Post-print: STM-Based Analysis of Traveler Preference Differences for Different Hotel Tiers

Authors: Yang Haixia, Wu Weifang, Sun Hanlin

Date: 2017-11-08T00:00:00+00:00

Abstract

【Objective】This study aims to investigate the differences in preferences for products or services provided by different categories of hotels among different types of travelers in the Web 2.0 era.

【Method】Hotels are categorized into luxury hotels and budget hotels, and travelers are divided into five travel types. The study analyzes the rating patterns of different travelers toward luxury and budget hotels, and employs the Structural Topic Model (STM) to conduct a detailed analysis of hotel online review texts, mining the topics of online reviews and analyzing the differences in service preferences of travelers from each travel type toward different categories of hotels.

【Results】Experimental results show that under all five travel types, travelers' average ratings for luxury hotels are higher than those for budget hotels; different types of travelers exhibit differences in preferences for products or services provided by different categories of hotels.

【Limitations】The experimental data is insufficient; factors such as gender and age that may influence the numerical ratings and textual content of online reviews are ignored.

【Conclusion】Analyzing the differences in preferences of different types of travelers toward different categories of hotels helps hotel managers formulate service supply strategies and assists consumers in making purchase decisions.

Full Text

Analyzing Travelers' Preferences for Different Hotel Categories Based on Structural Topic Model

Yang Haixia, Wu Weifang, Sun Hanlin
(School of Economics and Management, Wuhan University, Wuhan 430072,

China)

Abstract

[**Objective**] This study aims to detect preference differences among various types of travelers for products or services offered by different hotel categories in the Web 2.0 era. [**Methods**] We classified hotels into luxury and budget categories and travelers into five travel types, then analyzed the rating patterns of different travelers toward luxury and budget hotels. Using the Structural Topic Model (STM), we conducted a detailed analysis of hotel online review texts to mine review topics and examine service preference differences among travelers of various types toward different hotel categories. [**Results**] The experimental results show that across all five travel types, travelers' average ratings for luxury hotels were higher than those for budget hotels, and that different types of travelers exhibited distinct preferences for products or services offered by different hotel categories. [**Limitations**] The experimental data were insufficient, and factors such as gender and age that might influence numerical ratings and textual content of online reviews were ignored. [**Conclusions**] Analyzing preference differences among different types of travelers for different hotel categories can help hotel managers formulate service supply strategies and assist consumers in making purchase decisions.

Keywords: Online reviews; Rating patterns; Hotel category; Travel type; STM; Review topics
Classification Codes: F59; G350

1. Introduction

With the rapid development of the Web 2.0 era, an increasing number of users express their opinions on online platforms. Online review websites represent a highly popular Web 2.0 application [1], enabling users to publish their purchasing experiences regarding products or services at minimal cost [2], thereby generating User-Generated Content (UGC). Research indicates that UGC not only helps potential consumers make purchasing decisions but also assists relevant managers in improving product or service quality and gaining more revenue [3-5].

Existing studies show that travelers with different travel types have varying preferences for hotel services [6-9]. Meanwhile, different hotel categories have distinct positioning and target markets, leading to different consumer expectations and accommodation experiences [10]. However, when conducting textual analysis of hotel online reviews, previous scholars have primarily focused on topic extraction [11-13], with few considering consumers' travel types and hotel categories to analyze preference differences for hotel services. Therefore, this paper categorizes travelers into "business travel," "couple travel," "family travel," "friends travel," and "solo travel" based on their travel purposes [14-15], and divides hotels into luxury and budget categories according to star ratings

[16] to investigate rating differences and service preference variations among different traveler types toward different hotel categories.

This study conducts basic descriptive statistical analysis of numerical ratings from hotel online reviews and employs the Structural Topic Model (STM) [17-19] to extract topics from review texts. By incorporating "hotel category" as a covariate for both topic prevalence and topic content, we analyze service preference differences among various traveler types toward different hotel categories.

