Exploring Tourist Dining Preferences Based on Restaurant Reviews

EXPLORING TOURIST DINING PREFERENCES BASED ON RESTAURANT REVIEWS

Huy Quan Vu[1] , Gang Li[2] , Rob Law[3], and Yanchun Zhang[1]

1Center for Applied Informatics, Victoria University, Australia
2School of Information Technology, Deakin University, Australia
3School of Hotel & Tourism Management, The Hong Kong Polytechnic University, TST East, Kowloon, Hong Kong

ABSTRACT

Dining is an essential tourism component that attracts significant expenditure from tourists. Tourism practitioners need insights into the dining behaviors of tourists to support their strategic planning and decision making. Traditional surveys and questionnaires are time consuming and inefficient in capturing the complex dining behaviors of tourists at a large scale. Thus far, the understanding about the dining preferences and opinions of different tourist groups is limited. This paper aims to fill the void by presenting a method that utilizes online restaurant reviews and text processing techniques in analyzing the dining behaviors of tourists. The effectiveness of the proposed method is demonstrated in a case study on international tourists visiting Australia using a large-scale data set of more than 40,000 restaurant reviews made by tourists on 2,265 restaurants. The proposed method can help researchers gain comprehensive insights into the dining preferences of tourists.

Keywords: dining preference, restaurant review, text processing, cuisine, food

1. INTRODUCTION

Dining is one of the top five tourist activities on leisure trips (Pizam et al. 2004), and it plays a central role in travel experience, as all tourists need to eat when they travel. Food products also attract significant expenditure from tourists, accounting for a quarter to a third of their overall spending (Correia et al. 2008). Recently, interest on food experience in tourism research has been growing (Kim, Eves, and Scarles 2009), and tourism practitioners are demanding insightful understanding of the dining behaviors and preferences of tourists to improve decision making in areas such as marketing (Min and Lee 2014) and customer relationship management (Lee, Lambert, and Law 2016).

Early studies on the dining preferences of tourists focus on wine consumption. For instance, Getz and Brown (2006) examined the levels and characteristics of the demand for long-distance wine tourism among consumers living far from wine regions. Stewart, Bramble, and Ziraldo (2008) identified the key challenges in wine and culinary tourism to provide practical recommendations for the tourism development in Canada. Other studies investigate the perceptions of tourists on food in tourist destinations, such as the cases of Hispanics in Belgium (Verbeke and López 2005), Russian tourists in Finland (Mynttinen et al. 2015), or international tourists in Ghana (Bruwer, Lesschaeve, and Campbell 2012), India (Updhyay 2014), and Korea (Joo et al. 2015). Previous works study tourists' dining preferences by identifying their restaurant selection criteria, such as selection among restaurants in South Florida (Choi and Zhao 2010), floating restaurants in Egypt (Abdelhamied 2011), and quick-service restaurants in the metropolitan area of southwestern United States (Harrington, Ottenbacher, and Way 2016).

Despite such efforts, research on the food consumption behaviors and preferences of tourists remains in its infancy (Chang and Mak 2010). The nationality and culture of tourists are key factors influencing the selection and level of food consumed by tourists (Kivela and Crotts 2006; Torres 2002). Nevertheless, the scope of existing studies is limited, focusing only on few food types and specific groups of tourists; as a result, they fail to provide a comprehensive understanding of the preference differences among various nationalities and cultures. Tourism practitioners are interested not only in the food type preferences of tourists but also in the actual dishes they consumed for each type and for each meal of the day to develop focused marketing strategies. Unfortunately, prior studies do not provide such insightful and valuable knowledge. This limitation may be due to the adoption of traditional surveys and questionnaires in most existing studies and these traditional approaches are ineffective in extensively capturing the interests and dining behaviors of tourists.

Recent advances in Internet technology have allowed tourists to share their travel-related experiences on many online platforms, such as TripAdvisor, Airbnb, and Yelp. Tourism scholars have been shifting their attention to these online resources as a channel for capturing tourist preferences in an affordable, efficient, and nonintrusive manner (Li et al. 2015). Studies have utilized travel reviews in studying hotel selection and preferences (Chaves, Gomes, and Pedron 2012; Crotts, Peyton, and Davis 2009; Li et al. 2013; Li et al. 2015). Aside from hotel reviews, a sizeable proportion of reviews on the aforementioned platforms are food related, such as restaurant reviews. The reviews cover numerous restaurants in different tourist destinations worldwide. These reviews capture and provide rich information about the dining experiences of tourists, including the cuisines and dishes they prefer as well as their background. Restaurant reviews are valuable data resources for exploring the dining behaviors and preferences of tourists comprehensively and effectively. One issue with restaurant reviews is that they are in the form of natural text, which is unstructured and indirectly usable. Few works in the tourism literature report methods that can effectively process and analyze restaurant reviews for comprehensive insights into tourists' dining preferences. Researchers and tourism practitioners continue to face difficulty in answering the following fundamental questions about tourist dining: *What are the preferred cuisines of tourists from each country? What are their preferred dishes? What food do tourists prefer for different meals of the day? What are the differences in the preferences for restaurant features among tourists? What are tourist subjective opinions toward their visited restaurants?*

In this paper, we address the shortcomings in previous dining behavior analysis by employing large-scale restaurant reviews available on online platforms. Our objectives are

- to present a general framework that can process and analyze restaurant reviews for providing comprehensive insights into the dining preferences and subjective opinions of tourists and
- to demonstrate the effectiveness of the proposed method through a case study of a tourism destination, such as Australia.

In this paper, *dining preference* is used to refer to what tourists like in relation to dining activities, such as cuisines, dishes, meals, and restaurant features. *Subjective opinions* refer to the sentiments (*positive*, *negative*) that tourists expressed in the review comments about their experience at the visited restaurants. We selected Australia for our case study because Australia is a popular destination for food tourism (Kivela and Crotts 2006). Australian cuisine is a global cuisine, a by-product of the arrival of migrants from various countries (Tourism Australia 2010). Nearly 30% of the Australian populations were born overseas (Australia Bureau of Statistics 2016). When migrants left their homelands, they brought with them their foods, flavors, and cooking styles. As a result, tourists visiting Australia are provided great gastronomic opportunity (Chang and Mak 2010). Although Australia is a large country, majority of its population are located in several capital cities. Among these cities, Melbourne and Sydney are the two most populated cities whose local residents are mostly migrants from different countries (Australia Bureau of Statistics 2015). We selected Melbourne and Sydney as the hubs of our data collection. The analysis of the case study results, which are based on more than 40,000 reviews posted by international tourists for 2,265 restaurants in the two cities, can provide insights into the general behaviors and preferences of tourists. The proposed method can help researchers and tourism practitioners gain comprehensive insights into the dining preferences of tourists by utilizing online restaurant reviews as alternative data sources to traditional surveys and questionnaires approaches.

The rest of the paper is organized as follows. The second section provides a review of the relevant literature on the dining preferences of tourists and an overview of dining behavior studies in the context of Australian tourism. Limitations of the existing studies are also highlighted. The third section introduces the method used for extracting and processing restaurant reviews for dining preference analysis. The fourth section presents the results of a case study on international tourists visiting Australia and an analysis thereof, followed by a discussion of the practical implications of the research outcomes. The final section concludes the paper and presents future research directions.

2. LITERATURE REVIEW

2.1 Dining Preference Studies

The majority of tourism literature on dining preference focuses on investigating the cultural differences in relation to food preferences. An early attempt by Pizam and Sussmann (1995) focuses on American, Japanese, French, and Italian tourists to determine if nationality affects tourist dining behavior. Cohen and Avieli (2004) studied the willingness of tourists to try foods from other cultures and found that Asian tourists were less likely to try than Western tourists. Verbeke and López (2005) studied the attitudes and behaviors of Belgians toward Latin-American ethnic foods and those of Hispanics residing in Belgium toward Belgian food. Tse and Crotts (2005) confirmed that national culture is one of the four factors associated with the culinary choices of tourists in a case study in Hong Kong. Recent studies have extended their scope to examine the dining preferences of tourists from many countries and different continents (Eric, Ramos, and Amuquandoh 2014; Updhyay 2014). These studies have mainly employed survey and questionnaire approaches. Their scales are thus limited to a small number of food items and tourist groups. Food availability in the destination also significantly influences the food choices and preferences of tourists (Eric et al. 2014). For instance, tourists from Asia may prefer Asian cuisine. However, when they travel to a European destination with a small number of Asian restaurants, they have limited options and thus may decide to eat what the local restaurants offer. The analysis of the dining behaviors of tourists in destinations with limited food choices may not reflect the true preferences of tourists.

Restaurant features apart from food are also studied to determine their influences on customer satisfaction, which can help restaurant managers to develop specific strategies that cater to the needs and expectations of different customer groups according to restaurant type. For instance, Choi and Zhao (2010) examined various influencing factors on restaurant selection, such as *service*, *cleanliness*, *ambiance*, *health issues*, and *price*. Abdelhamied (2011) studied the relations between parking spaces, healthy dishes, local dishes, and restroom cleanliness to customer revisit intentions. Liu et al. (2014) developed an effective restaurant rating scale based on four restaurant features, namely, *service*, *ambience*, *meals*, and *value for price*. Kim Bergman, and Raab (2016) examined the factors that affect mature tourists' choices when dining out in different restaurant sectors or types, particularly fine dining, buffet, and family/casual dining restaurants. Rhee, Yang, and Kim (2016) found that *food*, *value*, *atmosphere*, and *service* were considered substantially important criteria in selecting restaurants. However, these studies are also carried out for a small number of tourist groups probably due to the limitation of the existing data collection and analysis approach.

