Journal of Travel & Tourism Marketing

ISSN: 1054-8408 (Print) 1540-7306 (Online) Journal homepage: https://www.tandfonline.com/loi/wttm20

Topic modelling for theme park online reviews: analysis of Disneyland

Jian Ming Luo, Huy Quan Vu, Gang Li & Rob Law

JOURNAL OF TRAVEL & TOURISM MARKETING
2020, VOL. 37, NO. 2, 272–285
https://doi.org/10.1080/10548408.2020.1740138

ARTICLE

Topic modelling for theme park online reviews: analysis of Disneyland

Jian Ming Luo [a], Huy Quan Vu [b], Gang Li [c,d] and Rob Law [e]

[a]Faculty of International Tourism and Management, City University of Macau, Macao, Macao; [b]Department of Information Systems and Business Analytics, Deakin University, Melbourne, Australia; [c]School of Information Technology, Deakin University, Melbourne, Australia; [d]Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, Xinjiang, China; [e]School of Hotel & Tourism Management, The Hong Kong Polytechnic University, Hong Kong

ABSTRACT
This study introduces topic modelling into the analysis of theme park online reviews to determine visitor behaviour and experiences. An exploratory analysis involving major Disneyland theme parks is presented using a large-scale review data set. A comprehensive list of the topics discussed by visitors when visiting Disneyland theme parks is constructed. Insights into the interests and concerns of various visitor groups across theme parks are revealed. The proposed approach and findings are beneficial to support theme park managers in understanding visitors' perception, through which effective marketing and improvement plans can be developed to attract and retain future customers.

ARTICLE HISTORY
Received 23 October 2019
Revised 13 January 2020
Accepted 2 March 2020

KEYWORDS
Theme park; Disneyland; TripAdvisor; topic modelling; latent Dirichlet allocation

Introduction

The tourism and hospitality industry is one of the pillars of the world's economy. Particularly, the hedonic characteristics of theme parks have resulted in the growth of this segment (Dong & Siu, 2013). In 2016, consumer expenditure in theme parks reached 43.1 billion, which represented a 5.6% growth rate. The expected consumer expenditure in 2021 is approximately 56.5 billion (IAAPA, 2018). The trend on theme parks will continue because they provide new and diverse vacation experiences and offer the convenience of on-site accommodation, food service, recreation, shopping, and other tourist services in addition to the core recreational and entertainment activities (A. Milman, 2001).

Studies have attempted to explore the behaviour of tourists at theme parks because of the increasing importance of this aspect in the tourism and hospitality industry (Nuryyev & Achyldurdyyeva, 2015). Existing studies have focused on exploring tourists' expected experiences (Chen & Lamberti, 2013), participation experience (Cabanas, 2019; Dong & Siu, 2013; Wei et al., 2019), movement pattern (Tsai & Chung, 2012; Zhang et al., 2017) and behavioural intention (Bigné et al., 2005). Evidently, understanding visitor behaviour is important for theme park managers to market their theme parks and design day rundown (Fotiadis, 2016). The majority of the previous studies have used qualitative, quantitative, or mixed methods to study visitor behaviour. However, studies that have employed any of the aforementioned methods face a trade-off between sample collection cost and representativeness of estimation (Guo et al., 2017).

Many tourists examine their preferred destinations before making decisions (Xie et al., 2014). Current online consumer reviews are amongst the most important sources of information because they convey product experience and visitors' recommendation of tourism products (Guo et al., 2017). Additionally, online reviews can reduce typical asymmetric information (i.e. consumers have no or limited information regarding product quality) by providing first-hand experiences of other consumers. Hence, previous studies have employed online consumer reviews to understand consumers' preferences (Guo et al., 2017) in a wide range of topics (Schuckert et al., 2015). Online consumer reviews were employed to study tourist opinions on hotels (Calheiros et al., 2017; Sanchez-Franco et al., 2019; Ye et al., 2018), airlines (Korfiatis et al., 2019), restaurants (Dickinger et al., 2017; H. Li et al., 2019) and attractions (Dickinger et al., 2017). However, the majority of the existing research has mainly focused on the hotel sector (Schuckert et al., 2015). To the best of our knowledge, online consumer reviews have not been used to examine tourist behaviour in theme parks despite their immense potential. Theme park managers continue to face challenges in understanding visitors' behaviour and experiences for effective management and marketing of theme parks. The following important questions have yet to be thoroughly answered: (1) *What are the key topics that tourists express in online platforms?* (2) *What*

CONTACT Gang Li gang.li@deakin.edu.au School of Information Technology, Deakin University, Melbourne 3125, Australia

is the heterogeneity of perception among groups of tourists across these topics? and (3) *What are the important aspects influencing tourists' experience?*

This study aims to fill in the preceding research gap and presents an approach for analysing the behaviour and experiences of theme park visitors based on online reviews. Our approach utilises topic modelling, which is a group of data mining techniques that aim to discover hidden semantic structures in textual data, such as documents and reviews. Topic modelling has been employed in some recent studies in the tourism and hospitality field (Bi et al., 2019; Calheiros et al., 2017; Guo et al., 2017). However, its performance in exploring tourists' behaviour at theme parks has not been evaluated. The current research demonstrates the effectiveness of topic modelling in a case study of visitors' experience at Disneyland, which is a popular theme park to tourists worldwide (Fung & Lee, 2009). The analysis is based on a large-scale online review of various Disneyland theme parks. Detailed information on popular activities and discussed topics for each visitor group is revealed. This study is beneficial for tourism researchers and theme park managers using online reviews to gain insights into tourists' behaviour at theme parks.

The remainder of this paper is organised as follows. "Literature review" section discusses the theme parks used in this study and a review of the existing literature on employing online reviews and topic modelling. "Methodology" section describes our methodology for textual data processing and topic discovery. "Case study" section presents the case study with the result analysis and practical implications. Lastly, "Conclusion" section provides the conclusion and direction for further research.

Literature review

Theme park industry and Disneyland

Theme parks are facilities with a theme. The goal of theme parks is to provide an interesting experience to visitors (Heo & Lee, 2009). The World Columbian Exposition was the first amusement park and opened in Chicago in 1893 (Weinstein, 1992). The first theme park in the US opened in the late 19th century (Graft, 1986). Since then, theme parks have spread endemically throughout the US in the 1960s and 1970s. Eventually, these theme parks became destination parks in the 1980s and 1990s (Heo & Lee, 2009). Currently, the theme park industry worldwide continues to expand and attracts thousands of people throughout the year (Alexander et al., 2012). In 2017, the top 10 theme park groups served 475.8 million visitors, which represented an 8.6% increase from 2016 (Themed Entertainment Association [TEA], 2018) (see Table 1).

Table 1. Top 10 theme park groups worldwide.

Rank	Group Name	Attendance (2017)	Attendance (2016)	Change %
1	Walt Disney Attractions	150,014,000	140,483,000	6.8%
2	Merlin Entertainment Group	66,000,000	61,200,000	7.8%
3	Universal Parks and Resorts	49,458,000	47,356,000	4.4%
4	Oct Parks China	42,880,000	32,270,000	32.9%
5	Fantawild	38,495,000	31,639,000	21.7%
6	Chimelong Group	31,031,000	27,362,000	13.4%
7	Six Flags, Inc.	30,789,000	30,108,000	2.3%
8	Cedar Fair Entertainment Company	25,700,000	25,104,000	2.4%
9	Seaworld Parks & Entertainment	20,800,000	22,000,000	−5.5%
10	Parques Reunidos	20,600,000	20,825,000	−1.1%
Total		475,767,000	438,267,000	8.6%

(Source: Themed Entertainment Association, 2018).

Disneyland is a pioneer in theme parks. The first Disneyland "theme park" was introduced in 1955 in Anaheim, California. The original goal was to provide a safe, clean, aesthetically appealing, and imaginative entertainment facility to all family members (Milman, 1991). The park received huge success and achieved numerous records, such as reaching 1 million visitors in the first two months (Trahan, 2005). The experience in the park and the products consumed brought Disney's stories, characters, and franchises to life. The themes are experienced at the parks, resorts, cruises and vacation experiences, as well as through consumer products, such as toys, apparel, books, and video games (The Walt Disney Company, 2019a). The company owns and operates various theme parks worldwide, including Walt Disney World Resort in Florida, Disneyland Resort in California, Disneyland Paris, Tokyo Disney Resort, Aulani-a Disney Resort, and Spa in Hawaii, Disney Vacation Club, Disney Cruise Line, and Adventures by Disney. Moreover, the company manages and has effective ownership interests of 47% in Hong Kong Disneyland Resort and 43% in Shanghai Disney Resort (The Walt Disney Company, 2019b). In 2018, the global revenue of the parks increased by 10% (equivalent to 1.8 billion) to 20.3 billion. The increase in revenue can be further broken down to 6% increase in guest spending and 2% increase in the number of visitors. The increase in guest spending can be particularly explained by increased ticket price, food, beverage and merchandise spending, and average room rates (The Walt Disney Company, 2019b).

