Extracting Value from Social Big Data: Empirical Studies on Online Customer Reviews and Managerial Responses

Wenjie Fan
Extracting Value from Social Big Data: Empirical Studies on Online Customer Reviews and Managerial Responses

Wenjie Fan
Supervising professor
Associate Professor Yong Liu, Aalto University School of Business, Finland

Supervising professor
Professor Virpi Kristiina Tuunainen, Aalto University School of Business, Finland

Preliminary examiners
Professor Reima Suomi, University of Turku, Finland
Professor Ilias Pappas, University of Agder, Norway

Opponents
Professor Reima Suomi, University of Turku, Finland
Professor Ilias Pappas, University of Agder, Norway

Aalto University publication series
DOCTORAL THESES 57/2022

© 2022 Wenjie Fan

ISBN 978-952-64-0776-0 (printed)
ISBN 978-952-64-0777-7 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Unigrafia Oy
Helsinki 2022

Finland

Done with the support of HSE Foundation, Jenny and Antti Wihuri Foundation, The Foundation for Economic Education, and The Marcus Wallenberg Foundation
Abstract

The advance of information technology has significantly digitalized our economy and activities and brought people into a big data era. The flourishing of social media has enabled massive-scale user-generated information sharing, which makes social big data available. In the context of e-commerce, consumers face a higher level of uncertainty and greater risk when purchasing online. As a result, many users utilize online customer review (OCR), as a novel Information Systems (IS) artifact, to alleviate their perceived uncertainty and the risks that hamper their online purchasing decisions. OCRs can influence consumer attitude and business performance, which drives companies to proactively intervene in the OCRs through the use of the managerial response (MR) function. An enormous amount of social big data in the form of OCRs and MRs has been accumulated online so far and presents both opportunities and challenges to researchers and stakeholders to use them as a rich source of business insights.

The objective of this dissertation is to offer new knowledge on the value of social big data in the form of OCRs and MRs for different stakeholders in the context of the tourism industry from three perspectives: the consumer perspective, the company perspective, and the industry perspective. Such new knowledge is derived from reflections on four previous papers that I had authored and co-authored.

From the angle of consumers, Paper 1 synthesizes literature on the helpfulness of OCRs and provides an integrated understanding of the determinants of OCR helpfulness through a systematic literature review. Paper 2 analyzes big data of OCRs in light of the attribution theory to investigate the impact of review content structures on OCR helpfulness and to demonstrate the important moderating effects of the reviewer reputation and review sentiments. Paper 3 focuses on the impacts of the MR function on company performance, specifically utilizing Kano’s theory of attractive quality to investigate the effect of dissipating benefits of IS service availability. This paper shows that whereas companies offering MRs gain constant advantages over those not employing the MR function, the ability of the MR function to improve business performance dissipates over time among the companies adopting it. Paper 4 quantifies the detrimental effect of air pollution on the revisit behaviors of foreign tourists by analyzing large volume of OCRs, which introduced a novel approach to examining collective consumer behavior using social big data from the industry perspective. This dissertation contributes to the scholarly discussion of social big data in the form of OCRs and MRs and concretely demonstrates the value of social big data to various stakeholders. In addition, this work can benefit IS researchers and stakeholders striving to exploit phenomena connected to IS artifacts and consumer behavior in the big data era.

Keywords Big Data, Social Media, Online Customer Review, Online Review Helpfulness, Review Content Structures, Managerial Response, Consumer Behavior, Business Performance
Acknowledgments

I am beyond lucky to have had the freedom and choice to work when I decided to apply my knowledge to solving real-life problems and to go back to school to study when I felt like I needed to expand my knowledge and pursue my PhD. The completion of this dissertation commemorates the end of my PhD journey at Aalto University. This unforgettable experience has been both intellectually stimulating and academically rewarding. Now, it is time for me to thank those who helped me along the way.

First of all, I would like to express my special gratitude to my supervisors, Associate Professor Yong Liu and Professor Virpi Tuunainen, for their excellent guidance throughout my research work. This dissertation would not have been possible without you. Thanks for the exceptional level of commitment and attention you have given me. I have benefited a lot from our collaboration. I also enjoyed very much our conversations, not only about research but also about life.

I also thank the preliminary examiners and opponents, Professor Reima Suomi and Professor Ilias Pappas. Their constructive comments have been extremely valuable and greatly helped me to improve the presentation of this dissertation. I am also honored to have both excellent scholars as my opponents for my public defense of this dissertation, which was an invaluable experience for me.

Next, I feel privileged to work as a researcher in a department with a friendly atmosphere and to work with immensely creative and capable people who are amazing both as colleagues and as friends: Yanqing, Yuting, Bikesh, Jiale, Venkata, Elena, Xun, Weimu, Yuan, Sheng, Timo, Merja, Johanna, Sampsa, and Yaping. I am especially grateful to Professor Virpi Tuunainen and Professor Matti Rossi for organizing all the courses and seminars that offered us opportunities to learn from and collaborate with renowned international scholars. I also want to sincerely thank my co-authors for their significant contributions to this dissertation, especially Professor Chee-Wee Tan, Professor Virpi Tuunainen, Associate Professor Yong Liu, and Associate Professor Hongxiu Li, with whom I have had the pleasure of co-authoring research papers, both for their insights as collaborators and their wisdom as friends. Working with you guys has been a tremendously enjoyable learning experience.

Over the years, I have also received a great deal of research and travel grants from various organizations that deserve sincere appreciation. I send my warm thanks to the HSE Foundation, the Foundation for Economic Education, the
Jenny and Antti Wihuri Foundation, the Marcus Wallenberg Foundation, and the Matti Lehti Fund for the generous financial support they have awarded to my dissertation project.

Last but not least, I wish to dedicate this dissertation to my family for always encouraging me to pursue my dreams and for all the love and support you have given me along the journey. I am grateful to my parents, Guanxin and She, for doing a commendable job in raising and educating me and providing me with enormous support and wisdom. Moreover, I greatly thank my wife, Wei, for accompanying me in this whole journey with unconditional love and support, as you had always done in our past journeys. Finally, thank you, Charlotte and Aina, my dearest daughters, for bringing new purpose to my life and for being the source of my impetus. Just as Bertrand Russell said, “The good life is one inspired by love and guided by knowledge,” thank you for bringing such magical ingredients into this journey and for sharing all this with me.

Espoo, 12 January 2022
Wenjie Fan
Contents

Acknowledgments ........................................................................................ i

Contents ...................................................................................................... iii

List of Abbreviations and Symbols............................................................. v

List of Publications.................................................................................... vii

Author’s Contribution.............................................................................. viii

Part 1: Summary .......................................................................................... 1

1. Introduction ............................................................................................. 3

   1.1 Objective and Research Questions .............................................. 7

   1.2 Thesis Structure ............................................................................ 8

2. Positioning the Study ........................................................................... 9

   2.1 Big Data in the Tourism Industry ................................................ 9

      2.1.1 User Data ................................................................................. 10

      2.1.2 Device Data ............................................................................... 11

      2.1.3 Transaction Data...................................................................... 13

   2.2 Online Customer Reviews .......................................................... 15

   2.3 Managerial Responses .................................................................17

      2.3.1 Effects of Managerial Responses on Customer Satisfaction 18

      2.3.2 Effects of Managerial Responses on Business Performance 20

3. Synthesizing Summary of the Papers ........................................... 23

   3.1 Consumer Perspective................................................................. 24


              Helpfulness: Synthesizing the Past to Guide Future Research ....24

      3.1.2 Paper 2. Quantifying the Effects of Online Review Content

              Structures on Hotel Review Helpfulness ...................................26

   3.2 Company Perspective .................................................................. 28

      3.2.1 Paper 3. The Timing Effect of IS Service Availability: The Case

                      of Managerial Response Service Usage in the Hospitality Industry ......28

   3.3 Industry Perspective...................................................................... 30

      3.3.1 Paper 4. Quantifying the Adverse Effect of Air Pollution on the

              Tourism Industry in China ..........................................................30
4. Discussion and Conclusions ......................................................... 33
  4.1 Theoretical Contributions....................................................... 34
  4.2 Practical Contributions............................................................ 38
  4.3 Limitations and Avenues for Future Research ......................... 39
  4.4 Conclusions......................................................................... 40

References ..................................................................................... 43

Part 2: Original Research Papers....................................................... 53
  Paper 1...................................................................................... 55
  Paper 2...................................................................................... 73
  Paper 3...................................................................................... 105
  Paper 4...................................................................................... 137
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BS</td>
<td>base station</td>
</tr>
<tr>
<td>DID</td>
<td>difference-in-differences analysis</td>
</tr>
<tr>
<td>ECT</td>
<td>expectation confirmation theory</td>
</tr>
<tr>
<td>e-commerce</td>
<td>electronic commerce</td>
</tr>
<tr>
<td>e-WOM</td>
<td>electronic word of mouth</td>
</tr>
<tr>
<td>GPS</td>
<td>global positioning system</td>
</tr>
<tr>
<td>IS</td>
<td>information systems</td>
</tr>
<tr>
<td>MR</td>
<td>managerial response</td>
</tr>
<tr>
<td>MRF</td>
<td>managerial response function</td>
</tr>
<tr>
<td>NLP</td>
<td>natural language processing</td>
</tr>
<tr>
<td>OCR</td>
<td>online customer review</td>
</tr>
<tr>
<td>PSM</td>
<td>propensity score matching</td>
</tr>
<tr>
<td>RevPAR</td>
<td>revenue per available room</td>
</tr>
<tr>
<td>RFID</td>
<td>radio-frequency identification</td>
</tr>
<tr>
<td>SEM</td>
<td>search engine marketing</td>
</tr>
<tr>
<td>UGC</td>
<td>user-generated content</td>
</tr>
<tr>
<td>WOM</td>
<td>word of mouth</td>
</tr>
</tbody>
</table>
List of Publications

This doctoral dissertation consists of a summary and of the following research papers which are referred to in the text by their numerals.


Author’s Contribution

**Paper 1:** What Makes Consumer Perception of Online Review Helpfulness: Synthesizing the Past to Guide Future Research

The author of this dissertation is the sole author of this publication.

**Paper 2:** Quantifying the Effects of Online Review Content Structures on Hotel Review Helpfulness

The author of this dissertation is the first author of this publication and contributed to developing the research design, reviewing the literature, processing the data, conducting the modeling, analyzing and discussing the findings, and composing the paper.

**Paper 3:** The Timing Effect of IS Service Availability: The Case of Managerial Response Service Usage in the Hospitality Industry

The author of this dissertation is the first author of this paper. He and the supervisor developed the research design. He was responsible for the literature review, the data collection and processing, and the data analysis. Under the guidance of the supervisors, he also visualized the results and was responsible for the writing of the paper.

**Paper 4:** Big Data for Big Insights: Quantifying the Adverse Effect of Air Pollution on the Tourism Industry in China

The author of this dissertation is the first author of this publication. He and the supervisor developed the research design. He contributed to reviewing the literature, collecting and processing the data, running the data analysis, analyzing and discussing the findings, and composing the paper.
Part 1: Summary
1. Introduction

Science and technology have advanced by leaps and bounds in recent decades. Information technologies are among the “mega trends” that have significantly transformed our daily lives. The Internet now connects almost everyone in the world and has accelerated information sharing and knowledge accumulation. Massive-scale structured, semi-structured, and unstructured data are generated, collected, and accumulated, forming the “big data” that have opened up a brand-new era (Kambatla et al. 2014).

With the help of digital information flowing at the speed of light down to the network cables, mankind has created the largest global market in human history. From two decades ago until now, electronic commerce (e-commerce) has continued to play an increasingly prominent role in our daily lives. Almost everyone in the world can join this free market, thus allowing us to reach the highest level of abundance in history. This phenomenon has proven the theory that social division of labor and free exchange can create more social wealth (Smith 1977). Because of the amplifying effect of human knowledge accumulation on the benefits of global division of labor and exchange, the integration of modern technology and the free market has led to the long-term compound growth of the modern economy (Li 2020). The modern technology evolution has become a change agent in the ways we conduct business and social interaction, which brings both opportunities and challenges.

Online shopping has become ubiquitous as it bridges the temporal and spatial constraints between merchants and consumers, making it more convenient than in-store shopping (Luo et al. 2012). Moreover, costs are lower not only because of the minimal overhead costs but also because of the economies of scale that can be achieved on the Internet. Recently, the coronavirus pandemic has made online shopping grow like never before due to the lockdowns and physical constraints. Thus, the rationale behind embracing online services is no longer how cool the technology is but its necessity. However, the temporal and spatial separation on the Internet make it difficult for consumers to experience products or services prior to purchasing or even consuming them. Therefore, consumers face a higher level of uncertainty and greater risk when shopping online than when shopping at brick-and-mortar stores (Hong et al. 2017).