2.2 Dining in Australian Tourism

The world-class nature and lifestyle of Australia are key attributes that motivate people to visit the country. In recent years, an increasing number of people have realized that food and wine also play an important role in their tourism experience while they are in Australia (Tourism Australia 2010). The restaurant, café, and catering sector is the largest contributor to the tourism industry in Australia, which generated AUD$24.3 billion (equivalent to approximately USD$19 billion) in turnover as reported by Restaurant and Catering Australia (2016). Employment growth in this sector is projected at 16.9% and expected to reach 647,900 people by 2019 (Restaurant and Catering Australia 2016). Cuisine-related tourism has become an essential component of Australia's marketing strategies for the international market (Hall and Mitchell 2002). "Great food, wine, and local cuisine" is currently a major factor in the holiday decision making of tourists visiting Australia, ranking third (at 38%) ahead of "beauty and natural environments" (at 37%) (Tourism Australia 2016). The industry enjoys strong support to maintain funding for Tourism Australia and initiatives that promote Australian food and wine experiences to international tourists (Restaurant and Catering Australia 2016).

Despite the potential of Australian dining tourism, studies on tourist dining behavior have been limited in recent years. Laesser and Crouch (2006) performed market segmentation to profile tourists according to expenditure patterns and found that tourists are likely to spend more in Australia if their motivation is to enjoy food and beverage than other reasons. Chang and Mak (2010) proposed a typology that describes and categorizes the dining attitudes, motivations, and behaviors of Chinese tourists. Subsequently, they identified attributes that influence the assessment of travel dining experience in Australia among Chinese, Taiwanese, and Hong Kong tourist groups (Chang and Mak 2011). These attributes include the food culture of the tourists, the contextual factor of the dining experience, the variety and diversity of food, the perception of the destination and service encounter, and the performance of the tour guide. The role of local food and beverage has been investigated to support the regional tourism development in Australia (Alonso and Northcote 2010; Anderson and Law 2012). Getz and Robinson (2014) examined the tendency of Australian food lovers, instead of international tourists, to travel both domestically and internationally for food-related experiences. Robinson and Getz (2014) also provided the demographic and socioeconomic profiles of the sample and their behavioral and travel preferences. Most studies conducted in Australia are at a small scale, using survey and questionnaire approaches for data collection. An overall picture about the dining preferences of international tourists in Australia has yet to be obtained.

2.3 Restaurant Review Analysis

A major factor limiting research on dining preferences is the reliance on traditional data collection approaches, such as survey and questionnaire. These approaches are costly, time consuming, and limited in terms of the number of responses (Li et al. 2013, 2015). Studies based on these approaches fail to provide comprehensive insights into the dining behaviors and preferences of tourists. The availability of online travel reviews, particularly restaurant reviews, presents opportunities for capturing the dining behaviors of tourists. To the best of our knowledge, only a few studies in the tourism literature attempt to use restaurant reviews at large scales. For instance, Zhang Zhang, and Law (2014) investigated the relationship between the attribute performance of a restaurant and the positive and negative word of mouth of customers. Gan et al. (2016) used review data on Yelp to examine the structure of reviews and the influence of review attributes and sentiments on restaurant ratings. Zhang et al. (2017) used restaurant review in a case study to establish a decision support model to assist tourists in selecting restaurants. These studies only make use of user rating of restaurant attributes and other social information. The actual review comments with rich information about tourists' behavior and preference are not utilized. No study attempts to establish a dining preference profile of international tourists that can accommodate the aforementioned questions. This research gap is probably due to the challenge in processing reviews, which are unstructured and unsuitable for traditional statistical methods. This paper introduces a general framework for large-scale restaurant review analysis, specifically for studying tourist dining preferences and sentiments.

3. METHODOLOGY

This section presents our method of exploring tourists' dining preference based on restaurant reviews. We adopt core steps in knowledge discovery in database (KDD) (Tan, Steinbach, and Kumar 2006) as the foundation of our approach. KDD is the overall process of converting raw data into useful information, which usually comprises *selection*, *preprocessing*, *transformation*, *data mining*, and *interpretation*. In our case, data are collected from online review platforms rather than from traditional databases, and our first step is *data crawling*. For simplicity, we merge *preprocessing* and *transformation* in a single step, that is, *data preprocessing*. In this step, unstructured review content and metadata are transformed into suitable formats for analysis. The final step is *exploratory data analysis*, which includes both *data mining* and *interpretation*. A series of analyses are carried out to provide comprehensive insights into tourists' dining preferences. Figure 1 presents our restaurant review analysis framework; the details are presented in the subsequent sections.

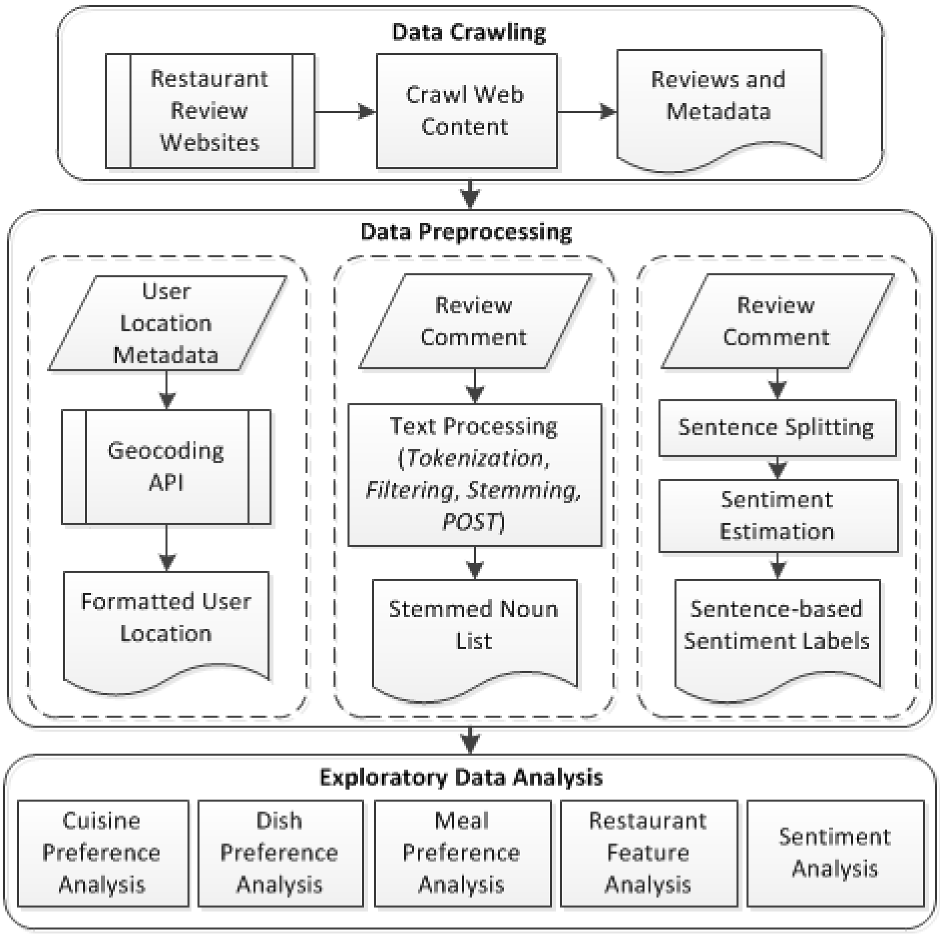


Figure 1. Restaurant Review Analysis Framework

3.1 Data Crawling

The first step in our framework is to crawl the data from restaurant review websites. Web crawler software is developed to navigate through the review websites and extract their content automatically. A typical restaurant review website contains *restaurant name*, *restaurant cuisine*, *review text*, *user ID*, *user location*, and *date visited.*

In this paper, we use TripAdvisor (www.tripadvisor.com) as a demonstration. TripAdvisor is a popular travel review site, which features a comprehensive database of restaurants with millions of reviews by tourists from all over the world. TripAdvisor is also used as the data resource in existing hospitality and tourism studies in different areas (Chaves et al. 2012; Crotts et al. 2009; Li et al. 2013, 2015).

Among the available data on TripAdvisor, *restaurant cuisine* is provided by the restaurant manager, which indicates the types of food served in the restaurant. One restaurant may serve food of more than one cuisine, for instance, Chinese and Japanese food. *User location* helps in identifying tourists' location of origin and reflects his/her cultural background, which is one of the key factors influencing dining preference (Kivela and Crotts 2006). *Review text* is the actual comment provided by reviewers after visiting restaurants. It contains rich information about a tourist's behavior, such as his/her consumed dishes or the relevant features of a restaurant. Data from *user location* and *review text* play an essential role in the analysis, but they are usually available in unstructured format and cannot be directly analyzed. We address this challenge with data preprocessing.

3.2 Data Preprocessing

3.2.1 User Location

The issue on *user location* is that users usually enter the information according to their own preferences. For example, some users may enter the country name in different forms, such as *USA*, *US*, *United States*, or *America*; other users may provide only the names of state (e.g., *Washington* or *WA*) and city (e.g., *Seattle*, *Olympia*, or *Tacoma*). Manually converting the location information of every available review into a unique form for each country is time consuming. We propose to adopt geography application programming interfaces (API) to automatically identify and return the full geographical information of a location. For instance, given a location name in the textual form such as *Washington*, a geography API can identify the full information with country name as the *United States*.

Various geography APIs are available for both public use and commercial purpose. In this paper, we utilize Google Maps Geocoding API, a common and free map tool with sufficient quota. The full documentation of Geocoding API is available at http://developers.google.com/maps. The automatic mapping from user-provided locations to unified country information allows for the convenient identification of each tourist's country of origin.