Online reviews and topic modelling

Online information, such as online reviews and newsgroups, specifically those generated by consumers, are important to travellers for travel decision-making (Hwang et al., 2018; Sparks et al., 2016). Marketing practitioners and researchers have become increasingly interested in this phenomenon (Wirtz et al., 2006) because information from other consumers is widely known to be reliable (Park et al., 2007). As this online information becomes more common and reliable, these data have also been utilised in numerous studies (Cheng et al., 2019). Lu and Stepchenkova (2015) stated several advantages of using online data. Firstly, data are easily available regardless of the time and method required to collect the information. Secondly, this information is frequently subject to minimal human influence, which is deemed reliable for determining authentic tourist experiences. Online information, such as reviews, has become increasingly appealing to researchers (Goh et al., 2013) because of its extensive details (Liang et al., 2017) and reveals what the consumers like and dislike (Banerjee & Chua, 2016). This information is easily accessible and immediately available on many social platforms (Guo et al., 2017). Moreover, these platforms provide real and actual information on tourist experiences by combining facts, personal opinions, feelings, and impressions (Wilson et al., 2012). Reviewers can use the most appropriate expressions to describe their impressions on and feelings for products (Wattanacharoensil et al., 2017) because of the absence of particular frameworks, guidelines, and restrictions on providing reviews.

Online reviews are often rich-in-text, which is a type of unstructured data and cannot be analysed using traditional statistical methods (Guo et al., 2017; Lu & Stepchenkova, 2015). Studies have combined several existing methods, such as web crawling, computational linguistics, data mining, machine learning, and other statistical methods, to collect, analyse and interpret textual data for meaningful insights (Baka, 2016; Xiang et al., 2017). Particularly, topic modelling is often used because of its capability of discovering topics hidden in textual data. Papadimitriou et al. (2000) firstly introduced the concept of topic modelling called "latent semantic indexing". The modelling of topics can be performed using various mathematical frameworks, such as singular value decomposition (Dumais, 2004), non-negative matrix factorisation (Arora et al., 2012) or probability theory with latent semantic analysis (Hofmann, 1999) and latent Dirichlet allocation (LDA) (Blei et al., 2003). LDA is the most popularly used technique (Liu, 2013) owing to its utilisation of the parse Dirichlet prior distributions. Documents are assumed to cover only a few topics and several words present topics. The representation of topics and words is in the form of probability distribution, which can be easily interpreted and analysed. Topic modelling has been successfully applied in various fields, such as medical science (Jiang et al., 2012), geography (Eisenstein et al., 2010), political science (Greene & Cross, 2015) and marketing (Reisenbichler & Reutterer, 2019). However, this type of modelling was only employed in a few recent studies in the tourism and hospitality field (Bi et al., 2019; Calheiros et al., 2017; Dickinger et al., 2017; Guo et al., 2017; Kim et al., 2019; Lim & Lee, 2019; Mazanec, 2017; Vu, Li, Law et al., 2019). The current study further explores the capability of topic modelling using LDA to discover new insights into visitors' behaviour and experiences in theme parks (Disneyland) to support theme park managers in effective management and improvement.

Methodology

This section presents the method for theme park review analysis by topic modelling. Data are extracted from an online review platform and pre-processed using text processing techniques. Thereafter, topic modelling technique is applied to discover the discussed topics within the reviews. Statistical analysis is conducted to explore and understand the topics.

Theme park review collection and processing

Theme park reviews are available on various online review platforms, such as Expedia, Google Reviews, and TripAdvisor. This study utilises TripAdvisor as the data source because the reviews on this platform are proven reliable in determining tourists' opinions and perceptions and have been widely used in tourism research (G. Li et al., 2013, 2015; Vu, Li, Law et al., 2019). The majority of the prior studies have utilised reviews on this platform to study the hotel and dining preferences of tourists. However, extensive reviews of theme parks are also available on this platform, but it has not been utilised by prior studies to study visitors' behaviour in theme parks. An automatic data extraction software was used to extract reviews on theme parks from TripAdvisor. The program can browse through the review pages of theme parks and extract textual review comments and other associated data, such as user location of origin, rating, and review time. Notably, the reviews are in free-text form, which should be processed before further analysis.

We adopted standard text processing techniques to process the theme park reviews (G. Li et al., 2015). Each review is loaded into a text-tokenising algorithm, which breaks the stream of text into words, phrases, and symbols or other meaningful elements called tokens. Thereafter, the tokens undergo a regular expression text filter, which removes any numbers or symbols. For example, uppercase letters are normalised into lowercase. The remaining tokens undergo a stemming process that reduces words to the base or root form. We followed the natural assumption that the English vocabulary of noun types is commonly used to refer to entities, such as theme park features, characters, and facilities. Therefore, we identified and extracted only stemmed words in noun form on the basis of a lexicon that contains a comprehensive list of English vocabulary of nouns. After preprocessing, each review is represented as a bag of words reflecting the entities mentioned by visitors related to their visits to theme parks.

Probabilistic topic modelling

This stage applies topic modelling to discover hidden topics in the review comments. We used LDA to accomplish this task. LDA is a powerful technique for topic discovery, which has proven performance in various applications (Kim et al., 2019; Lim & Lee, 2019; Mazanec, 2017). Let $R = \{r_1, r_2, \ldots, r_M\}$ be collection reviews, where each review r_i has a length of N words. The values of N vary depending on the length of each review. LDA models the review collection over a set of T topics, where topic$t$ is characterised by a vector of word probabilities $\Phi^t = \{\phi_1^t, \phi_2^t, \ldots, \phi_V^t\}$ and V refers to the number of words in the vocabulary of the processed reviews. Therefore, $\sum_{v=1}^{V} \phi_v^t = 1$ because Φ^t is a probability distribution. Across topics, the probability distribution of words varies. LDA models each review i as a mixture of topics with probability distributions $\Theta^i = \{\theta_1^i, \theta_2^i, \ldots, \theta_T^i\}$, $\sum_{t=1}^{T} \theta_t^i = 1$. The topic mixtures Θ and word probabilities Φ are assumed to follow a Dirichlet distribution with parameters α and β, respectively (Foulds et al., 2013). Furthermore, parameters α and Θ require inference by a learning process of LDA, such as maximum likelihood estimation (Asuncion et al., 2009) or stochastic solver (Hoffman et al., 2013).

A *word* $\times$ *document* matrix is used as input into the LDA model, where each row is a vector of length V. The value of each element pertains to the count of a corresponding word in the review. After the learning process, the topic mixture distribution Θ^i of each review and word probabilities of topic Φ^t are extracted for further analysis.

Each topic typically has only a few words with high probability presenting the meaning of the corresponding topics because modelling LDA utilises the Dirichlet prior distributions. We examined the word and corresponding probabilities Φ^t to determine a suitable label for each topic. To determine the differences, the topic mixture probabilities Θ^i of the reviews can be examined with respect to the reviews made by various travel groups. The topics can also be analysed with respect to review ratings, thereby enabling managers to identify the topics that receive considerable positive or negative comments from visitors.

Case study

Data collection

We collected reviews posted on TripAdvisor for three popular Disneyland parks, namely, Hong Kong Disneyland, Disneyland Paris, and Disneyland California. The data collection was carried out in April 2019. These parks are located in different continents, thereby enabling us to gather the reviews of tourists from many countries. We only extracted the reviews in English because of language barrier. However, in the collected data set, individuals coming from countries where English is not the first language provided reviews written in English. Thus, the collected data set can capture the discussed topics from visitors across demographic backgrounds. We collected review comments, visit time, rating and reviewers' location of origin. Table 2 shows that 43,869 reviews were collected from the theme parks. Disneyland California has the most number of reviews because it is the oldest theme park amongst the three Disneyland parks. The number of reviews exceeds the number of users because visitors can visit a theme park more than once, thereby enabling them to post several review comments.

