Master Thesis (2021)

Analyzing the food experience of Chinese Tourists in Japan through online restaurant reviews: What destination marketers can learn from big data
（レストランのオンラインレビューに見られる訪日中国人旅行者の食体験の分析：ビッグデータからもたらされる観光地域マーケティングへの示唆）

Ting ZHANG

20854515

Department of Tourism Science

Graduate School of Urban Environmental Sciences

Tokyo Metropolitan University

Supervisor: Tetsuo Shimizu

Abstract

The promotion of overseas visitors to Japan is a major objective at the present stage. Even though the inbound industry has been severely impacted by the sharp decline in the number of foreign visitors to Japan as a result of the spread of the new coronavirus infection, inbound promotion and information dissemination activities should be prioritized and strengthened. Among these, Chinese tourists account for the largest proportion of foreign visitors to Japan, and one of the most enthusiastic experiences for Chinese tourists is Japanese food. However, the reality is that due to the difference in cultural backgrounds, food habits, etc., there are many unsatisfactory and disparate differences between tourists and restaurants that need to be resolved, even though Japanese restaurants offer high-quality ingredients and cuisine and the best service. On the other hand, the development of the Internet and communication technology in recent years have made people more inclined to use and refer to reviews and blogs on online platforms before, during, and after travel, and this has created great opportunities and challenges for local tourism organizations and institutions and various tourism stakeholders. Nevertheless, the effectiveness of the appeal of these promotions and whether they are of interest to Chinese tourists is still unknown. Hence, it is of high importance to understand tourists' preferences and needs, as well as to improve on their grievances and problems. The Japan National Tourism Organization (JNTO) and other destination marketers need to adjust or change their promotional strategies as a result of recognizing and understanding these factors, such as knowing where promotion is beneficial and effective while also knowing where the effect of promotion is minimal or even negative.

Therefore, this study aims to analyze online restaurant reviews to explore the experiences of Chinese tourists in Japan related to their food activities. Furthermore, the study identifies tourist culinary preferences and needs, as well as their dissatisfactions and problems. The findings of this study can assist DMOs in determining what promotion/marketing issues enhance and decrease tourist food and cuisine experiences, and thus their sentiments about Japanese cuisine.

This study combines a machine learning model and a sentiment lexicon approach in an innovative way to explore topics and identify sentiments in complex tourist food experiences. By updating Chinese sentiment lexicons about tourism and food activities, it overcomes the limitations of Chinese sentiment analysis in the field of gastronomic experiences. It also utilizes a novel method for comparing and studying gastronomic experiences in various dimensions. The largest online review site in China, Dianping.com, was crawled and 53,442 restaurant reviews of Tokyo tourist attractions were collected for this study. After that, the LDA model was used to extract potential review topics. For further analysis and discussion, 24 topics were manually examined and filtered for use in this study, and these topics were further classified into four dimensions based on their term content and characteristics, which are Food/cuisine, Location and environment, Dining/restaurant, Cultural activity and promotion. Then, using the extracted topics as a preliminary step, sentiment analysis was performed, expanding the previous library of the Chinese sentiment lexicon, and the positive and negative reviews for each topic were counted. The topic salience-valence analysis (TSVA) was according to dimensional categories in order to quantitatively analyze the significance and positivity of each topic more intuitively.

The findings of this study show that Chinese tourists visiting Japan are very interested in and positive about Japanese culture and sweets, but their sentiments toward online promotion vary depending on the content and means. While they appreciate the Japanese cuisine and attentive service, as well as the beautiful surroundings of the restaurants, there are concerns about food safety, queuing, reservation and payment methods, and menu communication. Consequently, firstly, JNTO needs to re-evaluate and review its original online promotion methods and contents. Secondly, it needs to continuously promote positive and good images about the disaster areas and the impact on food and ingredients to eliminate foreign tourists' worries and doubts and gradually restore their trust in food safety. Finally, the research reveals a high level of interest and positive emotions among Chinese tourists for Japanese culture and desserts, as well as yakiniku and seafood cuisine, gourmet food, and restaurant environments. As a result, when developing policies and

marketing strategies aimed at attracting Chinese tourists, the JNTO and other destination marketers can focus on promoting these aspects, which will not only attract tourists' attention and curiosity but also have positive marketing effects.

Keywords

Tourism online review, Food experience, Latent Dirichlet allocation, Destination marketing

Content

LIST OF FIGURES

LIST OF TABLES

Chapter 1. Introduction

With the implementation of the plan of "Visit Japan Campaign" (welcome to visit Japan) program in 2003, formulated the "Tourism Nation Promotion Basic Law" in 2006 which introduced the "Basic Plan for Tourism Nation Promotion" and formally established inbound tourism as a pillar industry to promote the country's economic development. Since the implementation of the plan, Japan's inbound tourism has achieved fruitful development, the number of Japan's inbound tourists has increased from more than 6 million in 2011 to more than 30 million in 2019 (An increase of nearly about four times) (Fig.1). In 2030, the number of foreign visitors to Japan is expected to reach 60 million, and the amount of consumption is expected to reach 15 trillion yen, but the amount of consumption is sluggish (4.8 trillion yen in 2019), and the measures and effectiveness of information transmission are also issues. However, with the spread of the new coronavirus epidemic, inbound tourism has been hit hard (Fig.1), so it is becoming more vital at this stage to seize the opportunity to promote and disseminate information online, and more effectively to spread the appeal of traveling to Japan.

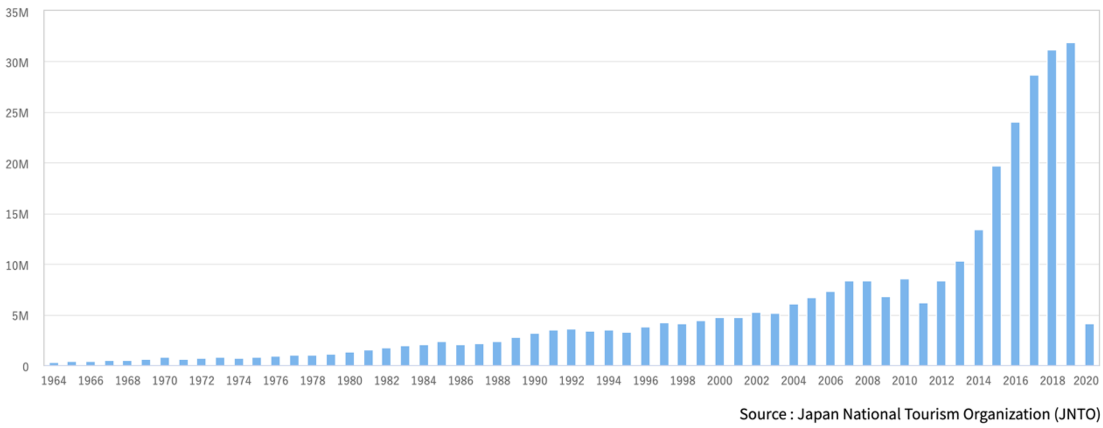

Figure 1 Overseas Residents' Visits to Japan by year

Meanwhile, Chinese visitors to Japan have grown significantly in recent years, with Chinese tourists accounting for the largest proportion of foreign visitors to Japan in terms of spending (Fig.2), and according to relevant questionnaires, Chinese tourists have shown a stronger willingness to visit Japan after Covid-19. Therefore, it is necessary for Japanese tourism-related organizations to understand the preferences and needs of Chinese tourists visiting Japan, to further enhance the attractiveness of tourist attractions, to further develop and revise policies to attract tourists, and to more effectively refine marketing strategies and information promotion on the Internet.

On the other hand, there is a Chinese proverb that says, "Food is the key to the people's livelihood," thus tasting Japanese food is a very important part of Chinese tourists' sightseeing. However, the reality is that due to the difference in cultural backgrounds, food culture, and habits, there are many unsatisfactory and disparate differences between tourists and restaurants that need to be resolved, even though Japanese restaurants offer high-quality ingredients and cuisine, and the best service and care. For example, some parts of the cuisine will be incomprehensible, and tourists will have negative feelings. Also, on the restaurant side, it may bring some uncivilized and impolite behaviors that are not in line with Japanese manners, as well as actions that disturb other customers.

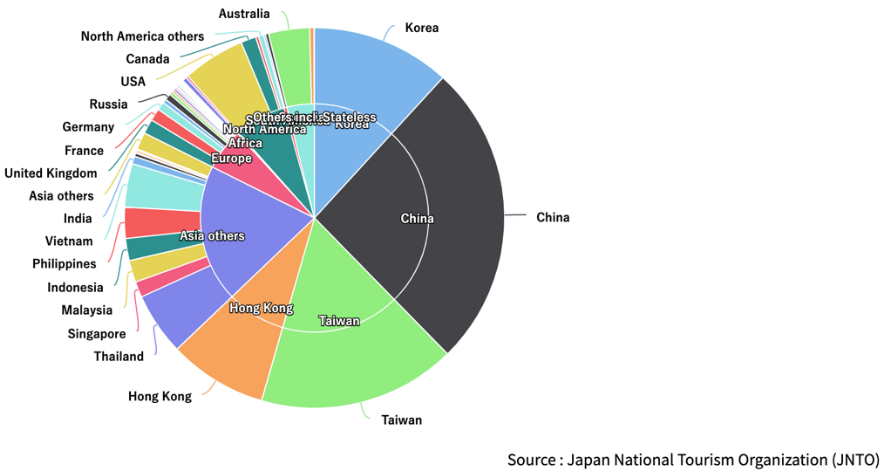

Figure 2 Overseas Residents' Visits to Japan by country, 2017

These days, with the rapid development of mobile and communication technologies, tourists are increasingly using a variety of technological tools to improve their travel experience: for example, through applications for mobile devices, virtual and augmented reality content, and interactive environments. Among these, online review sites play a crucial role. Tourists use online review sites to get a more convenient and comprehensive view of their destinations and become increasingly self-aware and independent when making travel decisions (Simeon et al., 2017). Customers are able to share their travel or purchasing experiences, feelings, and reviews online. These online reviews play a crucial role in accessing travel-related services and also have a significant impact on the decision-making behavior of other users (Gretzel & Yoo, 2008). The development of information and communication technologies has had a significant impact on traveler behavior and tourism development (Buhalis & Law, 2008). Studies have also shown that people visit about 26 websites and spend about two hours looking for reasonably priced attractions when browsing for travel-related information. Online reviews can be considered as a form of Internet communication exchange facilitated by different Internet applications, websites, review and rating sites, social networking sites (SNS), and blogs (Shafqat & Byun, 2019).

Nevertheless, the current situation is that destination markets do not understand and effectively use the information contained in these reviews, and they do not have timely and accurate access to the real voice of international visitors, their feelings and preferences, making it difficult to make use of a large amount of online review data. On the other hand, the effectiveness of destination market strategies and policies is also unknown. It is possible to determine whether policies are attracting more visitors, which strategies are more effective and beneficial, and which have minimal or negative effects by studying the reviews of international visitors. As a result, there is a demand for simple and effective methods and models for those studying destination markets.

In the current research, the author uses latent Dirichlet allocation (LDA) which is an appropriate machine learning technique, and sentiment analysis that can be used to answer the following research questions (RQs).

(1) What are the topics of online restaurant reviews? The LDA model was used to group together several terms (or words) found in online reviews on Dianping.com. Those groups of terms represent the identifiable topics of reviews.

(2) Which topics can be studied and how many dimensions can they be classified into? The current research divides topics with similar characteristics into different dimensions to study and compare them in order to explore the potential information.

(3) How positive are the reviews for each topic? The authors assess the food experience of tourists with respect to each topic. Positive and negative ratings for each topic are counted through sentiment analysis.

(4) How frequent and positive are the topics of each dimension? To obtain a deeper understanding of the food experience, this study explores their underlying topics. The differences between the expected and the observed numbers of positive reviews in each topic are used to represent the valence of each topic. The destinations will gain a competitive advantage if they can successfully turn discontent into delight (Baka, 2016).

(5) What are the findings of this research that can be used to inform and advise JNTO and other destination marketers, as well as identify problems and limitations in current strategies and policies?

Thus, the purpose of this study is ① Identify Chinese tourist culinary preferences and needs, as well as their dissatisfactions and problems. ② Assist JNTO/DMOs in determining what promotion/marketing issues enhance and decrease tourist food and cuisine experiences, and thus their sentiments about Japanese cuisine.

Chapter 2 presents relevant research and literature in four areas: eating experiences, online reviews and machine learning, review prices, and Chinese outbound travel.

Chapter 3 describes the data and methods used in this study.

Chapter 4 shows the results of the LDA model and sentiment analysis.

Chapter 5 provides further analysis and discussion of the results, showing the practical and theoretical implications of the study.

Chapter 6 discusses the limitations and future perspectives of this study, then summarize the thesis.