3.2.2 Review Text

We adopt a number of text mining techniques to process textual review data. Suppose a data set *R* with *m* review comments, $R = \{r_1, r_2, \ldots, r_m\}$. Text processing is conducted as follows. Each review, r_i, is first loaded into a text tokenization algorithm, in which the stream of text is broken into words, phrases, symbols, or other meaningful elements called "tokens." A filter is applied to the tokens to normalize all letters to lower case and remove symbols and numbers. The remaining tokens are inputted into the stemming module to reduce inflected words to their stem, base, or root form. For instance, a stemming algorithm can reduce the words “pizzas” to “pizza” or “eating” and “eaten” to “eat.” The stemmed words then go through part of speech tagging (POST), where each word is tagged with its corresponding type, such as *noun*, *verb*, or *adjective*. In our framework, we adopt the English lexicon available in GATE (http://gate.ac.uk/) for POST. GATE is a widely-used package for text processing with a large vocabulary database. We assume that the English vocabulary of noun type is used to refer to entities of interests, such as dishes or restaurant features. We keep words of noun type in a stemmed noun list $N = \{n_1, n_2, \ldots, n_o\}$ for further processing and discard others. The popularity of each noun, n_j, is then measured using a support value, $supp(n_j) = |n_j|/m$, where n_j is the number of reviews containing noun n_j, and *m* is the total number of reviews in the data collection. Nouns with $supp(n_j)$ greater than a user-defined threshold ($supp(n_j) \geq \delta$) are considered frequent and retained for further analysis; otherwise, they are discarded.

The advantage of this text-processing technique is that frequent nouns mentioned by tourists can be automatically collected. The researcher can inspect this list and select the nouns describing dishes or restaurant features for further analysis. This technique facilitates the exploration of all possible dining-related aspects that tourists are interested in or concerned about.

3.2.3 Sentiment Estimation

Analyzing tourist sentiment expressed in review comments is important for restaurant managers to gain insightful understanding about the experience and subjective opinions of tourists toward the foods and restaurant services. Restaurant review platforms, such as TripAdvisor, usually provide a rating function, which reflects the overall sentiment of reviewers. However, detailed sentiment information about various aspects of restaurants is unavailable. Tourists often comment on various aspects about their dining experience in their reviews. Relying on the overall rating of a comment or predefined features is inadequate for an insightful understanding about tourist experiences. Therefore, we propose to analyze the sentiment of tourists from review comments at sentence level for detailed insights.

We first break each review comment into sentences using sentence splitters. Review comments are broken into sentences based on a list of abbreviations or hand-coded rules to identify the end of a sentence. Examples of sentence-end indicators are full stop (.), exclamation mark (!), and question mark (?). A period followed by an upper-case letter usually ends a sentence; except for special cases, such as a period as part of an abbreviated title (e.g., Mr., Mrs., and Ms.).

We subsequently estimate the sentiment reflected in the sentences. Sentiment analysis requires a model to be trained on large corpus of textual data with true sentiment labels (Pang and Lee 2008). The trained model is then applied to new textual data for estimating the sentiment contained. However, obtaining training data and training a model are time consuming. Owing to the growing demand for sentiment analysis, various open source and commercial APIs with pre-trained sentiment analysis models are available for public use, such as Stanford Sentiment Analysis (https://nlp.stanford.edu/sentiment), Rosette Text Analytics (https://www.rosette.com), and SentiStrength (http://sentistrength.wlv.ac.uk/).

We utilize the SentiStrength sentiment estimation tool in our case study, as it exhibits near-human accuracy for short text in English (Thelwall 2017), and performs better than other sentiment analysis tools on different domains (Abbasi, Hassan and Dhar 2014). A sentence has three possible sentiment labels (*positive*, *negative*, and *neutral*). Table 1 lists examples of restaurant review sentences and their sentiment labels. S_1 and S_2 express positive sentiment of reviewers toward restaurant features, such as atmosphere, service, and food items. S_3 and S_4 express negative sentiment toward restaurant food. S_5 and S_6 are labeled as neutral because they mainly include facts rather than express any subjective opinion.

Table 1. Examples of review sentences and sentiment labels

ID	Comment	Sentiment Label
S_1	"*Capriccio has a warm and nice atmosphere, great service but especially amazing high quality Italian food*."	positive
S_2	"*We tried quite a few things and Haryali Chops, Paneer, Chicken Makhni and Prawn Vendaki were outstanding preparations*."	positive
S_3	"*They have tried to make the look and feel similar to Dhaba but I was not very satisfied with food which I will say below average*."	negative
S_4	"*You are just lowering the standards by these certificates when given to restaurants with such awful food.*"	negative
S_5	"*We were in and out in an hour after 2 courses*."	neutral
S_6	"*The polenta chips came with no aioli, or chili oil, so we asked for some*."	neutral

3.3 Exploratory Data Analysis

With the preprocessed data set, we first use a descriptive statistic to describe the data according to the country of origin as indicated in *user location* and rank the tourist groups according to their proportions. Only the groups with a significant number of reviews are selected. We then conduct a series of analysis on tourists' dining preference, which includes the following:

Cuisine Preference: We explore the cuisine preferences of tourists from different countries by a joint proportion analysis on *user location* and *restaurant cuisine*. The popularity of a *restaurant cuisine* in a tourist group is computed and then compared with those of other tourist groups to determine cuisine preference.

Dish Preference: We explore the dishes most frequently consumed by tourists for each examined cuisine on the basis of processed *review text* and *user location*.

We compute the support values of the nouns in the stemmed noun list and then select the nouns describing the dishes with high support values. This process can be done simultaneously for each cuisine to identify the top representative dishes.

Meal Preference: We examine the dining behaviors of different tourist groups, based on the meals of the day, including *breakfast*, *lunch*, and *dinner*. The analysis aims to answer the following questions: Which tourist groups usually go to restaurants for breakfast, lunch, and/or dinner? Which types of cuisine do tourist groups prefer at different meals? What dishes do tourist groups usually consume? The only issue with the temporal information is that the *date visited* is unhelpful in identifying the exact time. Thus, we naturally assume that if a review contains the keywords *breakfast*, *lunch*, or *dinner*, the tourist is likely to have visited the restaurant at a time corresponding to the meal. Reviews containing such keywords can be grouped based on the different meals for analysis.

Restaurant Feature Preference: We focus on restaurant features (other than food) that concern tourists. Restaurant features can be identified by the noun terms describing specific features (e.g., *staff*, *service*, *value*). The support of such noun terms is computed for different tourist groups and then compared to determine restaurant feature preference.

Sentiment Analysis: We examine the food items and restaurant features with respect to the sentiment labels of the sentences. A food item or restaurant feature receives a positive sentiment if its keyword is mentioned in a sentence with positive sentiment, and vice versa. The analysis can be carried out for different tourist groups and specific restaurants or cuisines.

In the abovementioned contrast analyses, statistical tests for proportions, such as a *z*-test or a Chi-square test, can be used to verify significant differences among tourist groups (Kanji 2006). In the fourth section, we demonstrate the application of the proposed method through a case study on the dining preferences of tourists in Australia.

4. CASE STUDY

4.1 Data Collection

The data were collected from TripAdvisor with the data extraction and web crawler software described in the methodology section. We focused our data collection on restaurants located in Melbourne and Sydney, the two most populated cities in Australia with a wide range of cuisine and dining options (Australia Bureau of Statistics 2015). The software navigated through the listed restaurants in Melbourne and Sydney to extract the reviews and associated information about the reviewers and restaurants. Among the available information, *restaurant cuisine* and *user location* are essential in establishing the dining preference profiles of tourists. However, not all listed restaurants have cuisine labels, and several reviewers did not provide their location information. Restaurants and reviews with missing essential information were excluded.

The data extraction software was utilized in April 2016, and navigating through the web sites and collecting approximately 170,000 review comments took two days. *User location* was then converted into a uniform format for country based on Google Maps Geocoding API. A large proportion of the collected reviews were made by local residents of Australia, with Australia indicated as the country of resident. TripAdvisor is an open platform, in which any registered user can post review comments, and local residents are more than international tourists in the collected data due to the unlimited time constraint for restaurant visits. We removed reviews posted by local residents, given that international tourists are the focus of our analysis. We were left with 40,948 reviews posted by international tourists for 2,265 restaurants (Table 2). Sydney shows slightly more restaurants, more reviews, and more reviews per restaurant than Melbourne. A number of tourists might have visited some restaurants more than once but posted only one review. One review is sufficient to indicate the tourist’s preference for the restaurant. Therefore, cuisine popularity can be measured based on the number of reviews in the subsequent analysis.

Table 2. Restaurant Review Data Collection

Destination	# of Restaurants	# of Reviews	# of Reviews/Restaurant
Melbourne	969	15,525	16.02
Sydney	1,296	25,423	19.62
Total	2,265	40,948	18.08

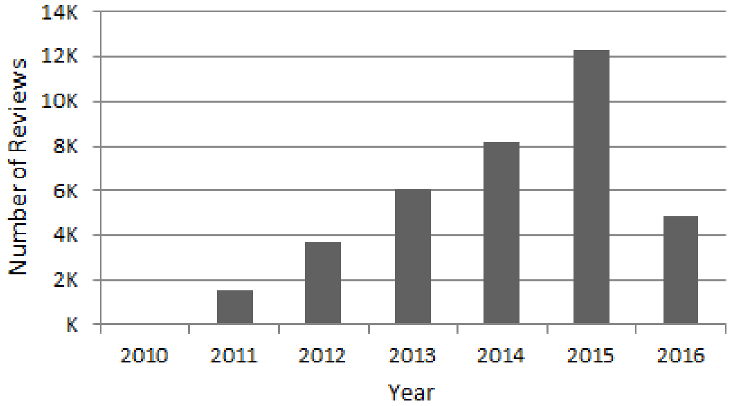


Figure 2. Distribution of Restaurant Reviews by Year

Figure 2 shows the distribution of restaurant reviews by year based on *date visited*. However, the review data for the entire 2016 were unavailable at the time of data collection. Most of the restaurant reviews were posted since 2010, and the number of reviews has been increasing in recent years. The fast growing number of restaurant reviews is probably due to the availability of review websites and the change in tourist behavior in information search and sharing about travel (Rong et al 2012). Given that the data set is relatively new, we consider the entire data set in this case study.