A solution to these challenges has emerged with the development of social media. These Internet channels for consumers to voice their opinions have enabled the utilization of user-generated content (UGC), such as online customer re-
views (OCR), to alleviate the uncertainty and risks that hamper online purchasing decisions. With consumers as primary sources of diversified information on these channels, social media have turned into a rich mine of the so-called “social big data,” which can provide relevant and even sensitive knowledge to social media users or companies (Bello-Orgaz et al. 2016).

OCR, often viewed as “electronic word of mouth” (e-WOM), can impact a company’s sales performance (Berger et al. 2010; Chevalier and Mayzlin 2006; Liu 2006) by increasing consumer awareness of products or services (Godes and Mayzlin 2004; Liu 2006) or influencing consumer attitudes towards them (Wang et al. 2015). Hence, many businesses are using OCR platforms as new channels for marketing (Chen and Xie 2008) and customer engagement (Chen et al. 2019). OCRs are seen as valuable information-based assets (Hlee et al. 2018). They function as novel information system (IS) artifacts based on social media, and their characteristics differ from those of traditional IT artifacts developed and used in organizations. A social media technology is “an ensemble IS artefact composed of technical, informational, and relational subsystems that interact distinctly according to the context of use” (Wakefield and Wakefield 2016, p. 140). Rather than being controlled by the organization that developed and provided the OCR functionality, OCR contents are created by all kinds of users and made instantly publicly available. Unlike surveys, OCRs, in the form of ratings and text comments, reflect the true perceptions of consumers at a massive scale. OCR functions can be used by consumers and companies to communicate about products and services and may influence customers’ purchase decisions, companies’ strategies, and even industry development. Therefore, by leveraging the sheer power of OCRs archived at social media sites, insights on various levels of stakeholders, from individual consumers to companies and even a whole industry, can be generated.

Noticeably, the production and consumption of OCRs are influenced, either directly or indirectly, by multiple stakeholders—specifically, reviewers, review readers, companies, platforms, and other customers (Figure 1). For instance, reviewers influence review readers by sharing the e-WOM message and information about their own characteristics (e.g., identity-relevant information, demographics, and preferences). Review readers’ level of motivation and characteristics (e.g., experience and personality) can impact their processing and evaluation of OCRs. Platforms can manage the review presentation (e.g., review order, summary statistics), provide reviewers options for personal information disclosure, and design various functionalities to other stakeholders. Other customers can vote on the helpfulness of OCRs. Companies can use marketing mix to alter the impact of OCRs and may even offer incentives to stimulate OCRs.

Moreover, the high persuasiveness and the broad accessibility of OCRs have also led to the inception of the managerial response function (MRF). This IS artifact was introduced as a consequence of the prevalence of OCRs. Companies are eager to proactively manage this form of customer-to-customer communication (Reimer and Benkenstein 2016). Therefore, MRF appears to be an OCR-related IS artifact, which is a natural response to the use of OCRs on the enter-
prise side. Managerial response (MR) is expected to enhance company relationships with their satisfied customers while reducing the possible damages of negative customer reviews (Willemsen et al. 2013; Xie et al. 2016). Meanwhile, potential customers may also regard companies that respond to OCRs as benevolent and trustworthy. MRF can help customers to better estimate the validity and consequence of OCRs. Hence, MRF may influence consumers’ decision-making and companies’ business performance.

Figure 1. Key stakeholders involved in the online customer review exchange.

A closer look at social big data in the form of OCRs and MRF allows their use as rich sources of information and insights. The role and use of OCRs in consumer decision-making and in business operations and management have been widely discussed in IS research. The usefulness of MRF has attracted increasing attention. However, scholars are still far from reaching unanimous conclusions on these important topics. Due to the fact that each stakeholder may have their own interests in the OCR exchange (Kozinets et al. 2010), it is worth investigating the value of OCRs from different stakeholders’ perspectives, as I will discuss below.

From the consumer perspective, they are increasingly relying on OCRs to support their purchase decisions, making the review system an important feature of online platforms. Although the availability of abundant OCRs provides more information to customers to improve their purchase decisions, the huge number of OCRs may also cause information overload, which would make customers feel too exhausted to identify the most meaningful reviews in support of their decision-making process. In the meantime, both the anonymous nature of OCRs and the conflicting opinions expressed hamper consumers’ full appreciation of OCRs. Therefore, OCR helpfulness plays an important role in identifying the
most useful reviews for supporting decision-making, but many studies on online review helpfulness have presented inconsistent findings on various determinants of the helpfulness of OCRs. This is the first research gap that this dissertation aims to address.

From the company perspective, OCRs have posed challenges to modern enterprises in interacting with their consumers due to consumers’ increasing reliance on OCRs for brand evaluation and purchase decision-making and due to the difficulty for companies of controlling the message that consumers receive about their brands. To manage companies’ online reputation, it is imperative for them to have a method of directly intervening in OCRs. Therefore, OCR platforms provide companies with MRF to let them proactively engage with their customers. However, although the business cases for online reputation management have been increasingly scrutinized in recent years, it is unclear whether the use of MRF helps improve companies’ business performance. This is the second knowledge gap to be addressed in this dissertation.

From the industry perspective, OCRs offer a useful instrument for understanding the performance of different service sectors and the factors that affect their performance. OCRs could serve as proxies of collective customer behavior, such that an analysis of all the reviews of an industry may offer a way to understand a change in customer behavior in the industry at the macro level. Novel approaches that tap into social big data and extract business values provide possibilities to efficiently answer questions confounded by complex factors. Taking the tourism industry as an example, a good environment with favorable meteorological conditions is critical for its development. While the prevalence of air pollution over recent years has afflicted many tourism destinations, recent statistics surprisingly seemed to violate the long-standing view that air pollution undermines inbound tourism (Becken et al. 2017). According to international tourism statistics, countries with the most severe air pollution in recent years received an increasing amount of inbound tourists (World Bank 2019). This implies a possibly spurious association of air pollution to the tourism market. In this vein, OCRs offer a possible alternative instrument for observing the collective behavior of large consumer groups and to understand various phenomena at the industry level. This dissertation also strives to fill this research void by demonstrating the effectiveness of OCRs in capturing tourists’ collective travel behavior.

In this dissertation, new information technologies, such as big data analytics and natural language processing (NLP), are used to more deeply comprehend unstructured social big data and to gain in-depth insights on the usefulness of the adoption of new service features. Meanwhile, the availability of social big data generated by a slew of users offers opportunities to analyze phenomena from higher dimensions. Hence, this dissertation addresses the abovementioned knowledge gaps, reconciles contradictory findings in literature, and develops new insights on multilevel stakeholders by leveraging the potential value of social big data focusing on OCRs and MRFs.
1.1 Objective and Research Questions

The objective of this dissertation is to provide new knowledge for understanding the value of social big data in the form of OCRs and MRFs from three perspectives: the consumer perspective, the company perspective, and the industry perspective. In general, this dissertation aims to explore the following overarching research question:

**RQ. How can social big data benefit stakeholders in the hospitality industry?**

![Figure 2. Overview of this dissertation.](image)

To address this question, I investigated the usage and effect of such IS artifacts as OCRs and MRFs, focusing on better understanding of how technologies come to be used to benefit various stakeholders in particular ways. Specifically, I break down this question and discuss it at the consumer level, the company level, and the industry level (Figure 2).

From the viewpoint of consumers, I reviewed literature on the helpfulness of online reviews and investigated the effects of review content structures on OCR helpfulness. I focused on two sub-research questions:

**SubRQ1. What is the role of review content structures in shaping OCR helpfulness?**

**SubRQ2. How do reviewer reputation and review sentiment moderate the effects of review content structures on OCR helpfulness?**

In this dissertation, I address these sub-questions in Paper 1 and Paper 2. In Paper 1, I systematically reviewed literature on OCR helpfulness to summarize the determinants of OCR helpfulness. This study deciphered both consistent and conflicting results and discussed the reasons for these mixed findings to point out the direction of the subsequent study. In Paper 2, I quantified the impact of “review sidedness,” “information factuality,” and “emotional intensity at the beginning of a review” on OCR helpfulness to investigate the effects of review content structures on OCR helpfulness.
From the perspective of companies, I reviewed literature on the business value of adopting MRFs and explored the timing effect of the IS service availability on the business performance of companies. Specifically, the study focused on two sub-research questions:

**SubRQ3.** Can the use of MRF for OCRs, as a new service feature, enhance the business performance of companies?

**SubRQ4.** Can the timing of MRF adoption by a company affect the impact of MRF on business performance?

In this dissertation, Paper 3 addresses these sub-questions. Specifically, I investigated the effectiveness and extent to which MRF can improve business performance. By examining the extent to which the offering of a new IS service benefits later companies when the IS service is increasingly implemented by rival companies and becomes a standard feature for consumers, this study also attested to the effect of dissipating benefits of IS service availability.

From the perspective of a whole industry, I analyzed big data of OCRs pertaining to actual travel behavior in a non-laboratory setting to examine the impact of environmental factors on the tourism industry. The study focused on the following sub-research question:

**SubRQ5.** Can OCRs be used as instruments for understanding tourist behavior as affected by air pollution?

In this dissertation, this sub-question is addressed in Paper 4. This study quantified the detrimental effect of air pollution on the revisit behavior of foreign tourists by analyzing a large volume of OCRs, which introduced a novel approach to examining collective consumer behavior using social big data.

### 1.2 Thesis Structure

This dissertation is divided into two main parts. The first part provides a summary of the research conducted to address the research questions outlined previously. The corresponding structure is as follows. I will first position this study within the IS discipline in Chapter 2 by reviewing relevant literature to identify knowledge gaps in them. Next, I will summarize the four papers in Chapter 3, present the research methodology, and discuss the findings according to the specific research questions. Finally, in Chapter 4, I will elaborate on the implications for theory and practice, sum up the contributions of this dissertation, and present its conclusions and limitations.

The second part of this dissertation consists of the four original research papers that provide full accounts of the research projects conducted for this dissertation.
In this chapter, I position my dissertation in relation to extant literature to illustrate how my studies can advance the existing research on realizing value from social big data. First, I systematically review the literature on various types of big data within the tourism context. Then, I briefly review current literature on OCRs and factors of OCR helpfulness. Finally, I present a comprehensive and representative analysis of the literature on MRs to understand the benefit of its adoption.

2.1 Big Data in the Tourism Industry

The advance of information technologies is bringing about a rapidly expanding ecosystem of diverse sources of massive-scale data. Massive datasets in structured, semi-structured, and unstructured formats are generated, recorded, and accumulated, which form big data (Kambatla et al. 2014). All these big datasets are diamonds in the rough, with the potential to provide invaluable insights when organized and analyzed appropriately (Kambatla et al. 2014). In conjunction with conceptual and technological innovations, big data have been adopted in a variety of domains, including, but not limited to, science, engineering, finance, business, healthcare, and tourism (Hashem et al. 2015).

Even at an early stage, scholars have applied an assortment of big data to tourism and hospitality research that have yielded invaluable insights (Li et al. 2018). For instance, big data could provide a novel approach to understanding tourist demand by overcoming the issues of the limited sample size and the lack of representativeness of survey data (Yang et al. 2015). Big data analytics also expands understanding of consumer behavior by both academia and industries based on sufficient data that are not susceptible to sampling bias (Li et al. 2017). Moreover, big data analytics offers numerous opportunities to develop new knowledge of the tourism field, which could reshape the understanding of the tourism industry and facilitate the corresponding sense-making and decision-making (Xiang et al. 2015). With these advantages, big data have allowed an enhanced understanding of tourism demand, tourist behavior, tourist experience and satisfaction, and so forth (Li et al. 2018).

The diverse big data that have been applied to tourism research were mainly derived from three categories of data sources: users, devices, and operations (Li et al. 2018). UGC is the most popular data source because of the data availability.
In contrast, the transaction data generated in operations are relatively less applied to tourism research. The possible reason for researchers’ wider use of UGC data is the low cost of such data and their easy accessibility online, whereas the lower adoption of transaction data is mainly ascribed to the fact that majority of such data are private information controlled by organizations or government sectors (Li et al. 2018).

2.1.1 User Data

The tremendous growth of social media and UGC on the Internet has dramatically changed the tourism sector. With the bountiful online textual data and online photo data actively shared by social media users, researchers are using big data analytics to investigate and solve real-life problems.

Online textual data constitute the largest proportion of UGC data applied to tourism research. There are mainly two types of online textual data: review data and blog data (Li et al. 2018). OCR data reflect consumer attitudes toward tourism products. Hotels, restaurants, and attractions are the reviewed targets that are mostly focused on. Popular OCR data sources are TripAdvisor (Fang et al. 2016), Expedia (Xiang et al. 2015), Yelp (Park and Nicolau 2015), Booking (Xu and Li 2016), Dianping (Zhang et al. 2010), and Ctrip (Ye et al. 2009).