Chapter 2. Literature review

2.1. Food experience in tourism and hospitality research

Tourist destinations are attaching ever-greater importance to food given their ability to attract visitors, enhance travel experiences and achieve differential positioning as culinary destinations (Björk & Kauppinen-Räisänen, 2016). Food has been regarded as not only being a basic necessity for tourist consumption but also an essential element of regional culture (Jones & Jenkins, 2002). Since food has been proven to be an important means of selling the identity and culture of a destination, food consumption is regarded as one of the important factors in destination marketing development. The other reason is that food consumption enables local food producers to add value to their products by creating a tourist experience around the raw materials (Hjalager & Richards, 2003).

Moreover, it is important to recognize that food consumption is not only a means of generating revenues for a destination but also an important part of the tourist experience (Hjalager & Richards, 2003). Although there is still little literature on foods in tourism, there is much that can be borrowed from the literature on foods in unusual and non-daily contexts such as restaurants (Warde & Martens, 2000) and festivals. While occasion, ambiance, company, and celebration bring special meanings to food consumption, these unusual contexts make food experience a source of pleasure and enjoyment (Warde & Martens, 2000). Similarly, tourism can be seen as an unusual context in which food consumption gains special meanings and pleasure (Hjalager & Richards, 2003). No matter whether trying different kinds of food is the main purpose for tourists to travel, food can at least provide extra opportunities for tourists to be in a more memorable and enjoyable holiday atmosphere than they expected. Gastronomy is thus seen as an important source of marketable images and experiences for the tourist. Furthermore, the increasing complexity of gastronomic experiences requires a more holistic analytic approach, including more attention to relational and co-creational processes. Linking together different experience elements and experience phases requires more holistic and contextual research approaches. (Richards, 2021)

Food experiences are heterogeneous and complex, with different components and emphases of experience available in different contexts and for different target groups, despite the general recognition of the importance of studying destination cuisine in prior research. As a result, it's critical to comprehend how tourists interact with restaurants and cuisines, to capture their emotions and needs, to develop methods and classifications for studying food experiences, and to extract valuable data for destination market analysis.

2.2. Online reviews and Text mining with machine learning

Online reviews are becoming an invaluable source of tourism research and practice. They reflect tourists' delight and dissatisfaction with their experiences (Banerjee & Chua, 2016). Online reviews, spontaneous and insightful feedback provided by tourists, are freely available on review platforms (Guo et al., 2017) and are an authentic mix of facts, opinions, impressions, and emotions (Wilson et al., 2012). Given the growing importance of online reviews, researchers have conducted various analyses over the past decade to investigate online reviews in the hospitality and tourism sectors. The number of reviews analyzed can range from a few hundred to hundreds of thousands, as shown in the 13 examples of recent studies in Table 1. Researchers have used a number of analytic techniques, including content analysis; regression analysis; and machine learning methods such as LDA, Naïve Bayes classification, and neural network models. Eight of the studies investigated hotel and lodging reviews; four involved attractions or destinations, and only one involved cooking food-related during the trip. Unlike hotels with well-defined attributes (including room quality, service, location, cleanliness and value for money), the culinary experience is complex and contextually relevant. Therefore, in order to develop relevant and effective strategies to improve the visitor experience, it is necessary for destination marketing organizations to gain a deeper understanding of the culinary experience and its potential dimensions.

Table 1 Literature on Text Mining and Analysis of Online Reviews

Authors	Purpose	Methodology	Subject
Li, Ye, and Law (2013)	To identify the determinants of customer satisfaction in hospitality venues	Content analysis using ICTCLAS (Institute of Computing Technology, Chinese Lexical Analysis System)	42,668 reviews on Daodao of 774 hotels in Beijing, China
Barreda and Bilgihan (2013)	To identify the main themes that motivate consumers to evaluate hotel experiences in online environments.	Content analysis using NVivo 8	17,357 reviews on TripAdvisor of 3,124 hotels in the northeast of the US
Berezina et al. (2016)	To examine the underpinnings of satisfied and unsatisfied hotel customers	Word categorisation using PASW Modeler and text-link analysis	2,510 reviews on TripAdvisor of Hotels in Sarasota, Florida, US of Hotels in Sarasota, Florida, US
Fang, Ye, Kucukusta, and Law (2016)	To investigate the effects of reviewer characteristics inferred from properties of historical rating distribution	A negative binomial regression model and a Tobit regression model	41,061 reviews on TripAdvisor of Attractions in New Orleans, US
Wong and Qi (2017)	To investigate the evolution of the content of TripAdvisor online reviews	Content analysis using NVivo 10 and IBM_ManyEyes	8,007 reviews on TripAdvisor in Macau, China
Geetha, Singha, and Sinha (2017)	To establish a relationship between customer sentiments in online reviews and customer ratings for hotels	Sentiment analysis using Naïve Bayes classification algorithm and hierarchical cluster analysis	Reviews on TripAdvisor of 40 hotels in Goa, India

Guo et al. (2017)	To identify the key dimensions of customer service voiced by hotel visitors	Latent dirichlet allocation (LDA)	266,544 reviews on TripAdvisor of 25,670 hotels in 16 countries
Simeon et al. (2017)	To analyse online reviews to explore the experiences of tourists related to cultural attractions	Content analysis and principal component analysis	12,592 reviews on TripAdvisor of 58 cultural attractions in Naples, Italy
Su and Teng (2018)	To extract service quality dimensions of museums	Content analysis using NVivo 8	286 worst reviews on TripAdvisor of 15 museums in US, UK, France, the Netherlands, and others
Nakayama and Wan (2018)	To identify cultural differences between Western and Japanese restaurant customers	Text analysis and knowledge mining (TAKMI) using IBM Watson Explorer Content Analytics (WCA)	76,704 Western and 56,159 Japanese reviews on Yelp
Bi et al. (2019)	To conduct importance-performance analysis (IPA)	Latent dirichlet allocation (LDA), the improved one-vs-one strategy-based support vector machine (IOVO-SVM) and the ensemble neural network-based model (ENNM)	24,276 reviews on TripAdvisor of Two Five-Star Hotels
Cheng et al. (2019)	To investigate the effect of online review contents on potential guests' trust perception	Manual content analysis and convolutional neural network (CNN) model	1,485 and 10,000 reviews on Airbnb of Lodge listings in New York City, US

Taecharungroj (2019)	To infer the possible place brand identities from user-generated content	Content analysis using Leximancer	9,633 reviews on TripAdvisor and Google Maps of 2 famous metropolitan areas in Bangkok, Thailand

A large volume of textual data accumulated on social media (e.g., TripAdvisor, Yelp, and Dianping.com) has recently revealed a potential in exploration to gain insightful information that may be beyond the scope of a predefined survey. Sentiment analysis is a popular method for dealing with texts used primarily in the hospitality and tourism industries. Sentiment analysis is a technique for determining the sentiment orientation of a text (Sun et al., 2017). It is divided into two traditional approaches, lexicon and machine learning (Ma et al., 2018). Researchers can use the lexicon approach to calculate a sentiment score for each word using a predefined dictionary (e.g., Linguistic Inquiry and Word Count, Wordnet, SentiWordNet, SentiStrength, or a dictionary created by researchers) and summarize the representation of sentiment in a text (e.g., M. Lee et al., 2017; Philander & Zhong, 2016). The ML approach, on the other hand, develops a training model using a classification ML algorithm (e.g., Nave Bayes and support vector machine) to classify the document into either a positive or negative category (Duan et al., 2016). Customers' perceptions of service quality in a hotel setting were investigated by Duan et al. (2016).

Furthermore, researchers can now more accurately identify lament meanings from a large number of unstructured texts thanks to the evolution of NLP combined with machine learning (Evans & Aceves, 2016). "A theory-motivated range of computational techniques for the automatic analysis and representation of human language" is what NLP stands for (Cambria & White, 2014, p. 48). It has been discussed the advantages of ML-based NLP over other text mining methods (e.g., word frequency-based analysis). First, because machine learning-based NLP does not require the use of a pre-developed lexicon, the detection result is more contextually relevant. Contextual differences are handled by ML-based NLP by recognizing co-occurrence patterns. Second, machine learning-based NLP can train complex models on large datasets, and the results are usually more accurate than traditional text mining. In most cases, the applied ML algorithms outperformed various traditional text mining algorithms, according to Collobert et al. (2011).

Topic modeling (Blei et al., 2003) is a well-known NLP technique that excels at discovering semantic factors using machine learning. Customer reviews contain multifaceted opinions and semantic features, which have the methodologic potential for structured survey-based research (Evans & Aceves, 2016). Topic modeling identifies essential topics based on the co-occurrences of words without the use of external lexicons. Furthermore, researchers can assess the relative importance of topics to each document quantitatively (Blei et al., 2003). Topic modeling has been used in consumer behavioral research in a few studies in hospitality and tourism management. Through the topic modeling of online reviews on TripAdvisor, Guo et al. (2017) identified key dimensions of customer satisfaction. Vu et al. (2019) compared tourists' preferences for activities based on different destinations using topic modeling and ANOVA for topic probabilities.

Although the topic model has its own powerful functions, it is inadequate in dealing with the relationship between topics and sentiment analysis. Therefore, it is necessary to conduct manual iterations to filter and classify the topics with research significance. However, the Chinese text mining and sentiment analysis in tourism food activity are relatively empty, so the topic model should be improved to extract the topics and analyze the sentiment by the combination of another algorithm. Meanwhile, some Chinese words related to tourism and hospitality in the sentiment analysis library need to be improved.

2.3. Valence of reviews

While the sheer number of reviews on these platforms is beneficial, an overabundance of reviews can lead to confusion and misjudgment (O'Connor, 2010). Users often use the social proof heuristic in this situation, evaluating the positivity and negativity of the overall ratings – or valence. The valence of online reviews refers to their positive or negative tone based on the reviewers' personal experiences (Kusumasondjaja et al., 2012). Consumers' and tourists' responses and decisions are influenced by the overall valence of reviews in the form of ratings (Schuckert et al., 2015).

At the level of individual reviews, the valence of reviews is also significant. When tourists are satisfied with an experience, reviews that assign a high rating, or positive reviews (Li et al., 2013; Nieto et al., 2014), are considered a powerful tool to efficiently and effectively promote offerings and improve brand recall (Barreda & Bilgihan, 2013). Positive reviews are critical when users consider products and services for hedonic consumption (Hwang et al., 2018), as they lead to positive expectations for the product and service (X. Cheng et al., 2019). In contrast, when expectations are breached due to incompetence, inefficiency, irresponsible attitudes, behaviors, tactics, or strategies, or inferior products (Barreda & Bilgihan, 2013), sharing complaints, which are broadly defined as negative reviews or comments, occurs. However, online platform users often gravitate towards negative reviews to avoid potential risks and losses (Schuckert et al., 2015). Interestingly, research has found that users deem negative reviews more credible, helpful, diagnostic, informative, persuasive, trustworthy, and valuable than positive ones (Berezina et al., 2016; Fang et al., 2016; Kusumasondjaja et al., 2012; Schuckert et al., 2015; Taecharungroj & Mathayomchan, 2019). The reason for this is that people are more likely to avoid loss (loss aversion) than they are to seek gains (Mellinas et al., 2019). As a result, it's critical for destinations to pay close attention to negative reviews and take corrective and preventative measures to reduce them.

RQ3 is “How positive are the reviews for each topic?”; this RQ calculates the valence of each topic of the dimensions. Additionally, RQ4 asks, “How frequent and positive are the topics of each dimension?” to indicate the salience and valence of the underlying topics of each dimension. These two RQs delve deeper than observations of overall ratings or the valence of individual reviews to identify and investigate the topics of the food activity and their underlying factors. Therefore, the destination must pay close attention to negative reviews and take corrective and preventive measures to minimize them. At the same time, positive reviews can provide references and suggestions for destination management organizations, explaining what they can focus on and what is effective when promoting.

2.4. Chinese Outbound Tourism related literature

The rapid growth of outbound travel from China has also recently gained the attention of travel academics. The literature listed below is a collection of studies on Chinese outbound tourism, including research on traveler motivations and investigations of destination experiences and images. a study by Hua & Yoo, (2011) investigated the motivations of potential Chinese tourists who wanted to travel to the United States for leisure. The study found a significant relationship between travel motivation and socio-demographic factors such as gender, marital status and educational background. The authors suggest the need to diversify tourism products when designing them for Chinese tourists, who are becoming increasingly sophisticated and diverse.

A study on the perceived destination image of Chinese outbound tourists (Li & Stepchenkova, 2012) investigated past or potential Chinese long-distance outbound tourists' psychological impressions of the United States as a destination. The results of this study suggest that representatives of U.S. destination management organizations promote attractions that are not yet fully recognized. In addition, Stepchenkova and Li (2012) analyzed differences in image perceptions among four different groups of Chinese travelers, divided according to their previous travel experiences, showing that the more experienced group (with previous travel experience to the United States or outside of Asia) perceived destinations as rather abstract and intangible, such as open, democratic, and liberal societies. The less experienced group (with experience traveling within Asia or only within China but interested in traveling outside of Asia in the near future) associated destinations with more tangible things. Another study (Aramberri & Liang, 2012) analyzed images from three leading Chinese travel magazines covering 13 European destinations to find out the image of the destinations depicted. France, Italy and Germany were the three most represented countries in the selected magazines, with "pleasant everyday life" being considered the most descriptive theme. Other themes were also examined, such as different cultural expectations and visa issues from the perspective of cooperation between tour operators (e.g., those in China and the UK) and how these issues affect actual market development (de Sausmarez et al., 2012). Tse & Zhang (2013) explored the Chinese online travel experience sharing, focusing on blogs and Weibo posts about Hong Kong as a tourist destination. The findings confirm that by analyzing the opinions expressed in blogs and microblogs, opportunities for using social media for online communication and marketing of destinations can be identified.

