International Journal of Hospitality Management 44 (2015) 120–130

Contents lists available at ScienceDirect

International Journal of Hospitality Management

journal homepage: www.elsevier.com/locate/ijhosman

What can big data and text analytics tell us about hotel guest experience and satisfaction?

Zheng Xiang[a,*], Zvi Schwartz[b], John H. Gerdes Jr.[c], Muzaffer Uysal[a]

[a] Department of Hospitality and Tourism Management, Pamplin College of Business, Virginia Tech, Blacksburg, VA 24061, USA
[b] Department of Hotel, Restaurant & Institutional Management, University of Delaware, Newark, DE 19716, USA
[c] Department of Integrated Information Technology, College of Hospitality, Retail, & Sport Management, University of South Carolina, Columbia, SC, USA

ARTICLE INFO

Keywords:
Big data
Text analytics
Guest experience
Satisfaction
Hotel management

ABSTRACT

The tremendous growth of social media and consumer-generated content on the Internet has inspired the development of the so-called big data analytics to understand and solve real-life problems. However, while a handful of studies have employed new data sources to tackle important research problems in hospitality, there has not been a systematic application of big data analytic techniques in these studies. This study aims to explore and demonstrate the utility of big data analytics to better understand important hospitality issues, namely the relationship between hotel guest experience and satisfaction. Specifically, this study applies a text analytical approach to a large quantity of consumer reviews extracted from Expedia.com to deconstruct hotel guest experience and examine its association with satisfaction ratings. The findings reveal several dimensions of guest experience that carried varying weights and, more importantly, have novel, meaningful semantic compositions. The association between guest experience and satisfaction appears strong, suggesting that these two domains of consumer behavior are inherently connected. This study reveals that big data analytics can generate new insights into variables that have been extensively studied in existing hospitality literature. In addition, implications for theory and practice as well as directions for future research are discussed.

1. Introduction

Social media and consumer-generated content on the Internet continue to grow and impact the hospitality industry (Browning et al., 2013; Xiang and Gretzel, 2010). The tremendous growth of these data-generating sources has inspired the development of new approaches to understanding social/economic phenomena in a variety of disciplines (Wood et al., 2013; George et al., 2014). The so-called big data analytics approach emphasizes and leverages the capacity to collect and analyze data with an unprecedented breadth, depth, and scale to solve real-life problems (Ginsberg et al., 2009; Manyika et al., 2011; Mayer-Schönberger and Cukier, 2013). In the hospitality field there is a growing interest in utilizing user-generated data to gain insights into research problems that have not been well understood by conventional methods (e.g., Ye et al., 2009a; Yang et al., 2013). Indeed, big data analytics opens the door to numerous opportunities to develop new knowledge to reshape our understanding of the field and to support decision making in the hospitality industry. However, while a handful of studies have employed new data sources to tackle important research problems, they were conducted on an *ad hoc* basis and the application of the big data analytics approach in hospitality is yet to be well developed and established.

The goal of this study is to explore and demonstrate the utility of big data analytics by using it to study core hospitality management variables that have been extensively studied in past decades. Specifically, hotel guest experience and satisfaction have long been a topic of interest because it is widely recognized that they contribute to customer loyalty, repeat purchases, favorable word-of-mouth, and ultimately higher profitability (see Oh and Parks, 1997). Particularly, the hotel industry is highly competitive in that hotel firms offer essentially homogeneous products and services, which drive the desire of hotels to distinguish themselves among their competitors. As such, guest satisfaction has become one of the key measures of a hotel's effectiveness in outperforming others. Since the 1970s a plethora of studies has been conducted with the aim to understand the components and antecedents of guest satisfaction (e.g., Choi and Chu, 2001; Hunt, 1975; Mattila and O'Neill, 2003; Oh, 1999; Pizam et al., 1982; Su, 2004; Wu and Liang, 2009).

* Corresponding author. Tel.: +1 5402313262.
E-mail addresses: philxz@vt.edu (Z. Xiang), zvi@UDel.edu (Z. Schwartz), gerdes@mailbox.sc.edu (J.H. Gerdes Jr.), samil@vt.edu (M. Uysal).

http://dx.doi.org/10.1016/j.ijhm.2014.10.013
0278-4319/

While this line of research offers a variety of perspectives on guest satisfaction, the vast majority of existing studies primarily relied upon conventional research techniques such as consumer surveys or focus group interviews to gauge what leads to guest satisfaction. As such, whether we can develop novel and meaningful insights into these building blocks of hospitality management using big data analytics becomes an intriguing research question.

This study employed one of the most important types of consumer-generated content, i.e., online customer reviews of hotel properties, to understand hotel guest experience and its relationships with guest satisfaction. Text analytics was applied to first deconstruct a large quantity of customer reviews collected from Expedia.com and then examine its association with hotel satisfaction ratings. Thus, the analytics approach aimed to gain insights into the nature and structure of guest experience expressed when a customer gave a specific satisfaction rating for the hotel he/she has stayed in. This paper is organized as follows: following the introduction, the subsequent section reviews literature on the big data analytics approach and hotel guest experience and satisfaction. Research questions are formulated with the focus on using online customer reviews to enrich our understanding of these constructs. The methodology section details data collection and the text analytical approach utilized to answer the research questions. Findings are then presented and discussed. Finally, the study's contributions to literature and practice as well as directions for future research are discussed.