4.2 Data Description

In this section, we provide an overview of the collected data set to serve as reference for the subsequent analysis. The locations of reviewers in the collected data set were first examined, and out of 195 countries worldwide, 167 countries were identified as *user location*. Table 3 shows the top 14 countries whose proportions are greater than 1%, accounting for more than 86% of the total reviews in the collected data set. The countries are listed in descending order of their proportions. Each country is denoted by the three-letter country code published by the International Organization for Standardization (ISO) in ISO 3166 (www.iso.org/iso/country_codes) for convenience of presentation. The *United Kingdom* and the *United States* rank the highest, with proportions of at least 23%. Interestingly, *New Zealand*, despite being the closest to Australia, only assumes the third position in terms of dining activity.

Table 3. Data Distribution by User Location

Location	# of Reviews	Proportion (%)	Label
United Kingdom	10,698	26.13	GBR
United States	9,420	23.00	USA
New Zealand	5,275	12.88	NZL
Singapore	2,211	5.40	SGP
Canada	1,680	4.10	CAN
Malaysia	955	2.33	MYS
India	818	2.00	IND
Italy	761	1.86	ITA
Hong Kong	735	1.79	HKG
Indonesia	646	1.58	IDN
France	617	1.51	FRA
Ireland	567	1.38	IRL
Switzerland	528	1.29	CHE
Germany	493	1.20	DEU
Others (153 countries)	5,545	13.54	

Subsequently, we examined the data distribution by restaurant cuisines, and cuisine labels were grouped into different categories, such as *country* and *restaurant type* (Table 4). Among 13 cuisines by country, *Italian* is the most popular, with a proportion of 19.38%, which is nearly twice the proportion of the second most popular cuisine, i.e., *Chinese*. *German*, *British*, and *Irish* are the least popular cuisines in the collected data set, each of which shows proportions less than 1%. Given that restaurants can serve multiple cuisines, they can possibly have multiple labels, such as *Chinese* and *Japanese*. If a review belongs to a *Chinese* and *Japanese* restaurant, then we consider such a review for both *Chinese* and *Japanese* cuisines because the tourists may order dishes from both cuisines. *Pub* and *café* are two special restaurant types. Different from an ordinary restaurant, a pub is an establishment licensed to serve alcoholic drinks. Aside from food and drink, it provides services such as entertainment venue and basic accommodation. A café is a type of restaurant that usually serves coffee and snacks.

We notice that some restaurants label their cuisine by food types (e.g., *seafood*, *steakhouse*, *barbecue*, and *vegetarian*) or actual dishes (e.g., *pizza* and *sushi*). However, these food types are usually labeled together with cuisine by country. For instance, *pizza* and *Italian* are usually labeled together for the same restaurant, given that *pizza* is a part of the *Italian* cuisine. In this case, conducting a separate cuisine analysis for countries and food types is unnecessary. Similarly, restaurants label their cuisine based on regions, such as *Asian*, *Mediterranean*, or *European*, which are redundant with country labels. Thus, the analysis focuses on cuisines based on country labels to obtain in-depth insights.

For the overview of tourist and cuisine backgrounds, a proportion of 1% or higher (approximately 410 reviews) is considered sufficient, given that the collected data comprise a large-scale data set. We selected 14 countries listed in Table 3 as the target tourist groups. The top 11 cuisines by country in Table 4, with proportions greater than 1%, are considered in a subsequent analysis.

Table 4. Data Distribution by Restaurant Cuisine

Cuisine	**# of Restaurants**	**# of Reviews**	**Proportion (%)**
By Country			
Italian	451	7,936	19.38
Chinese	272	4,506	11.00
Japanese	233	3,134	7.65
Thai	181	2,929	7.15
Spanish	115	2,478	6.05
American	108	2,437	5.95
French	95	1,685	4.11
Indian	93	1,388	3.39
Greek	24	1,189	2.90
Vietnamese	72	725	1.77
Mexican	44	600	1.47
German	63	354	0.86
British	19	341	0.83
Irish	9	266	0.65
Restaurant Type			
Pub	210	3,474	8.48
Café	88	1,669	4.08

4.3 Results and Analysis

4.3.1 Cuisine Preference

In this section, cuisine preference is analyzed with respect to the nationality of tourist groups. We compute the proportion of reviews posted by each tourist group according to restaurant cuisine. For instance, among the 10,698 reviews posted by tourists from the *United Kingdom*, 2,170 reviews are for *Italian* restaurants, accounting for 20.28%. The proportion reflects the likelihood of the tourist group under consideration to select a particular cuisine, implicitly reflecting the group’s preference. A high proportion represents high preference.

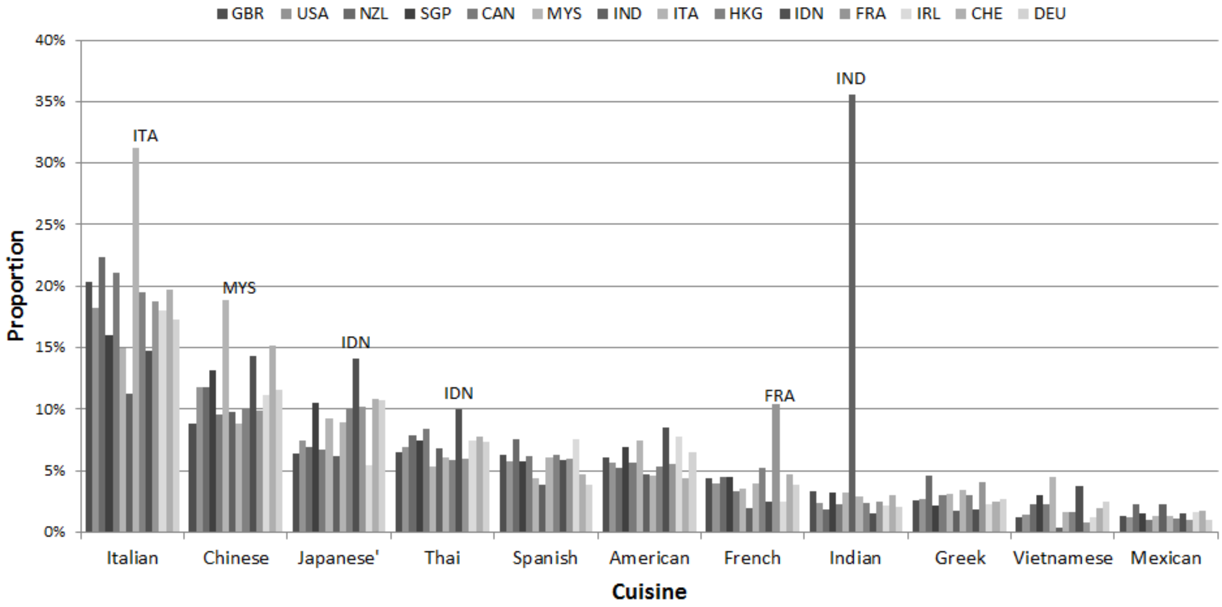


Figure 3. Cuisine Popularity by User Location

Figure 3 shows cuisine popularity by the nationality of tourists. The popularity of a cuisine is consistent with the number of available restaurants for that cuisine (Table 3). *Italian*, *Chinese*, and *Japanese* are the three most popular cuisines, as shown by the high proportion values based on nationality. However, considerable differences in the preferences exist among different tourist groups for certain cuisines. For instance, tourists from *Italy* (ITA) and *India* (IND) have the strongest preferences for their own cuisines, as evidenced by the significantly higher proportions for *Italian* and *Indian* cuisines, respectively, among other tourist groups. Specifically for Indian cuisine, *Indian* tourists represent the highest proportion, i.e., more than 35%, among other countries whose proportions are each less than 5%. Tourists from *France* (FRA) possess a relatively stronger preference for *French* cuisine than other tourist groups. In addition, certain tourist groups show strong preference for cuisines from other countries. For example, *Chinese* cuisine is most preferred by tourists from *Malaysia* (MYS), and *Japanese* and *Thai* cuisines are most preferred by tourists from *Indonesia* (IDN). Cuisine popularity, as shown in Figure 3, reflects the cuisine preferences of tourists, because tourists frequently visit restaurants with a particular cuisine. Tourists might harbor negative feeling after visiting restaurants due to the low quality of food or service. Tourists might not like a particular restaurant after visiting, but such a dislike does not necessary extend to the cuisine. We examine tourists' subjective opinions through sentiment analysis in a later section.