Previous studies have focused primarily on the geographic regions of Europe and the United States, focusing on the overall image as well as the experience of the region, nevertheless, this study begins to cover a gap in the research field by addressing the topic of Chinese experiences and experiences of Japanese cuisine and demonstrating the distribution of positive and negative feelings towards restaurant reviews posted on online review platforms. And this paper mainly explores the food preferences as well as needs of Chinese tourists visiting Japan, unpicking them and subdividing them into topics in different dimensions to better benefit the local tourism market and stakeholders to get some references and suggestions.

2.5. The novelty of the research

First, the tourism food experiences are very difficult to assess and evaluate because they are inherently experiential, intangible, and heterogeneous. This study will explore the food experiences of Chinese tourists visiting Japan. Unlike the dimensions of food experiences classified by previous researchers, this study will categorize the food experiences according to the extracted topics and explore the food experiences hidden in the online reviews from different perspectives. In the field of tourism and hospitality, most of the online text mining and analysis has been focused on accommodation, journeys, and attractions, leaving a gap in the field of food experiences.

Second, in terms of research methodology, this study will creatively combine machine learning analysis approaches with the Chinese sentiment lexicon to make up for the original topic model's shortcomings. Although the LDA model is capable of processing large amounts of text data, it is unable to analyze topic relationships or sentiment polarity. Most existing studies combine other machine learning methods to perform sentiment analysis, but they are unable to perform sentiment analysis on reviews that are not accompanied by explicit ratings. As a result, this study will update the original Chinese sentiment word lexicon to overcome the deficiency of sentiment analysis in Chinese reviews and to fill the gap and deficiency of sentiment analysis in the field of food experience.

Chapter 3. Methodology

3.1. Data collection

This research studies online restaurant reviews on Dianping.com, which is the largest food navigation site in China while more than 55% of Chinese tourists to Japan use it.

The scope of this study is the top 10 most popular/visited tourist attractions in Tokyo among Chinese tourists. To select the attractions, the author based on survey data from Baidu Analytics, an online survey of 74,800 Chinese who have visited or want to visit Tokyo was conducted to select the top 10 most popular/must-see Tokyo attractions in China. And combined with the top 10 most popular Tokyo attractions as rated by China's largest travel review site, Dianping.com, as the subject of the study.

Reviews in Chinese from January 2014 to January 2019 were collected. All the reviews, along with the ratings and the dates, were downloaded from Dianping.com using an automatic program. Because all these data have been well organized on the website, it is feasible to create a program that visits one restaurant after another, browses every page (each page contains 20 reviews), and copies each review. To ensure objectivity and reduce bias in the data collected, the maximum number of reviews that can be collected per restaurant is set at 100. In total, 53442 online reviews were initially included in this research.

Table 2 Number of reviews of attraction

Tourism attraction	Number of reviews	Number of restaurants	Number of words
Tsukiji Market	4770	42	166273
Ginza	5569	48	190894
Akihabara	5201	39	139593
Roppongi	4374	45	141220
Asakusa	5547	51	161309
Shibuya & Omotesando	7229	79	216559
Imperial Palace & Nihonbashi	4956	60	143881
Odaiba	4173	49	129870
Ueno	5729	61	155823
Shinjuku	5894	66	172820
Total	**53442**	**540**	**1618242**

3.2. Data pre-processing

Step1. Word extraction. Each sentence in a review was broken into separate words using Jieba 0.39, a set of Chinese words segmentation utilities (The Python community, 2017). Normally, a Chinese word contains one to four Chinese characters, and the common words have already been lexicalized in Jieba. Therefore, by comparing the segments of a sentence with the lexicon, the words therein are identifiable. Subsequently, the extracted words together formed a word list.

Step2. Word filtering. For words that have a frequency of below 1% (Büschken & Allenby, 2016), that is, each of these words appears in less than 1% of all reviews, they were removed from the word list. Also removed were words that have high frequency but contain no explaining power (Netzer et al., 2012), such as “I,” “one,” and “today.”

3.3. Data analysis: Latent Dirichlet allocation model

Topic modeling is one of the primary tasks of natural language processing. The most suitable topic modeling approach considers each document as a collection of different topics. The most common approach to performing feature extraction or topic modeling is LDA modeling (Guo et al., 2017), which is a "generative probabilistic model of the corpus" and is very effective in managing Big Data (Blei, Ng, and Jordan, 2003; Tirunillai and Tellis, 2014). The basic idea is that the LDA model assumes the existence of a hidden structure that consists of a set of dimensions in the entire corpus of reviews. The model helps to reveal a mixture of topics from an enormous number of comments. It uses the co-occurrence of terms (or words) in reviews to infer these topics (Blei, n.d.; Guo et al., 2017)

3.3.1. LDA analysis steps

The authors followed these six steps to complete the LDA analysis and to answer the first two RQs:

1. Determine the number of topics
2. Change the parameter and train the model
3. Extract the topics
4. Filter and classify the topics into different dimensions
5. Name the topics
6. Draw a chart comparing the topics

3.3.2. Simplified example of topic extraction for LDA model

The LDA model assumes the word generation in a document to be a two-stage process:

(1) Randomly choose a distribution of topics.

(2) For each word in the document,

(a) Randomly choose a topic from the distribution of topics in (1).

(b) Randomly choose a word from the corresponding distribution of the vocabulary.

In real situations, neither the distribution of topics over documents nor the distribution of words over topics is known a priori; only the documents are observed. For example, suppose one has the following simple documents.

(1). The waitress is very helpful.

(2). Our waiter speaks good English.

(3). The price is reasonable.

(4). We had an expensive meal, but we thought it worth the price.

(5). The waiter is impressive, and the price is OK.

The LDA might produce something like the following.

Sentence (1) and (2): 100% Topic A.

Sentence (3) and (4): 100% Topic B.

Sentence (5): 50% Topic A, 50% Topic B.

Topic A: 30% waiter, 20%, waitress, 10% helpful, 5% English, ... (at which point one would interpret Topic A to be about waiter/waitress).

Topic B: 40% price, 30% reasonable, 15% worth, 5% expensive, ... (at which point one would interpret Topic B to be about price).

3.3.3. Mathematical interpretation of LDA model

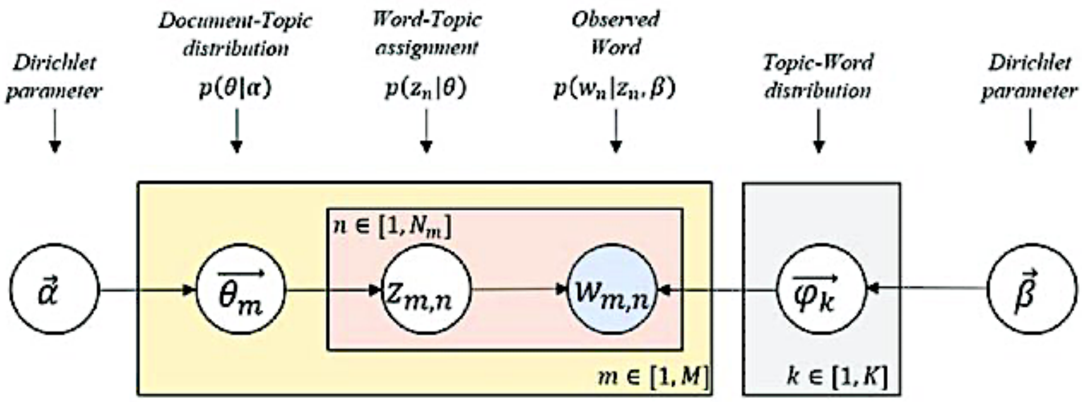

Figure 3 Graphical model of latent Dirichlet allocation (LDA). (J. Lee et al., 2018)

Mathematically, the connection between hidden and observed variables is the joint distribution expressed using Equation (1).

$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D}) = \prod_{i=1}^{K} p(\beta_i) \cdot \prod_{d=1}^{D} p(\theta_d) \cdot \sum_{n=1}^{N} p(z_{d,n} \mid \theta_d) p(w_{d,n} \mid \beta_{1:K}, z_{d,n}) \quad (1)$$

B_i: distribution of word in topic i, altogether K topics;

θ_d: proportions of topics in document d, altogether D documents;

z_d: topic assignment in document d;

$z_{d,n}$: topic assignment for the nth word in document d, altogether N words;

w_d: observed words for document d;

$w_{d,n}$: the nth word for document d.

The identification of topics and words is thus a posteriori estimation (Equation 2) using Gibbs sampling (Griffiths & Steyvers, 2004), without the trouble of creating a user dictionary in advance (Haselmayer & Jenny, 2017).

$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D} \mid w_{1:D}) = \frac{p(\beta_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D})}{p(w_{1:D})} \quad (2)$$

3.3.4. Evaluate the number of topics

In general, the number of topics should be evaluated based on the explanatory power of the model. To determine this, use Perplexity (Equation 3) which represents the number of branches or alternatives and is defined as the inverse of the probability. The smaller the perplexity, the higher the explanatory ability of the model.

In this study, we used 30,000 of the 53,442 reviews to train the LDA models, and then calculated the perplexity of each model across the remaining 23,442 documents. The number of topics used was varied: 2, 5, 10, 20, and then 100 in steps of 10.

$$\text{Perplexity}(\boldsymbol{W} \mid M) = \left(-\frac{\sum_{d=1}^{D} \log P(W_d \mid M)}{\sum_{d=1}^{D} N_d} \right) \quad (3)$$

3.4. Topic classification and sentiment Analysis

3.4.1 Topic filter and classification

The mechanical analysis' results are jumbled and unusable without manual processing, so the author went through all the topics and their high-frequency words manually and iteratively to filter out the ones that were relevant to the research. After that, related topics with similar characteristics are grouped into one dimension based on term features and patterns, and the words in each topic are manually and initially separated by positive and negative sentiments. Therefore, we chose the manual approach for better identification results as well as more detailed managerial suggestions.

3.4.2 Sentiment analysis

Since the review data collected in this study tended to have high ratings (out of 50,000 reviews, only 4645 reviews had ratings below 4, with negative reviews accounting for 8.6% of the total number of reviews), it was not possible to use some traditional methods and models for sentiment analysis, such as Naive Bayes and Support Vector Machines.

Therefore, the new library Cnsenti for Chinese sentiment analysis (Deng, X. 2019) is chosen to perform sentiment analysis on the reviews in the extracted classified topics. On the other hand, the previous sentiment lexicon of Cnsenti is mainly oriented to shopping-related, which needs to be supplemented by the author for tourism and cuisine-related words.

Step1. To improve the library, added travel, hospitality and food-related, frequently occurring nouns, adjectives, and adverbs to the dictionary, and used the Wordexpansion library to improve the speed of building the dictionary.

Step2. Calculate the number of positive and negative reviews in each topic for the subsequent analysis of topic salience-valence

The method for expanding the sentiment lexicon mentioned in the first step used seed sentiment words and SO-PMI. SO_PMI is an algorithm in the computer science field, which means pointwise mutual information (PMI) can be used to assess the semantic orientation (SO) of the words to be analyzed. SO_PMI can expand the emotional lexicon in specialized areas based on the most basic emotional seed words. Simply put, it means that sentiment words are not completely independent of each other, often things are grouped together and words with similar emotional polarity will have a certain degree of similarity and regularity. In this way, it may take only a hundred seed words to associate tens of thousands of similar words added to the sentiment lexicon. By setting a small number of initial positive seed words and initial negative seed words, the program will find many candidates with a high probability of positive sentiment and many candidates with a high probability of negative sentiment based on the initial positive seed words. This process consists of four steps, i: Segmenting documents to build a corpus (list of seed sentiment words). ii: Construct co-occurrence word combinations. iii: Calculate SO-PMI score. iv: Save positive and negative sentiment candidates.

The following used a review to illustrate how Cnsenti approaches reviews.

'*The restaurant has a beautiful view, and the service is very attentive, but the food doesn't taste very good.*'

At this point, the Cnsenti library identifies two positive sentiment words, beautiful view and meticulous service, and one negative sentiment word, bad taste, and finally marks the review as Pos-review. Through repeating the process, the sentiment polarity of the reviews within each topic can be identified and then counted.

3.5. Topic Salience-Valence analysis

Topic Salience-Valence analysis (TSVA) provides an effective and more intuitive understanding of the frequency and importance of each topic in each dimension of the food experience, as well as the degree of positivity (positive and negative sentiment) for each topic. TSVA analysis provides not only a qualitative understanding of the importance and emotional preference of each topic but also a quantitative approach to comparing the importance and positivity of the topics.