Hotel online reviews are experience evaluations of hotels generated by travelers on hotel websites or third-party review platforms [2-5]. Online reviews can attract potential consumers, increase visit duration, and enhance seller-buyer relationship quality. Existing research on online reviews has primarily focused on how variables such as review quantity, rating scores, review valence, review polarity, and rating variance affect product sales or purchase intentions, with these variables mostly derived from numerical ratings [20-21], rarely involving review textual content.

When writing hotel reviews on TripAdvisor.com, travelers are typically asked, "What sort of trip was this?" and must select one of five travel types (business, couple, family, friends, and solo travel) that applies to their trip [14-15]. Previous research indicates that different travelers have different travel purposes and expectations [22-23], which leads to varying preferences for hotel facilities or services such as location, cleanliness, and value for money. Rivers et al. [6] found that business travelers are highly concerned with hotel location convenience and internet availability, and less concerned with price. Lee et al. [7] demonstrated that couple travelers' satisfaction is easily influenced by destination climate changes and romantic atmosphere. Lai et al. [8] showed that compared to solo travelers, those traveling with family or friends place greater emphasis on safety.

Domestic research on tourist preferences started later than international research. In 1987, Bao Jigang [24] analyzed the relationship between tourist preferences, travel motivations, tourist behavior, and spatial interaction in tourism destinations. In 2006, Yang Rui [25] used regression analysis to examine the travel preferences, demand characteristics, and consumption patterns of university student tourists in Xi' an, comparing them with those of general tourists. In 2010, Yue Dongju et al. [26] used questionnaire surveys to analyze the travel preferences of domestic tourists in Xi' an. In 2011, Lei Lili [27] conducted preference analysis of tourist motivations, experiences, and products in Dalian, comparing travel behavior characteristics across different travel types.

However, current research lacks studies on the relationship between hotel category and consumer preferences. Although Gu Xiuling [28] considered tourists' evaluations of different star-rated hotels in an empirical study on consumer behavior of four customer types at Taihu Lake resort hotels, the study did not involve review textual content. Topic discovery analysis of hotel online review texts [29-32] can effectively identify preference differences among different traveler types. Therefore, this paper employs the unsupervised STM to analyze preference differences among various traveler types for different hotel categories.

As a member of topic models [33-34], STM assumes that documents are mixtures of topics and that topics are mixtures of words. We use STM to analyze differences in textual reviews of hotels across different travel types. Additionally, STM allows researchers to incorporate control variables as covariates for comparative analysis. This study primarily includes "hotel category" as a covariate to analyze topic prevalence and content, detecting travelers' preferences for different hotel categories across various travel types.

Like other topic models [33-34], STM is a generative model, meaning it defines a data generation process for each document and uses data to find the parameters that best fit the model. In the STM generative process, documents (D1, D2, ···, Dn), topics (T1, T2, ···, Tn), and topic-word distributions (w1, w2, ···, wn) are generated in connection with document metadata Xd (where d denotes a document). Similar to LDA (Latent Dirichlet Allocation) [34] and other topic models, topics are defined as mixtures of words, with each word belonging to a topic with a certain probability; documents are mixtures of topics, meaning a review can consist of multiple topics. Therefore, the sum of probabilities of all topics in a review equals 1, and the sum of probabilities of a word across all topics equals 1.

During document generation, both topic prevalence and topic content can be expressed as functions of document metadata. Topic prevalence indicates how much content in a review is related to a topic, while topic content is represented by words in the topic. Thus, topic prevalence covariates indicate topic prevalence, and topic content covariates explain topic content. It is important to note that STM can include only topic prevalence covariates, only topic content covariates, or no covariates at all.

In an STM model with k topics, the generation process for each review (document d) can be described as follows:

(1) Generate document-topic proportions d | X, ~ Logistic(γ), where γ represents the coefficients.

(2) Generate word distributions representing each topic (k) using baseline word distribution (m), topic deviations (Kk), covariate deviations (Kg), and their interaction deviations Ki = (kgd): d,kβ exp(m + Ki(kgd)).