Figure 3 shows interesting differences in the cuisine preferences of different tourist groups. The significance of the differences in proportions, however, needs verification through statistical tests. For each cuisine (*Italian*, *Chinese*, *Japanese*, *Thai*, *French*, or *Indian*), tourists from the country with the highest proportion are treated as a group, whereas tourists from other countries are assigned to another group. The proportions of the reviews are computed for both groups, and then *z-test* with a significance level of $p \leq 0.05$ is applied. The results in Table 5 confirm the statistical significance of the differences in the proportions. For example, for the *Italian* cuisine, tourists from *Italy* are treated as a group with a proportion of 31.27%; tourists from other countries are treated as another group with a proportion of 19.21%. The difference between the proportions of the two groups is 12.07%, which is significant at a *p*-value of less than 0.05 and *z*-score of 8.319. Among the examined cuisines, tourists from *India* have the strongest preference for *Indian* cuisine, exactly 13.42 times higher than the cumulative preference of tourists from other countries. For demonstration purpose, the analysis on Figure 3 is performed only for the cuisines and nationalities with outstanding difference between the largest proportion and the rest. Detailed analyses and statistical tests can be similarly conducted between any two nationalities depending on practical requirements.

Table 5. *z*-Test Results for Cuisine Popularity among Selected Groups

Cuisine	**Country**	**Proportion (%)**	**Difference (%)**	**Ratio**	***z*-Score**	***p*-Value***
Italian	Italy (ITA)	31.27	12.07	1.63	8.319	*0.000*
	Others	19.21				
Chinese	Malaysia (MYS)	18.85	8.12	1.76	7.934	*0.000*
	Others	10.72				
Japanese	Indonesia (IDN)	14.11	6.72	1.91	6.420	*0.000*
	Others	7.39				
Thai	Indonesia (IDN)	9.92	2.95	1.42	2.906	*0.004*
	Others	6.97				
French	France (FRA)	10.37	6.30	2.55	7.747	*0.000*
	Others	4.07				
Indian	India (IND)	35.57	32.92	13.42	51.902	*0.000*
	Others	2.65				

*Significant at $p \leq 0.05$

In addition to cuisine preference, we examined the preferences of tourists between two typical restaurant types, i.e., *pub* and *café*. We highlighted the influence of cultural factors, besides nationality, on the preference for these restaurant types. The top 14 countries, with at least a data percentage of 1% as shown in Table 3, are analyzed. The countries are mainly in North America, Europe, and Asia. Tourists from North America and Europe tend to have similar backgrounds that differ from those of tourists from Asia (Vu et al. 2015). Accordingly, we treat the tourists from Europe and North America as the *Western* group and the tourists from Asia as the *Asian* group. We noted that New Zealand is in Oceania, but the dominant ethnic group of New Zealand population is European (Statistics New Zealand 2014). Therefore, tourists from New Zealand were treated as in the *Western* group. *Chinese* users, one of the major inbound tourist groups, are limited in the collected data, which is probably because TripAdvisor is not widely used by Chinese, as other social network platforms (e.g., Dianping.com) are more popular for Chinese users.

Figure 4 illustrates the proportions of reviews for the tourist groups, and the dark and light bars represent the *Western* and *Asian* groups, respectively. Most countries in the *Western* group, except for *Switzerland* (CHE), have higher proportions for *pub* than *Asian* countries (Figure 4a). This result may probably because pubs originate from the *United Kingdom*, and many countries in Europe and North America are under the influence of British culture (Johnson, 2016). *Western* tourists are more likely to visit pubs than *Asian* tourists do. The *z*-test with a significance level of $p \leq 0.05$ verify the statistically significant difference between these two groups in terms of the preference for *pub*, with *z*-score = 10.9553 and $p < 0.001$. By contrast, no noticeable difference exists in the preferences for *café* between the *Western* and *Asian* groups (Figure 3b), although the *Indian* group is least likely to visit cafés. The number of reviews posted by tourists from the *United Kingdom* and *United States* is larger than those posted by tourists from other countries in the *Western* group. The preference for visiting pubs of *Western* group in the above comparison can be biased toward these two countries. Detailed comparison can be similarly conducted between any two nationalities depending on practical requirements.

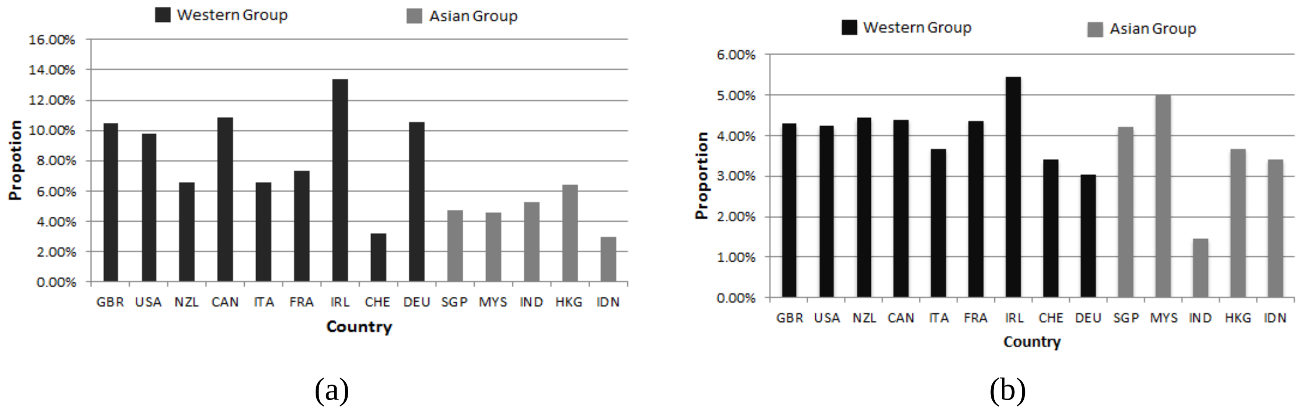


Figure 4. Popularity Levels of (a) *Pub* and (b) *Café* between *Western* and *Asian* Groups

4.3.2 Dish Preference

This section identifies the popular dishes, including food and drink, consumed by international tourists when visiting *Australia*. The review data for each cuisine and restaurant type are inputted into the text-processing algorithm presented in the methodology section. The support values are computed for the nouns appearing in the data sets. A support threshold of $\delta = 0.05$ is used to filter out infrequent nouns, and the frequent ones are retained for further analysis. Support values are commonly used for evaluating items of interest in a data set (Law et al. 2011). Furthermore, the number of returned items is small, which is manageable for manual inspection (Li et al. 2015). Thirty-three nouns are selected and referred to as *food items* for convenient presentation. Some food items are common across different restaurant types, whereas others are unique to specific restaurants. In Figure 5, these food items are listed in descending order of their counts based on the number of restaurant types offering them. Owing to space limitation, we only show the first five letters of the cuisine names in the column headings. Cells containing support values of 0.1 or higher are highlighted in gray background and bold text. We represent the cuisines according to their cultural origins, that is, *Western* and *Asian* cuisines.

	Restaurant Type													
	Western						Asian					Others		
Food Item	Itali	Spani	Ameri	Frenc	Greek	Mexic	Chine	Japan	Thai	India	Viet	Pub	Café	Count
Wine	**0.152**	**0.198**	0.055	**0.210**	**0.127**		0.054	0.084	0.065	0.054	0.050	0.077		11
Chicken			0.056		0.051	0.062	0.075		0.068	**0.120**	0.090			7
Pork			0.069			0.052	0.076		0.081		**0.108**			5
Salad	0.061				0.080				0.053		0.068			4
Dessert	0.055			0.069	0.058								0.050	4
Rice							0.074		0.077	0.081	**0.119**			4
Coffee	0.057			0.064									**0.267**	3
Beef							0.054		0.058		**0.130**			3
Burger			**0.195**									0.064	0.051	3
Steak			0.060	0.093								0.091		3
Chip			0.051			0.055						0.070		3
Fish								0.067				0.050		2
Curry									**0.105**	**0.122**				2
Beer			0.057									**0.285**		2
Lamb					**0.174**					0.051				2
Sauce						0.055					0.052			2
Pizza	**0.199**													1
Pasta	**0.136**													1
Duck							0.071							1
Sushi								**0.144**						1
Seafood								0.053						1
Sake								0.050						1
Rib			0.089											1
Pie			0.053											1
Bread					0.058									1
Roll											**0.156**			1
Soup											0.084			1
Taco						**0.172**								1
Corn						0.090								1
Cake													0.096	1
Egg													0.080	1
Chocolate													0.073	1
Toast													0.052	1
Count	6	1	9	4	6	6	6	5	7	5	9	6	7	

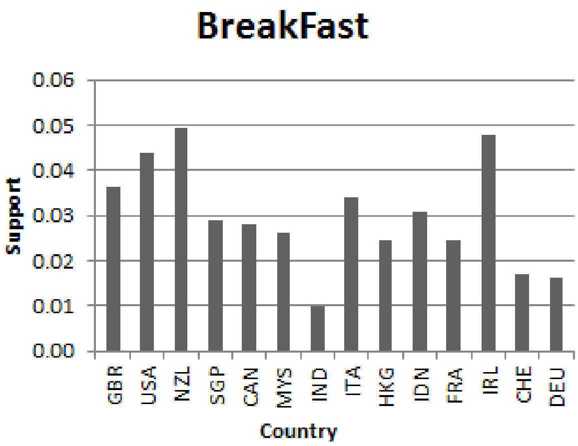
Figure 5. Popular Food Items Consumed by Tourists

Wine, *chicken*, and *pork* are widely consumed by tourists on the basis of the restaurant types offering them. *Wine* is consumed in 11 out of 13 restaurant types, especially for restaurant serving *Western* cuisines, such as *Italian*, *Spanish*, *French*, and *Greek*, for which the support values are higher than 0.1. This result is consistent with the fact that Australia is famous for its wine production, and wine tourism shows growing popularity in Australian tourism (Tourism Australia 2010). Tourists are interested in tasting wine as a part of their dining experience. *Desert*, *coffee*, *burger*, *steak*, and *chip* are popular in *Western* cuisines, whereas, *rice*, *beef*, *fish*, and *curry* are popular in *Asian* cuisines. Although *beef* and *steak* refer to the same kind of meat, their support values are different between *Western* and *Asian* cuisines, probably because the difference in cooking styles and cultures create distinct dishes.