The number of reviews on each topic was counted. The proportion of a topic's number of reviews to the dimensions' total number of reviews is regarded as the salience of that topic (Equation 5), which is used in subsequent analysis. To assess the valence of each topic, a cross-tabulation analysis between the valence and the topic of the reviews was conducted. The cross-tabulation analysis showed the expected numbers and the observed numbers of positive reviews of each topic. The valence of the topics (Equation 6) is the difference between the observed and expected numbers divided by the total number of reviews in that dimension. Finally, a chart of each topic of dimension was created to illustrate the salience and the valence of the topics. The charts are called the topic salience-valence analysis (TSVA).

Table 3 Process of calculating the number of expected positive reviews

Dimension	**Topic**	**Observed Positive Reviews**	**Expected Positive Reviews**	**Total Number of Reviews**
Dimension1	Topic 1	O1	E1	T1
	Topic 2	O2	E2	T2
	Topic 3	O3	E3	T3
	Topic 4	O4	E4	T4
	Topic 5	O5	E5	T5
	Total	**O**		**T**

The process of evaluating the valence:

Step1. Count the number of reviews in each topic (T1, T2…).

Step2. Conduct cross-tab analysis (Table 3) between observed positive reviews and total reviews, then calculate the difference between expected and Observed positive reviews to determine the value of valence.

$$\text{Expected positive reviews} = \frac{\text{Topic's Total Reviews * Dimemsion's Total Observed Positive Reviews}}{\text{Dimension's Total Reviews}} \qquad (4)$$

(e.g.: Topic 1's Expected positive reviews E1 = T1 * O / T)

It is difficult to visually and clearly observe the positivity of a topic by simply subtracting the number of positive reviews from the number of negative reviews to study its positivity. For example, if out of 1000 reviews on a topic, 700 are positive, 100 are negative, and 200 are neutral, the difference between positive and negative reviews is the same as if there were 800 positive reviews and 200 negative reviews, and it is impossible to determine the difference in the degree of positivity of a topic.

In contrast, by calculating the number of expected positive reviews, the value of valence can be used to determine the level of positivity of a topic by focusing not only on the number of positive and negative reviews, but also by taking into account the impact of the total number of reviews. Furthermore, the difference in the level of positivity of the topics in the dimensions for comparison can be better observed, and the level of positivity of the topics can be better understood with a more intuitive percentage value.

Thus, when the valence (Equation 6) is greater than 0, that is, the observed positive reviews exist more than the expected positive reviews, we can consider the topic as positive, and the larger the valence the higher the positive degree. On the contrary, when the valence is negative, that is, the observed positive reviews are less than the expected positive reviews, the topic can be considered as negative.

$$\text{Salience} = \frac{\text{Topic's No.of Reviews}}{\text{Total reviews of the dimension}} \tag{5}$$

$$\text{Valence} = \frac{\text{Observed No.of Positive Reviews-Expected No.of Positive Reviews}}{\text{Total reviews of the topic}} \tag{6}$$

Chapter 4. Findings

4.1. Topic extraction and classification result

4.1.1. Determine the topic number

A good indicator of the right number of topics from a computational standpoint is the number with which the model best predicts the data. This is comparable to statistical model goodness-of-fit measures. Perplexity is a commonly used indicator for topic models like LDA, where a lower perplexity indicates a better prediction (Blei, n.d.). To calculate the perplexity, we first use a portion of the data to train an LDA model. The model is then evaluated using the data that was left out. This procedure is repeated for models with varying numbers of topics, until it is clear which number results in the least amount of perplexity.

Figure 4 depicts the perplexity of different models for our data. We used 30,000 of the 53,442 reviews to train the LDA models, and then calculated the perplexity of each model across the remaining 23,442 documents. The number of topics used was varied: 2, 5, 10, 20, and then 100 in steps of 10. The findings show that as the number of topics increases, perplexity decreases, implying that a model with more topics is better at predicting the held-out data. The decrease in perplexity for additional topics becomes noticeably less between 30 and 70 topics. This is one way to interpret the appropriate number of topics, similar to how the elbow in a factor analysis' scree plot is interpreted. Another option is to count the number of topics with the least amount of perplexity. We can see that this point will be between 65 and 70, so we choose 67 for this study to extract the number of topics. However, in a mathematical sense, having the "right" number of topics says nothing about the interpretability of the topics produced. Therefore, manual inspection of these topics and their terms are required to determine the number of topics to be used for subsequent research.

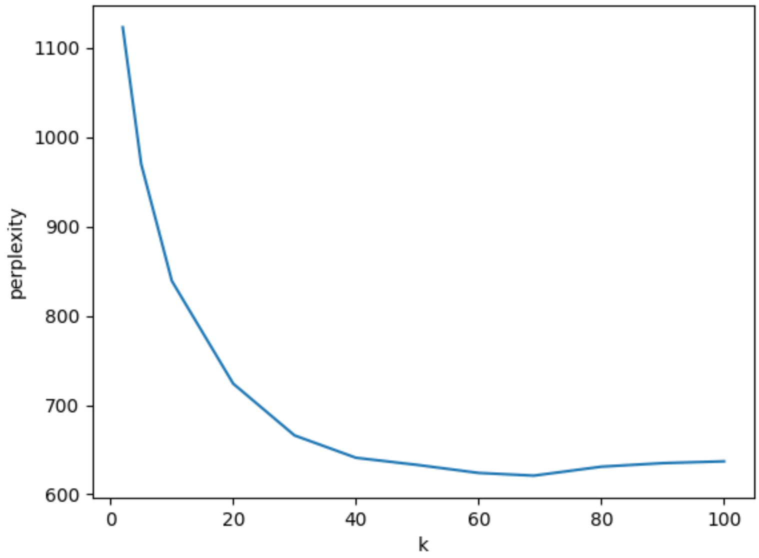

Figure 4 Perplexity of LDA models with different numbers of topics

4.1.2 The results of classifying topics into four dimensions

In fact, mathematical goodness-of-fit measures and human interpretation may lead to contradictory conclusions on the best number of topics, because the computer algorithm may find nuances that are not semantically meaningful to humans, especially with a large number of topics (Chang et al., 2009). Furthermore, because we use topics to answer substantive questions about the reviews we study, it's critical that the topics that emerge from the analysis contribute to answering these questions rather than simply providing the best data prediction. When deciding on the number of topics to use for the analysis presented in this article, we considered both perplexity and interpretability.

By reviewing and checking all the extracted topics and the terms in them several times, the author removed some topics with more repetitions of each other's contents, and some topics with more terms that have no practical meaning, such as topics with only a large number of culinary nouns or degree adverbs. Finally, from 67 topics, this study selected 24 topics of research and analysis significance. Then these topics were categorized into 4 dimensions: Food/cuisine, Location and environment, Dining/restaurant, Cultural activity and promotion. (Table 4)

The LDA algorithm was used to extract a specified number of topics from the reviews. Using Gensim Analytics, LDA returned the desired number of topics, their underlying terms, and the proportion of each term for the respective topic.

Table 4 Filtered topics of four dimensions

Dimension	Topic
Cultural activity and promotion	Japanese culture
	Wagashi
	Feeling
	Online influencer promotion
	Online promotion about cafe
Food/cuisine	Sashimi
	Yakiniku
	Seafood
	Sushi
	Food safety
	Dissatisfactory
	Expectation and disappointment
Location and environment	Tsukiji market nearby
	Restaurant surroundings
	Location
	Ueno nearby
	Omotesandou nearby
Dining / restaurant	High-class restaurant
	Service impression
	Reservation service
	Payment method
	Explanation of food & culture
	Queue
	Language and communication

Table 5, Table 6, Table 7, Table 8 show the topics of four dimensions and prevalent underlying terms. The author named those topics according to their underlying terms. The proportion of this term among all the terms in this topic is represented by Prop. The more frequently this term appears, the higher the value of prop. For instance, take the first dimension's topic: Japanese culture. The most frequently occurring term in this topic is “limited”, which in Chinese means seasonal or regional limited, and its proportion is -0.196, meaning that if the total proportion of all terms in this topic is 1, this term's proportion in the entire topic is 0.196. Each topic's terms are listed in descending order of proportion, with the proportion decreasing as you go down.

4.1.2. Topic distribution and sentiment marks in Cultural activity and promotion dimension

Table 5 Topics of Cultural activity and promotion dimension

Topic 8: Japanese culture	Prop.	Topic 21: Wagashi	Prop.	Topic 27: Feelings	Prop.	Topic 30: Online influencer promotion	Prop.	Topic 69: Online promotion about cafe	Prop.
Limited	-0.196	Wagashi	-0.157	Picture	-0.069	go to an internet-famous place	-0.260	Coffee	-0.101
Three Stars	-0.052	Cake	-0.130	Japanese	-0.044	Ice cream	-0.189	Bread	-0.057
Egg yolk	-0.048	Strawberry	-0.034	Degree	-0.044	online influencer	-0.094	go to an internet-famous place	-0.057
Strawberry	-0.037	Too late	-0.032	Curious	-0.032	internet-famous restaurant	-0.081	cup	-0.051
inheritor	-0.037	Seasonal	-0.028	Not eaten before	-0.031	coffee shop	-0.036	online influencer	-0.050
Seasonal	-0.033	souvenir	-0.017	Casual	-0.027	disparity	-0.024	good drink	-0.028
Onsen	-0.030	No chance	-0.017	not come again	-0.025	toppings	-0.024	afternoon tea	-0.027
Various	-0.024	pretty good	-0.015	crowded	-0.022	confirmed	-0.021	beautiful scenery	-0.022
MATSURI	-0.024	serial	-0.014	must order	-0.021	unintentional	-0.016	Tokyo	-0.018
historical	-0.022	Sakura	-0.014	come again	-0.020	relationship	-0.013	creative	-0.017
Jelly	-0.022	red	-0.014	store	-0.018	peach	-0.012	delicate	-0.012
Japanese-style	-0.018	Hardness	-0.011	well worth	-0.018	thematic	-0.012	photo	-0.011
Classic	-0.017	Forest	-0.011	great	-0.017	unexpected	-0.011	bottle	-0.011
Series	-0.016	Too sweet	-0.010	cant read	-0.017	all of a sudden	-0.008	Suitable	-0.009
Sakura	-0.015	sweetness	-0.009	translate	-0.016	relaxed	-0.007	Good looking	-0.009
Ancient Way	-0.012	special edition	-0.009	influential	-0.015	vanilla-flavored	-0.007	around	-0.008
Well-proportioned	-0.010	Barista	-0.009	attractive	-0.014	travel	-0.007	long queue	-0.007
look at the picture	-0.010	Airport	-0.007	faintly	-0.014	Picky	-0.007	Comfortable	-0.007
sweet and creamy	-0.010	natural	-0.007	try	-0.014	American	-0.007	Drinking coffee	-0.006
fatty	0.009	too small	-0.007	conspicuous	-0.014	Fake	-0.007	Starbucks	-0.006

The cultural activity and promotion dimension (Table 5) includes five topics: Japanese culture, Wagashi, Feeling, Online influencer promotion, Online promotion about café. The Japanese culture-related topic has terms such as hot springs, Matsuri, Japanese, and traditional, and has more positive adjectives such as seasonal, classical, and heritage. The second topic is Japanese dessert, a confectionery, accompanying gift, and sweetness shape, etc. The description of Wagashi has both positive and negative feelings. The third topic is about feelings, the main term is in the expression of emotions and feelings, such as rare, crowded, attractive, etc. The fourth topic is about online influencer, and the representative terms are online attractions and online restaurants, etc., but there are more negative feelings. Meanwhile, the fifth dimension is also related to online promotion, the main terms are coffee, afternoon tea, beautiful scenery, photos, etc., but most of the emotions are positive.