Table 3 shows the number of reviews by reviewer location and the proportions of reviews by theme park with respect to traveller group. We find that the distribution of the reviews aligns with the approximate location of reviewers to the nearest theme parks. For example, Hong Kong Disneyland is located in Asia and attracts many visitors from neighbouring countries, such as the Philippines, Singapore, Malaysia, and Indonesia, as indicated by the high proportion (i.e. 70%) of reviews generated by people from these countries. Disneyland Paris attracts numerous visitors from European countries, such as the UK, Ireland, the Netherlands, and France. However, visitors from Australia and New Zealand are more likely to visit Disneyland California than Hong Kong Disneyland despite the latter being closely located to these countries. This tendency is probably caused by the preference of visitors from these countries.

Table 2. Theme park review data sets.

Theme Park	No. of Reviews	No. of Users
Hong Kong Disneyland	9,797	9,784
Disneyland Paris	14,088	14,087
Disneyland California	19,984	19,978
Total:	43,869	43,849

Table 3. Review distribution by country of origin.

Reviewer Location	Proportion (%) Hong Kong Disneyland	Disneyland Paris	Disneyland California	No. of Reviews
US	6.07	9.15	84.78	14,509
UK	7.59	81.96	10.45	9,744
Australia	34.96	12.72	52.32	4,679
Canada	9.17	8.41	82.41	2,235
India	71.67	19.39	8.93	1,511
Philippines	92.62	2.24	5.14	1,070
Singapore	82.16	8.87	8.97	1,037
New Zealand	22.09	8.20	69.71	756
Malaysia	86.22	8.67	5.10	588
Hong Kong	95.13	2.53	2.35	554
Indonesia	88.49	6.60	4.91	530
Ireland	4.31	88.30	7.39	487
United Arab Emirates	59.14	29.71	11.14	350
Netherlands	14.23	72.33	13.44	253
France	8.64	82.30	9.05	243
South Africa	53.72	35.54	10.74	242
Thailand	83.86	9.42	6.73	223
Germany	19.59	66.49	13.92	194
China	86.19	7.18	6.63	181
Spain	15.65	70.07	14.29	147
Others (142 countries)				4,336

Figure 1(a) shows the number of reviews by year for each theme park. The majority of the reviews were generated between 2015 and 2016. Only a few reviews have been generated in recent years. The two possible reasons are as follows. Firstly, the number of visitors in Hong Kong is decreasing. For example, the number of visitors of Hong Kong Disneyland decreased from 7.4 million in 2013 to 6.7 million in 2018 (Statista, 2019). Secondly, Disneyland released its official app in August 2015. This app enables visitors to change their communication platform from posting online comments to smartphone app. Weinberger (2016) reported that 60% of all Disneyland guests use apps on their visit. Nevertheless, the collected data set from TripAdvisor is sufficiently large and comprehensive for the purpose of this case study. Additionally, the method presented in this research is general and applicable to park reviews on other platforms. Figure 1(b) shows the number of reviews by month for each theme park. The peak time for Hong Kong Disneyland is in December, whereas that for Disneyland Paris is in August probably because this month is a summer holiday in European countries. Disneyland California is popular throughout the year, except for January, February, and November. These seasonal patterns can be explained by the local weather. California is sunny throughout the year except during January and February. The weather patterns in Hong Kong in December and Paris in August are also excellent for outdoor activities. Fair weather contributes to the number of visitors in theme parks because many of the activities in Disneyland involve outdoor facilities, such as rollercoasters. We perform several analyses in the subsequent sections to explore the theme park review data set.

Rating analysis

We perform an initial analysis on the rating behaviour of visitors to obtain an overview of their satisfaction towards theme parks. We only account for the top 10 visitor groups in our data set because they provide sufficient data for analysis. We count the number of reviews with respect to the rating values and normalised them to present the probability distribution of ratings for each visitor group. The rating distributions lean towards high ratings of 4 and 5 (see Figure 2), but variations exist amongst the ratings of the theme parks and amongst the travel groups. We adopt the Kullback–Leibler (KL) divergence to quantitatively model the shape of the distributions (Press et al., 2007). KL divergence measures how one probability distribution differs from a second distribution. We use the KL divergence to measure the difference between the probability distribution of rating for each group and a random distribution. A high KL value indicates that the distribution of rating is skewed towards the high rating value and vice versa. Figure 3 plots the KL distance values by travel group and theme park. We find that Disneyland Paris tends to

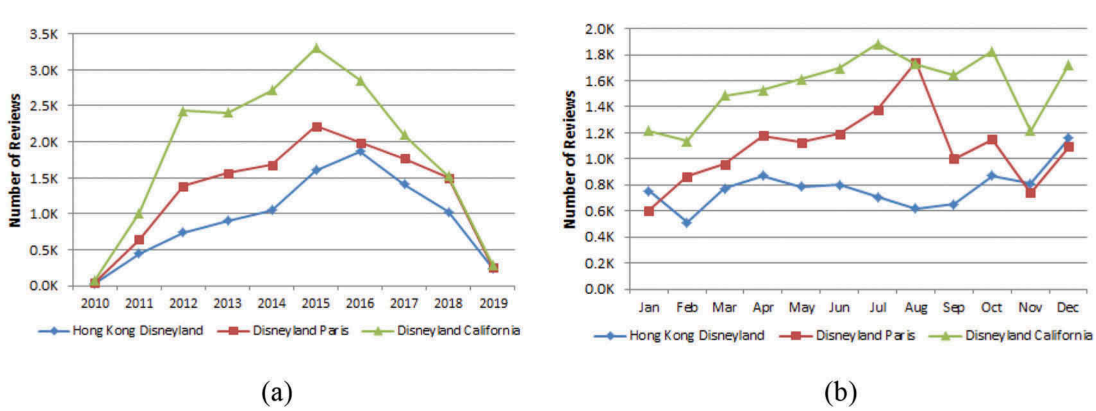

Figure 1. Number of reviews by (a) Year and (b) Month.

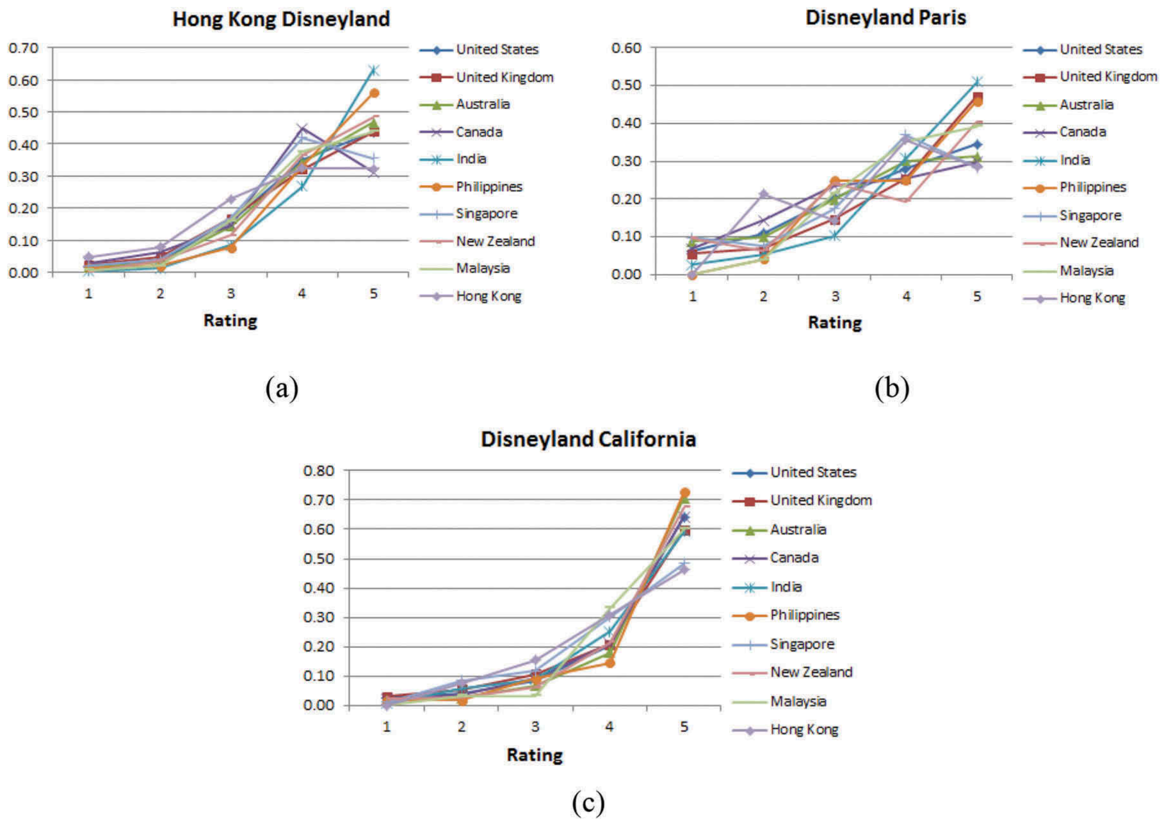

Figure 2. Rating distribution by reviewer group.