(3) For each word in the document, n (Λ): Assign the word to a topic based on document-specific topic distribution: z ~ Multinomial(d). Generate a word from the selected topic: w|z ~ Multinomial(βd,kz).

In this paper, we use "hotel category (luxury and budget hotels)"as a covariate in STM to estimate document-topic and topic-word probabilities from hotel online reviews, analyzing service preference differences among various traveler types for different hotel categories. To fit the model, we employ the Semi-Collapsed

Variational Expectation-Maximization Algorithm [17-19] to estimate model parameters based on convergence, thereby obtaining the topic distribution d for each review.

4. Empirical Analysis

4.1 Data Source TripAdvisor.com is a globally renowned travel review website with 350 million monthly unique visitors. In academic research, it is one of the most studied hotel review platforms, with many researchers selecting its data for analysis in recent years [2,35-37]. Therefore, this study' s use of TripAdvisor.com hotel online reviews for experimental analysis holds practical significance.

We collected hotel reviews from Las Vegas posted between January 1, 2012, and September 30, 2013, in October 2013. For each review, we collected the review ID, overall rating, review text, travel type profile, trip time, hotel star class, and hotel ID. Travel types included: Business, Couple, Family, Friends, and Solo. To analyze review text content, we removed non-English reviews, resulting in 101,846 reviews. Hotel star ratings in the collected data ranged from 1.5 to 5. We broadly divided hotels into two categories: luxury hotels with star rating ≥ 4 (4, 4.5, 5) and budget hotels with star rating < 4 (1.5, 2, 2.5, 3, 3.5) [16]. The final dataset contained 68,317 luxury hotel reviews and 33,529 budget hotel reviews.

4.2 Descriptive Statistical Analysis We conducted descriptive statistical analysis of ratings for each traveler type across different hotel categories, as shown in Table 1 .

Table 1 Descriptive Statistics of Hotel Online Review Ratings

(Note: N represents the number of reviews, Mean represents the average rating, and SD represents the standard deviation of ratings.)

The analysis reveals that the total number of online reviews for luxury hotels is significantly higher than that for budget hotels, and luxury hotels receive higher average ratings than budget hotels. Additionally, rating scores vary across different traveler types: couple travelers have the highest overall average rating, while business travelers have the lowest [2]. Considering hotel star categories, we find that business travelers give the lowest average ratings for both luxury and budget hotels; couple travelers give the highest average ratings for luxury hotels, while family travelers give the highest average ratings for budget hotels.

The average rating score of hotel online reviews influences potential consumers' hotel purchase decisions [2-5]. However, to improve a hotel' s average rating, it is necessary to understand why travelers choose a particular hotel category and which facilities or services they care about, thereby formulating appropriate product strategies to meet their actual needs.

4.3 Analysis of Travelers' Preference Differences for Different Hotel Categories To analyze hotel online review texts and mine review topics, we paired the textual analysis data to examine service preference differences for the same traveler type across different hotel categories. For each traveler type, we randomly selected 4,000 hotel online reviews (2,000 luxury and 2,000 budget). Using the stm package in R software [17], we preprocessed the review texts by converting all words to lowercase, removing stop words, eliminating numbers and punctuation, and applying stemming.

Before constructing the topic model, it is necessary to determine the number of topics—a model selection issue. Models with too few topics produce broad content, while those with too many generate numerous small subtopics that are difficult to interpret. Roberts et al. [18] argue that topic quality and interpretability should satisfy two criteria: (1) high internal coherence within topics (e.g., high-frequency words in a topic about "Internet" should include Internet, WiFi, Fee, etc.); and (2) distinct high-frequency words across topics (e.g., the high-frequency word "Staff" in the "Staff Service" topic should not appear in the "Location" topic).

We set the number of topics from 5 to 15 and found that 10 topics best fit our experimental model by comparing topic quality. Therefore, we used "hotel category" as a covariate in STM to analyze preference differences among various traveler types for luxury versus budget hotels.