4.1.3. Topic distribution and sentiment marks in Food/Cuisine dimension

Table 6 Topics of Food/Cuisine dimension

Topic 7: Sashimi	Prop.	Topic 14: Yakiniku	Prop.	Topic 17: Seafood	Prop.	Topic 29: Sushi	Prop.	Topic 31: Food safety	Prop.	Topic 38: Dissatisfactory	Prop.	Topic 44: Expectation gap	Prop.
Sashimi	-0.191	Happy	-0.089	Eel	-0.200	Sushi	-0.460	Fresh	-0.056	Safe	-0.087	Disappointment	-0.094
Crabmeat	-0.089	barbecue restaurant	-0.074	Tuna	-0.056	Honten	-0.030	Market	-0.035	Difficult to eat	-0.062	Crowded	-0.065
Tuna	-0.087	Affordable	-0.054	Second time	-0.029	French cuisine	-0.029	delicious	-0.035	Tired of shopping	-0.044	Small store	-0.064
Sea urchin	-0.083	pretty good	-0.044	Famous	0.017	Two stars	-0.025	seafood	-0.032	Crowded	-0.042	One star	-0.031
Fresh	-0.045	price	-0.041	soft	-0.014	Regret	-0.024	Sashimi	-0.032	seafood	-0.030	Because	-0.029
Hot pot	-0.044	not expensive	-0.033	expected	-0.014	One plate	-0.019	place of origin	-0.021	Nuclear radiation	-0.030	roadside	-0.027
soft and smooth	-0.037	eat one	-0.030	first	-0.013	archaic	-0.018	Ingredients	-0.019	suspicious	-0.024	confused	-0.026
Invincible	-0.029	first place	-0.026	bitter	-0.013	premier chef	-0.017	Platter	-0.017	Eat some	-0.021	signature dish	-0.024
early morning	-0.018	match	-0.025	too expensive	-0.012	Heaven	-0.012	trustless	-0.017	Try it	-0.018	brand	-0.023
Slippery	-0.014	alley	-0.023	fat	-0.011	Farewell House	-0.009	Tsukiji	-0.015	bad review	-0.018	a plate	-0.023
sign	-0.013	invincible	-0.023	one to go	-0.011	Documentary	-0.008	Crab meat	-0.014	cant walk	-0.010	exaggerated	-0.022
never again	-0.010	normal	-0.021	little expensive	-0.009	Experience	-0.008	radial	-0.012	too little	-0.009	expected	-0.021
afraid	-0.008	Explosion	-0.021	boyfriend	-0.009	fresh	-0.007	Recommended	-0.011	very early	-0.009	best friend	-0.017
snapper	-0.008	Wagyu beef	-0.020	dare not eat	-0.008	hidden	-0.007	doubtful	-0.011	insipid	0.008	reason	-0.017
practice	-0.008	sometimes	-0.016	Takashima	-0.008	narrow	-0.006	Special	-0.011	boring	-0.008	long queue	-0.017
feel it	-0.007	too fatty	-0.016	shrimp meat	-0.008	photo	-0.006	dare not	-0.010	custom	-0.007	worthless	-0.017
fatty	-0.007	conscience	-0.016	delicious	-0.008	time period	-0.006	like	-0.009	fast food restaurant	-0.007	revisit	-0.016
gold foil	-0.007	forward	-0.014	Grilled eel	-0.008	most famous	-0.006	Feeling	-0.009	expectation	-0.006	too expensive	-0.015
forced	-0.006	smooth	-0.014	Double-boiled	-0.008	average	-0.005	tasty	-0.009	world	-0.006	admired	-0.015
memory	-0.006	next door	-0.014	Tender	-0.008	hidden	-0.005	Super	-0.009	Winter	-0.005	disparity	-0.014

There are seven topics related to the Food/cuisine dimension (Table 6). Sashimi's terms are mainly crab meat, sea urchin, fresh, etc. Yakiniku mostly contains affordable, pretty, and beef terms. The term characteristics of seafood are composed of seafood type and soft and afraid to eat. The overall term sentiment about cuisine is positive. The fifth topic is related to food safety, and the obvious terms are negative words such as distrust, radiation, and worry. And the sixth and seventh topics are related to dissatisfaction and expectation deviation, and the main appearing terms are hard to swallow, crowded, queuing, etc.

4.1.4. Topic distribution and sentiment marks in Location and environment dimension

Table 7 Topics of Location and environment dimension

Topic 5: Tsukiji market nearby	Prop.	Topic 16: Restaurant surroundings	Prop.	Topic 34: Location	Prop.	Topic 45: Ueno nearby	Prop.	Topic 63: Omotesandou nearby	Prop.
Tsukiji	-0.080	Night View	-0.027	Hotel	-0.167	In the morning	-0.049	World	-0.043
Lovers	-0.034	chef	-0.024	Location	-0.033	Across the street	-0.042	Style	-0.028
seafood	-0.029	private room	-0.022	First-class	-0.030	Shinobazu Pond	-0.031	Park	-0.019
God	-0.023	Like	-0.021	Oversized	-0.030	comfortable	-0.027	Omotesandou	-0.017
Impressive	-0.023	Street	-0.018	family	-0.026	Global	-0.026	Unique	-0.014
America	-0.016	Ginza	-0.018	sheer	-0.023	Service charge	-0.021	Design	-0.014
Crowded	-0.016	illumination	-0.017	lost	-0.020	Park	-0.020	Product	-0.014
maybe	-0.014	Service	-0.016	sellable	-0.017	Sitting	-0.020	Landscape	-0.013
influence	-0.014	desired	-0.015	right	-0.016	road	-0.019	Decorated	-0.011
colorful	-0.013	Romantic	-0.014	takeaway	-0.015	Hanami	-0.019	Simplicity	-0.011
Sashimi	-0.013	Time	-0.013	traffic	-0.015	location	-0.018	Tired of shopping	-0.010
mood	-0.012	Although	-0.011	difficult	-0.013	all the way	-0.017	Attractive	-0.010
epidemic	-0.011	month	-0.011	confused	-0.013	chicken	-0.016	buildings	-0.010
Business	-0.011	They	-0.011	continuous	-0.012	Ueno	-0.015	Cherry Blossom	-0.010
Raw fish	-0.010	challenge	-0.011	back	-0.012	so beautiful	-0.014	By the window	-0.009
Memories	-0.009	market	-0.011	reason	-0.012	weekday	-0.014	Drinking coffee	-0.009
Regrets	-0.009	number	-0.010	healthy	-0.012	long ago	-0.013	good quality	-0.009
beautiful	-0.009	sparkling	-0.010	step	-0.011	material	-0.013	elegant	-0.009
too big	-0.009	price	-0.008	respect	-0.011	mother	-0.013	Located	-0.009
Journey	-0.009	sun	-0.008	peak	-0.011	during the day	-0.011	across the street	-0.009

Tsukiji market nearby, Ueno nearby, Omotesando nearby, Restaurant surrounding, and location are the five topics that make up the dimension of location and environment (Table 7), three of which are related to specific sightseeing spots, and the term also reflects the characteristics of each spot. Tsukiji market, for example, refers to both seafood and raw fish. Omotesando's terms, such as style, decorated, and elegant, are mostly related to fashion and design. The terms associated with tourist attractions are also more positive. Hotel, lost, traffic, and other terms are used in the fourth topic location, along with some negative emotions. The fifth topic is the restaurant's surroundings, and the key terms are night view, illumination, romance, and so on.

4.1.5. Topic distribution and sentiment marks in Dining and Restaurant dimension

Table 8 Topics of Dining and restaurant dimension

Topic 2: High-class restaurant	Prop.	Topic 24: Service impression	Prop.	Topic 33: Reservation service	Prop.	Topic 40: Payment method	Prop.	Topic 42: Explanation of food and culture	Prop.	Topic 47: Queue	Prop.	Topic 61: Language	Prop.
Restaurant	-0.062	Taste of Tea	-0.040	Appointment	-0.112	Alipay	-0.093	Sea urchin	-0.448	Queue	-0.082	Menu	-0.134
Michelin	-0.030	Impressive	-0.037	Advance	-0.095	Paypal	-0.050	Tuna	-0.045	Door	-0.023	Chinese	-0.107
Bread	-0.030	Onsen	-0.019	Service	-0.052	Coffee beans	-0.035	Explanation	-0.018	location	-0.018	Waiter	-0.044
Cuisine	-0.027	attentive	-0.015	reserved	-0.033	Payment	-0.023	Japanese culture	-0.012	specially	-0.018	English	-0.032
Tokyo	-0.026	good mood	-0.015	Helpful	-0.027	not guided	-0.021	Drinking tea	-0.011	tasty	-0.017	Shop assistant	-0.031
specialty	-0.017	Heartfelt	-0.010	three-star	-0.027	QR code	-0.021	traditional	-0.009	too long	-0.015	Japanese	-0.017
exquisite	-0.016	Premium ingredients	-0.010	Fully booked	-0.025	technology	-0.014	inexplicable	-0.008	a few hours	-0.015	Order food	-0.015
Japan	-0.016	As always	0.010	website	-0.024	acidity	-0.014	Picky	-0.007	tired	-0.014	Dessert store	-0.015
Ingredients	-0.015	historical	-0.010	waiter	-0.023	like	-0.013	curious	-0.007	incomprehensible	-0.014	translate	-0.014
Dessert	-0.013	unintelligible	-0.010	Time	-0.021	chaotic	-0.013	old man	-0.006	Waiting for a seat	-0.013	No need	-0.013
Experience	-0.011	responsible	-0.009	Dinner	-0.017	difficult	-0.012	authentic	-0.006	Akihabara	-0.013	unintelligible	-0.013
Dishes	-0.011	old man	-0.009	location	-0.016	Will come again	-0.012	marvelous	-0.005	Ginza	-0.011	picture	-0.013
Desserts	-0.011	meticulous	-0.009	confused	-0.015	must go	-0.011	dishes	-0.005	Why	-0.011	worry	-0.012
Gourmet	-0.010	enjoyable	-0.009	in place	-0.015	cash	-0.011	cultural connotation	-0.005	Nice	-0.010	confused	-0.010
Ginza	-0.009	professional	-0.009	photo	-0.014	microsoft	-0.011	salty	-0.005	exhausted	-0.010	cannot communicate	-0.010
world	-0.008	unforgettable	-0.008	plate	-0.012	hygiene	-0.011	seconds	-0.005	Self-service	-0.010	communication	-0.010
Perfect	-0.008	green	-0.008	how	-0.012	no support	-0.011	most awesome	-0.004	order	-0.008	attentive	-0.009
Premium	-0.008	excellent	-0.007	online	-0.012	information	-0.010	hidden	-0.004	narrow	-0.008	Order	-0.009
Delicious	-0.007	miss	-0.007	Active	-0.011	yen	-0.010	implicit	-0.004	find another	-0.008	guessing	-0.008
French cuisine	-0.007	Stay	-0.007	pretty nice	-0.010	Special	-0.010	cultural	-0.004	evening	-0.008	Explain	-0.008

Seven topics were extracted from the Dining and Restaurant dimensions (Table 8), namely High-class restaurants, Service impression, Reservation, Payment method, Explanation of food and culture, Queue, and Language. The main terms of High-class restaurants were Michelin, exquisite, premium, etc., with more positive emotions. etc. The terms related to reservation are appointment, fully booked, confused, etc. The fourth topic is related to payment methods, and there are some negative terms, such as not guided, chaotic. The fifth topic is the explanation of food and culture, including tea drinking, traditional, cultural connotation, and other terms, mainly for positive emotions. And the last two topics about queuing and ordering language communication, there are many negative emotions, like tired, exhausted, confused.

4.2. Topic Salience-Valence analysis result

4.2.1. TSVA result of Cultural activity and promotion

Table 9 Descriptive statistics of topic dimension - Cultural activity and promotion

Dimension	**Topic**	**Observed Positive Reviews**	**Expected Positive Reviews**	**Total Number of Reviews**	**Salience**	**Valence**
Cultural activity and promotion	Japanese culture	2547	2226.29	2871	23.40%	11.17%
	Wagashi	2376	2386.03	3077	34.77%	-0.33%
	Feeling	1,087	1061.58	1369	15.47%	1.86%
	Online influencer promotion	890	1194.62	1641	18.54%	-18.56%
	Online promotion about cafe	583	536.60	692	7.90%	6.70%
	Total	**7,483**		**9,650**		

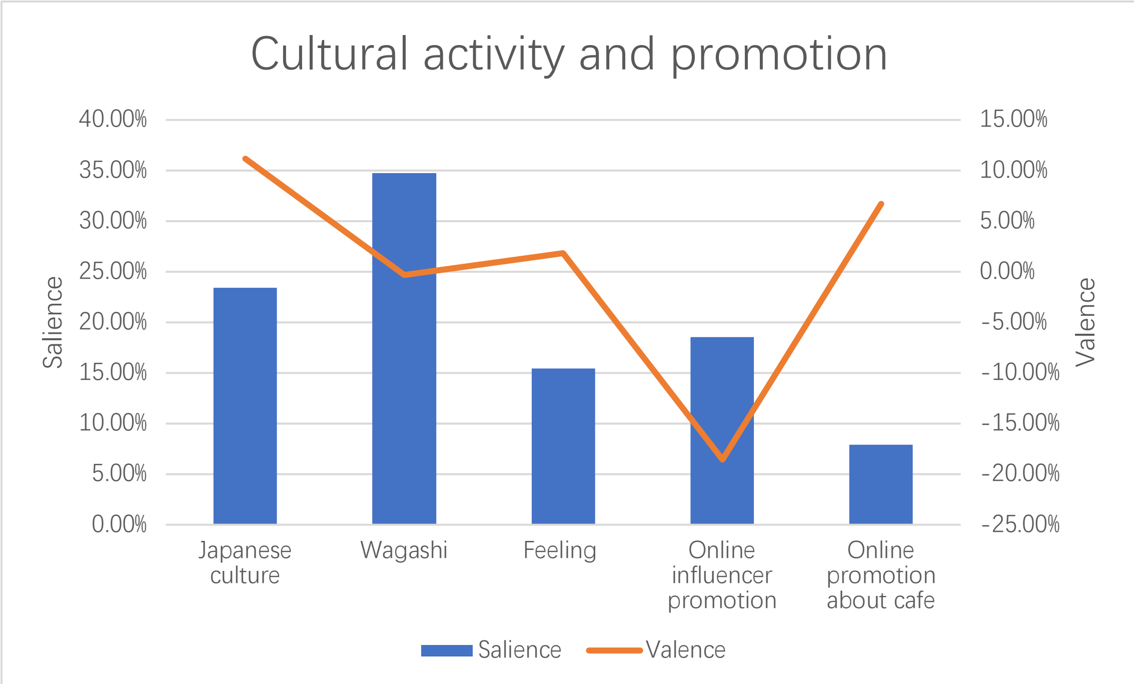

Figure 5 Topic salience-valence analysis of Cultural activity and promotion.