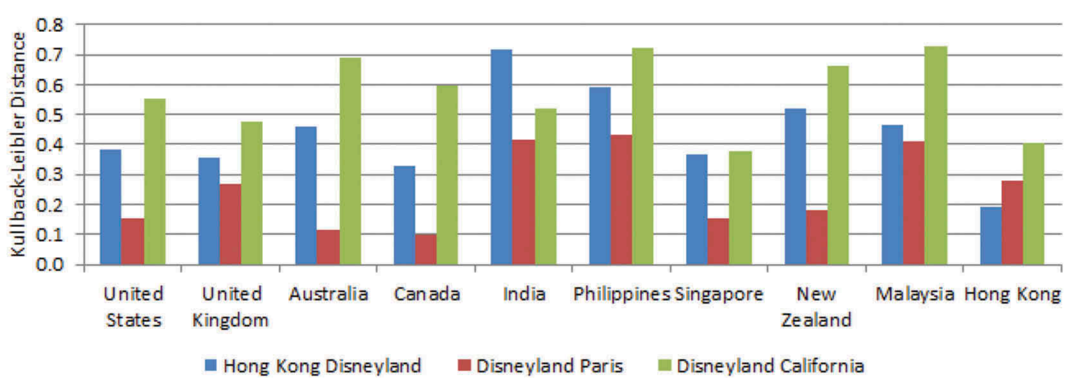

Figure 3. KL divergence between the rating distributions and random distribution.

have the lowest ratings because the KL distance values are the lowest for the majority of the travel groups. Disneyland California generally received the highest ratings amongst the theme parks. The difference in the rating distributions can reflect different levels of satisfaction when travellers visit these theme parks. Hong Kong citizens tend to give low ratings to Hong Kong Disneyland, whereas visitors from India tend to give this theme park high ratings. This tendency may be the result of the preferences and expectations amongst groups of visitors, such that their satisfaction levels differ even for the same theme park. Local visitors with annual passes visit the park multiple times, whereas others may only visit the park on special dates, such as birthdays and anniversaries (Fung & Lee, 2009). Therefore, the prior experiences of visitors play an important role in visitor satisfaction (Geissler & Rucks, 2011), thereby possibly influencing their ratings.

Topic discovery

In topic modelling using LDA, users should specify the value *T* for the number of topics to be learned from the review collection. The value of T can influence the quality of discovered topics. Thus, a suitable number T should be selected. We apply a common model evaluation strategy for LDA on the basis of an evaluation metric called perplexity (Hoffman et al., 2013). Perplexity reflects the goodness-of-fit of the learned LDA models when several topics are specified. A superior model would produce a low perplexity value. We adopt the 10-fold cross-validation approach (Kuhn & Johnson, 2013), in which an entire review data set is partitioned into 10. One part is for testing and the other nine parts are used as training data for LDA. The process is repeated 10 times for each testing part. Figure 4 presents the overall perplexity values with several topics. The perplexity decreased significantly when only a few topics are used, thereby indicating that the model can considerably determine review contents when numerous topics are used in the modelling process. However, the decreasing rate of perplexity values decreased to approximately 41 and fluctuated thereafter, thereby indicating that using a topic number above 41 does not considerably improve the model performance. By contrast, the model takes considerable time to compute with added topic numbers. Elbow method (KETCHEN Jr. & Shook, 1996) is often employed to identify a suitable topic number based on the experiment result. Therefore, we select $k = 41$, which is sufficient to model the topics in the text corpus with less computation time. The performance of the LDA model slightly varies across runs because initialising the topics is a random process. Therefore, we run the LDA model multiple times and select the model with the lowest perplexity as the final model for topic analysis. Three experienced researchers examined the returned topics and words in each topic to determine the topic labels.

Table 4 provides several topics and the popular words in each topic. Reviewers frequently use such words as *plan, tip, check, review,* and *map* when discussing trip planning and the related information. Frequent words in topics on dining options at theme parks include *meal, book hotel, dinner, lunch,* and *breakfast*. Other topics mentioned are *cooling facilities* under hot weather in summer or *non-human* and *human characters* in Disney stories. Figure 5 provides a list of 41 identified topics with labels and probability values.

We find that the most frequently mentioned topics are *happy experience, online information,* and *shopping options*. We group these topics into several categories for ease of interpretation (Table 5). Some topics are on

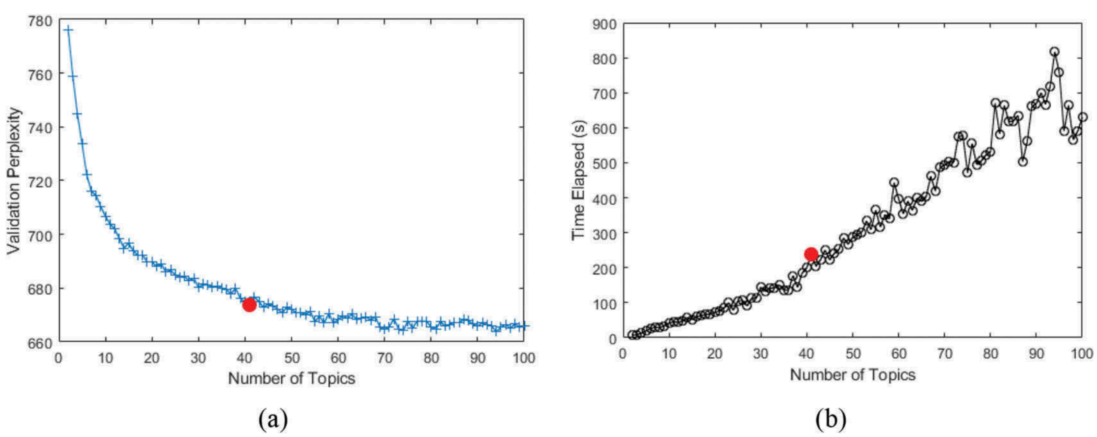

Figure 4. Validation results for several topics.

Table 4. Popular words and probability values in some example topics.

Planning Information		Dining Options		Cooling Facilities		Non-human Characters		Human Characters	
Plan	0.163	Meal	0.071	Hot	0.189	Toy	0.079	Pan	0.131
Tip	0.050	Book	0.069	Heat	0.081	Story	0.077	Peter	0.129
Check	0.040	Hotel	0.048	Summer	0.076	King	0.063	Buzz	0.065
Miss	0.037	Dinner	0.046	Sun	0.075	Lion	0.057	Pirate	0.054
Review	0.032	Lunch	0.042	Weather	0.061	Mickey	0.050	Snow	0.054
Map	0.030	Breakfast	0.038	Shade	0.054	Space	0.037	Thunder	0.049
Afternoon	0.030	Cafe	0.036	Air	0.045	Manor	0.036	White	0.049
Lunch	0.027	Table	0.032	Condition	0.041	Miss	0.027	Flight	0.043
Research	0.026	Voucher	0.030	Sit	0.034	Mine	0.025	Dumbo	0.029
Read	0.025	Buffet	0.029	Super	0.033	Cruise	0.025	Caribbean	0.026
Adventure	0.022	Half	0.026	Umbrella	0.029	River	0.024	Pinocchio	0.024
Schedule	0.021	Reservation	0.022	Fan	0.028	Gulch	0.023	Laser	0.024
Night	0.019	Board	0.022	Rest	0.017	Car	0.023	Space	0.021
Advance	0.018	Service	0.018	Drink	0.017	Fantasy	0.021	Castle	0.018

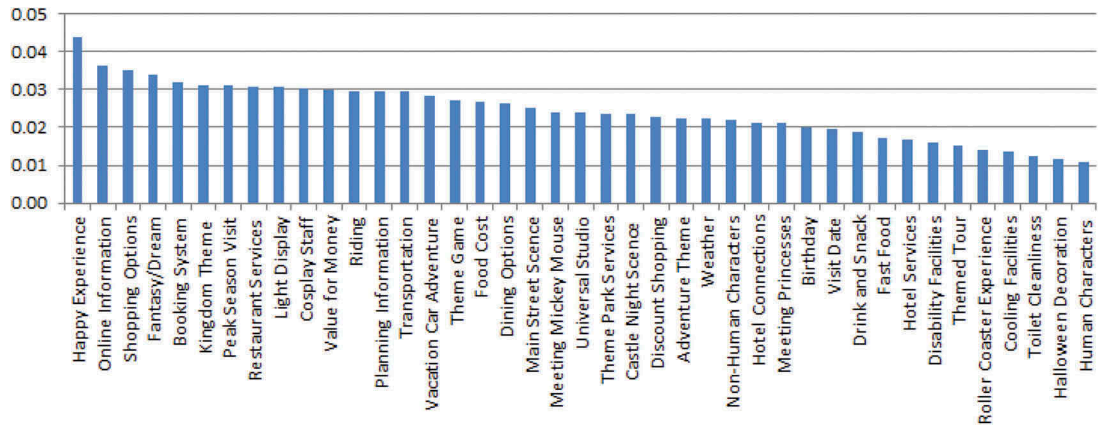

Figure 5. Probability distribution of topics.