(1) STM Modeling

In topic models, each review can be represented as a probability mixture of topics. For instance, based on business travelers' reviews of luxury hotels, we can calculate their attention proportion to each topic (A); similarly, based on their reviews of budget hotels, we can calculate C. For each travel type, using A - C reveals service preference differences for the same traveler type across different hotel categories. We define a difference as significant if the absolute value of the subtraction exceeds 0.03.

(2) Preference Difference Analysis

Business Travelers

As shown in Figure 1 [Figure 1: see original paper], compared to budget hotels, business travelers selecting luxury hotels also pay attention to scenery (Topic8); whereas compared to luxury hotels, those selecting budget hotels focus more on cleanliness (Topic3) and transportation (Topic4). Related research indicates two main reasons for business travelers choosing budget hotels: first, some business travelers must bear accommodation expenses themselves and have limited budgets, making them concerned about value for money; second, some are compelled to choose a particular budget hotel due to its convenient location [38].

Couple Travelers

As shown in Figure 2 [Figure 2: see original paper], compared to budget hotels, couple travelers selecting luxury hotels pay more attention to staff service (Topic6); whereas compared to luxury hotels, those selecting budget hotels focus more on cleanliness (Topic2). Related research shows that couple travel emphasizes mood and romance, with satisfaction easily influenced by perceived romantic atmosphere [7]. For high-end hotels, hardware facilities and services are relatively complete, making software services (such as staff service) crucial for couple travelers' satisfaction. For budget hotels, considering economic constraints, couple travelers focus more on cleanliness.

Family Travelers

As shown in Figure 3 [Figure 3: see original paper], compared to budget hotels, family travelers selecting luxury hotels pay more attention to hotel scenery (Topic8); whereas compared to luxury hotels, those selecting budget hotels focus more on transportation convenience (Topic6). This is because family travelers usually accompany children or elderly members [8], and even when accepting lower-category hotels due to budget constraints, they prioritize transportation convenience.

Friends Travelers

As shown in Figure 4 [Figure 4: see original paper], compared to budget hotels, friends travelers selecting luxury hotels pay more attention to hotel atmosphere (Topic6); whereas compared to luxury hotels, those selecting budget hotels focus more on bedroom features (Topic7).

Solo Travelers

As shown in Figure 5 [Figure 5: see original paper], compared to budget hotels, solo travelers selecting luxury hotels pay more attention to overall hotel feeling (Topic1), including location, staff, and cleanliness.

5. Discussion

5.1 Limitations and Future Research Directions While this study achieves certain breakthroughs and improvements based on existing research, it has limitations that provide directions for future research:

(1) The dataset was sourced from only one review website (TripAdvisor.com) and one city (Las Vegas), which may limit the representativeness of the results. Future research could collect data from multiple online review websites and include hotel review data from multiple cities to conduct textual analysis from a geographical perspective.

(2) This study only included "hotel category" as a covariate in the STM model, examining only travelers' preference differences for different hotel categories while ignoring other factors that might influence hotel service preferences (such as reviewer gender, age, etc.). Future research could incorporate additional covariates to detect preference differences for hotel services among different travelers.

5.2 Managerial Implications Hotel category influences review ratings, with luxury hotels receiving higher overall average ratings than budget hotels. Luxury hotels have complete facilities and well-developed hardware services. Therefore, to improve overall ratings, luxury hotels should strengthen staff service awareness to meet customers' psychological needs. Budget hotels, while maintaining lower operating costs and unable to provide extensive facilities, can improve service attitudes, enhance staff service quality—especially front desk check-in and check-out services—as customers' first and last impressions significantly impact ratings. Additionally, budget hotels should strengthen cleanliness management to reduce the probability of receiving low-score reviews.