Figure 5 shows that the topic with the highest salience is Wagashi (34.77%), but at the same time valence is negative (-0.33%). This indicates that Chinese tourists are enthusiastic about Japanese desserts, though there may be some dissatisfaction and issues. As a result, later in the discussion, it is necessary to check the specific dissatisfaction and negative evaluation of Wagashi.

The second highest salience (23.40 %) is associated with Japanese culture, and it also has the highest valence (11.17%). This shows that Chinese visitors are not only interested in Japanese culture, but also have positive and favorable feelings about it. This will serve as a guide for the Japan National Tourism Organization and destination marketers in selecting factors related to Japanese culture when promoting or attracting tourists, or in promoting the attractiveness of Japanese culture in order to bring more benefits.

Surprisingly, among the two topics about online promotion: online influencers and cafes, respectively the valence about online influencers is the lowest (-18.56%), in contrast, the valence about cafes is more positive (6.70%). This shows that if marketers decide to promote online, they should pay more attention to the promotion method and content, and re-evaluate its effectiveness.

4.2.2. TSVA result of Food/Cuisine

Table 10 Descriptive statistics of topic dimension – Food/Cuisine

Dimension	**Topic**	**Observed Positive Reviews**	**Expected Positive Reviews**	**Total Number of Reviews**	**Salience**	**Valence**
Food/cuisine	Sashimi	2901	2330.24	3985	12.61%	14.32%
	Yakiniku	4621	3097.43	5297	16.76%	28.76%
	Seafood	4080	3982.74	6811	21.55%	1.43%
	Sushi	3012	2116.80	3620	11.45%	24.73%
	Food safety	1059	2327.31	3980	12.59%	-31.87%
	Dissatisfactory	926	2216.79	3791	11.99%	-34.05%
	Expectation gap	1885	2412.69	4126	13.05%	-12.79%
	Total	**18,484**		**31,610**		

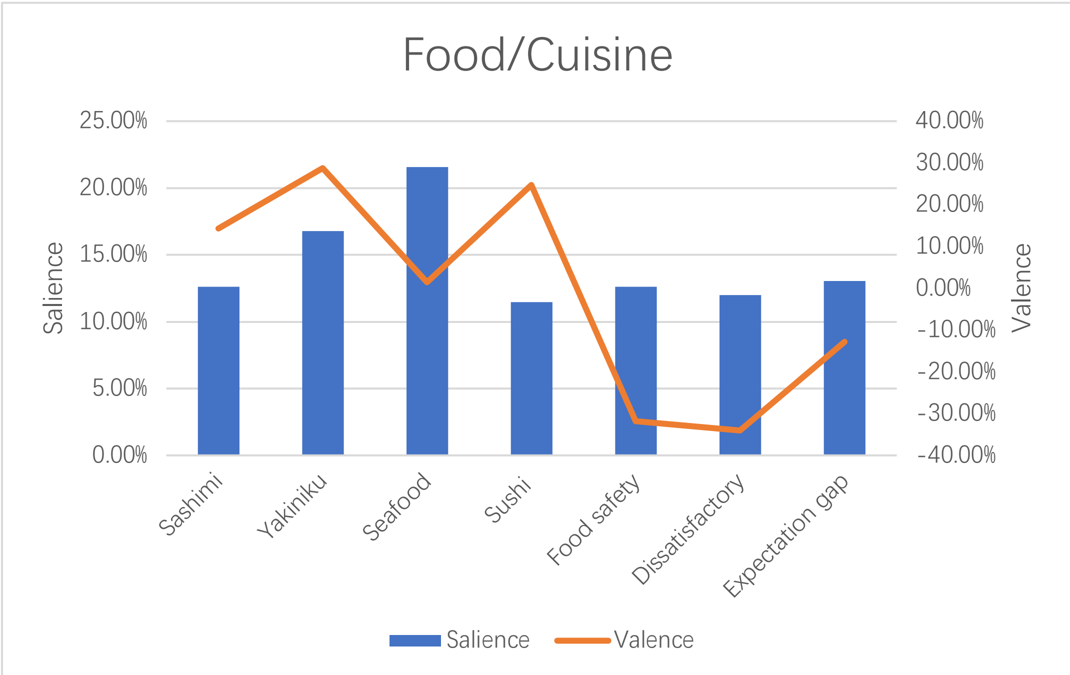

Figure 6 Topic salience-valence analysis of Food/Cuisine

As shown in Figure 6, the topic with the second lowest valence in the Food and Cuisine dimension is food safety (-31.87%). Returning to the extracted topic and term to investigate, we can see that the main concerns are ingredient distrust, radiation, etc. Therefore, JNTO must focus more on increasing consumer confidence in seafood and other ingredients, as well as attempting to address negative perceptions. Yakiniku (28.76%), sushi (24.73%), and sashimi (14.32%) among other food-related topics, all have a high valence, though seafood (1.43%) has a lower rating. It also suggests that Chinese tourists have favorable attitudes toward Japanese cuisine in general. The main reason for the lower valence could still be the negative emotions brought on by food trust and different eating habits. The topic of the expectation gap (-12.79%) is worth mentioning because it demonstrates that there is a gap between tourists' feelings and their actual experience, which both marketers and restaurants must be aware of and address.

4.2.3. TSVA result of Location and environment

Table 11 Descriptive statistics of topic dimension – Location and environment

Dimension	Topic	Observed Positive Reviews	Expected Positive Reviews	Total Number of Reviews	Salience	Valence
Location and environment	Tsukiji market nearby	2904	2703.36	4027	20.31%	4.98%
	Restaurant surroundings	3256	3223.62	4802	24.22%	0.67%
	Location	3129	3996.96	5954	30.03%	-14.58%
	Ueno nearby	2378	1958.88	2918	14.72%	14.36%
	Omotesandou nearby	1641	1425.19	2123	10.71%	10.17%
	Total	13308		19824		

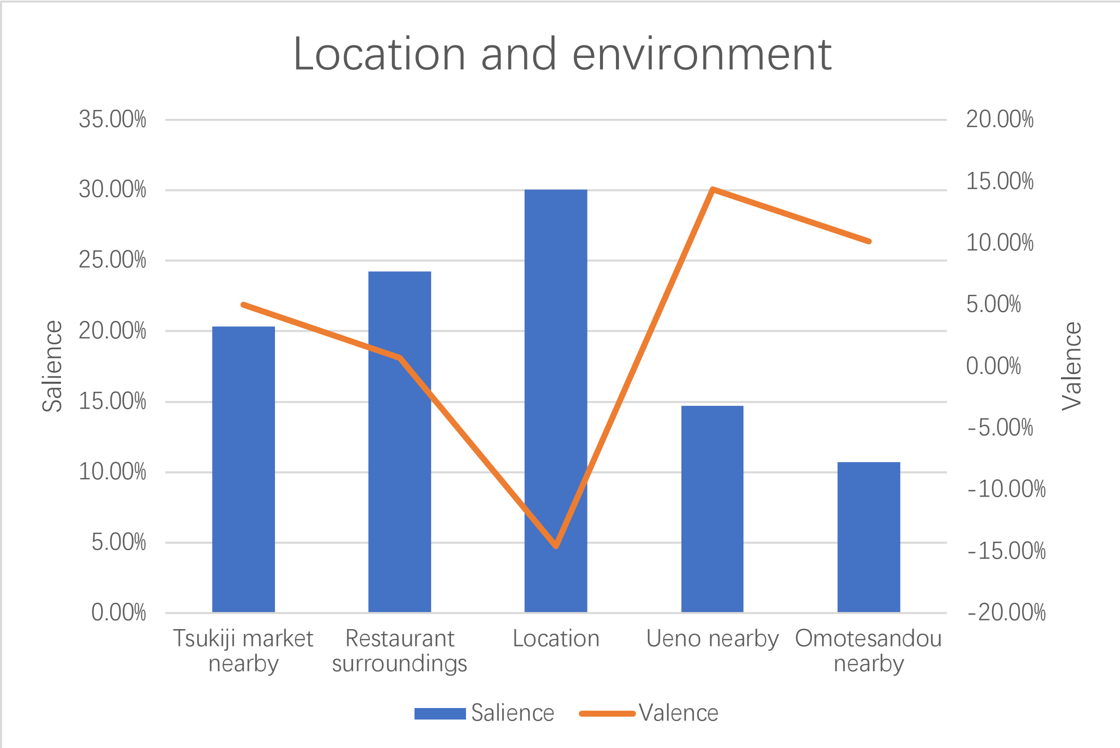

Figure 7 Topic salience-valence analysis of Location and environment

As shown in Figure 7, Among the topics appearing in the dimension of location and environment, there are three significant and frequent sightseeing spots, of which the positive valence is around Omotesando (10.17%) and Ueno (14.36%), while the lower valence is Tsukiji Market (4.98%).

The topic of Location has the highest salience (30.03%) but with the lowest value (-14.58%). The negative evaluation is generally due to the reasons of getting lost, confusion, traffic, and not being able to find the restaurant's location, based on the terms related to the topic. This can also provide suggestions and references for destination marketers, such as the location of some hidden restaurants and transportation route guidance that goes beyond the guidelines.

Not surprisingly, the Restaurant's immediate surroundings have a positive impact (0.67%) and are given a higher frequency (24.22%). It also demonstrates that, in terms of promotion, the environment has a positive and mutually reinforcing effect.

4.2.4. TSVA result of Dining and Restaurant

Table 12 Descriptive statistics of topic dimension – Dining and Restaurant

Dimension	**Topic**	**Observed Positive Reviews**	**Expected Positive Reviews**	**Total Number of Reviews**	**Salience**	**Valence**
Dining and restaurant	High-class restaurant	1208	1042.58	1530	6.84%	10.81%
	Service impression	3877	3803.72	5582	24.94%	1.31%
	Reservation service	2152	2051.77	3011	13.45%	3.33%
	Payment method	1891	1914.81	2810	12.56%	-0.85%
	Explanation of food & culture	2619	2526.73	3708	16.57%	2.49%
	Queue	1394	1623.84	2383	10.65%	-9.64%
	Language	2110	2287.55	3357	15.00%	-5.29%
	Total	15251		22381		

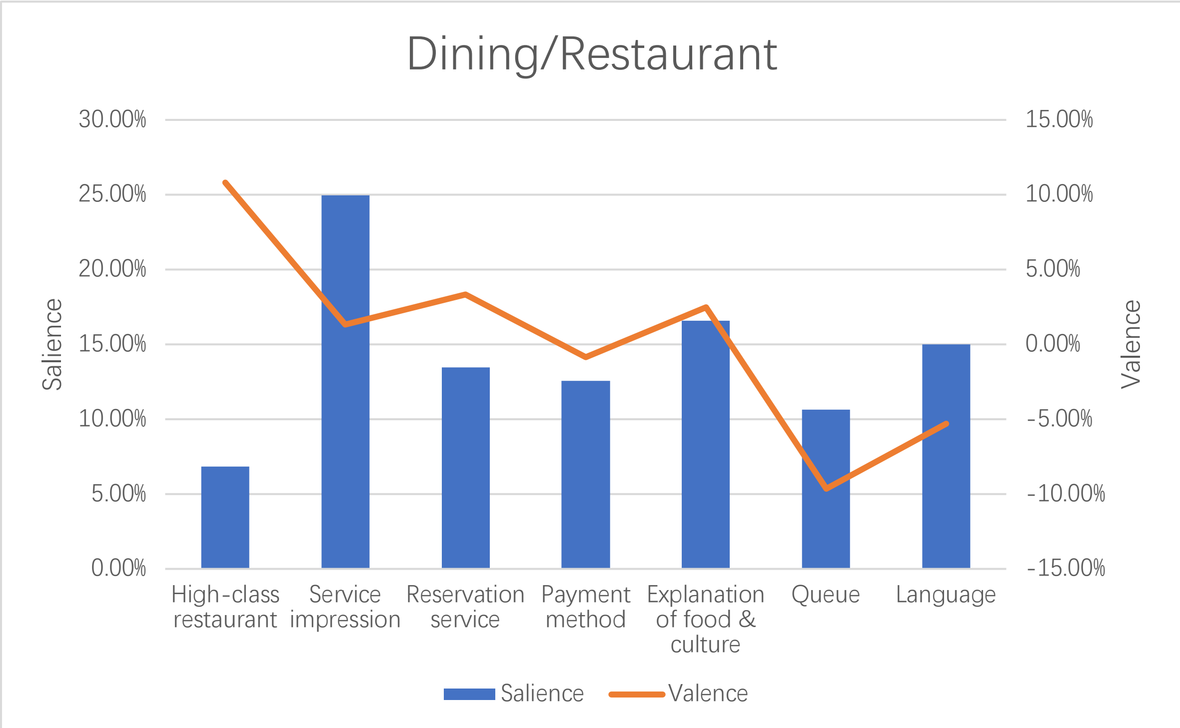

Figure 8 Topic salience-valence analysis of Dining and Restaurant

As shown in Figure 8, among the topics appearing in the dimension of Dining and restaurant, the highest salience is service impression (24.94%), and the valence is also positive (3.33%), indicating that careful service is the most impressive part of the dining process for Chinese tourists. Surprisingly, the valence of the topic of language is the second lowest (-5.29%). Based on common sense, the restaurant may assume that Chinese people can read Chinese characters, but focusing on negative words in the topic, demonstrates confusion and inability to order. JNTO must pay attention to the translation and provide the restaurant with relevant advice and guidance.

