of opinion mining and visualization of hotel reviews. In *Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services* (pp. 229-238). ACM.
- Blair-Goldensohn, S., Hannan, K., McDonald, R., Neylon, T., Reis, G. A., & Reynar, J. (2008, April). Building a sentiment summarizer for local service reviews. In *WWW workshop on NLP in the information explosion era* (Vol. 14, pp. 339-348).
- Blum, A. L., & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial intelligence*, 97(1-2), 245-271.
- Boiy, E., & Moens, M. F. (2009). A machine learning approach to sentiment analysis in multilingual Web texts. *Information retrieval*, 12(5), 526-558.
- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of computational science*, 2(1), 1-8.
- Bormann, E. G. (1972). Fantasy and rhetorical vision: The rhetorical criticism of social reality. *Quarterly journal of speech*, 58(4), 396-407.

- Boshoff, C., & Van Eeden, S. M. (2001). South African consumer sentiment towards marketing: a longitudinal analysis. *South African Journal of Business Management*, 32(2), 23-33.
- Botsman, R. (2015, May 27). Defining the sharing economy: What is collaborative consumption – And what isn't? Fast Company. Retrieved from <http://www.fastcoexist.com/3046119/defining-the-sharing-economy-what-is-collaborativeconsumption-and-what-isnt>.
- Botsman, R., & Rogers, R. (2010). Beyond zipcar: Collaborative consumption. *Harvard Business Review*, 88(10), 30.
- Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings* (pp. 1-45). Technical report C-1, the center for research in psychophysiology, University of Florida.
- Brand, R., & Rocchi, S. (2011). Rethinking value in a changing landscape: A model for strategic reflection and business transformation. Retrieved March 4, 2018, from http://www.design.philips.com/philips/shared/assets/design_assets/pdf/nvbD/april2011/paradigms.pdf
- Brochado, A., Troilo, M., & Shah, A. (2017). Airbnb customer experience: Evidence of convergence across three countries. *Annals of Tourism Research*, 63, 210-212.
- Brown, J. J., & Reingen, P. H. (1987). Social ties and word-of-mouth referral behavior. *Journal of Consumer research*, 14(3), 350-362.
- Bucur, C. (2015). Using Opinion Mining Techniques in Tourism. *Procedia Economics and Finance*, 23, 1666-1673.
- Buhalis, D., & Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research. *Tourism management*, 29(4), 609-623.
- Burgess, S., Sellitto, C., Cox, C., & Buultjens, J. (2009, June). User-generated content (UGC) in tourism: Benefits and concerns of online consumers. In *ECIS* (pp. 417-429).
- Casalo, L. V., Flavian, C., Guinaliu, M., & Ekinici, Y. (2015). Do online hotel rating schemes influence booking behaviors?. *International Journal of Hospitality Management*, 49, 28-36.
- Carlson, A. B. (2010). *Communication system*. Tata McGraw-Hill Education.
- Certify. (2015). Room for more: Business travelers embrace the sharing economy. *Certify.com*. Retrieved December 15, 2015, from <https://www.certify.com/Infograph-Sharing-Economy-Q2-2015.aspx>.

- Chafale, D., & Pimpalkar, A. (2014). Review on developing corpora for sentiment analysis using plutchik's wheel of emotions with fuzzy logic. *International Journal of Computer Sciences and Engineering*, 2(10), 14-18.
- Chan, N. L., & Guillet, B. D. (2011). Investigation of social media marketing: how does the hotel industry in Hong Kong perform in marketing on social media websites? *Journal of Travel & Tourism Marketing*, 28(4), 345-368.
- Chaovalit, P., & Zhou, L. (2005, January). Movie review mining: A comparison between supervised and unsupervised classification approaches. In *System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on* (pp. 112c-112c). IEEE.
- Chau, M., Shiu, B., Chan, I., & Chen, H. (2007). Redips: Backlink search and analysis on the Web for business intelligence analysis. *Journal of the Association for Information Science and Technology*, 58(3), 351-365.
- Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: from big data to big impact. *MIS quarterly*, 1165-1188.
- Chen, Y., & Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management science*, 54(3), 477-491.
- Cheung, C. M., Lee, M. K., & Rabjohn, N. (2008). The impact of electronic word-of-mouth: The adoption of online opinions in online customer communities. *Internet research*, 18(3), 229-247.
- Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of marketing research*, 43(3), 345-354.
- Chiang, C. F., & Jang, S. S. (2007). The effects of perceived price and brand image on value and purchase intention: Leisure travelers' attitudes toward online hotel booking. *Journal of Hospitality & Leisure Marketing*, 15(3), 49-69.
- Chung, J. Y., & Buhalis, D. (2008). Information needs in online social networks. *Information Technology & Tourism*, 10(4), 267-281.
- Clark, A., Fox, C., & Lappin, S. (Eds.). (2013). *The handbook of computational linguistics and natural language processing*. John Wiley & Sons.
- Clary, E. G., Snyder, M., Ridge, R. D., Copeland, J., Stukas, A. A., Haugen, J., & Miene, P. (1998). Understanding and assessing the motivations of volunteers: a functional approach. *Journal of personality and social psychology*, 74(6), 1516.
- Clemons, E. K., Gao, G. G., & Hitt, L. M. (2006). When online reviews meet hyperdifferentiation: A study of the craft beer industry. *Journal of management information systems*, 23(2), 149-171.