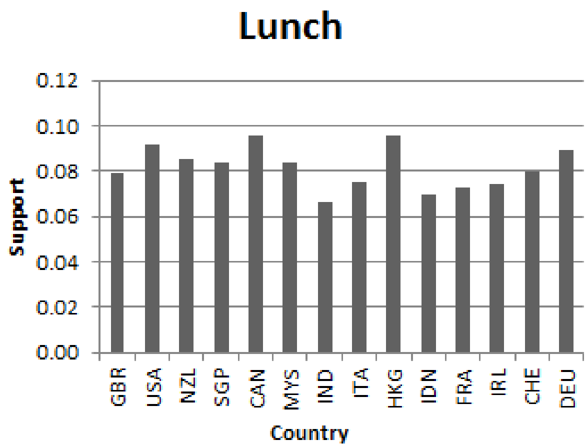
Aside from common dishes, unique popular food items are offered in different cuisines, such as *pizza* and *pasta* for *Italian* cuisine, *sushi* for *Japanese* cuisine, *roll* for Vietnamese cuisine, and *taco* for *Mexican* cuisine. The most popular item in *pub* is *beer*, and it shows a support value of 0.285, which is nearly four times higher than *wine*. Most cuisines have multiple popular food items consumed by tourists. For example, *Vietnamese* and *American* cuisines each have nine items. In particular, *Spanish* cuisine has only one but a highly popular item, *wine*, which has a support value of almost 0.2. *Café* is popular for light food items, such as *dessert*, *cake*, *chocolate*, and most importantly *coffee*.

4.3.3 Meal Preference

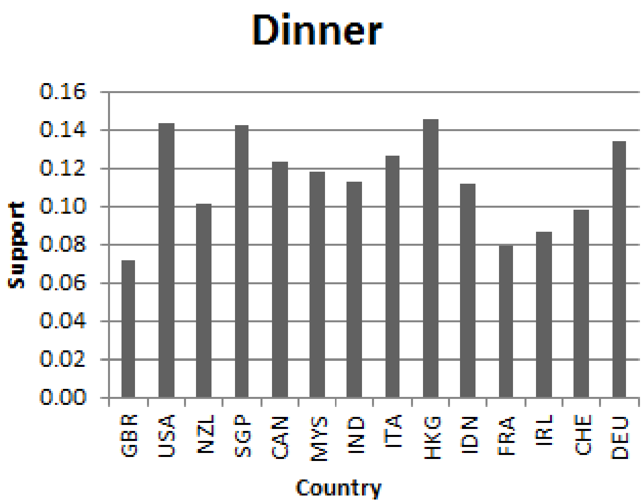
This section reveals the dining behaviors of tourists based on different meals of the day. We first examined the popularity of the reviews according to keywords *breakfast*, *lunch*, and *dinner*. The support values of each meal were calculated for each tourist group (Figure 6). Generally, the support values of *lunch* (Figure 6b) and *dinner* (Figure 6c) are higher than *breakfast* (Figure 6a) across all tourist groups. This result reflects the intuition that *lunch* and *dinner* are bigger meals; thus, they are more likely to be taken in restaurants. Tourists from the *United States* (USA), *New Zealand* (NZL), and *Ireland* (IRL) are more likely to have breakfast in restaurants than other groups, reflecting the preference differences.



(a)



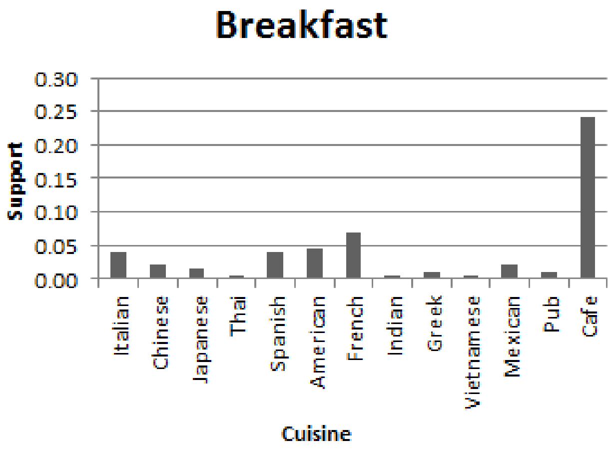
(b)



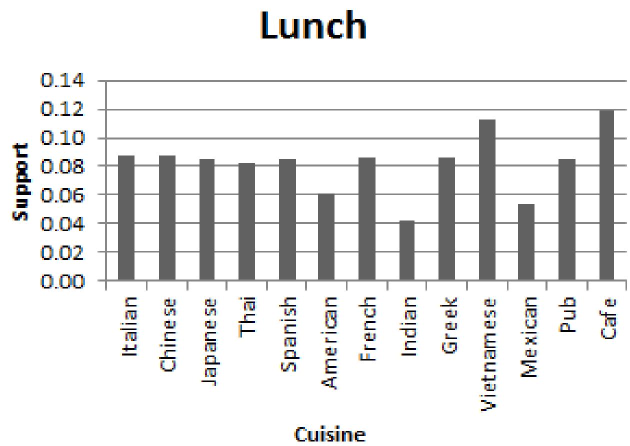
(c)

Figure 6. Meal Preference by Country

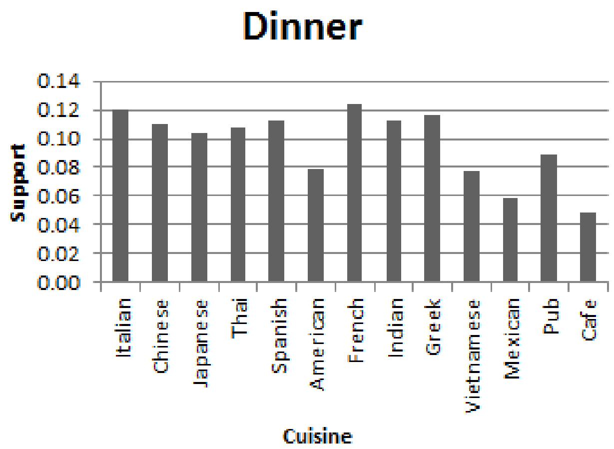
To determine where tourists go for each meal, we also computed the support values for each meal with respect to restaurant type (Figure 7). *Café* is the most popular place for breakfast, showing a significantly higher support value than all other restaurant types (Figure 7a). This result is expected, as people tend to have light food for breakfast, and a *café* is a good place to find such food. Most restaurants have similar support values for *lunch* (Figure 7b) and *dinner* (Figure 7c). Among them, *Indian* restaurants have the lowest support value for *lunch*, whereas *café* is the least popular for *dinner*.



(a)



(b)



(c)

Figure 7. Cuisine Preference for Meals

We further investigated the differences in the consumed food items among the three different meals for in-depth insights. *Café* is the most popular place for breakfast; this result may explain why *dessert*, *coffee*, *burger*, *cake*, *egg*, *chocolate*, and *toast* are usually consumed by tourists in cafés (Figure 5). Therefore, we only compared the popularity levels of food items in *lunch* and *dinner*. The food items with the most significant difference are shown in Table 6. The value in the *Difference* column is the gap between the support value of *dinner* and the support value of *lunch* for each food item. The ratio is computed by taking the higher value divided by the lower value. Tourists consumed more *wine*, *steak*, and *lamb* for *dinner* than for *lunch*, as indicated by the higher support values of the former. By contrast, tourists are more likely to have *salad*, *coffee*, *burger*, *chip*, *fish*, and *beer* for *lunch* than for *dinner*. The *z*-test with $p \leq 0.05$ was performed to verify the statistical significance of the differences.

Table 6. Popularity of Food items between *Lunch* and *Dinner*

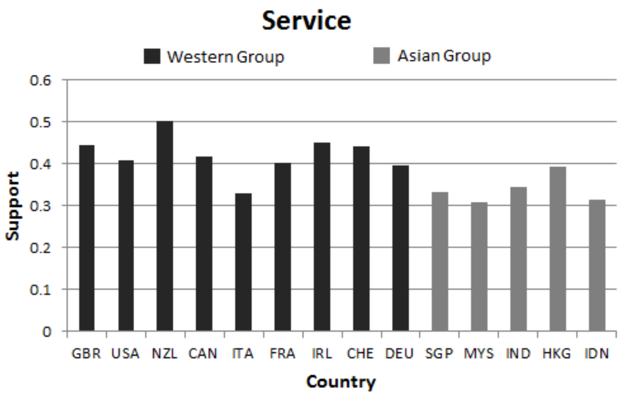
Dish	**Proportion (%)**		**Difference (%)**	**Ratio**	***z*-Score**	***p*-Value***
	Lunch	**Dinner**				
Wine	11.24	14.86	3.62	1.32	4.973	*0.000*
Steak	3.09	6.49	3.40	2.10	7.264	*0.000*
Lamb	2.28	3.64	1.36	1.60	3.692	*0.000*
Salad	7.34	5.23	−2.10	1.40	4.088	*0.000*
Coffee	4.38	1.99	−2.39	2.20	6.523	*0.000*
Burger	3.72	1.81	−1.91	2.06	5.573	*0.000*
Chip	4.06	2.11	−1.96	1.93	5.401	*0.000*
Fish	5.97	4.90	−1.07	1.22	2.226	*0.026*
Beer	5.55	3.82	−1.74	1.45	3.886	*0.000*

*Significant at $p \leq 0.05$

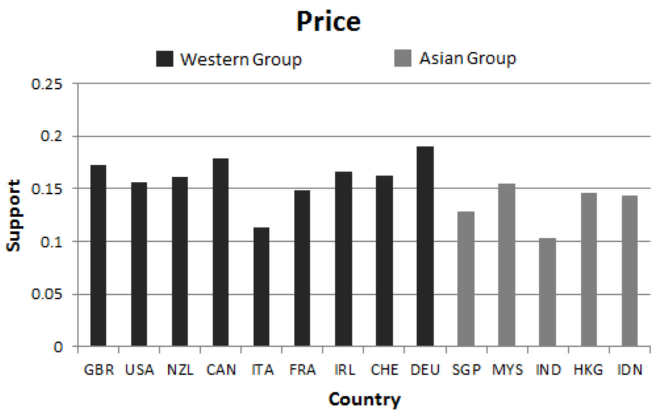
4.3.4 Restaurant Feature Preference

Restaurant features play important roles in the selection of tourists. This section examines the preferences for restaurant features across the different nationalities and cultures of tourists. By inspecting the nouns in the previous sections, *service*, *staff*, *price*, *value*, and *atmosphere* were used frequently as restaurant features. Restaurant features have approximate meanings, such as *service* and *staff* as well as *price* and *value*. Accordingly, we grouped reviews containing the aforementioned keywords into *service* and *price* features; as a result, three restaurant features were considered for further analysis. Figure 8 shows the support values of the restaurant features, and countries belonging to the *Western* and *Asian* groups are represented by dark and light bars, respectively.