Table 5. Topic grouping.

General Discussion	Themes	Dining/ shopping	Facilities	Services
Happy experience	Kingdom theme	Dining options	Online information	Restaurant services
Fantasy/ dream	Light display	Drink and snack	Booking system	Theme park services
Peak season visit	Theme game	Fastfood	Planning information	Hotel services
Cosplay staff	Main street scene	Shopping options	Transportation	Toilet
cleanliness				
Value for money	Meeting Mickey Mouse	Discount	shopping	Hotel
connections				
Riding	Universal Studios		Disability facilities	
Vacation car adventure	Castle night scene		Cooling facilities	
Food cost	Adventure theme			
Weather	Non-human characters			
Birthday	Meeting princesses			
Visit date	Themed tour			
Roller coaster experience	Halloween decoration			
	Human characters			

general discussions in relation to visitor experience and context of visit. Other topics mention the themes displayed in Disneyland, dining and shopping, facilities, and services. For detailed insights, the next section further analyses the distributions of the identified topics with respect to individual theme parks and visitor groups.

Topic distribution analysis

All theme parks in this study are of the same type, managed by the same company, located in different parts of the world and attract relatively different groups of visitors. Understanding the differences in topic popularity among the parks and visitor groups is beneficial for managers in various management and marketing tasks at each theme park. Figures 6 and 7 illustrate the computation and frequency of topic probability by theme parks and visitor group, respectively. The findings show that the majority of the topics have a similar probability to be mentioned in all theme parks, except for certain topics. Particularly, *vacation car adventure* and *adventure theme* are frequently mentioned in Disneyland California compared with the others. The topic on *non-human characters* is most frequently mentioned for Hong Kong Disneyland. The most frequently mentioned topics for Disneyland Paris are *food cost* and *dining options*. These differences could be attributed to the nature of theme parks or interests and preferences of visitors. We further examine the topic distribution by review groups for detailed insights (Figure 7).

We find that visitors from the US and Canada frequently mention *vacation car adventure*. This interesting finding indicates that visitors in these countries tend to visit Disneyland as part of their road trip during vacations. Travellers from the UK are generally concerned with *food cost* and *dining options* when visiting Disneyland Paris, as indicated by the highest probabilities for these topics. Figure 7 also shows that this group is most concerned with *hotel service* and *toilet cleanliness*. Additionally, the popularity of several topics considerably varies across groups, such as *non-human characters, happy experience, meeting Mickey Mouse, fantasy/dream,* and *discount shopping*. To verify the statistical significance, we perform ANOVA with a significance level of $p < 0.05$. The majority of the topics displayed significant differences except for *castle night scene*. Table 6 presents the result of certain topics with the largest differences attributed due to length limitation. Note that the values in the difference column reflect the differences between the highest and lowest groups, while the F-score reflects the overall variances between all groups. Topic with largest difference as

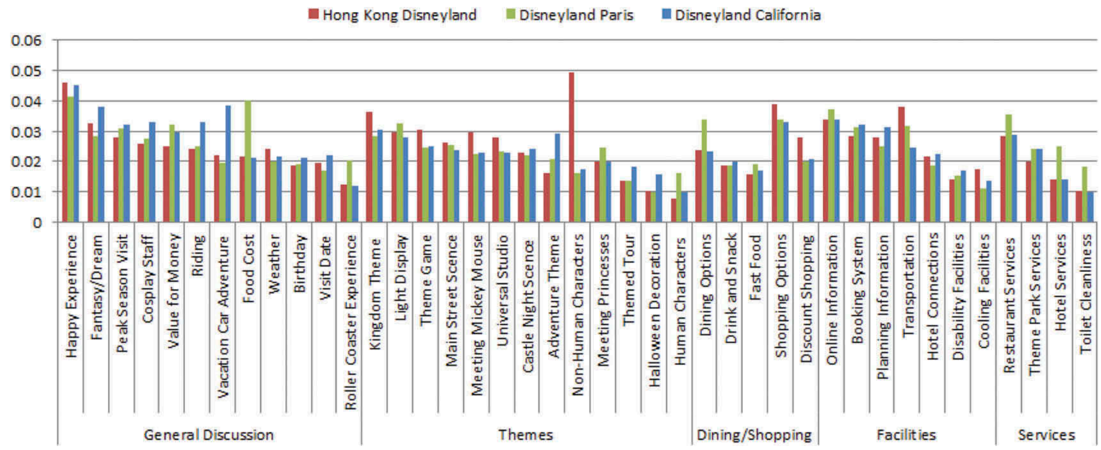

Figure 6. Topic popularity by theme park.

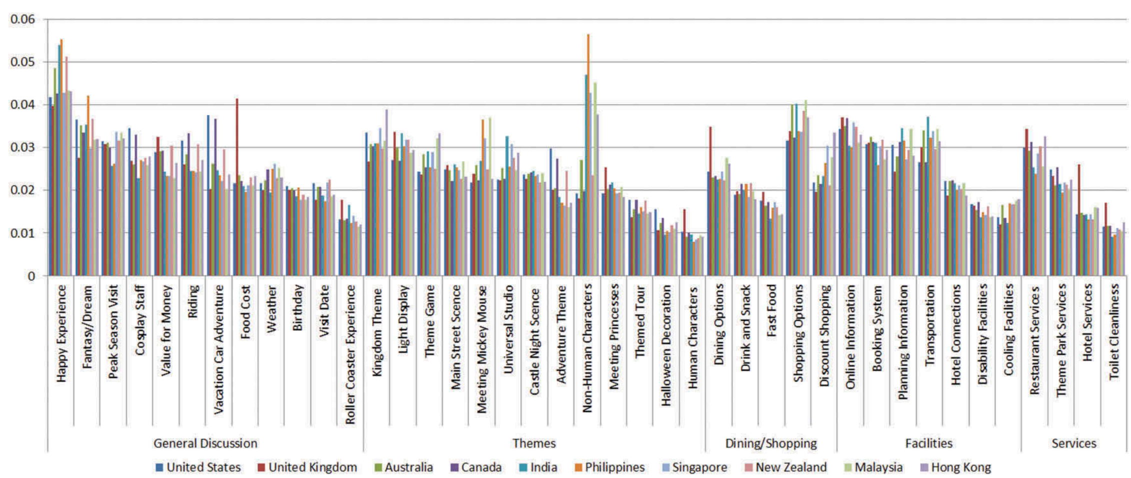

Figure 7. Topic popularity by reviewer group (this figure is best viewed in the electronic version of this paper).

Table 6. ANOVA of topics among the groups.

Reviewer Groups	Difference	F-statistic	p-value*
Non-human Characters	0.038	285.629	0.000
Food cost	0.022	265.815	0.000
Vacation car adventure	0.017	204.456	0.000
Happy experience	0.016	44.540	0.000
Meeting Mickey Mouse	0.015	48.929	0.000
Fantasy/dream	0.015	55.340	0.000
Discount shopping	0.014	36.422	0.000
Adventure theme	0.014	85.258	0.000
Hotel services	0.013	129.320	0.000
Dining options	0.012	78.389	0.000
Kingdom theme	0.012	30.495	0.000
Cosplay staff	0.012	58.990	0.000
Transportation	0.011	34.445	0.000
Restaurant services	0.011	25.960	0.000
Universal Studios	0.010	34.036	0.000
Planning information	0.010	36.997	0.000
Value for money	0.010	26.637	0.000
Theme game	0.010	25.792	0.000

*Significant at p-value $\leq$ 0.05

well as variances are *non-human characters, food cost, vacation car adventure*. Although *hotel services* are not among the topic with the highest differences, the variances between groups are high as indicated by high F-score.