- Cohen, B., & Kietzmann, J. (2014). Ride on! Mobility business models for the sharing economy. *Organization & Environment*, 27(3), 279-296
- Cohen, M., & Sundararajan, A. (2015). Self-regulation and innovation in the peer-to-peer sharing economy. *U. Chi. L. Rev. Dialogue*, 82, 116-134.
- Confente, I. (2015). Twenty-five years of word-of-mouth studies: A critical review of tourism research. *International Journal of Tourism Research*, 17(6), 613-624.
- Cox, C., Burgess, S., Sellitto, C., & Buultjens, J. (2009). The role of user-generated content in tourists' travel planning behavior. *Journal of Hospitality Marketing & Management*, 18(8), 743-764.
- Curhan, J. R., Elfenbein, H. A., & Eisenkraft, N. (2010). The objective value of subjective value: A multi-round negotiation study. *Journal of Applied Social Psychology*, 40(3), 690-709.
- Dao, N. L., & Vu, A. (2016). Customers' purchasing behaviour between Airbnb and hostels when travelling to Helsinki. Retrieved March 4, 2018, from: https://www.theseus.fi/bitstream/handle/10024/119462/Dao_Lien.pdf?sequence=1&isAllowed=y
- Daugherty, T., Eastin, M. S., & Bright, L. (2008). Exploring consumer motivations for creating user-generated content. *Journal of interactive advertising*, 8(2), 16-25.
- Dave, K., Lawrence, S., & Pennock, D. M. (2003, May). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web* (pp. 519-528). ACM.
- Davenport, T. H., Barth, P., & Bean, R. (2012). How big data is different. *MIT Sloan Management Review*, 54(1), 43.
- Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management science*, 49(10), 1407-1424.
- Dellarocas, C., Fan, M., & Wood, A. W. (2004). *Self-Interest, reciprocity, and participation in online reputation systems*. Working Paper. Accessed online from: http://ebusiness.mit.edu/research/papers/205_Dellarocas_EbayParticipation.pdf.
- Dellarocas, C., Zhang, X. M., & Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive marketing*, 21(4), 23-45.
- Denzin, N. K., & Lincoln, Y. S. (Eds.). (2011). *The Sage handbook of qualitative research*. Sage.
- Dey, L., & Haque, S. M. (2009). Opinion mining from noisy text data. *International Journal on Document Analysis and Recognition (IJDAR)*, 12(3), 205-226.

- Dhar, V., & Chang, E. A. (2009). Does chatter matter? The impact of user-generated content on music sales. *Journal of Interactive Marketing*, 23(4), 300-307.
- Ding, X., Liu, B., & Yu, P. S. (2008, February). A holistic lexicon-based approach to opinion mining. In *Proceedings of the 2008 international conference on web search and data mining* (pp. 231-240). ACM.
- Deokar, A. V., Gupta, A., Iyer, L. S., & Jones, M. C. (Eds.). (2017). *Analytics and Data Science: Advances in Research and Pedagogy*. Springer.
- Duan, W., Gu, B., & Whinston, A. B. (2008). Do online reviews matter? An empirical investigation of panel data. *Decision support systems*, 45(4), 1007-1016.
- Edelman, B. G., & Geradin, D. (2015). Efficiencies and regulatory shortcuts: How should we regulate companies like Airbnb and Uber. *Stan. Tech. L. Rev.*, 19, 293.
- Edelman, D. C. (2007). From the periphery to the core: As online strategy becomes overall strategy; marketing organizations and agencies will never be the same. *Journal of Advertising Research*, 47(2), 130-134.
- Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, 6(3-4), 169-200.
- Ert, E., Fleischer, A., & Magen, N. (2016). Trust and reputation in the sharing economy: The role of personal photos in Airbnb. *Tourism Management*, 55, 62-73.
- Esuli, A., & Sebastiani, F. (2007). SentiWordNet: a high-coverage lexical resource for opinion mining. *Evaluation*, 1-26.
- Fairclough, N. (2003). *Analysing discourse: Textual analysis for social research*. Psychology Press.
- Fairclough, N. (2003). *Analysing discourse: Textual analysis for social research*. Psychology Press.
- Fick, G. R., & Brent Ritchie, J. R. (1991). Measuring service quality in the travel and tourism industry. *Journal of Travel Research*, 30(2), 2-9.
- Filieri, R. (2016). What makes an online consumer review trustworthy?. *Annals of Tourism Research*, 58, 46-64.
- Fortenbaugh, W. W. (1970). Aristotle's Rhetoric on Emotions. *Archiv für Geschichte der Philosophie*, 52(1), 40-70.
- Fraiberger SP, Sundararajan A (2015) Peer-to-peer rental markets in the sharing economy. Research paper, Stern School of Business, New York University, New York.
- Frenken, K., Meelen, T., Arets, M. and Glind, P.V. (2015). Smarter regulation for the sharing economy, *The Guardian*, 20 May, available

- at: www.theguardian.com/science/political-science/2015/may/20/smarter-regulation-for-the-sharing-economy (accessed 30 October 2017).
- Frijda, N. H. (1986). *The emotions*. Cambridge University Press.
- Gentile, C., Spiller, N., & Noci, G. (2007). *How to sustain the customer experience: An overview of experience components that co-create value with the customer*. *European Management Journal*, 25, 395-410.
- Ghani, R., Probst, K., Liu, Y., Krema, M., & Fano, A. (2006). Text mining for product attribute extraction. *ACM SIGKDD Explorations Newsletter*, 8(1), 41-48.
- Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1(12).
- Godbole, N., Srinivasaiah, M., & Skiena, S. (2007). Large-Scale Sentiment Analysis for News and Blogs. *Icwsn*, 7(21), 219-222.
- Goh, K. Y., Heng, C. S., & Lin, Z. (2013). Social media brand community and consumer behavior: Quantifying the relative impact of user-and marketer-generated content. *Information Systems Research*, 24(1), 88-107.
- Gopalkrishnan, V., Steier, D., Lewis, H., & Guszczka, J. (2012, August). Big data, big business: bridging the gap. In *Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications* (pp. 7-11). ACM.
- Graneheim, U. H., & Lundman, B. (2004). Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse education today*, 24(2), 105-112.
- Green, A. E., & White, R. J. (2007). *Attachment to place: Social networks, mobility and prospects of young people*. York: Joseph Rowntree Foundation.
- Green, N. (2007). Functional polycentricity: A formal definition in terms of social network analysis. *Urban Studies*, 44(11), 2077-2103.
- Gretzel, U. (2006). Consumer generated content—trends and implications for branding. *E-review of Tourism Research*, 4(3), 9-11.
- Gretzel, U., & Yoo, K. H. (2008). Use and impact of online travel reviews. *Information and communication technologies in tourism 2008*, 35-46.
- Gretzel, U., Yoo, K. H., & Purifoy, M. (2007). Online Travel Review study: Role and Impact of Online Travel Reviews, Laboratory for intelligent systems in Tourism, Texas, A&M University (2007).
- Gruen, T. W., Osmonbekov, T., & Czaplewski, A. J. (2005). How e-communities extend the