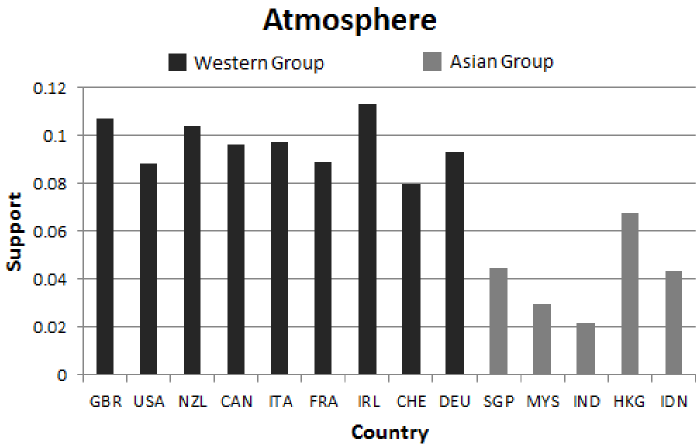
Among the three restaurant features, *service* is the most important to international tourists, with support values higher than 0.3 for all countries (Figure 8a). By contrast, support values are less than 0.2 for *price* (Figure 8b) and 0.12 for *atmosphere* (Figure 8c). Support values for countries in the *Western* group are generally higher than for those in the *Asian* group. A statistical test was conducted to verify the significance of the differences. The reviews were grouped based on the cultural groups (i.e., *Western* and *Asian*); then, the proportions containing the keywords reflecting restaurant features were computed. Table 7 shows the *z*-test results of the proportions (significant at $p \leq 0.05$). The *Western* group has significantly higher preferences for all examined restaurant features than the *Asian* group, with a sizable difference (10.17%) for the *service* feature. As the majority of reviews in the *Western* group are posted by tourists from the *United Kingdom* and *United States*, the differences in the proportional analysis can be biased toward these two countries.



(a)



(b)



(c)

Figure 8. Restaurant Feature Preference by Country

Table 7. Contrast Analysis of Restaurant Feature Preferences

Feature	**Proportion (%)**		**Difference (%)**	**Ratio**	***z*-Score**	***p*-Value***
	Western	**Asian**				
Service	43.75	33.58	-10.17	1.30	13.893	*0.000*
Price	16.36	13.37	-2.99	1.22	5.519	*0.000*
Atmosphere	9.88	4.16	-5.73	2.38	13.488	*0.000*

*Significant at $p \leq 0.05$

In addition, which restaurant types receive the most comments regarding their features is worth investigating. We counted the number of reviews containing at least one of the feature keywords and then computed their support values over the entire data collection. The tourists are concerned about the restaurant features most when visiting *Italian* and *pub* restaurants (Table 9). Restaurants serving Western cuisine receive more reviews on the considered features, whereas *Chinese*, *Thai*, and *Vietnamese* restaurants receive considerably few reviews on restaurant features.

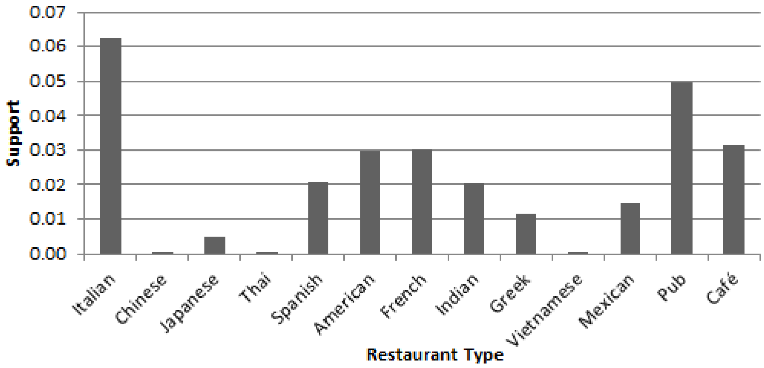


Figure 9. Feature Popularity by Restaurant Type

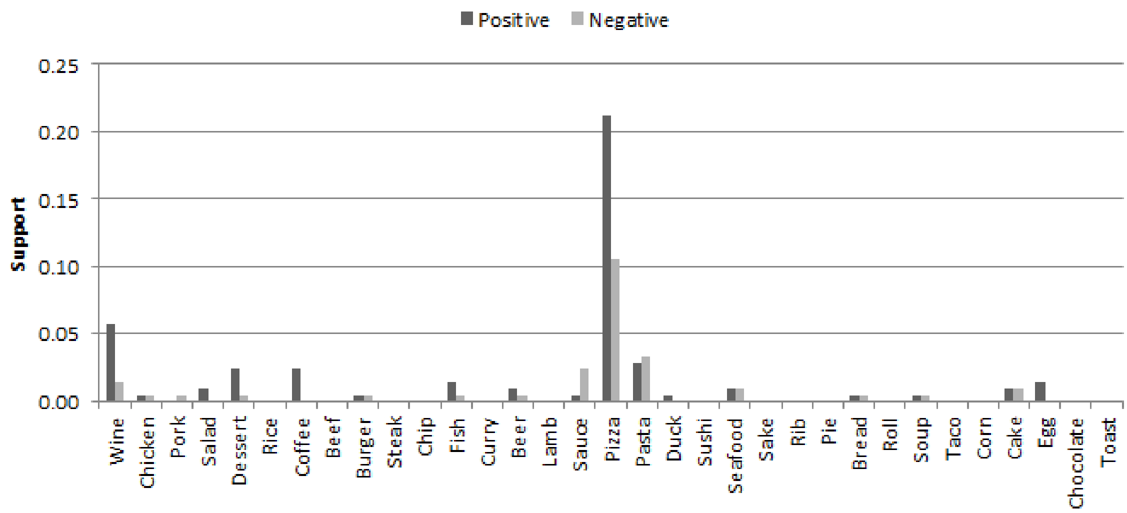
4.3.5 Sentiment Analysis

This section presents a sentiment analysis to examine tourists' subjective opinions after visiting restaurants. For demonstration, we examined review comments by *Italian* and *Indian* tourists toward food items at *Italian* and *Indian* restaurants, respectively. The review comments were broken into sentences and then sentiments were estimated using the trained model provided by the SentiStrength API. Keywords describing food items were identified from the sentences with sentiment labels before their support were computed. Only sentences with positive or negative sentiment label were accounted. A sentence with neutral label was not considered as it does not express a subjective opinion.

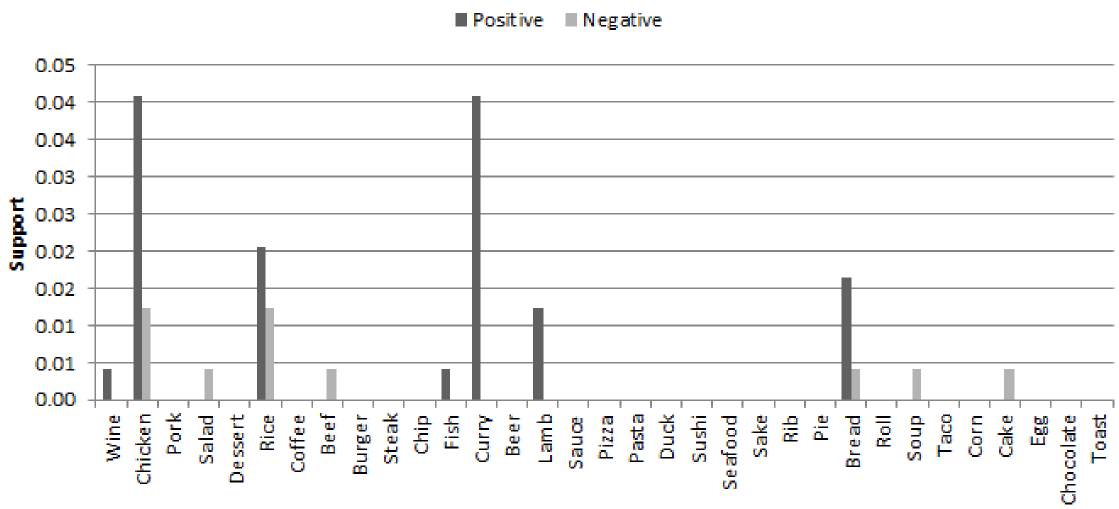
Figure 10a shows that *Italian* tourists express sentiment mostly on *pizza*; many reviews are positive, but considerable numbers of reviews are negative. Similar patterns are found for *wine*. More negative reviews than positive reviews are found for *sauce* and *pasta*.