We further examine the visitors' ratings with respect to the topic to identify their perceptions towards various aspects of the theme parks. Figure 8 shows the rating distributions with respect to the topic. We find that visitors tend to provide high ratings in reviews about *happy experience, fantasy,* and *dream*. However, they are generally disappointed with *value for money* and *food cost*, which is shown by the distributions oriented towards low ratings. They are happy with *shopping options* but tend to give low ratings when discussing *discount shopping*. The majority of the travellers tend to give high ratings for displayed themes (e.g. *light display, theme game, main street scene, castle night scene*, and *adventure theme*) at the parks. *Online information* and *booking system* are likely to receive low ratings. *Disability facilities, restaurant services, theme park services,* and *toilet cleanliness* frequently received low ratings.

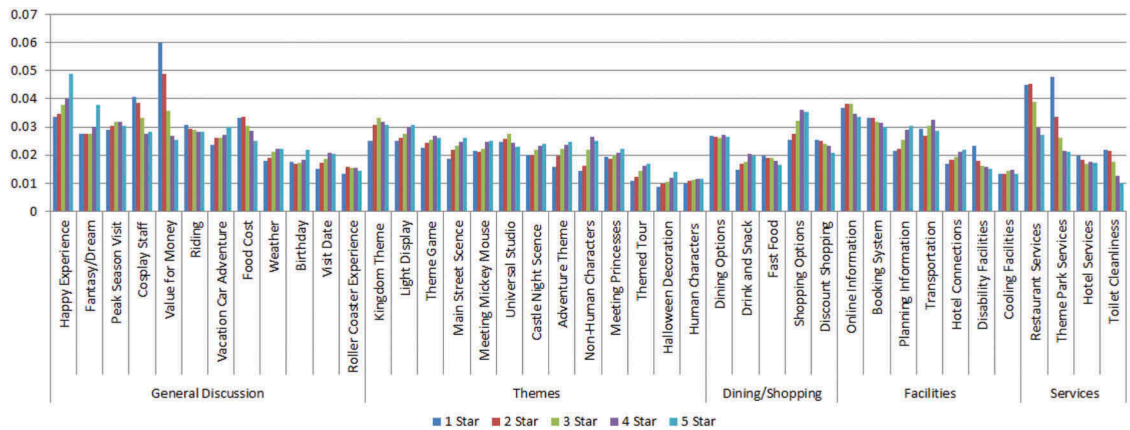

Figure 8. Rating distribution of topics.

Discussion

This study demonstrated the capability of topic modelling in discovering the behaviour of theme park visitors. A comprehensive list of topics discussed by visitors describing their theme park experiences was constructed. Unlike the majority of previous studies employing topic modelling (Bi et al., 2019; Guo et al., 2017; Kim et al., 2019; Mazanec, 2017), where the researchers predetermined the topic number, we evaluated the models to determine the optimal number of topics for modelling and subsequent analysis. The constructed model can effectively determine the underlying topics in the review data. The findings in the present study are consistent with and complement the findings in prior research. For example, *happy experience* is most important for all three Disneyland theme parks. Disneyland has been catering to the American dream and fantasy worldwide for a long time. Accordingly, this theme is one of the most important elements of the success of Disneyland (Hench, 2008). The analysis of rating distribution across topics (Figure 8) reveals that when visitors share their happy experiences, they tend to provide high ratings. Similarly, when visitors are sharing other experiences, such as *value for money, food price,* and *cleanliness*, they tend to provide low ratings. These results are consistent with the early findings of A. Milman (2009), Geissler and Rucks (2011), and Dong and Siu (2013). Nevertheless, the list of topics identified in our analysis (Table 5) is more comprehensive compared with those of previous studies on theme parks because we identified the topics from a large-scale review data set. Furthermore, our data were not limited to predetermined topics designed by previous researchers using the traditional survey and questionnaire methods. Theme parks are a service-orientated industry and should offer an interactive instead of a passive experience (A. Milman, 2001). Thus, the antecedents and consequences of visitors' experiences should be understood by theme park managers (Dong & Siu, 2013). Such an experience was captured in the identified topics, specifically those in the general discussion groups on *happy experience, fantasy/dream* and *roller coaster experience*. Information on the context and motivation for the visits is also captured. For example, people visit theme parks on *birthdays* of family members or as part of a *vacation car adventure.*

The analysis indicates that the popularity of topics varies amongst theme parks and groups of visitors (Figures 6 and 7). Thus, targeted marketing strategies can be developed to attract visitors to specific theme parks. For example, advertising materials that target American and Canadian residents can include information on *vacation car adventure* to attract them to Disneyland California. Non-human characters, such as those in *Toy Story, Lion King,* and *Mickey Mouse* can be incorporated into advertising materials to target visitors from the Philippines, India, Malaysia, and Singapore, as indicated by the high popularity of these topics. Information on *dining options* can be advertised to potential visitors from the UK. Additionally, we highlight that Disneyland across locations tends to exhibit different themes in relation to the location and destination image of Disneyland. Such destination images may influence people's perception of the corresponding Disneyland locations. For example, Hong Kong is a popular shopping destination, thereby making the availability of shopping important for visitors' experience. The results show that *shopping options* and *discounted shopping* are frequently mentioned in the reviews for Hong Kong Disneyland (Figure 6). Given that, France is famous for its cuisine; *dining options,* *food cost,* and *restaurant services* are frequently mentioned in the reviews of Disneyland Paris. Therefore, managers should incorporate the destination image of the corresponding destinations to promote the theme parks. Particularly, Hong Kong Disneyland should advertise and provide additional shopping options for potential visitors. It may collaborate with other popular brands, fashion, or luxury to enhance the shopping experience of visitors.

Another benefit of this study is that visitors' satisfaction was also analysed with respect to the topics, thereby providing valuable intelligence for managers in facilitating and improving theme park performance. For example, managers should considerably focus on improving the services at parks and restaurants, hotel and toilet areas. Additional effort should be exerted to improve visitors' perception towards *food cost* and *value for money* (i.e. providing high-quality food and service or creating flexible and combo deals for visitors). Theme park managers can use the corresponding results to improve service management and enhance the competitiveness of the company (Hou et al., 2019). Doing so will definitely attract many visitors and encourage spending as well. Moreover, we found queue time for *riding* an important factor that negatively affects visitors' experiences. When visitors spend several hours waiting to enter the attraction and rides, their visiting experience will be negatively affected. The effect will become severe when the park is filled with people and they rush into certain areas simultaneously (Brown et al., 2013). Park managers can provide recommendation for alternative rides and attractions to avoid overloading specific rides or attraction during peak hours.

Conclusion

Theme parks have become an increasingly important segment of the tourism and hospitality industry, which attracts visitors and generates revenues for tourism destinations. Insights into visitors' experiences are important for theme park managers in managing and promoting their businesses. However, theme parks, even if managed by the same group, possess different characteristics, thereby resulting in their visitors having varied expectations and concerns. Evidently, obtaining such a comprehensive understanding is a challenging task for theme park managers, specifically when traditional approaches often rely on surveys and questionnaires. Online reviews have emerged as a useful source of information for various studies on tourism and hospitality. However, their potential in supporting the management of theme parks has not been fully utilised to address the existing knowledge gaps. This study aimed to address these shortcomings and proposed to analyse theme park online reviews by topic modelling to capture and reveal comprehensive insights into visitor experience, concerns, and satisfaction. The case study involving Disneyland theme parks demonstrated the effectiveness of this approach and at the same time reveals useful findings with various practical implications for theme park management, especially for Disneyland, which is one of the most popular theme park groups worldwide. The presented approach is general, which can be applied to the online reviews of other theme park types and groups for comprehensive understanding.