- concept of exchange in marketing: An application of the motivation, opportunity, ability (MOA) theory. *Marketing Theory*, 5(1), 33-49.
- Guess, A. M. (2015). Fact-checking on Twitter: An examination of campaign 2014. *American Press Institute. Online. Viitattu*, 17, 2015.
- Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59, 467-483.
- Guttentag, D. (2015). Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector. *Current issues in Tourism*, 18(12), 1192-1217.
- Hamari, J., Sjöklint, M., & Ukkonen, A. (2016). The sharing economy: Why people participate in collaborative consumption. *Journal of the Association for Information Science and Technology*, 67(9), 2047-2059.
- Harwood, S. (2007). *Travelers trust user reviews over travel agent content*, Revolution UK.
- Havitz, M. E., & Dimanche, F. (1990). Propositions for testing the involvement construct in recreational and tourism contexts. *Leisure Sciences*, 12(2), 179-195.
- Hearst, M. (2003). *What is text mining*. SIMS, UC Berkeley.
- Heinrichs, H. (2013). Sharing economy: a potential new pathway to sustainability. *GAIA-Ecological Perspectives for Science and Society*, 22(4), 228-231.
- Hellwig, K., Morhart, F., Girardin, F., & Hauser, M. (2015). Exploring different types of sharing: A proposed segmentation of the market for “sharing” businesses. *Psychology & Marketing*, 32(9), 891-906.
- Helm, R. & Mark, A. 2007. Implications from cue utilisation theory and signalling theory for firm reputation and the marketing of new products. *International Journal of Product Development*, 4: 396–411.
- Hemmington, N. (2007). From service to experience: Understanding and defining the hospitality business. *The Service Industries Journal*, 27(6), 747-755.
- Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word of mouth via consumer opinion platforms: What motivates consumers to articulate themselves on the Internet?. *Journal of interactive marketing*, 18(1), 38-52.
- Heo, Y. (2016). Sharing economy and prospects in tourism research. *Annals of Tourism Research*, 58, 166-170.
- Hertzum, M., Andersen, H. H., Andersen, V., & Hansen, C. B. (2002). Trust in information sources: seeking information from people, documents, and virtual agents. *Interacting with computers*, 14(5), 575-599.

- Hollander, J. B., Graves, E., Renski, H., Foster-Karim, C., Wiley, A., & Das, D. (2016). A (short) history of social media sentiment analysis. In *Urban Social Listening* (pp. 15-25). Palgrave Macmillan, London.
- Holsing, C., & Olbrich, R. (2012, August). Effect of user-generated content on website stickiness: the case of social shopping communities. In *Proceedings of the 14th Annual International Conference on Electronic Commerce* (pp. 301-308). ACM.
- Holsti, L. R. (1969). Clinical experience with split-course radiotherapy: A randomized clinical trial. *Radiology*, 92(3), 591-596.
- Hong, L., & Davison, B. D. (2010, July). Empirical study of topic modeling in twitter. In *Proceedings of the first workshop on social media analytics* (pp. 80-88). ACM.
- Hsieh, J. K., Hsieh, Y. C., & Tang, Y. C. (2012). Exploring the disseminating behaviors of eWOM marketing: persuasion in online video. *Electronic Commerce Research*, 12(2), 201-224.
- Hsu, M. H., Ju, T. L., Yen, C. H., & Chang, C. M. (2007). Knowledge sharing behavior in virtual communities: The relationship between trust, self-efficacy, and outcome expectations. *International journal of human-computer studies*, 65(2), 153-169.
- Hu, M., & Liu, B. (2004, August). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168-177). ACM.
- Hu, N., Bose, I., Koh, N. S., & Liu, L. (2012). Manipulation of online reviews: An analysis of ratings, readability, and sentiments. *Decision Support Systems*, 52(3), 674-684.
- Hu, N., Liu, L., & Zhang, J. J. (2008). Do online reviews affect product sales? The role of reviewer characteristics and temporal effects. *Information Technology and Management*, 9(3), 201-214.
- Hu, N., Pavlou, P. A., & Zhang, J. (2006, June). Can online reviews reveal a product's true quality?: empirical findings and analytical modeling of Online word-of-mouth communication. In *Proceedings of the 7th ACM conference on Electronic commerce* (pp. 324-330). ACM.
- Hung, J. L., & Zhang, K. (2012). Examining mobile learning trends 2003–2008: A categorical meta-trend analysis using text mining techniques. *Journal of Computing in Higher education*, 24(1), 1-17.
- Hutto, C. J., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media*.