The explanation of food and culture ranks second in terms of salience (16.57%) and also has positive valence (1.31%). This relates to the previous dimension's culture-related topic, as Chinese tourists are very interested in Japanese culture. Destination marketers can increase the introduction and promotion of culture, then direct relevant restaurants to show and explain more about Japanese food culture, resulting in an increase in tourist numbers and satisfaction.

Chapter 5. Discussion and implications

5.1. Discussion

The results of the study in Chapter 4 demonstrated that Chinese tourists visiting Japan are very interested in and positive about Japanese culture and sweets, but their sentiments toward online promotion vary depending on the content and means. While they appreciate the Japanese cuisine and attentive service, as well as the beautiful surroundings of the restaurants, there are concerns about food safety, queuing, reservation and payment methods, and menu communication. As a result, the discussion in the following chapter should return to the original reviews to determine the specific reasons why these results emerged for further discussion. And the findings need to be discussed in depth, first about the emotional polarization of online promotion, then about the specific content and concerns about food safety, and finally about the specific content and reasons for dissatisfaction that emerged about queuing, reservation, and payment services.

5.1.1. Methods and contents of online promotion

Regarding the large positive divergence in online promotion, it is necessary to return to the comments to find the reasons for it.

To illuminate the issues, some examples of online reviews were selected.

"This restaurant has always been very popular, and many influencers have come to visit and highly recommend it. This time I finally got to taste it, but probably due to high expectations, the food was not as good as in the photos, and the taste was average, so I had a serious feeling of being cheated."

"The environment and taste are not that good, because there are many people recommended on Twitter, so I queued up for a long time to try it. But only full of disappointment…"

“Finally in the cherry blossom season to visit this popular cafe, the scenery is as beautiful as a fairy tale, completely beyond the previous expectations.”

These reviews show that what the internet promotions and what influencers recommend when it comes to cuisine and dishes may not be very good, because everyone has different tastes and preferences. However, if it is related to the environment and atmosphere, satisfaction will be higher.

In recent years, JNTO has been active in promoting and advertising online overseas, such as the SNS Posting Campaign from November 2021 to January 2022 (Figure 9), and 'Online Seminar for Resident Foreign Influencers' was held from November 21, 2020. These campaigns were first implemented by the plan of "Visit Japan Campaign" program in 2003. Nevertheless, the results of this study show that these marketing strategies for foreign visitors are indeed receiving attention and feedback from visitors, as evidenced by the presence of topics such as online promotion and influencer. However, the research findings revealed that not all of these marketing strategies had positive feedback, and even dissatisfaction was caused by the gap between expectations and actual experience caused by the campaign. As a result, JNTO needs to re-evaluate the direction and content of its online campaigns and adopt effective approaches and methods.

Figure 9 SNS Posting Campaign's promotional poster (source: https://www.jtbcom.co.jp/inbounders/jnto.mybestjapanmoments/cn/)

The results of several topics and TSVA analysis, on the other hand, demonstrate that Chinese tourists have a strong interest in and positive attitudes about Japanese traditions and food culture, but that communicating about the dishes is challenging. These findings imply that the JNTO should strengthen its promotion of Japanese culture and the food culture embedded in cuisine, while also providing guidance and assistance to restaurants that serve foreigners on how to better explain the meanings and backstory of cuisine and Japan's gourmet culture.

5.1.2. Trust building and negative impressions of food safety

In the food/cuisine dimension, the topic of food safety appears, and this is very noteworthy. Hence it is necessary to go to the raw reviews and check some relevant ones to find out the specific causes. The following are the relevant comments that were checked.

"At a very famous kaiseki restaurant, the chef was very considerate in serving the dishes one by one, and the food was very delicious, but I was afraid to try the sashimi

when I was going to eat it because I was really afraid that the seafood was polluted by radiation.”

“Because I love to eat eel, so I chose this old store, but when I know the place of origin I still hesitated, after all, it is possible that the ingredients are not safe.”

“Do not recommend this crab buffet restaurant, there are no other dishes except crab, and the crab is not good, it feels strange because I always think that it was irradiated by nuclear radiation.”

The raw reviews show that the negative effects of the nuclear leakage caused by the previous Great East Japan Earthquake are still affecting the trust of Chinese tourists in food. Even though the reviews show that tourists are satisfied with the deliciousness of the cuisine and the service of the restaurants, the concerns about food safety caused by the radiation are still a devastating issue. Most of the concerns and dissatisfaction are focused on seafood, sashimi and other raw food dishes, but also on the origin and ingredients. As the Tourism Nation Promotion Basic Plan reveals, tourism will play a great publicity role in recovering and enhancing this trust. Therefore, continued steady efforts will need to be made to restore the trust on the Japan brand by providing accurate information that is appropriate from the viewpoint of consumers, preventing the occurrence and/or expansion of damages due to harmful rumors, and letting travelers from overseas see Japan as it is. Meanwhile, mass and social media have strong effects on people’s perceptions of the disaster(J. W. Cheng et al., 2016).

The findings of this study show that, despite the fact that the disaster occurred ten years ago, JNTO and the destination still need to pay close attention and take continuous and effective measures to gradually reduce and eliminate negative impressions and concerns among foreign tourists, as well as to promote ingredients and origins in order to build tourists' trust and support in the disaster area.

5.1.3. Dissatisfaction due to cultural and dietary backgrounds

The findings of this study revealed many surprising topics, such as reservation services, queues, payment methods, etc. So, check out the related reviews to get more insights.

"It's really a pity that you have to make a reservation to taste this restaurant, I was totally unaware of this before coming."

"This grilled offal restaurant is very popular, there was already a long line outside the door, I didn't want to wait for a long time, but the Japanese people seemed to not care."

"This old restaurant is very good and the service is attentive, but I had some trouble with the checkout, not only could I not use my phone to pay, but I couldn't even use my credit card."

Surprisingly revealed by these reviews, Chinese tourists show confusion and dissatisfaction about reservation service and payment methods, and the main reason may be that in normal life in China, reservation is not required for restaurants in general. And with the rapid development of online payment in recent years, the third-party payment companies are turning China into a "cashless society". It requires the awareness of JNTO to provide more help and guidance to foreign tourists regarding reservation and means of payment, etc., on the one hand, and to help and guide restaurants so that they and foreign tourists can more clearly understand and embrace the differences arising from different national and cultural backgrounds. At the same time, it has been discovered that Chinese tourists do not know how to eat certain foods and are unaware of and concerned about raw food, which has an impact on the dining experience. These should also pique the interest of destination marketers. Because of the cultural differences in food, more efforts are needed to introduce and popularize the unique Japanese cuisine and serving methods in order to heighten tourists' interest and empathy.

The findings of this study also reveal a great opportunity that the topic related to high-class restaurants has the highest valence value in the Dining and restaurant dimension, although its salience is not significant, which also reminds the marketers of the destination to promote the high-class consumption of the middle-rich Chinese tourists visiting Japan by conducting some publicity and marketing about high-class restaurants.

5.2. Theoretical implications

This study innovatively combines a machine learning model with a sentiment lexicon approach to explore topics and identify sentiments in sophisticated tourists' culinary experiences. By overcoming the limitations of Chinese sentiment analysis on food experiences, this study fills a gap in the sentiment lexicon. It also utilizes a novel approach to categorizing food experiences into various dimensions for comparison and study.

First, this study re-examines the prevalence of tourists' food experiences. While past research on restaurants and food experiences typically consisted of five dimensions, Food Quality, Service, Price, Environment, and Location. This study demonstrates through topic extraction and sentiment analysis of online reviews that the food experience is multidimensional, with composite experience value factors. For instance, topics such as cultural influences and the implementation of promotions, while food experiences can also be influenced by cultural background and other social events and even disasters. In addition, this study reveals four dimensions and 24 topics among them in the food experience of Chinese tourists visiting Japan.

Second, this study developed a novel methodological approach by which researchers can obtain quantitative measures of food experience and grasp emotional polarity from a large number of unambiguously rated online visitor reviews. Topic modeling is one of the NLP techniques that has attracted the attention of many researchers due to its efficiency in text processing. In this study, a combination of topic modeling and sentiment lexicon method was used to deeply explore and analyze the component dimensions of food experience and the potential sentiment positivity among Chinese tourists visiting Japan. In addition, in order to overcome the inadequacy of the original LDA model, the parameters of the model are adjusted to increase the weight of sentiment words in the extracted topics. This approach brings new research opportunities for tourist experience related to trips where emotional polarity cannot be clearly determined, e.g., future researchers can use this approach to extract experiences and emotions for accommodation or other activities.

5.3. Practical implications

The findings from the present study provide practical implications as follows. First, 24 topics about Chinese tourists visiting Japan were extracted and divided into four dimensions containing the actual culinary experiences of tourists. The previous food experience dimensions are conceptual which destination practitioners may have difficulty in applying directly to management and services. Therefore, this study provides a way to extract specific topics and factors from real user reviews and feedback on online platforms to extract the actual culinary experience of visitors so that destination marketers and promoters can use it to evaluate the effectiveness of existing strategies and to improve on shortcomings.

Second, this study identifies many negative factors of tourists' food experience. Thus, practitioners need not only improve the positive factors but also mitigate the negative factors, given that they are related to customer satisfaction. factors but also mitigate the negative factors, given that they are related to customer satisfaction. Thus, how destination marketers perceive, and control negative factors is critical to the sustainability of the destination. However, these negatives are not easily obtained and detected, resulting in some negatives being overlooked as destination personnel may focus heavily on the easily seen aspects. The findings of this study therefore inform and suggest ways for destination marketers to re-examine existing approaches and develop appropriate and effective management and communication strategies for thereafter.

Chapter 6. Conclusion and future research

Online reviews are becoming important sources for tourism research and practices. They reflect visitors' delight in and discontent with their experiences (Banerjee & Chua, 2016). However, their potential in providing insights into travelers' activity experiences and preferences has not been utilized effectively. The food activities and experiences of a destination are often difficult to evaluate prior to consumption because they are experiential, intangible, and heterogeneous. Therefore, this study aims to explore and analyze the food activity experiences of Chinese tourists visiting Japan by combining a machine learning approach and a sentiment lexicon method. The topics extracted from the LDA model were divided into four dimensions according to similar characteristics for subsequent analysis. The sentiment analysis was also applied to reviews with no clear sentiment polarity by updating the Chinese sentiment dictionary with terms related to tourism and food activities. The findings revealed that Chinese tourists were interested in Japanese culture and Japanese desserts, and were satisfied with the yakiniku, seafood and sushi cuisines, as well as with the attentive service, surroundings and tourism attractions around the restaurant. On the contrary, there were negative feelings and dissatisfaction about online influencer, reservation and payment services, and queuing, along with distrust and concern about food safety.

In addition, by calculating the topic salience and valence within each dimension, the significance and positivity of the topic can be better understood and grasped. TSVA analysis shows that Chinese tourists' attitudes towards online promotion are emotionally polarized depending on the content or means. There are also significant concerns about the safety of the food as well as the ingredients and origin, which also affect the satisfaction with the culinary experience. Consequently, firstly, JNTO needs to re-evaluate and review the original online promotion methods and contents, and secondly, it needs to continuously promote positive and good images about the disaster areas and the impact on food and ingredients to eliminate foreign tourists' worries and doubts and gradually restore their trust in food safety. Finally, the analysis shows a high level of interest and positive emotions among Chinese tourists for Japanese culture and desserts, in cuisine is for yakiniku and seafood, as well as fine dining related and restaurant surroundings. Therefore, JNTO and related destination marketers can focus on promoting these aspects when developing policies and marketing strategies related to attracting Chinese tourists, which will not only attract the attention and curiosity of tourists but also bring positive marketing effects at the same time. As a result, the findings of this study can help DMOs figure out what promotion/marketing aspects improve and reduce tourist food and cuisine experiences, and therefore their sentiments toward Japanese cuisine.

However, the findings in this study are subject to a few limitations. Firstly, the reviews used in this study are from online review platform and do not include internet-wide coverage. Secondly, Since the extracted topics were manually classified by the author's judgment, it possibly influenced the results of TVSA. Consequently, future research is needed to cover more websites and further explore the differences between research findings stemming from online tourism reviews that are on different travel websites or different audience groups. Moreover, further research should shed more light on comparative analysis of multi-platform and multi-lingual data.