The satisfaction of customers had been the focus of studies that used OCR data. For instance, dimensions of tourist satisfaction had been identified from online hotel reviews (Guo et al. 2017); and using these dimensions, new insights had been offered on the heterogeneity of consumer perceptions across different demographic segments. The advantages of a large amount of OCRs had been leveraged and had improved understanding of the determinants of the satisfaction of hotel guests by discriminating among the languages in which the review comments were written (Liu et al. 2017). Also explored were the relationship between customer satisfaction and other relevant factors, such as service performance (Slevitch and Oh 2010), guest experience (Xiang et al. 2015) and competitive position (Crotts et al. 2009). The impact of review ratings on perceived usefulness and enjoyment had been assessed by analyzing OCRs of restaurants (Park and Nicolau 2015). Empirical application of a sample of online attraction reviews demonstrated that both the review readability and the reviewer characteristics affect the perceived value of OCRs (Fang et al. 2016). The corresponding research with OCR data can provide insightful implications for customers and practitioners.

Since OCRs pose an unprecedented challenge for companies to interact with their customers, many well-known online review platforms, such as TripAdvisor, Yelp and Ctrip, have implemented MRF for companies to get intervention in OCRs. Recently, MRs to OCRs have attracted extensive attention from researchers. A number of scholars analyzed MR data and argued for a positive impact of MRF use on brand reputation and business performance (Chen et al. 2019; Proserpio and Zervas 2017; Xie et al. 2016). In the meantime, some researchers warned of a possible negative effect (Mauri and Minazzi 2013; Xie et al. 2014). Moreover, the outcomes of MRF adoption were also found contingent on the conditions of OCRs (valence and volume) and on the MR strategy (Lee et al. 2018).
Therefore, the effectiveness of MR in various situations remains unclear and deserves further exploration.

As for blogs, they are commonly used to record travel stories and tourist feelings. Scholars mainly rely on blog data to analyze tourist sentiments and tourism recommendations. For instance, sentiment analysis has been applied to Twitter data to build a cost-effective application that captures the attitudes of hospitality customers in real time (Philander and Zhong 2016). A novel approach to extracting and summarizing popular information from massive tourism blog data and identifying hot locations had been proposed (Xu et al. 2015). Travel blog information had been explored to extract popular tourist locations and travel trajectories for better travel scheduling (Yuan et al. 2016).

Social media platforms allow users to post not only textual data but also photos. Popular photo-sharing platforms, namely, Flickr, Instagram, and Panoramio, have become the prime sources of the photo data applied to tourism research (Li et al. 2018). Online photo data usually convey additional metadata such as temporal information (e.g., the date taken and the date posted) and geographical information (e.g., the location and the latitude and longitude). Consequently, the online photo data published by tourists, which contain valuable information corresponding to the user, time, and location, provide a new avenue for investigating tourist behavior (Lu et al. 2017; Vu et al. 2015) and refine tourism recommendations (Kurashima et al. 2013; Lee et al. 2014) and marketing strategies (Deng and Li 2018).

2.1.2 Device Data

With the unprecedented advancement in the implementation of various tracking technologies, diverse devices have been developed and used to collect high-resolution data on tourist mobility, which incorporate mobile technologies ranging from the Global Positioning System (GPS) and the mobile base station (BS) to Wi-Fi and Bluetooth, as well as radio-frequency identification (RFID) technologies (Shoval and Ahas 2016). In parallel, thanks to the continuous advances in our computing abilities and particularly in the development of geographic information systems (Richardson et al. 2013), researchers see growing possibilities for mining the enormous spatial-temporal data produced by these tracking technologies.

GPS, a globally available system with sufficient accuracy, has been extensively applied in tourism research. Tourist behavior has been the main focus, ranging from spatial behavior and temporal behavior to combined spatial-temporal behavior (Li et al. 2018). Regarding spatial behavior, a study by Edwards and Griffin (2013) explored tourists’ movement patterns in cities by analyzing GPS tracking data. Moreover, a large group of researchers incline towards exploring spatial-temporal behavior by combining spatial data with temporal data. To elucidate the importance of temporal perspectives in understanding tourist activity, GPS has been used to track and record the spatial-temporal trajectories of theme park visitors to demonstrate temporal mass behavior patterns (Birenboim et al. 2013). In terms of tourism recommendations, GPS tracking
data can be used to predict the next destination of individual tourists (Zheng et al. 2017) and to propose a smart itinerary (Yoon et al. 2010).

Although GPS accounts for the largest volume of device data applied to tourism studies, other device data sources are also essential for tourism research. Nonetheless, regarding the data collected using mobile BS, RFID, Bluetooth, and Wi-Fi, the application of these types of data to tourism research has been far from sufficient in terms of scantily related studies and a limited research area (Li et al. 2018).

Mobile BSs, when communicating with mobile phones, automatically record logs in the memory files and databases of mobile network operators (Raun et al. 2016). Compared with GPS data, mobile positioning data retrieved from BSs show different data characteristics, such as cheaper data collection but lower spatial precision (Ahas et al. 2008). However, due to the restricted access to mobile positioning data by virtue of business secrets and concerns with privacy and surveillance, such data have not been widely applied to tourism research thus far (Ahas et al. 2008; Li et al. 2018). In the extant tourism literature, mobile positioning data have been used to analyze tourist flows (Raun et al. 2016), travel distances (Nilbe et al. 2014), space consumption (Ahas et al. 2007), destination loyalty (Ahas et al. 2007), and repeat visits (Kuusik et al. 2011) among other tourist behaviors. Recently, a methodological framework for generating cross-border tourism statistics was proposed using mobile roaming data for better planning and management of tourism and for transport optimization and tourism effects management (Saluveer et al. 2020).

Bluetooth data are collected by Bluetooth sensors preplaced in the target area. These sensors can detect other Bluetooth-enabled personal carry-on devices, which detect unannounced movements of massive individuals to monitor their positions and trajectories in a non-participatory manner (Li et al. 2018). However, due to the delimited monitoring range of Bluetooth, most studies that depended on Bluetooth data were confined to certain indoor places or planned event-tourism activities. For instance, a group of scholars applied Bluetooth tracking data obtained in one or several indoor places, such as a tribune or a shopping center (Stange et al. 2011), a museum (Yoshimura et al. 2014), and an attraction that consisted of a cathedral, a church, and an indoor market (Versichele et al. 2014). It is worth noting that Bluetooth circumvents the disadvantage of GPS technology of its inapplicability to indoor contexts. On the other hand, Bluetooth scanners have been used to analyze the spatiotemporal dynamics of human movement within and around the festival sites at a cultural and theater festivity event (Versichele et al. 2012). The behavioral patterns of visitors tracked by Bluetooth at a trade fair have been examined by analyzing spatiotemporal sequences (Delafontaine et al. 2012).

Wi-Fi, as an alternative to Bluetooth that allows involuntary tracking of tourist movements, is considered more convenient than Bluetooth due to its low cost and, more importantly, unobtrusive virtue (Bonné et al. 2013). Despite these advantages, however, studies that use Wi-Fi data are still relatively few. An involuntary Wi-Fi tracking system for capturing tourist behavior at mass events has been proposed and implemented by Bonne et al. (2013), which is deemed
the first attempt to introduce Wi-Fi tracking data to tourism research (Li et al. 2018). Techniques for cleaning up sets of raw Wi-Fi detection data into sets that can subsequently be used for crowd analytics have been explored by placing Wi-Fi scanners in the city center during a festival and gathering crowd localization data (Chilipirea et al. 2016). Nonetheless, Wi-Fi data have increasing prospects for tourism research as more people use smartphones.

The RFID system comprises a reader, antennas, and tags. Data are transmitted through radio waves from the tags, that are held by men or attached to objects, to the antenna-reader combination. In the tourism and hospitality contexts, the tags can be attached to, or incorporated into, mobile terminals (Wan 2009), tourist cards (Della Lucia 2013), loyalty cards (Capizzi and Ferguson 2005), casino chips (Gellatly 2005), and electronic passports, public transport cards, luggage, hotel items, and the like (Öztayşi et al. 2009). With an overview of the broad range of means of utilizing RFID as a tool for improving tourism processes in scenarios of hotels, cruise ships, resorts, and theme parks, the application of RFID data in tourism for better operations and marketing has been investigated (Hozak 2012). The possibility that RFID is being applied in the hospitality industry to improve service quality and customer satisfaction and loyalty, expand market share, and enhance profitability has been discussed (Öztayşi et al. 2009). Specifically, a lightweight RFID-based action tracking framework that gathers data on places visited by tourists during an event has been designed (Zeni et al. 2009). In addition, information collected with RFID technology is used to provide personalized tourism recommendations (Tsai and Chung 2012; Wan 2009) and guidance on the given location and the surrounding objects (Deka et al. 2016), and for tourism management and support (Prusty et al. 2020).

2.1.3 Transaction Data

Transaction data comprise a wide range of records concerning tourism-related operations and activities, such as web searches, website visits, bookings, card transactions, and sales. As a valuable type of big data for tourism research, transaction data are extensively used to finetune tourism prediction, to promote search engine marketing (SEM), and to understand tourist behavior (Li et al. 2018).

Web search data, which represent users’ information needs and searching behavior on the Internet, have become the most prevalent genres of transaction data. Given the increasing significance of information search in online travel planning, web search data have attracted considerable attention in tourism research. Ample studies have explored the possibilities of utilizing web search data as tools for tourism demand forecasting at the country and region levels (Bangwayo-Skeete and Skeete 2015; Park et al. 2017), the city level (Gunter and Önder 2016; Li et al. 2017), and the attraction level (Höpken et al. 2019; Huang et al. 2017 a), and have achieved excellent performance. In addition, SEM has become an important strategic tool for online tourism marketing, which can help businesses, organizations, and tourist destinations to gain visibility online
A successful SEM strategy demands a thorough comprehension of the dynamics of SEM in tourism. The literature on search engine use for travel-related purposes has been synthesized and a conceptual model that elucidates the evolving dynamics of SEM has been proposed (Pan et al. 2011). To gain insights on the behavioral aspect of the use of search engines, the patterns of the construction of travel queries have been investigated, and the commonalities and differences in travel queries across tourist destinations have been analyzed, which, in turn, offer perceptive implications of SEM for tourism destinations (Xiang and Pan 2011).

Apart from web search data, most transaction data are in the possession of tourism companies and organizations (e.g., travel agencies, hotels, and attraction operators) as well as government agencies (e.g., bureaus of statistics and tourism administrations). Even though few studies had managed to collect and utilize such data, they had contributed insightful findings to literature. For instance, the collection of webpage visiting data using a web analyzer program revealed that the website performance and potential online marketing effectiveness depends on their traffic source, namely, direct visits, referring site visits, and search engine visits (Plaza 2011). The findings indicate that direct visits are the most effective in generating return visits and nurturing visit duration, followed by search engine visits and then link-entries. Based on online booking data for several major Kyoto hotels offered by the National Institute of Informatics, the choice behavior of visitors was examined, the optimal room charge was obtained, and the expected sales were inferred (Saito et al. 2016). With a transaction data set of US hotel reservations that included sales prices and volumes over a three-month period, combined with a variety of UGC data, the economic values of diverse location and service characteristics of hotels were estimated, and a novel ranking system for hotel search was proposed based on the computation of the expected utility gain from each hotel (Ghose et al. 2012).

Using attraction sales data, the impact of daily weather variations on downhill ski lift ticket sales at ski resorts was quantified, which contributed to the modeling of the implications of climate change for this activity and for the industry sector (Shih et al. 2009). The impact of climate variability on winter tourism was examined by analyzing the association of the tourism demand to snow accumulation based on detailed panel data that included hotel stays at 28 Austrian ski resorts for a period of 20 years (Falk 2010). Moreover, an approach to using big data of hotel consumption (e.g., electricity and water consumption) to achieve the joint goals of reducing operating expenditure and sustainability has been proposed (Kahn and Liu 2016). By mining the express highway data acquired from a provincial communication department, the carbon emissions of self-driving tourism were gauged, and their spatial relationship with scenic spots was analyzed (Huang et al. 2017 b). It was argued that carbon emission flows from self-driving tours are associated with the development level of economies but are irrelevant to the grades of the scenic spots.

Last but not least, consumer card data captured and recorded as tourists making purchases had also attracted researchers’ attention. Consumer cards such as credit cards, reward cards, and payment cards have been explored as the means
by which corporations collect data from consumers and comprehend consumption-related behaviors and experiences (Weaver 2008). Moreover, the increasing availability of big data offers people novel opportunities to observe and understand social phenomena. For example, with the use of big data from bank card transactions as proxy for tourist mobility patterns for the construction of mobility networks, the impact of the nationality of tourists on their mobility behavior has been quantitatively examined (Sobolevsky et al. 2014). This way of constructing mobility patterns using big data from transactional records is transposable to various other datasets.