- Irie, G., Satou, T., Kojima, A., Yamasaki, T., & Aizawa, K. (2010). Affective audio-visual words and latent topic driving model for realizing movie affective scene classification. *IEEE Transactions on Multimedia*, 12(6), 523-535.
- Jalilvand, M. R., & Samiei, N. (2012). The impact of electronic word of mouth on a tourism destination choice: Testing the theory of planned behavior (TPB). *Internet Research: Electronic Networking Applications and Policy*, 22(5), 591-612.
- Jeong, M., & Mindy Jeon, M. (2008). Customer reviews of hotel experiences through consumer generated media (CGM). *Journal of Hospitality & Leisure Marketing*, 17(1-2), 121-138.
- Jiménez, F. R., & Mendoza, N. A. (2013). Too popular to ignore: The influence of online reviews on purchase intentions of search and experience products. *Journal of Interactive Marketing*, 27(3), 226-235.
- Jo, Y., & Oh, A. H. (2011, February). Aspect and sentiment unification model for online review analysis. In *Proceedings of the fourth ACM international conference on Web search and data mining* (pp. 815-824). ACM.
- John, N. A. (2013). The social logics of sharing. *The Communication Review*, 16(3), 113-131.
- Jung, J., Yoon, S., Kim, S., Park, S., Lee, K. P., & Lee, U. (2016, May). Social or Financial Goals?: Comparative Analysis of User Behaviors in Couchsurfing and Airbnb. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 2857-2863). ACM.
- Jurafsky, D. (2000). *Speech & language processing*. India: Pearson Education.
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons*, 53(1), 59-68.
- Kaplan, R. A., & Nadler, M. L. (2015). Airbnb: A case study in occupancy regulation and taxation. *U. Chi. L. Rev. Dialogue*, 82, 103-115.
- Kassan, J., & Orsi, J. (2012). The legal landscape of the sharing economy. *J. Envtl. L. & Litig.*, 27, 1-20.
- Kau, A. K., & Wan-Yiun Loh, E. (2006). The effects of service recovery on consumer satisfaction: a comparison between complainants and non-complainants. *Journal of Services Marketing*, 20(2), 101-111.
- Keng, C. J., Liao, T. H., & Yang, Y. I. (2012). The effects of sequential combinations of virtual experience, direct experience, and indirect experience: the moderating roles of need for touch and product involvement. *Electronic Commerce Research*, 12(2), 177-199.

- Kim, W. G., & Kim, D. J. (2004). Factors affecting online hotel reservation intention between online and non-online customers. *International Journal of Hospitality Management*, 23(4), 381-395.
- Knutson, B. J., & Beck, J. A. (2004). Identifying the dimensions of the experience construct: development of the model. *Journal of Quality Assurance in Hospitality & Tourism*, 4(3-4), 23-35.
- Knutson, B. J., Beck, J. A., Kim, S., & Cha, J. (2009). Identifying the dimensions of the guest's hotel experience. *Cornell Hospitality Quarterly*, 50(1), 44-55.
- Knutson, J. N. (1973). *Handbook of political psychology*. Indianapolis, IN: Jossey-Bass.
- Koopman, C., Mitchell, M., & Thierer, A. (2014). The sharing economy and consumer protection regulation: The case for policy change. *J. Bus. Entrepreneurship & L.*, 8(2), 529-545.
- Koster, M. (1994). *A standard for robot exclusion*. NEXOR.
- Kotler, P., Bowen, J. T., Makens, J. C., & Baloglu, S. (2006). *Marketing for hospitality and tourism*. Prentice-Hall, New Jersey.
- Krallinger, M., Valencia, A., & Hirschman, L. (2008). Linking genes to literature: text mining, information extraction, and retrieval applications for biology. *Genome biology*, 9(2), S8.
- Kurtz, K., Mak, P. K., & Werndorfer, S. (2006). *U.S. Patent Application No. 10/966,404*.
- Kwak, D. H., Kim, Y. K., & Zimmerman, M. H. (2010). User-versus mainstream-media-generated content: Media source, message valence, and team identification and sport consumers' response. *International Journal of Sport Communication*, 3(4), 402-421.
- Laboy, F., & Torchio, P. (2007). Web 2.0 for the travel marketer and consumer: a white paper. E-site Marketing and the International Association of Online Communicators. Retrieved March 4, 2018, from <http://www.esitemarketing.com/web2-travel-marketing.php>
- Laczniak, R. N., DeCarlo, T. E., & Ramaswami, S. N. (2001). Consumers' Responses to Negative Word-of-Mouth Communication: An Attribution Theory Perspective. *Journal of consumer Psychology*, 11(1), 57-73.
- Lak, P., & Turetken, O. (2017). The Impact of Sentiment Analysis Output on Decision Outcomes: An Empirical Evaluation. *AIS Transactions on Human-Computer Interaction*, 9(1), 1-22.
- Lee, H. A., Law, R., & Murphy, J. (2011). Helpful reviewers in TripAdvisor, an online travel community. *Journal of Travel & Tourism Marketing*, 28(7), 675-688.

- Lee, J., Park, D. H., & Han, I. (2011). The different effects of online consumer reviews on consumers' purchase intentions depending on trust in online shopping malls: An advertising perspective. *Internet research*, 21(2), 187-206.
- Lee, S., & Kim, D. Y. (2018). Brand personality of Airbnb: Application of user involvement and gender differences. *Journal of Travel & Tourism Marketing*, 35(1), 32-45.
- Lehr, D. D. (2015). An analysis of the changing competitive landscape in the hotel industry regarding Airbnb. Master thesis, Dominican University of California, San Rafael, CA.
- Li, J., Moreno, A., & Zhang, D. J. (2015). Agent behavior in the sharing economy: Evidence from Airbnb. *Ross School of Business Working Paper Series*, 1298.
- Lin, C., & He, Y. (2009, November). Joint sentiment/topic model for sentiment analysis. In *Proceedings of the 18th ACM conference on Information and knowledge management* (pp. 375-384). ACM.
- Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism management*, 29(3), 458-468.
- Liu, B. (2010). Sentiment Analysis and Subjectivity. *Handbook of natural language processing*, 2, 627-666.
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.
- Liu, J., & Seneff, S. (2009, August). Review sentiment scoring via a parse-and-paraphrase paradigm. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1* (pp. 161-169). Association for Computational Linguistics.
- Liu, S. Q., & Mattila, A. S. (2017). Airbnb: Online targeted advertising, sense of power, and consumer decisions. *International Journal of Hospitality Management*, 60, 33-41.
- Liu, Z., & Park, S. (2015). What makes a useful online review? Implication for travel product websites. *Tourism Management*, 47, 140-151.
- Lu, C., & Kandampully, J. (2016). What drives customers to use access-based sharing options in the hospitality industry?. *Research in Hospitality Management*, 6(2), 119-126.
- Lu, X., Ba, S., Huang, L., & Feng, Y. (2013). Promotional marketing or word-of-mouth? Evidence from online restaurant reviews. *Information Systems Research*, 24(3), 596-612.