References

Aramberri, J., & Liang, C. (2012). The Chinese Gaze: Imaging Europe in Travel Magazines. *Journal of China Tourism Research*, *8*(3), 284–301. https://doi.org/10.1080/19388160.2012.704248

Baka, V. (2016). The becoming of user-generated reviews: Looking at the past to understand the future of managing reputation in the travel sector. *Tourism Management*, *53*, 148–162. https://doi.org/10.1016/j.tourman.2015.09.004

Banerjee, S., & Chua, A. Y. K. (2016). In search of patterns among travellers' hotel ratings in TripAdvisor. *Tourism Management*, *53*, 125–131. https://doi.org/10.1016/j.tourman.2015.09.020

Barreda, A., & Bilgihan, A. (2013). An analysis of user-generated content for hotel experiences. *Journal of Hospitality and Tourism Technology*, *4*(3), 263–280. https://doi.org/10.1108/JHTT-01-2013-0001

Berezina, K., Bilgihan, A., Cobanoglu, C., & Okumus, F. (2016). Understanding Satisfied and Dissatisfied Hotel Customers: Text Mining of Online Hotel Reviews. *Journal of Hospitality Marketing & Management*, *25*(1), 1–24. https://doi.org/10.1080/19368623.2015.983631

Bi, J.-W., Liu, Y., Fan, Z.-P., & Zhang, J. (2019). Wisdom of crowds: Conducting importance-performance analysis (IPA) through online reviews. *Tourism Management*, *70*, 460–478. https://doi.org/10.1016/j.tourman.2018.09.010

Björk, P., & Kauppinen-Räisänen, H. (2016). Local food: A source for destination attraction. *International Journal of Contemporary Hospitality Management*, *28*(1), 177–194. https://doi.org/10.1108/IJCHM-05-2014-0214

Blei, D. M. (n.d.). *Latent Dirichlet Allocation*. 30.

Buhalis, D., & Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research. *Tourism Management*, *29*(4), 609–623. https://doi.org/10.1016/j.tourman.2008.01.005

Cambria, E., & White, B. (2014). Jumping NLP Curves: A Review of Natural Language Processing Research [Review Article]. *IEEE Computational Intelligence Magazine*, *9*(2), 48–57. https://doi.org/10.1109/MCI.2014.2307227

Chang, J., Gerrish, S., Wang, C., Boyd-graber, J., & Blei, D. (2009). Reading Tea Leaves: How Humans Interpret Topic Models. *Advances in Neural Information Processing Systems*, *22*. https://proceedings.neurips.cc/paper/2009/hash/f92586a25bb3145facd64ab20fd554ff-Abstract.html

Cheng, J. W., Mitomo, H., Otsuka, T., & Jeon, S. Y. (2016). Cultivation effects of mass and social media on perceptions and behavioural intentions in post-disaster recovery – The case of the 2011 Great East Japan Earthquake. *Telematics and Informatics*, *33*(3), 753–772. https://doi.org/10.1016/j.tele.2015.12.001

Cheng, X., Fu, S., Sun, J., Bilgihan, A., & Okumus, F. (2019). An investigation on online reviews in sharing economy driven hospitality platforms: A viewpoint of trust. *Tourism Management*, *71*, 366–377. https://doi.org/10.1016/j.tourman.2018.10.020

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (n.d.). Natural Language Processing (Almost) from Scratch. *NATURAL LANGUAGE PROCESSING*, 45.

de Sausmarez, N., Tao, H., & McGrath, P. (2012). Chinese Outbound Tourism to the United Kingdom: Issues for Chinese and British Tour Operators. *Journal of China Tourism Research*, *8*(3), 268–283. https://doi.org/10.1080/19388160.2012.704235

Duan, W., Yu, Y., Cao, Q., & Levy, S. (2016). Exploring the Impact of Social Media on Hotel Service Performance: A Sentimental Analysis Approach. *Cornell Hospitality Quarterly*, *57*(3), 282–296. https://doi.org/10.1177/1938965515620483

Evans, J. A., & Aceves, P. (2016). Machine Translation: Mining Text for Social Theory. *Annual Review of Sociology*, *42*(1), 21–50. https://doi.org/10.1146/annurev-soc-081715-074206

Fang, B., Ye, Q., Kucukusta, D., & Law, R. (2016). Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism Management*, *52*, 498–506. https://doi.org/10.1016/j.tourman.2015.07.018

Geetha, M., Singha, P., & Sinha, S. (2017). Relationship between customer sentiment and online customer ratings for hotels—An empirical analysis. *Tourism Management*, *61*, 43–54. https://doi.org/10.1016/j.tourman.2016.12.022

Gretzel, U., & Yoo, K.-H. (2008). Use and Impact of Online Travel Reviews. In *Information and Communication Technologies in Tourism 2008* (p. 46). https://doi.org/10.1007/978-3-211-77280-5_4

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, *59*, 467–483. https://doi.org/10.1016/j.tourman.2016.09.009

Hjalager, A.-M., & Richards, G. (2003). *Tourism and Gastronomy*. Routledge.

Hua, Y., & Yoo, J. J.-E. (2011). Travel Motivations of Mainland Chinese Travelers to the United States. *Journal of China Tourism Research*, *7*(4), 355–376. https://doi.org/10.1080/19388160.2011.627000

Hwang, J., Park, S., & Woo, M. (2018). Understanding user experiences of online travel review websites for hotel booking behaviours: An investigation of a dual motivation theory. *Asia Pacific Journal of Tourism Research*, *23*(4), 359–372. https://doi.org/10.1080/10941665.2018.1444648

Jones, A., & Jenkins, I. (2002). 'A Taste of Wales – Blas Ar Gymru': Institutional malaise in promoting Welsh food tourism products. In *Tourism and Gastronomy*. Routledge.

Kusumasondjaja, S., Shanka, T., & Marchegiani, C. (2012). Credibility of online reviews and initial trust The roles of reviewer's identity and review valence.

Journal of Vacation Marketing, *18*, 185–195. https://doi.org/10.1177/1356766712449365

Lee, J., Kang, J.-H., Jun, S., Lim, H., Jang, D., & Park, S. (2018). Ensemble Modeling for Sustainable Technology Transfer. *Sustainability*, *10*, 2278. https://doi.org/10.3390/su10072278

Lee, M., Jeong, M., & Lee, J. (2017). Roles of negative emotions in customers' perceived helpfulness of hotel reviews on a user-generated review website: A text mining approach. *International Journal of Contemporary Hospitality Management*, *29*(2), 762–783. https://doi.org/10.1108/IJCHM-10-2015-0626

Li, H., Ye, Q., & Law, R. (2013). Determinants of Customer Satisfaction in the Hotel Industry: An Application of Online Review Analysis. *Asia Pacific Journal of Tourism Research*, *18*(7), 784–802. https://doi.org/10.1080/10941665.2012.708351

Li, X. (Robert), & Stepchenkova, S. (2012). Chinese Outbound Tourists' Destination Image of America: Part I. *Journal of Travel Research*, *51*(3), 250–266. https://doi.org/10.1177/0047287511410349

Ma, E., Cheng, M., & Hsiao, A. (2018). Sentiment analysis – a review and agenda for future research in hospitality contexts. *International Journal of Contemporary Hospitality Management*, *30*(11), 3287–3308. https://doi.org/10.1108/IJCHM-10-2017-0704

Mellinas, J. P., Nicolau, J. L., & Park, S. (2019). Inconsistent behavior in online consumer reviews: The effects of hotel attribute ratings on location. *Tourism Management*, *71*, 421–427. https://doi.org/10.1016/j.tourman.2018.10.034

Nakayama, M., & Wan, Y. (2018). Is culture of origin associated with more expressions? An analysis of Yelp reviews on Japanese restaurants. *Tourism Management*, *66*, 329–338. https://doi.org/10.1016/j.tourman.2017.10.019

Nieto, J., Hernández-Maestro, R. M., & Muñoz-Gallego, P. A. (2014). Marketing decisions, customer reviews, and business performance: The use of the Toprural website by Spanish rural lodging establishments. *Tourism Management*, *45*, 115–123. https://doi.org/10.1016/j.tourman.2014.03.009

O'Connor, P. (2010). Managing a Hotel's Image on TripAdvisor. *Journal of Hospitality Marketing & Management*, *19*(7), 754–772. https://doi.org/10.1080/19368623.2010.508007

Philander, K., & Zhong, Y. (2016). Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. *International Journal of Hospitality Management*, *55*, 16–24. https://doi.org/10.1016/j.ijhm.2016.02.001

Richards, G. (2021). Evolving research perspectives on food and gastronomic experiences in tourism. *International Journal of Contemporary Hospitality Management*, *33*(3), 1037–1058. https://doi.org/10.1108/IJCHM-10-2020-1217

Schuckert, M., Liu, X., & Law, R. (2015). Hospitality and Tourism Online Reviews: Recent Trends and Future Directions. *Journal of Travel & Tourism Marketing*, *32*(5), 608–621. https://doi.org/10.1080/10548408.2014.933154

Shafqat, W., & Byun, Y.-C. (2019). A Recommendation Mechanism for Under-Emphasized Tourist Spots Using Topic Modeling and Sentiment Analysis. *Sustainability*, *12*(1), 320. https://doi.org/10.3390/su12010320

Simeon, M. I., Buonincontri, P., Cinquegrani, F., & Martone, A. (2017). Exploring tourists' cultural experiences in Naples through online reviews. *Journal of Hospitality and Tourism Technology*, *8*(2), 220–238. https://doi.org/10.1108/JHTT-10-2016-0067

Su, Y., & Teng, W. (2018). Contemplating museums' service failure: Extracting the service quality dimensions of museums from negative on-line reviews. *Tourism Management*, *69*, 214–222. https://doi.org/10.1016/j.tourman.2018.06.020

Sun, S., Luo, C., & Chen, J. (2017). A review of natural language processing techniques for opinion mining systems. *Information Fusion*, *36*, 10–25. https://doi.org/10.1016/j.inffus.2016.10.004

Taecharungroj, V. (2019). User-generated place brand identity: Harnessing the power of content on social media platforms. *Journal of Place Management and Development*, *12*(1), 39–70. https://doi.org/10.1108/JPMD-11-2017-0117

Taecharungroj, V., & Mathayomchan, B. (2019). Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. *Tourism Management*, *75*, 550–568. https://doi.org/10.1016/j.tourman.2019.06.020

Tse, T. S. M., & Zhang, E. Y. (2013). Analysis of Blogs and Microblogs: A Case Study of Chinese Bloggers Sharing Their Hong Kong Travel Experiences. *Asia Pacific Journal of Tourism Research*, *18*(4), 314–329. https://doi.org/10.1080/10941665.2012.658413

Vu, H. Q., Li, G., & Law, R. (2019). Discovering implicit activity preferences in travel itineraries by topic modeling. *Tourism Management*, *75*, 435–446. https://doi.org/10.1016/j.tourman.2019.06.011

Warde, A., & Martens, L. (2000). *Eating Out: Social Differentiation, Consumption and Pleasure*. Cambridge University Press.

Wilson, A., Murphy, H., & Fierro, J. C. (2012). Hospitality and Travel: The Nature and Implications of User-Generated Content. *Cornell Hospitality Quarterly*, *53*(3), 220–228. https://doi.org/10.1177/1938965512449317

Wong, C. U. I., & Qi, S. (2017). Tracking the evolution of a destination's image by text-mining online reviews—The case of Macau. *Tourism Management Perspectives*, *23*, 19–29. https://doi.org/10.1016/j.tmp.2017.03.009

Acknowledgment

The two years I spent as a master student at Tokyo Metropolitan University was extremely rewarding and enlightening. Despite the impact of the Corona epidemic, this will be a very memorable part of my life, and I will begin a new chapter with the knowledge and abilities I gained here. I would like to express my most sincere gratitude to all the professors, classmates, and family members who have helped and supported me at this time.

First, I would like to thank my supervisor, Professor Tetsuo Shimizu, for his profound knowledge and excellent character, which I have learned a lot from. With his careful guidance and help, I was able to complete my thesis successfully. There is an old Chinese proverb that says, " Once a teacher, a lifetime father". Shimizu sensei's guidance and influence will not only help me during my study abroad but will also inspire me in my future life. I would also thank to Professor Susumu Kawahara, Associate Professor U Okamura, Assistant Professor Yuki Ohira, Associate Professor Lingling Wu, Assistant Professor Mitsuru Noda for your help and comments on my study.

Second, I would like to express my appreciation to all of my seniors and peers in the Tourism and Mobility Planning Laboratory for their help and advice. During the seminar, I learned a lot of theories and knowledge that are not my major, and I gained both knowledge and friendship during the discussion.

Finally, I would like to thank my parents and grandparents for their support and help. It is because of their selfless love and encouragement that I have the opportunity to go to more places and see more scenery. I would also like to thank my friends for sticking by me and encouraging me through both good and bad times.