This study has several limitations despite the comprehensive analysis on a large-scale data set. An analysis was performed only for Disneyland, the results of which may be inapplicable to other types of theme parks (e.g. water parks and amusement parks) or theme park groups (see Table 1). The analysis focused on only three out of six worldwide Disneyland parks. The findings and recommendation may not be generalised to other Disneyland theme parks because they may have different characteristics and visitors' experiences compared with those included in the current study. Hence, Disneyland managers should conduct further study on other theme parks for appropriate practical applications. Moreover, we only used reviews written in English owing to the language barrier. The identified topics may not cover all possible visitor concerns and experiences, particularly those from countries where English is not the first language. Other text processing techniques can be considered to process non-English reviews for a comprehensive understanding. The authors determined the labels of the topics to describe the topic meanings. Thus, they may not be the best labels for the topics. However, these labels should have captured the general meanings of the topics because the wording of the labels is determined based on the most popular words that appear for each topic. The number of topics to be specified for the LDA model is subject to the nature of the collected data set. A larger data set may capture more topics, whereas a smaller data set may contain less topics. Nevertheless, determining the appropriate topic numbers should be evaluated using perplexity and cross-validation. The present approach is not meant to completely replace the traditional approaches (i.e. using survey and questionnaire) despite the major advantages of the former. Accordingly, this study used a complementary approach. Online platforms typically provide limited information on the demographics and background of reviewers, in which case, traditional survey approaches would still be required.

Future studies can investigate text processing techniques for languages other than English, such that more reviews from more visitors can be included to generate further results that are representative and comprehensive. Reviews from other review sites and platforms could also be considered in future studies. Additionally, we highlight that researchers should focus on popular, less popular, new, and emerging topics when analysing online reviews. Analysis of topic popularities and ratings can be carried out with respect to temporal dimension to examine how topics and visitors' experiences evolve over time. Rating distributions of topics can also be done with respect to nationalities or demographics of the visitors for insights into their experiences. New topics can cover new games, new technologies, or new types of services that visitors expect to experience in theme parks. In this manner, useful information can be provided to managers for proposing future development plans and maximising the potential benefit that theme parks can generate for the tourism and hospitality industry in tourism destinations.

Acknowledgments

The work was completed when Gang Li is on ASL in Chinese Academy of Sciences, and we also would like to thank Deakin University's ASL fund and Xinjiang research fund with Chinese Academy of Sciences.

Disclosure Statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Chinese Academy of Sciences [Xinjiang research fund]; Deakin University [ASL fund]; Hong Kong Polytechnic University [SHTM]; Macau Foundation [MF].

ORCID

Jian Ming Luo http://orcid.org/0000-0002-5779-1152
Huy Quan Vu http://orcid.org/0000-0003-1947-2879
Gang Li http://orcid.org/0000-0003-1583-641X
Rob Law http://orcid.org/0000-0001-7199-3757

References

Alexander, M., MacLaren, A., O'Gorman, K., & White, C. (2012). Priority queues: Where social justice and equity collide. *Tourism Management*, *33*(4), 875–884. https://doi.org/10.1016/j.tourman.2011.09.009

Arora, S., Ge, R., & Moitra, A. (2012, October). Learning topic models–going beyond SVD. *In 2012 IEEE 53rd annual symposium on foundations of computer science*,(vol. *2*, pp. 1–10). New Brunswick, NJ: IEEE.

Asuncion, A., Welling, M., Smyth, P., & Teh, Y. W. (2009, June). On smoothing and inference for topic models. *In Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence* (pp. 27–34). Montreal Quebec, Canada: AUAI Press.

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. (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

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

Bigné, J. E., Andreu, L., & Gnoth, J. (2005). The theme park experience: An analysis of pleasure, arousal and satisfaction. *Tourism Management*, *26*(6), 833–844. https://doi.org/10.1016/j.tourman.2004.05.006

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*(1), 993–1022. https://dl.acm.org/doi/10.5555/944919.944937

Brown, A., Kappes, J., & Marks, J. (2013). Mitigating theme park crowding with incentives and information on mobile devices. *Journal of Travel Research*, *52*(4), 426–436. https://doi.org/10.1177/0047287512475216

Cabanas, E. (2019). Experiencing designs and designing experiences: Emotions and theme parks from a symbolic interactionist perspective. *Journal of Destination Marketing & Management*. https://doi.org/10.1016/j.jdmm.2018.12.004

Calheiros, A. C., Moro, S., & Rita, P. (2017). Sentiment classification of consumer-generated online reviews using topic modeling. *Journal of Hospitality Marketing & Management*, *26*(7), 675–693. https://doi.org/10.1080/19368623.2017.1310075

Chen, S., & Lamberti, L. (2013). Segmenting Chinese tourists by the expected experience at theme parks. *International Journal of Engineering Business Management*, *5*, 5–22. https://doi.org/10.5772/56605

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

Dickinger, A., Lalicic, L., & Mazanec, J. (2017). Exploring the generalizability of discriminant word items and latent topics in online tourist reviews. *International Journal of Contemporary Hospitality Management*, *29*(2), 803–816. https://doi.org/10.1108/IJCHM-10-2015-0597

Dong, P., & Siu, N. Y. M. (2013). Servicescape elements, customer predispositions and service experience: The case of theme park visitors. *Tourism Management*, *36*, 541–551. https://doi.org/10.1016/j.tourman.2012.09.004

Dumais, S. T. (2004). Latent semantic analysis. *Annual Review of Information Science and Technology*, *38*(1), 188–230. https://doi.org/10.1002/aris.1440380105

Eisenstein, J., O'Connor, B., Smith, N. A., & Xing, E. P. (2010, October). A latent variable model for geographic lexical variation. *In Proceedings of the 2010 conference on empirical methods in natural language processing* (pp. 1277–1287). Cambridge, Massachusetts: Association for Computational Linguistics.

Fotiadis, A. K. (2016). Modifying and applying time and cost blocks: The case of E-da theme park, Kaohsiung, Taiwan.

Tourism Management, *54*, 34–42. https://doi.org/10.1016/j.tourman.2015.10.013

Foulds, J., Boyles, L., DuBois, C., Smyth, P., & Welling, M. (2013, August). Stochastic collapsed variational Bayesian inference for latent Dirichlet allocation. *In Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 446–454). ACM.

Fung, A., & Lee, M. (2009). Localizing a global amusement park: Hong Kong Disneyland. *Continuum*, *23*(2), 197–208. https://doi.org/10.1080/10304310802711973

Geissler, G. L., & Rucks, C. T. (2011). The overall theme park experience: A visitor satisfaction tracking study. *Journal of Vacation Marketing*, *17*(2), 127–138. https://doi.org/10.1177/1356766710392480

Goh, K. Y., Heng, C. S., & Lin, Z. (2013). Social media brand community and consumer behavior: Quantifying the relative impact of user-and marketer-generated content. *Information Systems Research*, *24*(1), 88–107. https://doi.org/10.1287/isre.1120.0469

Graft, J. H. (1986). The future of amusement parks and attraction industry. *Tourism Management*, *7*(1), 60–62. https://doi.org/10.1016/0261-5177(86)90059-2

Greene, D., & Cross, J. P. (2015, June). Unveiling the political agenda of the european parliament plenary: A topical analysis. *In Proceedings of the ACM web science conference* (p. 2). ACM. Oxford, United Kingdom.

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

Hench, J. (2008). *Designing Disney: Imagineering and the art of the show*. Disney Editions.

Heo, C. Y., & Lee, S. (2009). Application of revenue management practices to the theme park industry. *International Journal of Hospitality Management*, *28*(3), 446–453. https://doi.org/10.1016/j.ijhm.2009.02.001

Hoffman, M. D., Blei, D. M., Wang, C., & Paisley, J. (2013). Stochastic variational inference. *The Journal of Machine Learning Research*, *14*(1), 1303–1347. https://dl.acm.org/doi/10.5555/2567709.2502622

Hofmann, T. (1999). Probabilistic latent semantic indexing. *Proceedings of the 22nd annual international ACM SIGIR conference on research and development in information retrieval*. (pp. 50–57). Berkeley, CA, USA.

Hou, Z., Cui, F., Meng, Y., Lian, T., & Yu, C. (2019). Opinion mining from online travel reviews: A comparative analysis of Chinese major OTAs using semantic association analysis. *Tourism Management*, *74*, 276–289. https://doi.org/10.1016/j.tourman.2019.03.009

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

IAAPA. (2018). *Global theme and amusement park outlook: 2017-2021*. International Association of Amusement Parks and Attractions. Received 06 June, 2019 from https://www.iaapa.org/about-iaapa/2018-iaapa-annual-report

Jiang, Z., Zhou, X., Zhang, X., & Chen, S. (2012, October). Using link topic model to analyze traditional chinese medicine clinical symptom-herb regularities. *In 2012 IEEE 14th international conference on e-health networking, applications and services (Healthcom)* (pp. 15–18). Beijing, China: IEEE.