- Lu, Y., & Zhai, C. (2008, April). Opinion integration through semi-supervised topic modeling. In *Proceedings of the 17th international conference on World Wide Web* (pp. 121-130). ACM.
- Maedche, A., & Staab, S. (2001). Ontology learning for the semantic web. *IEEE Intelligent systems*, 16(2), 72-79.
- Malhotra, A. & Van Alstyne, M. (2014). The dark side of the sharing economy... and how to lighten it. *Communications of the ACM*, 57(11), 24-27.
- Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT press.
- Marine-Roig, E., & Clavé, S. A. (2015). Tourism analytics with massive user-generated content: A case study of Barcelona. *Journal of Destination Marketing & Management*, 4(3), 162-172.
- Matzler, K., Veider, V., & Kathan, W. (2015). Adapting to the sharing economy. *MIT Sloan Management Review*, 56(2), 71-81.
- McKee, A. (2003). *Textual analysis: A beginner's guide*. Newcastle upon Tyne, United Kingdom: Sage.
- McNamara, B. (2015). Airbnb: A not-so-safe resting place. *J. on Telecomm. & High Tech. L.*, 13, 149-161.
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093-1113.
- Meyer, C., & Schwager, A. (2007). Customer experience. *Harvard business review*, 85(2), 116-126.
- Meyer, D., Hornik, K., & Feinerer, I. (2008). Text mining infrastructure in R. *Journal of statistical software*, 25(5), 1-54.
- Middleton, V. T., Fyall, A., Morgan, M., & Ranchhod, A. (2009). *Marketing in travel and tourism*. London, United Kingdom: Routledge.
- Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International journal of lexicography*, 3(4), 235-244.
- Mo, P. K., Malik, S. H., & Coulson, N. S. (2009). Gender differences in computer-mediated communication: a systematic literature review of online health-related support groups. *Patient education and counseling*, 75(1), 16-24.
- Moe, W. W., & Trusov, M. (2011). The value of social dynamics in online product ratings forums. *Journal of Marketing Research*, 48(3), 444-456.

- Moens, M. F., Li, J., & Chua, T. S. (Eds.). (2014). *Mining user generated content*. Boca Raton, FL: CRC Press.
- Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3), 436-465.
- Möhlmann, M. (2015). Collaborative consumption: determinants of satisfaction and the likelihood of using a sharing economy option again. *Journal of Consumer Behaviour*, 14(3), 193-207.
- Moraes, R., Valiati, J. F., & Neto, W. P. G. (2013). Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Systems with Applications*, 40(2), 621-633.
- Moutinho, L., Ballantyne, R., & Rate, S. (2011). Consumer behaviour in tourism. *Strategic management in tourism*, 2(2), p83-126.
- Mudambi, S. M., & Schuff, D. (2010). Research note: What makes a helpful online review? A study of customer reviews on Amazon. com. *MIS quarterly*, 185-200.
- Mukherjee, A., & Liu, B. (2012, July). Aspect extraction through semi-supervised modeling. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1* (pp. 339-348). Association for Computational Linguistics.
- Munezero, M., Montero, C. S., Mozgovoy, M., & Sutinen, E. (2013, November). Exploiting sentiment analysis to track emotions in students' learning diaries. In *Proceedings of the 13th Koli Calling International Conference on Computing Education Research* (pp. 145-152). ACM.
- Nasukawa, T., & Yi, J. (2003, October). Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the 2nd international conference on Knowledge capture* (pp. 70-77). ACM.
- Neuhofer, B., Buhalis, D., & Ladkin, A. (2014). A typology of technology-enhanced tourism experiences. *International Journal of Tourism Research*, 16(4), 340-350.
- Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903-2909*.
- Nonnecke, B., Andrews, D., & Preece, J. (2006). Non-public and public online community participation: Needs, attitudes and behavior. *Electronic Commerce Research*, 6(1), 7-20.
- Nyer, P. U. (1997). A study of the relationships between cognitive appraisals and consumption emotions. *Journal of the Academy of Marketing Science*, 25(4), 296-304.

- O'Connor, P. (2008). User-generated content and travel: A case study on Tripadvisor.com. *Information and communication technologies in tourism 2008*, 47-58.
- Ögüt, H., & Onur Taş, B. K. (2012). The influence of internet customer reviews on the online sales and prices in hotel industry. *The Service Industries Journal*, 32(2), 197-214.
- Oh, C., & Sheng, O. (2011, December). Investigating Predictive Power of Stock Micro Blog Sentiment in Forecasting Future Stock Price Directional Movement. In *Icis*(pp. 1-19).
- Oh, H. (1999). Service quality, customer satisfaction, and customer value: A holistic perspective. *International Journal of Hospitality Management*, 18(1), 67-82.
- Ortigosa, A., Martín, J. M., & Carro, R. M. (2014). Sentiment analysis in Facebook and its application to e-learning. *Computers in Human Behavior*, 31, 527-541.
- Oskam, J., & Boswijk, A. (2016). Airbnb: the future of networked hospitality businesses. *Journal of Tourism Futures*, 2(1), 22-42.
- Otto, J. E., & Ritchie, J. B. (1996). The service experience in tourism. *Tourism management*, 17(3), 165-174.
- Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In *LREc* (Vol. 10, No. 2010).
- Pan, B., & Fesenmaier, D. R. (2006). Online information search: vacation planning process. *Annals of Tourism Research*, 33(3), 809-832.
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1-2), 1-135.
- Pang, B., Lee, L., & Vaithyanathan, S. (2002, July). Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*(pp. 79-86). Association for Computational Linguistics.
- Pannala, N. U., Nawarathna, C. P., Jayakody, J. T. K., Rupasinghe, L., & Krishnadeva, K. (2016, December). Supervised learning-based approach to aspect based sentiment analysis. In *Computer and Information Technology (CIT), 2016 IEEE International Conference on* (pp. 662-666). IEEE.
- Pantelidis, I. S. (2010). Electronic meal experience: A content analysis of online restaurant comments. *Cornell Hospitality Quarterly*, 51(4), 483-491.
- Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). Servqual: A multiple-item scale for measuring consumer perc. *Journal of retailing*, 64(1), 12.
- Park, H., Xiang, Z., Josiam, B., & Kim, H. (2014). Personal profile information as cues of credibility in online travel reviews. *Anatolia*, 25(1), 13-23.