In general, despite the extensive development and wide application of mobile communication technologies and of the RFID, Bluetooth, and Wi-Fi technologies, the potential of big data to enhance tourism research has not been sufficiently exploited yet. Similarly, transaction data, although informative and with their respective advantages, have failed to sufficiently contribute to tourism research, in terms of the scarce related studies and narrow research area. This may be because some kinds of data are mainly in the possession of organizations or government departments, which makes them difficult to access due to business competition and privacy concerns (Li et al. 2018). Thus, UGC remains the dominant data source for tourism research, thanks to its easy access and public availability, as well as its low privacy concern. This dissertation relies on UGC to address the earlier mentioned research questions by gaining a deeper understanding of a huge number of OCRs and exploring a novel approach to appreciating social big data. Particularly, in Paper 4, user-generated big social data were used to investigate travelers’ collective revisit behavior.

2.2 Online Customer Reviews

OCRs are increasingly being relied upon by consumers as a well-established reputation mechanism for encouraging trust in online markets with asymmetric information (Resnick and Zeckhauser 2002). Meanwhile, OCRs have also received substantial attention from both the business and academic communities.

On the one hand, industry research reports have demonstrated that consumers, when making purchase decisions online, trust OCRs posted by unknown former customers more than they trust traditional media (Nielsen 2010). On the other hand, scholars have found that UGC in the form of OCRs significantly influences online consumers’ purchase decisions by decreasing the associated uncertainty and risks (Filieri and McLeay 2014; Hong et al. 2017; Hu et al. 2008; Mudambi and Schuff 2010). Given the importance of OCRs, ample studies have tried to explore their impact and usefulness in various contexts.

A first set of papers largely explored the value of OCRs to companies and organizations across various industries by regarding OCRs as given and focusing on the association between OCRs and sales performance (Chevalier and Mayzlin 2006; Clemons et al. 2006; Dellarocas et al. 2004). For instance, book reviews from two leading online booksellers have been analyzed to examine the effect of OCRs on sales patterns of firms (Chevalier and Mayzlin 2006). The use of OCRs to evaluate the effectiveness of product differentiation strategies has also been
investigated (Clemons et al. 2006). In the study, the OCRs and sales data in the craft beer industry were analyzed to demonstrate the effects of OCRs on the success of new product launches. The reason for this is that OCRs influence consumer opinion and purchase intentions (Lee et al. 2008, 2011), which indicates the prevalence and wide acceptance of OCRs (Gao et al. 2006). Rather than relying merely on the aggregated summary statistics like average star ranking, consumers indeed read and respond to written reviews (Chevalier and Mayzlin 2006). The sales of digital microproducts at low and uniform prices have also been studied (Amblee and Bui 2011). This empirical study suggested that social discussion in the form of OCRs acts as a collective signal of the reputations of products and brands, which ultimately plays a significant role in customers’ buying decisions. Moreover, many researchers are devoted to forecasting companies’ sales or revenue by leveraging OCRs (Chong et al. 2016; Dellarocas et al. 2004).

Although relevant literature have established the benefit of OCRs in reducing the uncertainty of consumer purchase decisions, the problems of information overload and conflicting comments accumulated in the sheer volume of OCRs can make consumers less confident of OCRs and more confused with them (Park and Lee 2008). Wherever human beings make judgments, there are noises that corrupt them (Kahneman et al. 2021). Moreover, comment spamming in OCRs may decrease the efficiency of the consumer decision-making process (Lappas et al. 2016). OCR helpfulness rating has become an important mechanism for solving these problems. Therefore, understanding how online consumers perceive the helpfulness of OCRs has become imperative for researchers and e-commerce practitioners.

The helpfulness of OCRs, one of the major attributes of reviews, has a large section for literary contributions. There are two streams of research related to OCR helpfulness: predictive modeling of helpfulness and determinants of review helpfulness. A primary objective of this dissertation is to better understand the underlying mechanism of OCR appraisal. Therefore, the first two papers included in this dissertation focused on the determinants of OCR helpfulness. The drivers of OCR helpfulness were classified into three categories—review-related characteristics, product-related characteristics, and reviewer-related characteristics—and investigated in Paper 1. Specifically, the star rating, review length, review readability, and so on are classified as review-related characteristics. The product type and the total review number are considered product-related characteristics. The reviewer experience, reviewer expertise, information disclosure, and so on are reviewer-related characteristics.

It was found that scholars have reached a consensus on the positive effects of review properties, such as the review length, readability, and review picture, on the perceived helpfulness of OCRs. In addition, researchers have reached a consistent conclusion about the influences of the reviewer characteristics on OCR helpfulness. In the extant literature that examined the reviewer’s experience, information disclosure, expertise, and online attractiveness, each of these characteristics was found to enhance OCR helpfulness. Nonetheless, there are con-
troversies regarding the influence of the rest of the factors of the perceived helpfulness of OCRs, including of the star rating, review extremity, valence, and sidedness; and others (see Paper 1).

In this dissertation, I first conducted a systematic literature review of the determinants of OCR helpfulness to provide an overview of the extant OCR literature in Paper 1. Then, in Paper 2, I aimed to reconcile the inconsistent finding by using the NLP technique to capture the nuances embedded in the review contents and to examine the impacts of review content structures on OCR helpfulness. For instance, review sidedness was measured with a binary variable in past studies, referring to whether one- or two- sided arguments are presented in OCR content (Chen 2016). By quantifying the degree of co-presence for both positive and negative sentiments in an OCR, it is possible to measure the magnitude of review sidedness, which can better reflect the subtleties of the sentiments embedded in review text. Meanwhile, an OCR can simultaneously exhibit high degrees of information factuality and express strong emotions. Therefore, the review content structure that integrates information factuality and review emotion should be considered when examining the helpfulness of OCRs.

2.3 Managerial Responses

Companies face unprecedented customer service pressures as modern customers become more demanding, better informed, and capable of using Internet channels to instantly voice their displeasure or dissatisfaction (Dens et al. 2015; Hennig-Thurau et al. 2010). The increasingly prominent role of OCRs in brand evaluation and purchase decision-making compels companies to intervene with the influences of OCRs by enhancing their favorable effects and mitigating their unfavorable ones. Since customer engagement activities can impact firm performance directly or indirectly, service providers have proactively developed MR strategies. Hotels with various classes, ranks, online ratings, or popularity levels usually employ dissimilar MR strategies (Levy et al. 2013; Liu et al. 2015 a; Sparks and Bradley 2017). Not only are highly rated hotels more likely to proactively respond to customers on social media, but these hotels also have different engagement attitudes. Whereas many hotels express appreciation, apologize, and provide explanations in their MRs (Levy et al. 2013), others choose bolstering and enhancing postures (Liu et al. 2015 a). A “Triple A” topology representing “acknowledgments, accounts, and actions” (Sparks and Bradley 2017, p. 719) is proposed to help service providers make effective MRs.

MRs have become a fertile area for IS and tourism research but have also sparked heated debates. Particularly, the effectiveness of MRs has been an intriguing topic in terms of diversity and controversy. For the many studies claiming a beneficial effect of MR usage, there are counterstudies asserting either a detrimental or inconsequential effect. I examined relevant MR literature to find out what they have to say about the effectiveness of MR usage, and I grouped extant studies based on their outcome variables, namely, customer satisfaction and business performance.
2.3.1 Effects of Managerial Responses on Customer Satisfaction

Service failure is likely to result in unpleasant customer experiences and negative OCRs. Service failure is at times unavoidable due to machine malfunctions or human errors, but customer dissatisfaction may be avoided. Because negative OCRs are particularly detrimental to business, many companies proactively respond to them to restore customer satisfaction and influence online opinions. In response to poor service quality, effective service recovery is expected to raise levels of satisfaction and even re-patronage intentions (Min et al. 2015). However, inconclusive findings on the effects of MRs on customer satisfaction indicate that not every response increases customer satisfaction (see Table 1).

Table 1. Literature on effects of managerial responses on customer satisfaction.

<table>
<thead>
<tr>
<th>Study</th>
<th>Research Context</th>
<th>Data</th>
<th>Method</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pantelidis (2010)</td>
<td>Online restaurant forum</td>
<td>2,471 OCRs for 300 restaurants</td>
<td>Content analysis</td>
<td>Positive</td>
</tr>
<tr>
<td>Litvin and Hoffman (2012)</td>
<td>Experiment</td>
<td>263 Substantially complete surveys</td>
<td>ANOVA</td>
<td>Positive</td>
</tr>
<tr>
<td>Van Noort and Willemsen (2012)</td>
<td>Experiment</td>
<td>Survey with university students as respondents</td>
<td>ANCOVA</td>
<td>Mixed</td>
</tr>
<tr>
<td>Dens et al. (2015)</td>
<td>Experiment</td>
<td>973 Usable responses from a panel managed by a marketing research agency</td>
<td>MANOVA and univariate Scheffé post hoc tests</td>
<td>Mixed</td>
</tr>
<tr>
<td>Liu et al. (2015b)</td>
<td>TripAdvisor</td>
<td>88,786 OCRs for 187 hotels</td>
<td>Correlation analysis and linear regression</td>
<td>Positive</td>
</tr>
<tr>
<td>Ma et al. (2015)</td>
<td>Twitter</td>
<td>Twitter communications between a company and its 714 customers</td>
<td>Markov Chain Monte Carlo</td>
<td>Mixed</td>
</tr>
<tr>
<td>Min et al. (2015)</td>
<td>Experiment</td>
<td>Survey with 176 university students as respondents</td>
<td>Three-way ANOVA</td>
<td>Mixed</td>
</tr>
<tr>
<td>Rose and Blodgett (2016)</td>
<td>Experiment</td>
<td>255 Usable responses from students at several US universities in Study 1 and 133 additional surveys in Study 2</td>
<td>ANOVA</td>
<td>Mixed</td>
</tr>
<tr>
<td>Xie et al. (2016)</td>
<td>TripAdvisor</td>
<td>56,284 OCRs for 1,045 Texas hotels</td>
<td>Panel data models with fixed effects estimations</td>
<td>Positive</td>
</tr>
<tr>
<td>Proserpio and Zervas (2017)</td>
<td>TripAdvisor</td>
<td>552,051 OCRs for 2,697 Texas hotels</td>
<td>Cross-platform and cross-hotel DDD analysis</td>
<td>Positive</td>
</tr>
<tr>
<td>Chevalier et al. (2018)</td>
<td>TripAdvisor, Expedia, Hotels, Orbitz, and Priceline</td>
<td>1,843 US hotels in the upper midtier or higher categories</td>
<td>Multiple-platform DID analysis</td>
<td>Negative</td>
</tr>
<tr>
<td>Wang and Chaudhry (2018)</td>
<td>TripAdvisor, Expedia, Hotels, and Orbitz</td>
<td>20 Million OCRs for 65,099 hotels</td>
<td>Multiple-platform DID analysis</td>
<td>Mixed</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Platform</td>
<td>Study Sample</td>
<td>Methodology</td>
<td>Findings</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Zhang et al. (2020 b)</td>
<td>TripAdvisor</td>
<td>500 Texas hotels with 221,279 OCRs</td>
<td>Support vector machines, panel data models with fixed effect estimations</td>
<td>Positive</td>
</tr>
<tr>
<td>Sheng et al. (2021)</td>
<td>TripAdvisor</td>
<td>765,082 OCRs and 339,123 MRs</td>
<td>Text mining, multilevel model</td>
<td>Mixed</td>
</tr>
<tr>
<td>Liu et al. (2022)</td>
<td>Ctrip</td>
<td>4,888 Hotels with 2,102,376 OCRs and 1,194,486 MRs</td>
<td>Fixed effects regression</td>
<td>Positive</td>
</tr>
</tbody>
</table>

Note: ANOVA—Analysis of variance; ANCOVA—Analysis of covariance; MANOVA—Multivariate analysis of variance; DID—difference-in-differences analysis; DDD—difference-in-difference-in-differences analysis.

On the one hand, MRs to negative OCRs have been found to positively affect consumers’ attitudes toward the company (Litvin and Hoffman 2012). Furthermore, an experimental study affirmed that MRs not only have positive effects on potential consumers’ attitudes and subsequent purchase behavior but also alleviate the negative impacts of negative OCRs (Le and Ha 2021). Several studies that empirically measured the influence of MRs on customer satisfaction by analyzing a big number of OCRs gathered from TripAdvisor showed a positive relationship between MRs and OCR ratings (Liu et al. 2015 b; Proserpio and Zervas 2017; Xie et al. 2016). It was found that a higher incidence of matches of MR topics with OCR topics leads to an increase in subsequent e-WOM valence (Zhang et al. 2020 b). Moreover, it was claimed that a successful MR could turn a complaining customer into a loyal one (Pantelidis 2010).