Different traveler types exhibit different rating patterns across hotel categories. Business travelers give the lowest and most stringent ratings, while couple travelers give the highest and most lenient ratings. For luxury hotels, managers can adjust service supply strategies based on traveler types, such as assigning rooms with better internet signals to business travelers to meet their business needs, and rooms with better views to couple travelers to satisfy their scenic appreciation needs. Similarly, budget hotel managers can provide rooms with better internet to business travelers and more food options to family travelers. Additionally, hotels could consider offering free or low-cost airport pickup services to reduce customers' costs in locating the hotel.

References

[1] Buhalis D, Law R. Progress in Information Technology and Tourism Management: 20 Years on and 10 Years After the Internet—The State of eTourism Research [J]. Tourism Management, 2008, 29(4): 609-623.

[2] Banerjee S, Chua A Y K. In Search of Patterns Among Travellers' Hotel Ratings in TripAdvisor [J]. Tourism Management, 2016, 53: 125-131.

[3] Brown J, Broderick A J, Lee N. Word of Mouth Communication Within Online Communities: Conceptualizing the Online Social Network [J]. Journal of Interactive Marketing, 2007, 21(3): 2-20.

[4] Goldenberg J, Libai B, Muller E. Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth [J]. Marketing Letters, 2001, 12(3): 211-223.

[5] Ye Q, Law R, Gu B. The Impact of Online User Reviews on Hotel Room Sales [J]. International Journal of Hospitality Management, 2009, 28(1): 180-182.

[6] Rivers M J, Toh R S, Alaoui M. Frequent-stayer Programs: the Demographic, Behavioral, and Attitudinal Characteristics of Hotel Steady Sleepers[J]. Journal of Travel Research, 1991, 30(2): 41-45.

[7] Lee C F, Huang H I, Chen W C. The Determinants of Honeymoon Destination Choice—The Case of Taiwan [J]. Journal of Travel & Tourism Marketing, 2010, 27(7): 676-693.

[8] Lai L, Graefe A R. Identifying Market Potential and Destination Choice Factors of Taiwanese Overseas Travelers [J]. Journal of Hospitality & Leisure Marketing, 1999, 6(4): 45-65.

[9] Poston R S. Using and Fixing Biased Rating Schemes [J]. Communications of the ACM, 2008, 51(9): 105-109.

[10] Munar A M, Jacobsen J K S. Motivations for Sharing Tourism Experiences Through Social Media [J]. Tourism Management, 2014, 43: 46-54.

[11] Mukherjee A, Liu B. Aspect Extraction Through Semi-supervised Modeling [C]. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics. 2012: 339-348.

[12] Wang H, Lu Y, Zhai C. Latent Aspect Rating Analysis on Review Text Data: A Rating Regression Approach [C]. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA. 2010: 783-792.

[13] Zhao Y, Dong S, Yang J. Effect Research of Aspects Extraction for Chinese Hotel Reviews Based on Machine Learning Method [J]. International Journal of Smart Home, 2015, 9(3): 23-34.

[14] Dolnicar S. Business Travellers' Hotel Expectations and Disappointments: A Different Perspective to Hotel Attribute Importance Investigation [J]. Asia Pacific Journal of Tourism Research, 2002, 7(1): 29-35.

[15] O' Connor P. User-generated Content and Travel: A Case Study on Tripadvisor.com [A]. // Information and Communication Technologies in Tourism 2008[M]. Springer Vienna, 2008.

[16] Gao Baojun, Sun Hanlin, Wang Hanning. Influence of Online Reviews on Hotels' Full-occupancy Rates[J]. Tourism Tribune, 2016, 31(4): 109-117.

[17] Roberts M E, Stewart B M, Tingley D. STM: R Package for Structural Topic Models [J]. General Information, 2014, 57(1): 445-460.

[18] Roberts M E, Stewart B M, Tingley D, et al. Structural Topic Models for Open-Ended Survey Responses [J]. American Journal of Political Science, 2014, 58(4): 1064-1082.

[19] Lucas C, Nielsen R A, Roberts M E, et al. Computer-Assisted Text Analysis for Comparative Politics [J]. Political Analysis, 2015, 23(2): 254-277.