- Park, Y. A., & Gretzel, U. (2007). Success factors for destination marketing web sites: A qualitative meta-analysis. *Journal of travel research*, 46(1), 46-63.
- Passmore, L., Goodside, J., Hamel, L., Gonzalez, L., Silberstein, T. A. L. I., & Trimarchi, J. A. M. E. S. (2003). Assessing decision tree models for clinical in-vitro fertilization data. *Dept. of Computer Science and Statistics University of Rhode Island, Technical Report TR03-296*.
- Piatetsky-Shapiro, G. (1996). *Advances in knowledge discovery and data mining* (Vol. 21). U. M. Fayyad, P. Smyth, & R. Uthurusamy (Eds.). Menlo Park: AAAI press.
- Piatetsky-Shapiro, G. (1996). *Advances in knowledge discovery and data mining* (Vol. 21). U. M. Fayyad, P. Smyth, & R. Uthurusamy (Eds.). Menlo Park: AAAI press.
- Pine, B. J., & Gilmore, J. H. (2011). *The experience economy*. Harvard Business Press.
- Plank, A. (2016). The hidden risk in user-generated content: An investigation of ski tourers' revealed risk-taking behavior on an online outdoor sports platform. *Tourism Management*, 55, 289-296.
- Plutchik, R. (1980). *Emotion: A psychoevolutionary synthesis*. Harpercollins College Division.
- Plutchik, R. (1994). *The psychology and biology of emotion*. New York, NY, US: HarperCollins College Publishers.
- Prasad, K., Wirtz, P. W., & Yu, L. (2014). Measuring hotel guest satisfaction by using an online quality management system. *Journal of Hospitality Marketing & Management*, 23(4), 445-463.
- Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health information science and systems*, 2(1), 3-13.
- Rambocas, M., & Gama, J. (2013). *Marketing research: The role of sentiment analysis* (No. 489). Universidade do Porto, Faculdade de Economia do Porto.
- Ranchordás, S. (2015). Does sharing mean caring: Regulating innovation in the sharing economy. *Minn. JL Sci. & Tech.*, 16, 414-475.
- Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The effects of Twitter sentiment on stock price returns. *PloS one*, 10(9), 1-21.
- Ren, L., Qiu, H., Wang, P., & Lin, P. M. (2016). Exploring customer experience with budget hotels: Dimensionality and satisfaction. *International Journal of Hospitality Management*, 52, 13-23.

- Reza Jalilvand, M., & Samiei, N. (2012). The effect of electronic word of mouth on brand image and purchase intention: An empirical study in the automobile industry in Iran. *Marketing Intelligence & Planning*, 30(4), 460-476.
- Richardson, A. (2010). Using customer journey maps to improve customer experience. *Harvard Business Review*, 15(1), 2-5.
- Rifkin, J. (2014). *The zero marginal cost society: The internet of things, the collaborative commons, and the eclipse of capitalism*. New York, NY: St. Martin's Press.
- Robinson, R., Goh, T. T., & Zhang, R. (2012). Textual factors in online product reviews: a foundation for a more influential approach to opinion mining. *Electronic Commerce Research*, 12(3), 301-330.
- Rochet, J. C., & Tirole, J. (2004). *Defining two-sided markets*. mimeo, IDEI, Toulouse, France, January.
- Rong, J., Li, G., & Law, R. (2009). A contrast analysis of online hotel web service purchasers and browsers. *International Journal of Hospitality Management*, 28(3), 466-478.
- Rothkopf, E. (2014). CCTP-725: remix and dialogic culture. Retrieved September 20, 2017 from: <https://blogs.commonsgorgetown.edu/cctp-725-fall2014/2014/01/27/transitioning-to-post-postmodernism-a-look-at-twitterairbnb/>
- Rothkopf, E. Transitioning to Post-Postmodernism: A Look at Twitter & Airbnb. Retrieved March 4, 2018, from <https://blogs.commonsgorgetown.edu/cctp-725-fall2014/2014/01/27/transitioning-to-post-postmodernism-a-look-at-twitter-airbnb/>
- Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological bulletin*, 115(1), 102-141.
- Ryu, K., Han, H., & Jang, S. (2010). Relationships among hedonic and utilitarian values, satisfaction and behavioral intentions in the fast-casual restaurant industry. *International Journal of Contemporary Hospitality Management*, 22(3), 416-432.
- Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 3(3), 210-229.
- Savage, N. (2014). Big data versus the big C. *Nature*, 509(7502), 1-2.
- Savolainen, R. (2007). Information source horizons and source preferences of environmental activists: A social phenomenological approach. *Journal of the Association for Information Science and Technology*, 58(12), 1709-1719.
- Schaefer, T., Lawson, S. J., & Kukar-Kinney, M. (2016). How the burdens of ownership promote consumer usage of access-based services. *Marketing Letters*, 27(3), 569-577.