2. Research background

2.1. Big data analytics and business intelligence

Big data is being generated through many sources including Internet traffic (e.g., clickstreams), mobile transactions, user-generated content, and social media as well as purposefully captured content through sensor networks, business transactions, and many other operational domains such as bioinformatics, healthcare, and finance (George et al., 2014). Big data analytics aims to generate new insights that can meaningfully and, oftentimes in real time, complement traditional statistics, surveys, and archival data sources that remain largely static. The classic example of big data analytics is the pioneer study using Google search queries to detect epidemic diseases in the society (Ginsberg et al., 2009). As demonstrated by the study, big data analytics leads to a profound epistemological change that reframes key questions about the constitution of knowledge, the processes of research, how we should engage with information, and the nature and the categorization of reality (Boyd and Crawford, 2012). As such, big data analytics can be seen as a new research paradigm, rather than a uniform method, that may utilize a diverse set of analytical tools to make inferences about reality using large data. Importantly, although big data analytics does not preclude hypothesis testing, it is often applied to explore novel patterns or predict future trends from the data (Aiden and Michel, 2014). While it is widely accepted as a new approach to knowledge creation, there has been recently voice of concerns about the potential pitfall of spurious correlations and thus calls for theory-based approaches to big data analytics (Boyd and Crawford, 2012; Marcus and Davis, 2014)

One of the application areas of growing importance is the so-called business intelligence in that big data analytics can be used to understand customers, competitors, market characteristics, products, business environment, impact of technologies, and strategic stakeholders such as alliance and suppliers. Many examples and cases have been cited to illustrate the applications of big data analytics to discover and solve business problems (Mayer-Schönberger and Cukier, 2013). Mining social media and consumer-generated content has attracted much attention for their value as public and community data (George et al., 2014). For instance, research has demonstrated that online consumer reviews can be used to predict product quality (Finch, 1999), stock market volatility (Antweiler and Frank, 2004; Schumaker and Chen, 2009) and box office sales in the movie industry (Duan et al., 2008). It has been found that online news postings have sufficient linguistic content to be predictive of a firm's earnings and stock returns (Tetlock et al., 2008). More recently, Ghose and Ipeirotis (2011) used text content and reviewer characteristics to estimate the helpfulness and economic impact of online hotel product reviews. Abrahams et al. (2012) devised a technique to detect automobile defects through online consumer discussion forums. Moreover, it has been shown that marketing tools such as product recommender systems can be developed based upon the mining of consumer-generated content in combination with other data sources (Ghose et al., 2012).

Due to the volume and unstructured nature of social media and consumer generated content, opinion mining and sentiment analysis, i.e., the so-called text analytics, plays an important role in big data analytics. Indeed, opinion mining and sentiment analysis is considered well-suited to various types of market intelligence applications (Pang and Lee, 2008). Sentiment-analysis technologies for extracting opinions from unstructured human-authored documents can be excellent tools for handling many business intelligence tasks including reputation management, public relations, tracking public viewpoints, as well as market trend prediction. Broadly speaking, sentiment analysis and opinion mining denote the same techniques that are derived from and based upon natural language processing (NLP), information retrieval (IR), information extraction (IE), and artificial intelligence (AI). Typical tasks of sentiment analysis include (1) finding documents relevant for a specific topic or purpose; (2) pre-processing collected documents, e.g., tokenizing documents into single words and extracting relevant information from them; and (3) identifying the sentiment surrounding the product or company (Schmunk et al., 2013). In comparison with the broader scope of text mining approach, sentiment analysis may be considered a special type of text mining with the focus on identification of subjective statements and contained opinions and sentiments, particularly in consumer-generated content on the Internet.

2.2. Hotel guest experience and satisfaction

Hotel guest satisfaction is a complex human experience within a hospitality service setting. The study of guest satisfaction was initiated as early as the 1970s. Different definitions of guest satisfaction have emerged. Hunt (1975) considers satisfaction as an evaluation on which the customers have experienced with the services is at least as good as it is supposed to be, while others (e.g., Oliver, 1981) define customer satisfaction as an emotional response to the use of a product or service. Oh and Parks (1997) postulate that satisfaction involves cognitive and affective processes, as well as other psychological and physiological influences. A commonly used definition of customer satisfaction adopts a disconfirmation perspective of consumer satisfaction/dissatisfaction, suggesting that satisfaction is the result of the interaction between a consumer's pre-purchase expectation and post-purchase evaluation (Engel et al., 1990). In the tourism literature, various perspectives have been employed to conceptualize the concept of tourist satisfaction including the expectation/disconfirmation paradigm, the equity view, the norm view, as well as the perceived overall performance (see Yoon and Uysal, 2005 for a comprehensive review).

From the managerial point of view, it is, perhaps, more important to understand the components or antecedents of hotel guest satisfaction. For example, it has been conceptualized that the hotel product consists of several levels. That is, the core product, i.e.,