On the other hand, while a number of authors found a positive association of MR with customer satisfaction (e.g., Pantelidis 2010; Xie et al. 2016), others obtained mixed findings and argued that the effectiveness of MR on customer satisfaction may be moderated by factors such as the platform type (Van Noort and Willemsen 2012), the customer type (Gu and Ye 2014), the MR strategy (Min et al. 2015), and the e-WOM status and market position of the firm (Sheng et al. 2021). In particular, the effect of “webcare” interventions in mitigating the impairment of unfavorable OCRs has been examined (Van Noort and Willemsen 2012). The findings indicated that even though MRs induced more favorable brand evaluations, its effectiveness depends upon platform types (company-owned versus third-party) and response strategy (reactive versus proactive). Discernment between complaining customers and other observing customers revealed that MRs improve the satisfaction level of complaining customers but have limited influence on observing customers (Gu and Ye 2014). Furthermore, the cause of the service failure (controllable versus uncontrollable factors) (Rose and Blodgett 2016) and the response contents (Min et al. 2015) have also been identified as moderators of the effects of MRs on customer satisfaction. In general, the positive influence of company engagement in the online review platform is likely to be shaped by the goodwill signaled by MRs, but an increase in customer satisfaction is largely associated with the response strategy, such as
the intensity, promptness, and sentiment of MRs (Sheng et al. 2021). Recently, a saturation effect was identified in MRs, which indicates that MR’s marginal effect diminishes even though it can still enhance the rating valence of a hotel (Liu et al. 2022).

Although MRs can improve customers’ relationships with a company, it has also been warned that MRs can have an opposite effect of raising customer expectations and inciting more complaints (Ma et al. 2015). Similarly, it has been suggested that companies be alert to an unexpected negative outcome of MRs’ stimulation of reviewing activity through more and longer reviews—the triggering of the posting of more impactful negative OCRs and the decreasing of subsequent ratings (Chevalier et al. 2018). As the review set balance plays a pivotal role in forming potential customers’ evaluation of a company and influence the aftereffect of MRs, companies should choose an appropriate MR strategy for intervening with OCRs that impacts customer satisfaction and customers’ patronage intentions (Dens et al. 2015).

### 2.3.2 Effects of Managerial Responses on Business Performance

A cohort of studies have acknowledged the effects of OCRs on companies’ sales efforts and financial performance (Berger et al. 2010; Chevalier and Mayzlin 2006; Liu 2006). To strengthen the beneficial effects and alleviate the adverse effects of OCRs, proactive online engagement with customers via social media, such as by using MRF, has recently become an imperative for business operators. It is believed that the presence of MRs affects consideration sets, customer attitudes, customer satisfaction and trust, and future reviews, which can ultimately impact company sales (Chen et al. 2019; Proserpio and Zervas 2017; Sparks et al. 2016). Nevertheless, the results of previous studies on the effects of MRs on business performance have been inconsistent and contradictory (see Table 2).

<table>
<thead>
<tr>
<th>Study</th>
<th>Research Context</th>
<th>Data</th>
<th>Method</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mauri and Minazzi (2013)</td>
<td>Experiment</td>
<td>Online questionnaires with 349 respondents</td>
<td>Correlation analysis</td>
<td>Negative</td>
</tr>
<tr>
<td>Xie et al. (2014)</td>
<td>TripAdvisor</td>
<td>843 Texas hotels and 4,994 quarterly-level observations</td>
<td>Panel data analysis</td>
<td>Negative</td>
</tr>
<tr>
<td>Kim et al. (2015)</td>
<td>TripAdvisor</td>
<td>12 Months of OCR data for 128 US hotels and performance records</td>
<td>Multiple regression</td>
<td>Positive</td>
</tr>
<tr>
<td>Lee et al. (2016)</td>
<td>TripAdvisor</td>
<td>3,763 MRs to 28,443 OCRs for 730 hotels in southern U.S.</td>
<td>Instrumental variable-fixed effects regression and multi-level mixed effects regression</td>
<td>Mixed</td>
</tr>
<tr>
<td>Xie et al. (2016)</td>
<td>TripAdvisor</td>
<td>56,284 OCRs for 1,045 Texas hotels</td>
<td>Panel data models with fixed effects estimations</td>
<td>Positive</td>
</tr>
</tbody>
</table>
A number of studies have reported that MRs, as alternatives to service failure recovery, support more positive business performance by allowing companies to respond to queries and concerns of their dissatisfied customers (e.g., Kim et al. 2015; Kumar et al. 2018; Proserpio and Zervas 2017). Data from diverse sources have demonstrated MRs’ association with business performance. Several studies thus far have chosen the OCR volume or the check-in volume as a proxy for the customer volume or sales, which has linked MRF adoption to appreciable improvement in customer volume and business performance (Kumar et al. 2018; Proserpio and Zervas 2017; Xie et al. 2016; Ye et al. 2008). For instance, by matching hotels across two travel agency sites and comparing the difference between MR and non-MR hotels, it was found that MRs significantly and positively impact the volume of subsequent OCRs (Chen et al. 2019). Furthermore, several studies that investigated the association of MRF with business performance improvement using actual hotel performance data, specifically the average daily rate and the revenue per available room (RevPAR), confirmed such association (Kim et al. 2015; Xie et al. 2016; Xie et al. 2017a). Indeed, ample studies had offered empirical support for the benefit of MR use on companies’ business performance. Major OCR platforms have implemented MRF to facilitate the management of OCRs.

Nonetheless, even though appropriate MRs to customer complaints may lead to favorable outcomes, not all MRs are guaranteed to appease unpleasant cus-
tomers and favor the company. Considering that companies’ proactive engagement may be perceived by certain customers as unsolicited or intrusive and thus, may weaken their purchase intention and consequently, the companies’ financial performance, some researchers have opposed the use of MRF (Mauri and Minazzi 2013; Xie et al. 2014). A survey of 349 consumers in an experimental study revealed a decline in their purchase intentions after the appearance of MRs to OCRs (Mauri and Minazzi 2013). In the same vein, a decrease in hotels’ RevPAR was observed after MRF was adopted by Texas hotels (Xie et al. 2014).

Moreover, the mixed results of some studies indicate that MRs can either synergize customer relationships or compound service failures, and the outcomes of MRF adoption are contingent on the conditions of OCRs (valence and volume) and on the MR strategy (Lee et al. 2016; Xie et al. 2017 b). Depending on the trustworthiness of a platform, customers may treat MRs on a less trustworthy platform as acts of review manipulation instead of positive efforts by companies (Xu et al. 2020). Therefore, companies should use MRF strategically to complement review valence, improve trust in them, and ultimately boost their business performance.

Collectively, the advice offered by extant literature on the usefulness of adopting MRs for boosting business performance has been inconsistent and contradictory due to various limitations in data and methodology. A combination of robust statistical methods and actual business performance data may yield interesting findings. In addition, past studies on MRs mainly investigated a single market or region. Although in-depth perspectives of the effect of MRs were seen, findings from a single region may not be generalizable to other markets. Paper 3, to offer a more complete understanding of the effect of MRF use, analyzed cross-market data by integrating the statistical methods of propensity score matching (PSM) and difference-in-differences (DID) analysis and identified a timing effect of MRF use on business performance. Moreover, estimations were made on actual hotel performance data to confirm the validity of the findings.
3. Synthesizing Summary of the Papers

The empirical part of my thesis comprises four separate but interconnected research papers that strove to extract the value of OCRs and MRs from the points of view of consumers, companies, and industries. Although the details of the papers vary, they also have some shared elements. Excluding the literature review paper, the papers followed the positivist paradigm and used a quantitative methodology with a deductive approach, informed by theories widely used to explain user and organization behaviors in IS research. In addition, all the papers are empirical and inspect the value of OCRs and MRs.

Quantitative research is data-oriented and uses a range of methods geared towards systematically exploring and investigating social phenomena. Quantitative methods emphasize objective measurements by gathering quantifiable data and analyzing them statistically, mathematically, or numerically (Babbie 2010). A quantitative approach is best “if the problem is identifying factors that influence an outcome, the utility of an intervention, or understanding the best predictors in outcomes” (Creswell 2003, p. 23). A research method can be defined as “simply a technique for collecting data” (Bell et al. 2018, p. 45), such as primary and secondary data. Instead of relying on already existing data, primary data collection has the distinct property of focusing on gathering data directly from main sources through interviews, surveys, experiments, and so on (Brandin and Abrishami 2021). In contrast, secondary data collection involves utilizing existing data (the so-called “secondary data”), from existing data sources such as the Internet, government resources, libraries, and research reports. By summarizing and collating existing data, secondary quantitative research can aid in the validation and exploration of the data that were collected from primary quantitative research. With the extensive Internet penetration today, it has become increasingly common for quantitative research to be conducted using Internet data sources. In this study, the quantitative secondary research method was adopted. The statistical methods employed in this dissertation are introduced in the following subsections.

This chapter presents a summary of each of the four papers, which elaborates on where the studies reported in the papers fit into the conceptual research framework discussed in previous chapters. The following subsections detail their study design—their objective, theoretical background, level of analysis, analysis method and findings. That is, it demonstrates the positioning described in Chapter 2: the first two papers focus on the helpfulness of OCRs for individual consumers. The third paper discusses whether and to what extent the use of the
MRF helps improve companies’ business performance. The last paper demonstrates the analysis of user-generated social big data to obtain insights into collective consumer behavior patterns for an industry.

Thus, each of the four papers is a report on a standalone study with a distinct research context. Several theories and research methods were applied depending on the research questions. The following subsections summarize each paper by briefly introducing the study it reports on and then describing the data and methods used and outlining the main findings.

3.1 Consumer Perspective

As discussed in Chapter 1, by comprehending OCRs from those who have experienced a product or service, consumers benefit from reduced uncertainty and risk associated with online purchase (Ye et al. 2011). On the other hand, the influx of OCRs may cause information overload (Baek et al. 2012; Yin et al. 2014), which will likely lead to consumer exhaustion in the act of identifying the most relevant review for their purchase decision.

To address this challenge, OCR platforms enable users to indicate the informativeness of OCRs by rating reviews as “helpful” or “unhelpful” (Mudambi and Schuff 2010). “OCR helpfulness” refers to the perceived value of the information included in a review (Li et al. 2013). It can also be viewed as “a reflection of review diagnosticity,” as it measures the extent to which an OCR helps a reader make informed purchase decisions (Mudambi and Schuff 2010, p. 186). Thereafter, OCR helpfulness has become an important instrument in customers’ decision-making (Ullah et al. 2015).


For a better understanding of the appraisal mechanism for OCR helpfulness, ample studies have been conducted. Such studies identified the determinants of OCR helpfulness. However, inconsistent results were reported by the past studies that used varied research methods and distinctive datasets. Thus, it is necessary to review extant publications to synthesize existing findings. Paper 1 was written in response to this problem, to enable a better understanding of the drivers of the perception of OCR helpfulness through a systematic literature review that covered a large volume of academic publications.

**Data Collection.** Following the systematic and structured approach proposed by Webster and Watson (2002), I conducted manual keyword searches on the four most widely visited research databases and search engines for academic research (Buhalis and Law 2008), namely, Science Direct, EBSCOHost, ProQuest, and Google Scholar. I searched for journal articles related to OCR helpfulness using the following keywords: “helpfulness/usefulness/trustworthiness,” “online (customer/consumer) review,” “word of mouth,” “WOM,” and “eWOM.” In addition, I traversed the references cited in the retrieved journal articles to ensure the extensiveness of our literature dataset and that no major OCR helpfulness articles would be overlooked. After applying the inclusion and
exclusion criteria and browsing the abstract of each article in the initial review sample, 68 journal articles published between 2007 and 2020 were identified. I then extracted relevant information from these selected articles for further analysis, including information on the research method, theoretical ground, studied factors, operationalization, and findings.

**Method.** *Systematic Literature Review*: In this systematic literature review, the research models and findings of these identified articles were consolidated and classified. These 68 papers in our dataset pointed to 18 determinants of OCR helpfulness. A trend map of OCR helpfulness drivers was produced to offer a straightforward overview of the factors investigated in the extant literature. This visualization displays the transitions of the focuses of past research on precursors of OCR helpfulness, and more importantly, their relationship. The factors that had been extensively investigated were categorized into three groups: (1) review-related characteristics, (2) product-related characteristics, and (3) reviewer-related characteristics. The definitions and operationalizations of the abovementioned factors, along with their effects on the perceived helpfulness of OCRs, were consolidated and summarized. Finally, a structured analysis and consolidation of both consistent and conflicting findings were carried out, which depicted the status quo, helped capture the emerging trends, and pointed out future research avenues.