- Schifferstein, H. N., & Desmet, P. M. (2007). The effects of sensory impairments on product experience and personal well-being. *Ergonomics*, 50(12), 2026-2048.
- Schindler, R. M., & Bickart, B. (2005). Published word of mouth: Referable, consumer-generated information on the Internet. *Online consumer psychology: Understanding and influencing consumer behavior in the virtual world*, 32, 35-61.
- Schmunk, S., Höpken, W., Fuchs, M., & Lexhagen, M. (2013). Sentiment analysis: Extracting decision-relevant knowledge from UGC. In *Information and Communication Technologies in Tourism 2014* (pp. 253-265). Springer, Cham.
- Schor, J. (2016). Debating the sharing economy. *Journal of Self-Governance & Management Economics*, 4(3), 1-15.
- Schor, J. B., & Fitzmaurice, C. J. (2015). 26. Collaborating and connecting: the emergence of the sharing economy. *Handbook of research on sustainable consumption*, 410.
- Schor, J., Walker, E., Lee, C., Parigi, P., & Cook, K. (2015). Viewpoints on the sharing economy. *Contexts*, 14, 12-19.
- Schouten, K., & Frasincar, F. (2016). Survey on aspect-level sentiment analysis. *IEEE Transactions on Knowledge and Data Engineering*, 28(3), 813-830.
- Schuckert, M., Liu, X., & Law, R. (2016). Insights into suspicious online ratings: direct evidence from TripAdvisor. *Asia Pacific Journal of Tourism Research*, 21(3), 259-272.
- Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM computing surveys (CSUR)*, 34(1), 1-47.
- See-To, E. W., & Ho, K. K. (2014). Value co-creation and purchase intention in social network sites: The role of electronic Word-of-Mouth and trust—A theoretical analysis. *Computers in Human Behavior*, 31, 182-189.
- Senecal, S., & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of retailing*, 80(2), 159-169.
- Shi, H. X., & Li, X. J. (2011, July). A sentiment analysis model for hotel reviews based on supervised learning. In *Machine Learning and Cybernetics (ICMLC), 2011 International Conference on* (Vol. 3, pp. 950-954). IEEE.
- Sigala, M., Christou, E., & Gretzel, U. (Eds.). (2012). *Social media in travel, tourism and hospitality: Theory, practice and cases*. Farnham, United Kingdom: Ashgate Publishing, Ltd.
- Soderland, S. (1999). Learning information extraction rules for semi-structured and free text. *Machine learning*, 34(1-3), 233-272.

- Söderlund, M., & Rosengren, S. (2007). Receiving word-of-mouth from the service customer: An emotion-based effectiveness assessment. *Journal of retailing and consumer services*, 14(2), 123-136.
- Sotiriadis, M. D., & Van Zyl, C. (2013). Electronic word-of-mouth and online reviews in tourism services: the use of twitter by tourists. *Electronic Commerce Research*, 13(1), 103-124.
- Sparks, B. A., & Browning, V. (2011). The impact of online reviews on hotel booking intentions and perception of trust. *Tourism management*, 32(6), 1310-1323.
- Sparks, B. A., Perkins, H. E., & Buckley, R. (2013). Online travel reviews as persuasive communication: The effects of content type, source, and certification logos on consumer behavior. *Tourism Management*, 39, 1-9.
- Statistic Brain. (2017). *Internet Travel & Hotel Booking Statistics*. Retrieved from <https://www.statisticbrain.com/internet-travel-hotel-booking-statistics/>
- Stephany, A. (2015). *The business of sharing: Making it in the new sharing economy*. Springer.
- Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. *Handbook of latent semantic analysis*, 427(7), 424-440.
- Streitfeld, D. (2014). Airbnb listings mostly illegal, New York state contends. *International New York Times*.
- Stringam, B. B., & Gerdes Jr, J. (2010). An analysis of word-of-mouth ratings and guest comments of online hotel distribution sites. *Journal of Hospitality Marketing & Management*, 19(7), 773-796.
- Sundaram, D. S., & Webster, C. (1999). The role of brand familiarity on the impact of word-of-mouth communication on brand evaluations. *ACR North American Advances*.
- Sundararajan, A. (2014). Peer-to-peer businesses and the sharing (collaborative) economy: Overview, economic effects and regulatory issues. *Written testimony for the hearing titled The Power of Connection: Peer to Peer Businesses*, Retrieved March 4, 2018, from http://smallbusiness.house.gov/uploadedfiles/1-15-2014_revised_sundararajan_testimony.pdf
- Surowiecki, J. (2005). *The wisdom of crowds*. New York, NY: Anchor.
- Tan, A. (1999). Text mining: Promises and challenges. In *Proceedings south east Asia research computer confederation (SEARCC99)*, Singapore City, Singapore.
- Tang, T., Fang, E., & Wang, F. (2014). Is neutral really neutral? The effects of neutral user-generated content on product sales. *Journal of Marketing*, 78(4), 41-58.