[20] Liu Y. Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue [J]. Journal of Marketing, 2006, 70(3): 74-89.

[21] Sun M. How Does the Variance of Product Ratings Matter? [J]. Management Science, 2012, 58(4): 696-707.

[22] Ariffin A A M, Maghzi A. A Preliminary Study on Customer Expectations of Hotel Hospitality: Influences of Personal and Hotel Factors [J]. International Journal of Hospitality Management, 2012, 31(1): 191-198.

[23] Liu S, Law R, Rong J, et al. Analyzing Changes in Hotel Customers' Expectations by Trip Mode [J]. International Journal of Hospitality Management, 2013, 34(1): 359-371.

[24] Bao Jigang. Research on Travelers' Behavior[J]. Social Scientist, 1987(6): 19-22.

[25] Yang Rui. The Study on the University Student' s Travel Behavior Pattern in Xi' an [D]. Xi' an: Shaanxi Normal University, 2006.

[26] Yue Dongju, Yang Yuan. An Empirical Analysis of Domestic Tourists' Traveling Preferences in Xi'an [J]. Journal of Xi'an University of Arts & Science: Natural Science Edition, 2010, 13(2): 116-119.

[27] Lei Lili. Research on Tourism Preference and Behavior Characteristics of Domestic Tourist in Dalian [D]. Dalian: Dongbei University of Finance & Economics, 2011.

[28] Gu Xiuling. An Empirical Research of Consumers' Behavior[J]. Journal of Mudanjiang Normal University: Philosophy Social Sciences Edition, 2014(5): 35-38.

[29] Zhao W X, Jiang J, Yan H, et al. Jointly Modeling Aspects and Opinions with a MaxEnt-LDA Hybrid [C]. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. 2010: 56-65.

[30] Titov I, McDonald R. Modeling Online Reviews with Multi-grain Topic Models [C]. In: Proceedings of the 17th International Conference on World Wide Web. ACM, 2008: 111-120.

[31] Kasper W, Vela M. Sentiment Analysis for Hotel Reviews[C]. In: Proceedings of Computational Linguistics Applications Conference. 2012.

[32] Chen Y S, Chen L H, Takama Y. Proposal of LDA-Based Sentiment Visualization of Hotel Reviews [C]. In: Proceedings of the 2015 IEEE International Conference on Data Mining. 2015: 687-693.

[33] Blei D M. Probabilistic Topic Models [J]. Communications of the ACM, 2012, 55(4): 77-84.

[34] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet Allocation [J]. Journal of Machine Learning Research, 2003, 3: 993-1022.

[35] Yoo K H, Gretzel U. What Motivates Consumers to Write Online Travel Reviews? [J]. Information Technology & Tourism, 2008, 10(4): 283-295.

[36] Duan H, Zirn C. Can We Identify Manipulative Behavior and the Corresponding Suspects on Review Websites Using Supervised Learning? [A]. // Secure IT Systems [M]. Springer Berlin Heidelberg, 2012: 215-230.

[37] Wu G, Greene D, Cunningham P. Merging Multiple Criteria to Identify Suspicious Reviews [C]. In: Proceedings of the 4th ACM Conference on Recommender Systems. ACM, 2010: 241-244.

[38] Blanke J, Chiesa T. The Travel & Tourism Competitiveness Index 2011: Assessing Industry Drivers in the Wake of the Crisis [R]. World Economic Forum, 2011.

Author Contributions

Yang Haixia: Conceived the research idea, designed the study, conducted experiments, drafted and revised the manuscript.
Wu Weifang: Revised the manuscript.
Sun Hanlin: Collected, cleaned, and analyzed data; revised the manuscript.

Conflict of Interest Statement

All authors declare no conflict of interest.

Supporting Data

Supporting data are self-archived by the authors, E-mail: haixiayang@whu.edu.cn.

Received: June 12, 2016
Revised: July 16, 2016

Note: Figure translations are in progress. See original paper for figures.