![Figure 3. Framework of OCR helpfulness.](image)

**Findings.** There is a consensus among researchers that three review- and product-related characteristics, namely, the review length, readability, and review picture, have positive effects on the perceived helpfulness of OCRs. The relationships between the reviewer characteristics and OCR helpfulness have also been widely investigated and have yielded clear conclusions. The reviewer characteristics, including their information disclosure, experience, expertise, and online attractiveness, were analyzed in 39 studies, and were found to serve a significant role in enhancing OCR helpfulness.
Regarding the relationships between the remaining factors and the perceived helpfulness of OCRs, mixed or contradictory findings were seen in the extant literature. These factors included star ratings and review extremity, valence, sidedness, and so on.

This study proffered three possible explanations for the controversial findings. First, the variations in the product type may be the factor that led to the emergence of mixed or contradictory results concerning the influence of the review rating and its quadratic term, review extremity, and review valence. Another reason for the inconsistent outcomes may be the disparities in the measurement of the determinants of OCR helpfulness. Finally, the lack of persistent findings may also be the result of the varied research settings, selection bias, or even misuse of statistical analysis.

Moreover, recent advances in the NLP and text mining techniques have made emotions and linguistic cues popular and widely studied as factors that affect OCR helpfulness. Nonetheless, this study still underscored the insufficiency of the exploration of the influences of emotions and linguistic factors on OCR helpfulness, which led to the follow-up study in Paper 2.

3.1.2 Paper 2. Quantifying the Effects of Online Review Content Structures on Hotel Review Helpfulness

Consumers are increasingly relying on OCRs to support their purchase decisions, making the review system an important feature of online platforms. However, many studies on OCR helpfulness have presented inconsistent findings on the influence of review sentiments on the perceived helpfulness of online reviews. Additionally, the impact of review content structures—such as the use of sentiment-free sentences in the review text—has been mostly ignored. Drawing on the attribution theory, this study seeks to remedy these problems by analyzing the impacts of review content structures on OCR helpfulness. More specifically, this study focuses on three variables pertinent to review content structures: review sidedness, information factuality, and emotional intensity at the beginning of a review. Additionally, the moderating effects of the reviewer reputation and the review sentiment are examined.

**Theoretical Background.** The attribution theory is defined as dealing with “how the social perceiver uses information to arrive at causal explanations for events. It examines what information is gathered and how it is combined to form a causal judgment” (Fiske and Taylor 1991, p. 23). These causal inferences, called “attributions,” mainly manifested in two ways: as dispositional attribution and as situational attribution (Heider 1958). Dispositional attribution allots the cause of a behavior to the internal characteristics of a person, whereas situational attribution accredits the cause of behavior to environmental factors outside the control of a person.

In the OCR context, the attribution theory has been used to elucidate how consumers evaluate OCR helpfulness. Following the logic of the attribution theory, review readers assess the helpfulness of OCRs based on that to which they attribute OCRs, such as the reviewer’s motivation to post the review (Kelley 1973,
Synthesizing Summary of the Papers

1987). For instance, the causal inference for an OCR, whether the cause of a review is attributed to the dispositional characteristics of the reviewer or to the performance of the product or service reviewed (cf. Chen et al. 2020; Lee and Youn 2009), is disposed to influence readers’ judgments of the helpfulness of this review (Chen and Farn 2020) and their purchase decisions (Sen and Lerman 2007).

Moreover, leveraging the discounting principle of the attribution theory, the presence of alternative reasons or causes may discount the role of a given cause in exerting a certain effect (Chen and Farn 2020). When it is suspected that the endorsement in a review was caused by an incentive from the company or by other non-stimulus factors, a review reader may regard the reviewer as biased and may be unconvinced by the review content (cf. Kelley 1973). The discounting principle of the attribution theory had been used in previous studies to identify the impact of the source characteristics on the OCR helpfulness (Lee and Youn 2009; Senecal and Nantel 2004).

In line with earlier research that used the attribution theory as a theoretical lens to study the influence of review sources and text content on OCR helpfulness, the attribution theory is appropriate in explaining the impacts of review content structures on the perceived helpfulness of OC\hspace{1pt}Rs.

**Data Collection.** OCR data crawled from TripAdvisor were used in the analysis. The dataset comprised 144,982 English reviews for over 1,200 hotels in China, Finland, and Germany, which had been posted between June 2002 and February 2016. The NLP technique was employed to gauge the review sentiment of each OCR at the sentence level while accounting for the presence of adverbs and negative terms. The degrees of review sidedness, information factuality, and emotional intensity at the beginning of a review, which are the three focal variables pertinent to review content structures, were quantified.

**Method.** Zero-inflated Negative Binomial Regression: Negative binomial regression is commonly used to model count outcome variables, specifically, overdispersed data. Being a generalization of Poisson regression, negative binomial regression loosens the restrictive assumption of “equidispersion” made by the Poisson model (Maxwell et al. 2018, p. 179). Overdispersion leads to a deflated standard error and inflated test statistics and occurs when the success outcome is not rare. Among various potential causes of overdispersion, that which this study faced is the excessive zeros in the outcome variables.

To handle such a situation, zero-inflated negative binomial regression is commonly recommended (Greene 1994, 2012). The review helpfulness is a count variable that takes on only positive integer values, but it is obvious that many reviews did not receive votes. By virtue of the abovementioned superiorities of zero-inflated negative binomial regression, this approach was deemed appropriate because of the excessive zeros in the dependent variable as well as its skewed distribution (Liu and Park 2015; März et al. 2017).

**Findings.** Our results revealed that review sidedness exerts a negative influence on perceived OCR helpfulness, which indicates that the more two-sided an OCR is, the less helpful it is perceived to be by the review readers, while the
reviewer reputation mitigates the negative influence of two-sidedness. Information factuality enhances the perceived helpfulness of OCRs, and positive sentiments amplify this effect, which indicates that OCRs with positive sentiments appearing alongside an objective description are perceived to be more helpful. Furthermore, an OCR that begins with a highly emotional assertion tends to be evaluated as less helpful, whereas an OCR that begins with a fact-based narrative may create a better first impression of the reviewer’s authenticity and may be perceived as more helpful.

3.2 Company Perspective

Company operations require balance between profit and cost. These two factors must be considered when making any strategic decision. The prevalence of OCRs has posed a significant challenge to companies. On the one hand, consumers rely on OCRs to support their purchase decisions, which, in turn, affects the sales and revenue of the corresponding companies. On the other hand, most OCRs are posted on third-party platforms, which are beyond the companies’ control. When MRF is made available to companies, they have a means of intervening in and reacting to the OCRs. However, the question of whether to invest human resources and time in the use of MRF to proactively respond to OCRs lies with the company operators.

The effectiveness and extent to which MRF can improve business performance is the main question to address. The following study aimed to solve this problem by analyzing the social big data created by both customers and companies.

3.2.1 Paper 3. The Timing Effect of IS Service Availability: The Case of Managerial Response Service Usage in the Hospitality Industry

In this study, the investigation of the usefulness of MRF usage was abstracted as the impact of IS availability on companies. In addition, drawing on the life cycle view of Kano’s theory of attractive quality (Kano 2001; Kano et al. 1984), this study offered a theoretical angle to explain the timing effect of the availability of an IS artifact for customer service on companies’ business performance. In other words, this study verified the effectiveness of using MRF for OCRs on business performance improvement in the context of the tourism and hospitality sector. Meanwhile, the timing effect of MRF adoption was investigated by measuring the change in the company performance against the timing of the MRF availability.

Theoretical Background. Kano’s theory of attractive quality (Kano 2001; Kano et al. 1984) articulated that the (non)availability of a product attribute could affect the overall customer satisfaction, and the magnitude of this effect differs in accordance with the timing of the attribute availability. If a service or service feature improves business, firms offering a novel IS service attribute will constantly have a higher level of customer satisfaction than those that do not. Kano’s theory, a seminal piece of work in service quality, accentuated that an innovative service or service feature, when initially offered, often exhumes an
attractive quality that consumers do not expect but are excited to have. Therefore, offering an innovative IS attribute surprises and delights customers, which makes them more likely to be loyal to the brand. If the IS attribute is a pioneer offering, it would bring about a high level of return, even if no dissatisfaction will be caused by not offering the attribute. When customers get used to the IS attribute, its availability can make them feel contented; whereas they will start to feel dissatisfied when the attribute is not available. Thus, the IS attribute is considered a one-dimensional quality. Over time, repeated usage of the IS attribute will make it a must-have quality whose availability will not increase customer satisfaction but whose absence will culminate in strong customer dissatisfaction. This study expanded Kano’s theory to explain the effect of IS availability on customer satisfaction and company performance by proposing an effect of dissipating benefits of a novel IS service. In the context of the hospitality industry, the availability of MRF was expected to enhance business performance. Moreover, the business performance benefit of adopting the MRF mainly appeared at the early stage of MRF use in pioneering companies and vanished over time.

Data Collection. The empirical data for the first study were collected from TripAdvisor, which consisted of OCRs of hotels and corresponding MRs of hotel operators. Specifically, all the reviews posted before February 2016 on Finnish hotels, hotels in two major German cities (Berlin and Hamburg), and hotels in 11 Chinese cities (e.g., Beijing and Shanghai) were collected—more than 968,000 reviews of more than 27,000 hotels. Furthermore, in this data set, 200,632 (20.71%) reviews received MRs from 1,960 (7.24%) hotels.

In order to validate the results, I also accessed the business performance data of Finnish hotels between January 2009 and February 2016 via Statistics Finland and conducted a second study with this real sales data. I first manually matched the Finnish hotels’ business ID with their TripAdvisor ID. Statistics Finland linked the TripAdvisor data and the actual hotel performance data, and then pseudonymized the data. After the hotels that operate seasonally (such as those open only in winter in northern Finland near the Arctic) were removed, the business performance data of 258 Finnish hotels were used for further analysis, and 139 of them used the MRF during the period of investigation.

Methods. Propensity Score Matching and Difference-in-Differences Analysis: PSM is a statistical method that is used to “reduce the impact of treatment-selection bias in the estimation of causal treatment effects using observational data” (Austin 2007, p. 1128). Treatment-selection bias is a commonly observed bias caused by confounding variables that affect the likelihood of an observation belonging to the treatment group or the control group. In case this bias is not controlled, the estimated effect of a treatment is often biased to the internal differences between the treatment group and the control group. The propensity score is defined as the probability that a subject will receive a specific treatment on the basis of the observed covariates (Austin 2008). By matching the propensity scores of individual observations between the treatment and control groups using proper algorithms, the method effectively rules out the possible treatment-selection bias.
**DID analysis** is an instrument for estimating a treatment effect on the treatment and control groups by analyzing their differences at the pre- and post-treatment stages. Specifically, it measures the effect of a treatment on an outcome by comparing the change in the outcome variable (before and after the treatment) for the treatment group with the change in the outcome variable for the control group, which diminishes the effects of extraneous factors to achieve precise results (Rishika et al. 2013). By comparing the pre- and post-difference of the treatment group with the pre- and post-difference of the control group, precise results can be obtained by ruling out the possible bias caused by extraneous factors.

**Findings.** In Study 1, The results of DID analysis indicated that hotels using the MRF tend to have a significantly higher volume of reviews than their counterparts, whereas an insignificant difference in the hotel ratings was observed between the treatment and control groups. Furthermore, a fixed effects regression analysis that controlled for the country difference revealed the positive relationship of more frequent use of the MRF to an increase in the customer volume. Meanwhile, the moderating effect of the start time of MRF use on the abovementioned effect showed that the start time of MRF use moderates the impact of the response volume on the increase in the customer volume.

Study 2 triangulated the abovementioned findings using real hotel sales data. The hotels that used MRF outperformed their counterparts in the period after the MRF use in terms of the RevPar, the occupancy, and revenue of the hotels. Similarly, the analysis showed that the changes in the hotel performance were affected by the frequency and timing of MRF use. This result further showed that the benefits of MRF use dissipated over time.

### 3.3 Industry Perspective

The increasing availability of big data offers people novel aspects of social phenomena to observe and understand. Social big data usually consist of massive records generated by a huge number of users. Therefore, social big data, such as large amount of OCRs, can reflect collective user behavior patterns. The following study proposed a novel approach to extracting travelers’ revisit patterns using OCR data and examined the detrimental effect of air pollution on the tourism industry.