Ketchen Jr., D. J., & Shook, C. L. (1996). The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management Journal*, *17*(6), 441–458. https://doi.org/10.1002/()1097-0266

Kim, K., Park, O., Barr, J., & Yun, H. (2019). Tourists' shifting perceptions of UNESCO heritage sites: Lessons from Jeju Island-South Korea. *Tourism Review*, *74*(1), 20–29. https://doi.org/10.1108/TR-09-2017-0140

Korfiatis, N., Stamolampros, P., Kourouthanassis, P., & Sagiadinos, V. (2019). Measuring service quality from unstructured data: A topic modeling application on airline passengers' online reviews. *Expert Systems with Applications*, *116*, 472–486. https://doi.org/10.1016/j.eswa.2018.09.037

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (vol. 26). Springer.

Li, G., Law, R., Vu, H. Q., & Rong, J. (2013). Discovering the hotel selection preferences of Hong Kong inbound travelers using the Choquet Integral. *Tourism Management*, *36*, 321–330. https://doi.org/10.1016/j.tourman.2012.10.017

Li, G., Law, R., Vu, H. Q., Rong, J., & Zhao, X. R. (2015). Identifying emerging hotel preferences using emerging pattern mining technique. *Tourism Management*, *46*, 311–321. https://doi.org/10.1016/j.tourman.2014.06.015

Li, H., Wang, C. R., Meng, F., & Zhang, Z. (2019). Making restaurant reviews useful and/or enjoyable? The impacts of temporal, explanatory, and sensory cues. *International Journal of Hospitality Management*, *83*, 257–265. https://doi.org/10.1016/j.ijhm.2018.11.002

Liang, S., Schuckert, M., Law, R., & Chen, C. C. (2017). Be a "Superhost": The importance of badge systems for peer-to-peer rental accommodations. *Tourism Management*, *60*, 454–465. 10.1016/j.tourman.2017.01.007

Lim, J., & Lee, H.C.A. 2019. Comparisons of service quality perceptions between full service carriers and low cost carriers in airline travel. *Current Issues in Tourism*, 1–16. (in press). https://doi.org/10.1080/13683500.2019.1604638.

Liu, Z. (2013). *High performance latent dirichlet allocation for text mining*. (Doctoral dissertation, Brunel University School of Engineering and Design PhD Theses). 10.1094/PDIS-02-13-0213-PDN

Lu, W., & Stepchenkova, S. (2015). User-generated content as a research mode in tourism and hospitality applications: Topics, methods, and software. *Journal of Hospitality Marketing & Management*, *24*(2), 119–154. https://doi.org/10.1080/19368623.2014.907758

Mazanec, J. A. (2017). Determining long-term change in tourism research language with text-mining methods. *Tourism Analysis*, *22*(1), 75–83. 10.3727/108354217X14828625279771

Milman, A. (1991). The role of theme parks as a leisure activity for local communities. *Journal of Travel Research*, *29*(3), 11–16. https://doi.org/10.1177/004728759102900302

Milman, A. (2001). The future of the theme park and attraction industry: A management perspective. *Journal of Travel Research*, *40*(2), 139–147. 10.1177/004728750104000204

Milman, A. (2009). Evaluating the guest experience at theme parks: An empirical investigation of key attributes. *International Journal of Tourism Research*, *11*(4), 373–387. https://doi.org/10.1002/jtr.v11:4

Nuryyev, G., & Achyldurdyyeva, J. (2015). Visitor behaviour and profitability of Turkmenbashi World of Fairytales in Turkmenistan. *Journal of Hospitality and Tourism Technology*, *6*(1), 73–88. https://doi.org/10.1108/JHTT-01-2015-0006

Papadimitriou, C. H., Raghavan, P., Tamaki, H., & Vempala, S. (2000). Latent semantic indexing: A probabilistic analysis. *Journal of Computer and System Sciences*, *61*(2), 217–235. https://doi.org/10.1006/jcss.2000.1711

Park, D. H., Lee, J., & Han, I. (2007). The effect of on-line consumer reviews on consumer purchasing intention: The moderating role of involvement. *International Journal of Electronic Commerce*, *11*(4), 125–148. https://doi.org/10.2753/JEC1086-4415110405

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2007). *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge university press.

Reisenbichler, M., & Reutterer, T. (2019). Topic modeling in marketing: Recent advances and research opportunities. *Journal of Business Economics*, *89*(3), 327–356. https://doi.org/10.1007/s11573-018-0915-7

Sanchez-Franco, M. J., Cepeda-Carrion, G., & Roldán, J. L. (2019). Understanding relationship quality in hospitality services: A study based on text analytics and partial least squares. *Internet Research*, *29*(3), 478–503. https://doi.org/10.1108/IntR-12-2017-0531

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

Sparks, B. A., So, K. K. F., & Bradley, G. L. (2016). Responding to negative online reviews: The effects of hotel responses on customer inferences of trust and concern. *Tourism Management*, *53*, 74–85. https://doi.org/10.1016/j.tourman.2015.09.011

Statista. (2019). *Attendance at the Hong Kong Disneyland theme park from 2009 to 2018*. Received 06 June, 2019 from. https://www.statista.com/statistics/236183/attendance-at-the-hong-kong-disneyland-theme-park/

Themed Entertainment Association. (2018). *2018 TEA/AECOM theme index and museum index*. Received 06 June, 2019 from. http://www.teaconnect.org/Resources/Theme-Index/index.cfm

Trahan, K. (2005). *Disneyland detective: An independent guide to discovering Disney's legend, lore, and magic*. PermaGrin.

Tsai, C. Y., & Chung, S. H. (2012). A personalized route recommendation service for theme parks using RFID information and tourist behavior. *Decision Support Systems*, *52*(2), 514–527. https://doi.org/10.1016/j.dss.2011.10.013

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

Vu, H. Q., Li, G., Law, R., & Zhang, Y. (2019). Exploring tourist dining preferences based on restaurant reviews. *Journal of Travel Research*, *58*(1), 149–167. https://doi.org/10.1177/0047287517744672

The Walt Disney Company. (2019a). *About the Walt Disney company*. Received 06 June, 2019 from. https://www.thewaltdisneycompany.com/about/

The Walt Disney Company. (2019b). *Fiscal year 2018 annual financial report*. Received 01 July, 2019, from. https://www.thewaltdisneycompany.com/investor-relations/#reports

Wattanacharoensil, W., Schuckert, M., Graham, A., & Dean, A. (2017). An analysis of the airport experience from an air traveler perspective. *Journal of Hospitality and Tourism Management*, *32*, 124–135. https://doi.org/10.1016/j.jhtm.2017.06.003

Wei, W., Qi, R., & Zhang, L. (2019). Effects of virtual reality on theme park visitors' experience and behaviors: A presence perspective. *Tourism Management*, *71*, 282–293. https://doi.org/10.1016/j.tourman.2018.10.024

Weinberger, M. (2016). *How Disneyland is taking cues from Uber and Apple to make sure everybody enjoys their vacation more. Business Insider Australia*. Received 01 July, 2019, from. https://www.businessinsider.com/disneyland-disney-app-2016-10

Weinstein, R. M. (1992). Disneyland and Coney Island: Reflections on the evolution of the modern amusement park. *Journal of Popular Culture*, *26*(1), 131. https://doi.org/10.1111/j.0022-3840.1992.00131.x

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

Wirtz, J., Brah, S., Goetzinger, L., Park, J. K., & Widdows, R. (2006). E-customers' third party complaining and complimenting behavior. *International Journal of Service Industry Management*, *17*(2), 193–206. https://doi.org/10.1108/09564230610656999

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, *58*, 51–65. https://doi.org/10.1016/j.tourman.2016.10.001

Xie, K. L., Zhang, Z., & Zhang, Z. (2014). The business value of online consumer reviews and management response to hotel performance. *International Journal of Hospitality Management*, *43*, 1–12. https://doi.org/10.1016/j.ijhm.2014.07.007

Ye, B. H., Luo, J. M., & Vu, H. Q. (2018). Spatial and temporal analysis of accommodation preference based on online reviews. *Journal of Destination Marketing & Management*, *9*, 288–299. https://doi.org/10.1016/j.jdmm.2018.03.001

Zhang, Y., Li, X. R., & Su, Q. (2017). Does spatial layout matter to theme park tourism carrying capacity? *Tourism Management*, *61*, 82–95. https://doi.org/10.1016/j.tourman.2017.01.020