- Tesnière, L. (1965). *Eléments de syntaxe structurale*. Editions Klincksieck.
- Teubner, T. (2014). Thoughts on the sharing economy. In *Proceedings of the International Conference on e-Commerce* (Vol. 11, pp. 322-326).
- Teubner, T., Adam, M. T., Camacho, S., & Hassanein, K. (2014). Understanding resource sharing in C2C platforms: the role of picture humanization. *ACIS*.
- Thelwall, M. (2001). A web crawler design for data mining. *Journal of Information Science*, 27(5), 319-325.
- Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social web. *Journal of the Association for Information Science and Technology*, 63(1), 163-173.
- Thet, T. T., Na, J. C., & Khoo, C. S. (2010). Aspect-based sentiment analysis of movie reviews on discussion boards. *Journal of information science*, 36(6), 823-848.
- Titov, I., & McDonald, R. (2008). A joint model of text and aspect ratings for sentiment summarization. *proceedings of ACL-08: HLT*, 308-316.
- Trivett, V., & Staff, S. (2013). *What the sharing economy means to the future of travel*. New York, NY: Skift.
- Tsao, W. C., Hsieh, M. T., Shih, L. W., & Lin, T. M. (2015). Compliance with eWOM: The influence of hotel reviews on booking intention from the perspective of consumer conformity. *International Journal of Hospitality Management*, 46, 99-111.
- Türkmen, H., Ekinci, E., & Omurca, S. İ. (2016, September). A Novel Method for Extracting Feature Opinion Pairs for Turkish. In *International Conference on Artificial Intelligence: Methodology, Systems, and Applications* (pp. 162-171). Springer, Cham.
- Verhoef, P. C., Lemon, K. N., Parasuraman, A., Roggeveen, A., Tsiros, M., & Schlesinger, L. A. (2009). Customer experience creation: Determinants, dynamics and management strategies. *Journal of retailing*, 85(1), 31-41.
- Vermeulen, I. E., & Seegers, D. (2009). Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism management*, 30(1), 123-127.
- Wallach, H. M. (2006, June). Topic modeling: beyond bag-of-words. In *Proceedings of the 23rd international conference on Machine learning* (pp. 977-984). ACM.
- Walls, A. R. (2013). A cross-sectional examination of hotel consumer experience and relative effects on consumer values. *International Journal of Hospitality Management*, 32, 179-192.

- Wang, D., & Nicolau, J. L. (2017). Price determinants of sharing economy-based accommodation rental: A study of listings from 33 cities on Airbnb.com. *International Journal of Hospitality Management*, 62, 120-131.
- Wang, H., Lu, Y., & Zhai, C. (2010, July). Latent aspect rating analysis on review text data: a rating regression approach. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 783-792). ACM.
- Wang, P. Y. A. (2013). # Irony or# Sarcasm—A Quantitative and Qualitative Study Based on Twitter. In *Proceedings of the 27th Pacific Asia Conference on Language, Information, and Computation (PACLIC 27)* (pp. 349-356).
- Werthner, H., & Ricci, F. (2004). E-commerce and tourism. *Communications of the ACM*, 47(12), 101-105.
- Westbrook, R. A. (1987). Product/consumption-based affective responses and postpurchase processes. *Journal of marketing research*, 258-270.
- Williamson, S., Wang, C., Heller, K., & Blei, D. (2010). The IBP compound Dirichlet process and its application to focused topic modeling.
- Wilson, T., Wiebe, J., & Hoffmann, P. (2005, October). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing* (pp. 347-354). Association for Computational Linguistics.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Burlington, MA: Morgan Kaufmann.
- Wixom, B., Ariyachandra, T., Douglas, D. E., Goul, M., Gupta, B., Iyer, L. S., ... & Turetken, O. (2014). The current state of business intelligence in academia: The arrival of big data. *CAIS*, 34, 1.
- Wunnava, S. (2015). Mobile commerce usage: Application of theory of reasoned action (TRA) and technology acceptance model (TAM). *World*, 5(2).
- Xiang, Z., Schwartz, Z., Gerdes Jr, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction?. *International Journal of Hospitality Management*, 44, 120-130.
- Yannopoulou, N., Moufahim, M., & Bian, X. (2013). User-generated brands and social media: Couchsurfing and AirBnb. *Contemporary Management Research*, 9(1), 85-90.
- Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, 28(1), 180-182.

- Ye, Q., Law, R., Gu, B., & Chen, W. (2011). The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. *Computers in Human behavior*, 27(2), 634-639.
- Ye, Q., Li, H., Wang, Z., & Law, R. (2014). The influence of hotel price on perceived service quality and value in e-tourism: An empirical investigation based on online traveler reviews. *Journal of Hospitality & Tourism Research*, 38(1), 23-39.
- Ye, Q., Zhang, Z., & Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert systems with applications*, 36(3), 6527-6535.
- Yeh, Y. H., & Choi, S. M. (2011). MINI-lovers, maxi-mouths: An investigation of antecedents to eWOM intention among brand community members. *Journal of Marketing Communications*, 17(3), 145-162.
- Yessenalina, A., Yue, Y., & Cardie, C. (2010, October). Multi-level structured models for document-level sentiment classification. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing* (pp. 1046-1056). Association for Computational Linguistics.
- Yi, J., Nasukawa, T., Bunescu, R., & Niblack, W. (2003, November). Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In *Data Mining, 2003. ICDM 2003. Third IEEE International Conference on* (pp. 427-434). IEEE.
- Yoo, K. H., & Gretzel, U. (2008). What motivates consumers to write online travel reviews?. *Information Technology & Tourism*, 10(4), 283-295.
- Zervas, G., Proserpio, D., & Byers, J. W. (2014). The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry. Boston U. School of Management Research Paper (2014). Retrieved March 4, 2018, from <http://people.bu.edu/zg/publications/airbnb.pdf>
- Zhang, C., Zeng, D., Li, J., Wang, F. Y., & Zuo, W. (2009). Sentiment analysis of Chinese documents: From sentence to document level. *Journal of the Association for Information Science and Technology*, 60(12), 2474-2487.
- Zhang, D., Mei, Q., & Zhai, C. (2010, July). Cross-lingual latent topic extraction. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* (pp. 1128-1137). Association for Computational Linguistics.
- Zhang, Z., Ye, Q., Law, R., & Li, Y. (2010). The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews. *International Journal of Hospitality Management*, 29(4), 694-700.

- Zhou, L., Ye, S., Pearce, P. L., & Wu, M. Y. (2014). Refreshing hotel satisfaction studies by reconfiguring customer review data. *International Journal of Hospitality Management*, 38(2014), 1-10.
- Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of marketing*, 74(2), 133-148.
- Zhuang, L., Jing, F., & Zhu, X. Y. (2006, November). Movie review mining and summarization. In *Proceedings of the 15th ACM international conference on Information and knowledge management* (pp. 43-50). ACM.