#### 3.3.1 Paper 4. Quantifying the Adverse Effect of Air Pollution on the Tourism Industry in China

As the tourism industry is closely interconnected with climate and weather (Matzarakis 2006), ample studies have acknowledged its unambiguous relationship with meteorological factors such as rainfall, sunshine, and temperature (Agnew and Palutikof 2006; Álvarez-Díaz and Rosselló-Nadal 2010; Rosselló-Nadal et al. 2011). In recent years, more and more studies have argued that the increasing occurrences of adverse meteorological factors, especially extreme weather and toxic smog, may severely afflict the tourism sector (Parkin 2019; Ross 2019; Saksornchai 2019).
However, recent international tourism data from the World Bank (2019) contravened the long-held assumption that air pollution decreases inbound tourists (Becken et al. 2017). Unexpectedly, countries with dreadful air pollution in recent years had the most prosperous inbound tourism market, with an increasing amount of tourists (World Bank 2019), which implies a possible spurious relationship between air pollution and tourism volume. This study used social big data to reconstruct consumer behavior and to elucidate the influence of air pollution on the revisit behavior of travelers. Specifically, this study strived to quantify the adverse impact of air pollution on foreign tourists’ revisiting behaviors to China by analyzing large numbers of OCRs.

**Theoretical Background.** “The concept ‘image’ can be applied to a political candidate, a product, or a country. It describes no individual traits or qualities, but the total impression an entity makes on the minds of others” (Dichter 1985, p. 75). The destination image is defined as people’s overall evaluative representation of a destination, which expresses to what extent a destination is liked or disliked (Frías et al. 2008). It has been pointed out that tourists’ decision to visit a tourist destination is driven by their predilection related to the particular destination (Josiassen et al. 2016). Being the most frequently used destination predisposition in tourism, the destination image is of great importance because of its profound influence on tourist choice behavior (Josiassen et al. 2016).

It has also been pointed out that beliefs about a country and its people are directly relevant to tourists’ destination beliefs and indirectly relevant to their touristic intentions through the evaluation of the destination and the desired associations with the country (Nadeau et al. 2008). Therefore, tourists’ intentions were situated in the greater context of the host destination and broader country image beliefs, evaluations, interests, and intentions (Nadeau et al. 2008).

The horn effect, also called the “reverse halo effect” and the “devil effect,” suggests that an unfavorable reputation can induce negative assumptions and, in turn, lead to further image damage (Coombs and Holladay 2002; MacDougall et al. 2008). This contagious process occurs because of the tendency of individuals to maintain cognitive consistency (Freedman 1968; Holbrook 1983). Due to the horn effect, a negative first impression of a certain entity—for example, of a product, a brand, or a destination—can affect the assessment of, or the opinion on, similar or associated entities and can overshadow the excellence of other attributes (Dodds 2017; Nicolau et al. 2020). Therefore, in accordance with the horn effect, damage to a destination’s image may portend impairment of the country image.

**Data Collection.** This study drew on a dataset of OCRs from TripAdvisor posted before December 2019, which included 269,847 reported trip experiences posted by 181,698 travelers on 5,142 hotels. These hotels are located in 15 major Chinese cities, which are at a relatively higher level of economic development and are geographically representative of different regions in China. Similar to the study of Sobolevsky et al. (2014), this study utilized a huge amount of OCR data to construct travelers’ visit patterns and to investigate their trip experience and revisit behavior. To ascertain the quality of the treatment group data,
which represented travelers’ encountered air pollution issue, I first queried the
database with keywords that included “smog,” “smoggy,” “haze,” “pollution,”
and “air quality,” and then I manually went through all these selected OCRs to
ensure that each review that was kept was related to air pollution.

Methods. **Propensity Score Matching** and **T-test**: PSM, which is based
on the nearest-neighbor one-to-one matching method without replacement,
was used to generate a comparable sample. As mentioned in Chapter 3.2.1, PSM
can rule out selection bias and improve the robustness of the subsequent anal-
ysis by identifying comparable samples (Caliendo and Kopeinig 2008).

The t-test is a widely used hypothesis testing tool. This inferential statistic can
determine whether a significant difference exists between the means of two
groups of observations. The t-test allows testing of the applicability of an as-
sumption to a population.

Findings. The results of the analysis indicate that compared to the travelers
without air pollution-affected experiences, their counterparts who encountered
air pollution problems during their previous trip were much less likely to revisit
the city and the country, with their revisiting likelihoods decreasing by over 90
percent. The evidence presented in this study suggests that air pollution ham-
pers tourists’ revisits to a city and even a country. Furthermore, the results in-
dicate a possible extending effect of air pollution. That is, the trip experience to
a city may affect tourists’ future visit behavior to other cities in the country.
This dissertation aims to generate insights into how the massive-scale social big data available from tourists can benefit different stakeholders in the context of the tourism industry. I set out to achieve this objective through four empirical research papers on the different facets of the research question.

Unlike traditional records stored in various legacy systems, data collected from social media are often less structured yet contain rich user opinions and behavioral information (Chen et al. 2012). In the context of the tourism industry, the sheer volume of data that tourists generate on social media pertaining to their travel experiences can be a rich mine of business insights. Social big data in tourism have the potential to add value in understanding the decision-making processes, assisting in customer engagement activities, conducting quick experiments, and facilitating product or service innovations. Therefore, social big data represent a new source of relevant insights for improving the effectiveness and efficiency of sense-making and decision-making (Chen et al. 2012). Specifically, this dissertation focuses on the value of social big data in the form of OCRs and MRs.

OCRs and MRs have been continuously evolving. OCR functions were first developed as a new IS artifact to alleviate the uncertainty in online shopping and booking brought about by the temporal and spatial separation. Although purchasing online does not allow physically visiting a business or touching a product, comments published by fellow consumers ameliorate decision-making conditions. The emergence of OCRs helps potential customers to judge the trustworthiness of a business and the quality of a product or service.

However, with more OCRs accumulated for a given product or service, information overload results and makes it more difficult for review readers to quickly identify useful reviews to facilitate their decision-making. Moreover, noises are inevitable in OCRs, which cause overloaded review readers to hesitate to make purchase decisions. Therefore, the functionality of helpfulness votes was developed to alleviate the difficulties in accessing useful OCRs caused by the bewildering variety of entries (Chua and Banerjee 2016).

Except for the influence of OCRs on consumers, the pervasiveness of OCRs leads to paradigm shifts among other stakeholders. On the one hand, OCRs put intense customer service pressure on service providers. Equipped with the Internet and social media, modern customers have channels to instantly broadcast their displeasure or dissatisfaction (Hennig-Thurau et al. 2010). Consumers are becoming better informed, more assertive, and more demanding (Dens et al.
Because OCRs affect consumer attitudes and sales (Blal and Sturman 2014; Wang et al. 2015), companies must engage customers online and intervene in the OCRs. As a result, the prevalence of OCRs elicits the provision of MRF as an artifact integral to OCRs. One must also recognize that MRs can initiate a chain reaction among the reviewer, the company, and potential customers, which impact firm performance directly or indirectly (Wei et al. 2013). On the other hand, the sheer volume of OCRs that reflect consumer experience and behavior provides industries opportunities to capture the dynamics of massive consumers instantly at a low cost.

From the consumer perspective, I systematically reviewed the literature on OCR helpfulness and conducted an empirical study to offer a novel assessment of the effects of review content structures on OCR helpfulness. Specifically, I examined the impacts of review sidedness, information factuality, and emotional intensity on OCR helpfulness by evaluating online review sentiments at the sentence level. The moderating effects of the reviewer reputation and the review sentiment were also investigated.

From the company perspective, the timing effect of the MRF adoption in the hospitality industry was empirically investigated. It was found that MRF gives a company implementing it an advantage over other companies in terms of improved business performance, but the capability of MRF to improve the company’s business performance dissipates over time.

From the industry perspective, I studied the Chinese tourism market and quantified the adverse effect of air pollution on foreign tourists’ revisit behavior to China as reflected in the big volume of OCRs. Initially, the study identified travelers who encountered air pollution issues by keyword filtering and manual screening of their reviews. Then, the PSM technique was used to discern a matching group of travelers with homogenous characteristics who did not report an air pollution-affected experience in their reviews. By estimating their respective likelihoods of revisiting, the empirical evidence suggested that travelers who had encountered air pollution during their trips are significantly less likely to revisit the specific city, nor to revisit China.

In the following subsections, the contributions of this dissertation for academics and practitioners interested in using social big data are presented. Finally, this dissertation is concluded by recognizing the limitations of the study and presenting suggestions for future research and a summary of conclusions.

4.1 Theoretical Contributions

In general, all the papers included in this dissertation demonstrated how different stakeholders in the tourism industry—consumers, companies, and policymakers—can better appreciate the value of social big data, in terms of OCRs and MRs. As discussed, OCR and MR are ever-evolving and -interacting artifacts. The progression of the OCR and MR phenomena follows the law of development of the Internet world, in terms of innovation, scale, and efficiency, accompanied by the spiral of problem appearance and resolution. By breaking temporal and
spatial limitations, innovative e-commerce improves the efficiency of the economy but creates uncertainty. OCRs appear and offer users information aplenty to reduce the uncertainty and risk of online purchasing. The proliferation of OCRs produces massive-scale information that overwhelms users, which in turn undermines efficiency. OCR helpfulness assessment arises from this information overload issue. Meanwhile, the vigorous development of OCRs poses challenges to companies, which has led to the availability of MRF. Furthermore, during the process, much more information has become available online, enabling industry practitioners and government sectors to gain in-depth insights into the industry. It is obvious that the information revolution on the tourism industry extensively influences all its stakeholders. Relying on social big data, this dissertation therefore set out to contribute to literature by exploring the phenomena of OCRs and MRs using four separate but interlinked studies.

The first two studies focused on the helpfulness of OCRs and identified the role of review content structures in shaping OCR helpfulness. Meanwhile, the moderating effect of the reviewer reputation and the review sentiment on the relationship between the review content structures and OCR helpfulness was untangled, which reconciled the conflicting findings in the literature on OCR helpfulness. Adhering to the attribution theory, the findings of the study from the consumer perspective showed that consumers’ assessments of OCR helpfulness are determined by the review content structures. Specifically, it was elucidated that these structures encompass review sidedness, information factuality, and emotional intensity at the beginning of the review. Adopting the attribution theory as the primary theoretical lens, Paper 2 proffered a novel and useful perspective for comprehending the determinants of OCR helpfulness by concurrently analyzing the emotional and factual contents of OCRs. The findings of the study also provided further evidence of the important role of information factuality in determining OCR helpfulness or diagnosticity (Filieri et al., 2018a). These empirical findings accord with marketing studies that recognized informational appeal as an essential factor of advertisements in influencing product sales (e.g., Teichert et al. 2018).

On the other hand, some researchers have claimed that the inclusion of a two-sided argument in an OCR enhances its perceived helpfulness since the review may be perceived as less biased (Cheung et al., 2012; Filieri et al., 2018b). On the contrary, because two-sided OCRs may contain ambiguous information and offer unclear suggestions, it has been argued that they are not always helpful and may even be less persuasive than one-sided OCRs (Chen 2016; März et al. 2017; Pentina et al. 2018; Schlosser 2011). Given the previous inconsistent and contradictory findings on how review sidedness affects OCR helpfulness, Paper 2 presented important evidence of a negative effect of two-sidedness in this respect by applying the NLP technique to the analysis of social big data to detect the review sentiment, rather than using the more limited lexicon-based method. This finding further supports previous research on this area, which held the view that consumers prefer one-sided reviews to two-sided reviews to facilitate purchase decision-making (März et al. 2017; Pentina et al. 2018). Moreover, this
study is among the first to demonstrate how the reviewer reputation moderates the impact of review sidedness on OCR helpfulness.

Second, to intervene with regard to the influence of OCRs, companies adopt MRF to respond to consumers’ comments. The study from the company perspective contributes to the discussion on whether MRF use affects business performance. Based on two studies and the integrated use of the PSM and DID methods to account for the influence of extraneous factors and treatment-selection bias, the results offer two concrete and mutually corroborated proofs of the positive impact of MRF use on hotels’ business performance. On the one hand, the results of Paper 3 showed that the hotels that used the MRF performed significantly better than the hotels that did not use the function. On the other hand, this study demonstrated that the frequency of MRF use is positively related to the level of business performance improvement among hotels using MRF. Moreover, by collecting and verifying actual hotel sales data, this study responded to the recent call for “future research to obtain sales data on a large scale to assess the influence of managerial response on future reviews through sales” (Chen et al. 2019, p. 14).

Furthermore, the timing effect of companies’ MRF adoption on their business performance was investigated. Kano’s theory of attractive quality, which says that the benefits obtained from a new service offering taper off over time, was also found to be applicable to IS contexts. That is, companies that pioneered the implementation of a new IS service receive significant competitive advantages, but the companies that implement the service later gain fewer benefits. Eventually, the offering of the IS service becomes a competitive necessity when consumers consider the service offering to be a prerequisite to selecting a brand. This theory offers a novel lens for understanding customer satisfaction and business performance related to the availability of an IS, which varies from mainstream IS satisfaction theories built on the expectation confirmation theory (ECT) (Benlian et al. 2011; Bharadwaj 2000). ECT asserts that customer satisfaction is determined after a customer evaluates product performance, by forming a (dis)confirmation judgement about such performance that corresponds to the customer’s original expectations. Regarding the (non)availability of an IS system, ECT makes no assumption of it in relation to customer satisfaction. By drawing on Kano’s theory of attractive quality, this study examined a dynamic impact of the availability of IS attributes on overall service satisfaction and business performance. Specifically, based on two studies on MRF use in the hospitality sector, empirical support was obtained for the moderating effect of timing on the benefits of MRF adoption.

The results of Paper 3 also suggest a possible placebo effect of using the MRF to improve the user experience. Based on two different studies, MRF use was found to have had no impact on hotels’ overall ratings, as there was no improvement in the customer experience even after the hotels made MRs to customer reviews. This result contradicts previous studies. In other words, even if hotel managers read and responded to OCRs, their efforts did not bring about sufficient change to their overall service. As a result, the user experience exhibited
neither improvement nor deterioration after the hotels used the MRF. Noticeably, this result that the MRF did not exert a significant impact on customer ratings for hotels does not negate Kano’s theory concerning the impact of the attractive quality on satisfaction. The MRF has an effect when consumers select a hotel, by creating a more favorable attitude toward the hotels that offer MRs, therefore increasing the customer volume of those hotels. Nonetheless, the hotel reservation service is virtually separated from the actual lodging service, so MRF use has little impact on how consumers later evaluate a hotel.

Third, from the tourism industry perspective, the last study demonstrated that OCRs can be used as instruments for understanding tourist behavior that is affected by air pollution. This substantiates the effectiveness of social big data in reflecting consumer behavior, as it helped reveal the spurious correlation between air pollution and inbound tourism growth.

Although previous research has acknowledged well the impact of environmental factors on the tourism industry and recognized air quality as an integral aspect of the environment, little is known about the influence of air pollution on people’s travel habits (cf. Zhang et al. 2020 a). In addition, revisiting customers have been identified as offering much more business value than new customers, since they are more habitual in visits as well as more destination-loyal (Oppermann 1998, 2000). However, there is a dearth of studies on the impact of air pollution on the revisit behavior of travelers. This research contributed to filling this gap by analyzing large-scale, publicly available OCR data. Specifically, by quantifying the adverse impact of air pollution on foreign tourists’ revisit behaviors to China, Paper 3 confirmed a significant deterrent effect of the air pollution issue on traveler’s revisiting tendency. Supplemetning previous research that showed that air pollution hinders travelers from initiating a trip (e.g., Becken et al. 2017), this study focused on air pollution’s adverse effect on travelers’ revisit behavior. Moreover, the evidence presented in this study revealed a possible confounding effect of economic growth. Despite the increasing number of inbound travelers, it was found that air pollution has a negative effect on inbound tourism.

Furthermore, relying on the analysis of social big data, this study contributed to a more comprehensive understanding of the impact of air pollution on travelers’ actual behavior. The advantage of the presented approach is its capability to identify the revisiting pattern from a huge user base at a low cost but with high efficiency. Compared to the traditional data collection approach, the proposed approach is beneficial to researchers who wish to reach a large number of people and to deepen insights into their travel behaviors. To the best of my knowledge, this study is among the first attempts to examine the influence of environmental factors by analyzing social big data pertaining to actual travel behavior in non-laboratory settings. Additionally, responding to the call to use big data to further understand the tourism industry (Bramwell et al. 2017), this study demonstrated the incorporation of diverse analytic methods with social big data to obtain new insights for tourism research.
4.2 Practical Contributions

From the practical aspect, the prevalence of social big data makes it imperative for all stakeholders of an industry to cope with it and to exploit value from it. The studies that constitute this dissertation offer implications for individual consumers, companies, and even the entire tourism industry.

The findings of the first two studies can help potential customers, as reviewer readers, to better assess OCR helpfulness. Moreover, the evidence presented offers practical guidance for online reviewers in composing more helpful OCRs. For example, the second study suggested that novice reviewers should provide only one-sided reviews, whereas expert reviewers may also post two-sided reviews. Besides, it was recommended that reviewers start writing their reviews with objective narratives instead of emotional opinions. OCRs composed according to this approach may induce review readers to attribute review content to focal services or products themselves, which will enhance their perception of the trustworthiness and helpfulness of OCRs.

The findings of the second study are not limited to users of OCR platforms. Design guidelines for OCR platforms are proffered as well. Specifically, the study demonstrated the effectiveness of applying NLP techniques to discern the magnitudes of review sidedness and the levels of information factuality. It is believed that OCR platforms may benefit from proactively providing such augmented information, which can help consumers efficiently identify useful reviews and make purchase decisions, which will in turn increase their user satisfaction.

The study on MRF offered companies the following practical suggestions. First, the findings encourage companies to be pioneers in offering new IS services to consumers because this will engender greater competitive advantages for them than for followers, but to also consider the possible risks and costs. Second, the timing effect of the IS service availability shows that enterprises that lag in offering a favorable IS service may lose an opportunity to improve their business performance, especially after the IS service becomes a common feature in the market. Third, especially for the hospitality sector, hotel operators that have not adopted the MRF are encouraged to use the function to proactively respond to customer reviews online. The study showed that frequent responses, rather than occasional responses, to OCRs are the key to benefitting from MRF use. Fourth, statistical analysis indicates that hotels mainly use the MRF for OCR management but fail to use the online feedbacks of customers to improve their service and the customer experience. In other words, responding to OCRs appears to be an isolated business activity for most hotel managers, thus resulting in little impact on hotel operations improvement. As OCRs may reveal customers’ service needs and expectations that are trending, hotels can improve their competitiveness by fulfilling such needs and expectations before their competitors do. Therefore, hotel operators are encouraged to take action to improve their service quality by using social big data.

The study from the industry perspective yielded several practical implications for tourism practitioners. By analyzing social big data generated by travelers on OCR platforms, individual-level user behavior data are consolidated to reflect
industry-level customer behavior patterns. For hotel managers, the empirical results of the study highlighted that the travelers who have stayed at a hotel during the period when there was severe air pollution are much less likely to revisit the country nor re-patronize the hotel. Fewer repeat visitors may result in huge losses from a reliable revenue stream and from word-of-mouth channels that bring new customers (Reid and Reid 1994). Measures that mitigate the damage caused by adversarial environmental problems must be adopted, such as price reduction and other interventions to attract revisits (Atzori et al. 2018). Besides reducing pricing during periods of heavy smog, hotels may also provide discounted offers for revisiting tourists whose previous visit was affected by air pollution. Moreover, hotel operators should aim to provide enhanced indoor air quality by equipping their hotels with air conditioners and air purifiers, since customers are very likely expecting these during smoggy days. In their daily operations, hotels should try to improve their customer care by informing tourists about the air quality, offering advice to help travelers better organize their local itinerary, and inviting customers to revisit during good seasons. Besides, offering more indoor entertainment facilities and activities can be a good option that should boost customer satisfaction during periods of smoggy weather.

Finally, it is important for the tourism industry to take note of the detrimental impact of air pollution on inbound tourism. Although pollution-associated economic development may bring about a short-term increase in the number of inbound tourists, policymakers should be aware that these tourists are much less likely to become repeat tourists. Less valuable repeat tourists can ultimately lead to a long-term loss for the tourism market. Therefore, confining the air pollution problem and pursuing an environment-friendly economic development approach are recommended for regulators, which, in turn, should achieve balanced tourism and economic development in both the short and long terms.

4.3 Limitations and Avenues for Future Research

Despite the comprehensive investigation and analysis performed in this dissertation, it was not without limitations. The limitations of each study are as follows.

First, this literature review study summarized the determinants of OCR helpfulness from academic works published over a decade to consolidate the consistent results, decipher contradictory findings, and discuss the possible explanations for these mixed findings. However, only journal articles were included in the review sample. Conference publications are also worth reviewing so as to identify emergent research trends.

Second, Paper 2 and Paper 4 focused only on reviews written in English on only one OCR website. One should be cautious about generalizing the conclusions to consumers with different cultures. Specifically, for Paper 2, I assumed that at least a certain portion of the OCRs in the sample had been written by non-native English speakers. These reviewers may lack the ability to express their emotions as well as they could do in their native language. Furthermore, cultural differences between reviewers or readers were not accounted for. The
diversity of cultural norms and values might have influenced how the content of an OCR (particularly its sentiments) was written and interpreted. Therefore, future studies should compare reviews written in different languages and posted on multiple platforms while accounting for cultural differences. Moreover, information overload, which is a growing issue hampering consumers’ appreciation of OCRs, was not considered in this study. Future studies should incorporate the effect of perceived information overload to confirm the validity of the findings of this study.

Third, although the research design of Paper 3 did not suffer from a language constraint, some limitations of this study must be highlighted. Since the contents of the OCRs and MRs were not the focus of this study, the MR styles (e.g., polite versus impolite responses) were assumed to have been generally polite (Sparks et al. 2016; Zhang and Vásquez 2014). The psychological process of how MRs affect consumer decision-making was also not investigated. In this regard, conducting content analysis and performing survey or experiment-based studies in the future would help address this problem. Because only the benefits of IS use were focused on, this study did not consider the risks and costs of IS implementation. This study investigated only one IS service, namely, the MRF. As there are many other diverse types of IS service, I believe testing the effect of the dissipating benefits of IS in the context of other IS services would offer valuable insights.

Fourth, Paper 4 constructed a sample of the treatment group that included travelers whose previous trips were influenced by air pollution issues, using keyword matching. Although I did my best to construct a representative keyword list, it is possible that travelers who had encountered polluted air did not mention this issue in their reviews. These travelers could have been assigned to the control group, thereby potentially reducing the variance between the control and treatment groups. Since the estimation could have been relatively conservative, the actual effect of air pollution on tourism may be even more severe.

Finally, even though the PSM technique is helpful in controlling the issues caused by selection bias and endogeneity by constructing statistically equivalent treatment and control groups from observational data, it still has the limitation of accounting only for observable covariates (Garrido et al. 2014). Thus, there could have been unobserved differences between the treated and comparison groups in this study.

### 4.4 Conclusions

This dissertation was motivated by the desire to use the rich mine of social big data generated online to improve consumers’ decision-making process, enhance and enrich customer experience, birth new business models, and develop new products and services. However, the sheer volume of social big data overwhelms business stakeholders and makes the data difficult to comprehend. Controversial research findings impede various stakeholders from fully appreciating the bountiful benefits of social big data. Therefore, in this dissertation, I took the tourism and hospitality industry as a focus point and set out to discover how
social big data can benefit stakeholders in the industry. This overarching research question was broken down into sub-questions and answered from the perspective of different stakeholders: the individual consumers, companies, and industries.

In the consumer-level study, I applied NLP on social big data to bridge qualitative texts and statistical analysis. Drawing on the attribution theory, this study empirically explored the associations of emotional and factual content with the evaluation of OCR helpfulness using a large set of OCR data. By quantifying review content structures in terms of review sidedness, information factuality, and review sentiment at the beginning of a review, I obtained evidence that two-sidedness reduces OCR helpfulness, while the reviewer reputation alleviates this negative impact. The findings of this study also suggest that information factuality increases OCR helpfulness, and the appearance of positive sentiments alongside fact-based information strengthens this effect. Finally, OCRs that begin with an objective narrative tended to receive higher helpfulness ratings. This study contributes to the growing body of literature on OCR helpfulness by confirming the importance of review content structures from the attribution theory perspective. This study also contributes to the literature by reconciling contradictory findings and by introducing an approach to quantifying review content structures to investigate their effects on OCR helpfulness.

Based on a life cycle view of IS availability, the company-level study investigated the timing effect of the companies’ MRF usage in the hospitality industry. Through the lens of Kano’s theory of attractive quality, this study attested to the effect of dissipating benefits of MRF adoption by analyzing a big dataset of OCRs and MRs. Among the companies that used MRF, the capability of the MRF to improve business performance dissipated over time, although these companies gained constant advantages over the companies that did not adopt MRF.

The industry-level study introduced a novel approach to appreciating social big data by investigating how air pollution hampers tourists from revisiting a city or a country. A large set of OCR data pertaining to actual travel behavior was used to understand the influence of environmental factors on collective traveler behavior patterns in non-laboratory settings. Furthermore, by using the horn effect as the primary theoretical lens and incorporating the destination image and the country image, this study investigated a possible extending effect of air pollution: the trip experience to a city may influence tourists’ future visit behavior to other cities in the country. Understanding this extending effect would enrich the country and destination image literature in the tourism domain and unveil novel implications for the tourism industry.