Figure 10b indicates that Indian tourists are mostly satisfied with food items, such as *wine*, *fish*, *curry*, and *lamb* at *Indian* restaurants. Other food items, such as *chicken*, *rice*, and *bread*, also receive positive comments, but several *Indian* tourists are unsatisfied as indicated by their negative comments. Mostly negative comments are found for *salad*, *beef*, *soup*, and *cake.*

Restaurants that provide exceptional food and service may receive mostly positive comments, but others may not. Figure 10 only reflects the overall sentiments of *Italian* and *Indian* tourists toward the restaurants serving *Italian* and *Indian* cuisine in the data collection as case demonstration. Detailed insights into the sentiment of different tourist groups toward specific restaurant can be obtained by more fine-grained analyses in future studies.



(a)



(b)

Figure 10. Positive and negative sentiments toward food items of (a) *Italian* at *Italian* Restaurants and (b) *Indian* at *Indian* Restaurants

4.4 Discussion and Implications

The case study demonstrates the capability of large sale online restaurant reviews in capturing comprehensive information about tourists' dining preference. Rather than using existing approaches to data collection (e.g., survey and questionnaire), this study employs online restaurant review data with support from text processing techniques. The analysis is carried out on a large-scale data set to provide a comprehensive understanding of tourists' dining preferences in terms of cuisine, dishes, meal, and restaurant features.

The findings bear potential implications for the tourism industry in Australia. Certain tourist groups have strong preference for their own national cuisine, such as *Italian*, *Indian*, and *French* (Figure 2). Tourism practitioners should pay special attention to these groups in their strategic planning and decision making, such as highlighting the availability of diverse *Indian* food options in marketing materials to attract *Indian* tourists. Travel package developers can provide dining options to make *Japanese* and *Thai* cuisines available to *Indonesian* tourists and *Chinese* cuisine available to *Malaysian* tourists for improving the dining experience and satisfaction of these tourist groups. Given that *Western* tourists are more interested in going to pubs (Figure 3a); tourism practitioners can introduce pub-dining experience to tourists from this cultural group. Based on the comprehensive list of food items in Figure 5, tourism practitioners should provide dining recommendations to tourists on where to go for each kind of food. The findings of the current study also confirm those of existing studies that *wine* is a key component of tourist dining experience in Australia, as indicated by popular in different restaurant types. Meal preference analysis (Figure 7a) indicates that tourists are more likely to have breakfast in *cafés* than in other restaurant types. Aside from tourist preference, the opening hours of Australian restaurants can be a factor for the result. Ordinary restaurants do not open early, thus limiting the options of tourists for breakfast. *Café* managers may consider offering more light food items in their breakfast menu to provide tourist with more options and consequently enrich the dining experience. In terms of specific food items (Table 6), restaurant managers can offer more *wine* options, *steak* or *lamb* dishes for dinner, whereas more *salad*, *burger*, *chip*, *fish*, and *beer* can be reserved for lunch because tourists are likely to consume more of these at different meals. *Western* tourists are more concerned about restaurant features than *Asian* tourists, as discussed in the restaurant feature preference analysis section. Tourism practitioners may highlight restaurant features, particularly the quality of *service*, when introducing dining options to Western tourists. Managers of *Italian* restaurants and *pubs* should pay attention to their restaurant features for improving tourist satisfaction. By contrast, managers of *Chinese*, *Thai*, and *Vietnamese* restaurants should direct their limited resources to improve food quality, as tourists are less concerned about other features. In addition, the subjective opinion of tourists can be obtained by sentiment analysis in our proposed framework. Restaurant managers can focus their effort to improve the quality of *pizza* at *Italian* restaurants because *Italian* tourists show preference for their own cuisine, but many of them are not satisfied as indicated by the negative comments (Figure 10a). *Indian* tourists can be suggested to try *curry* and *lamb* at *Indian* restaurants, which are likely to satisfy them, as indicated by the positive comments (Figure 10b).

Despite the comprehensive analysis conducted in this study, it is not without limitations. Data were collected solely from TripAdvisor, which is not the most widely used travel review platform in all countries. For example, most Chinese tourists use other platforms, such as Dianping.com. Hence, tourists from popular source markets of Australian tourism, such as *China* and *Taiwan* (Tourism Research Australia 2016), were not included in the analysis. Data were collected for restaurants in Melbourne and Sydney, which are populated cities offering various restaurants and cuisines, and, thus, the results may not be generalized well with other cities in regional areas. The data collection procedure could be biased to web users, because the data were collected only from the online review platform. Demographic profiles and characteristics of web users could be considered for specific applications of the results. Although a list of food items was constructed, which could provide tourism practitioners with specific information about a consumed food or particular food item, *salad*, for example, may be prepared differently between *Italian* and *Vietnamese* restaurants. Different tourist groups may prefer different cooking styles for the same dishes, such as *beefsteak* (Cox, Cunial, and Winter 2016). The current study does not consider such differences. The sentiment analysis results may have potential bias as some restaurants may have more reviews than others. Analysis for individual restaurant can be performed in future studies for specific practical applications.

Nevertheless, the proposed approach provides more insights into the dining behavior and opinions of tourists than prior works using restaurant reviews (Gan et al. 2016; Zhang et al. 2014; Zhang et al. 2017) because textual reviews are analyzed. The results do not rely on user rating of restaurant attributes and other social information. Detailed information about the subjective opinions of tourists can be explored by sentiment analysis, which is effective in assessing tourist satisfaction and identifying shortcomings for future improvement. The presented text processing and sentiment analysis methods can be applied not only to comments posted on websites designated to travel reviews, but also to short text discussed about dining on various mobile social media platforms, such as Twitter, Facebook and Foursquare. For instance, Twitter provides Streaming API, which allows for real time extraction of tweets text via keywords specified by users (Vu et al. 2017). Tourism practitioners can develop applications to monitor tourist preferences for food and their sentiment towards food providers in real-time. Mobile recommendation applications can also be developed to provide tourist with suitable dining offer while traveling (Kotiloglu et al. 2017).

In terms of technical limitation, different websites feature distinct representations and structures. Web crawler software setting needs adjustment when extracting data from other websites. The free account of Google Maps Geocoding API has daily quota limits; higher quota can be requested with a small fee or geography API from other service providers can be utilized. The text-processing framework is mainly designed for reviews in English. Specific text-processing techniques and language lexicon should be further investigated to process reviews from other languages. Although SentiStrength supports many other languages (e.g., German, Spanish, Italian), its best performance is reached with review comments in English (Abbasi et al. 2014; Thelwall 2017). Future studies can investigate other tools to better support sentiment analysis of review comments in other languages. It is worth mentioning that review websites, such as TripAdvisor, do provide API for direct access to their data. However, the data collected via TripAdvisor API are subject to its own terms and conditions (https://developer-tripadvisor.com/content-api/terms-and-conditions/), which may have restrictions on data analysis and result publication. Researchers and business managers are suggested to get themselves familiar with the terms and conditions before carrying out their research using TripAdvisor API, and be familiar with potential restrictions for using the alternative web crawling approach presented in this study.

5. CONCLUSIONS

Dining is an essential component of tourism, and tourism practitioners need to know the dining preferences of tourists to enhance their strategic planning and decision making. The dining behaviors of tourists are complex and varied across different nationalities and cultures. Specific information on consumed food and preference differences among meals of the day is also important to tourism practitioners and restaurant managers for devising their marketing and management plans. Restaurant features, other than food, are also of interest to restaurant managers because enhancing these features can enrich the dining experience of tourists and consequently improve their satisfaction. Previous studies are unable to address the needs of tourism practitioners and restaurant managers comprehensively because of the limitations of traditional data collection and analysis approaches.

To address the existing barriers, this paper presents a method that employs online restaurant reviews to study the dining behavior of tourists. The advantage of the proposed method is the capability of directly analyzing the online textual review comments to capture and explore rich information they contain. Fundamental and important insights into tourists' dining behavior in terms of preferred cuisines, dishes, meals of the day, restaurant features, and their subjective opinions can be obtained efficiently. The proposed method can be applied to big data set for large-scale studies owing to the automation of data collection and utilization of software tools and APIs for data preprocessing, which allows for convenient subsequent statistical analysis. The effectiveness of the proposed method is demonstrated in a case study on Australian tourism by using a large-scale data set. The analyses conducted involve international tourists from 14 different countries and 13 restaurant types, which are more comprehensive than prior studies on tourists' dining behavior. The presented method is beneficial to researchers who wish to deepen insights into the dining behaviors of tourists at a low cost but high efficiency.

The proposed method is a general analysis framework applicable to data from review websites other than TripAdvisor. Future research can consider combining review data from multiple travel platforms to obtain representative results and analysis. Text processing tools and sentiment analysis API for languages other than English can be integrated into our framework for restaurant reviews written by tourists from non-English speaking countries. The analyses are carried out mainly using statistics to demonstrate the capability of restaurant reviews in capturing rich information. Future study can investigate other sophisticated text mining techniques to extract patterns according to specific needs from the processed data resulted from our framework. Apart from text, photos of foods and restaurants can provide insights into tourists' own experience and specific interests. Future research can focus on developing techniques for analyzing photos together with the review comments for in-depth understanding. In this work, the main focus was on tourists who traveled to a tourism destination. Therefore, we examined the dining behavior of people as reflected in restaurant reviews at the destination, rather than local residents staying in their own country. Future studies can apply the proposed method to examine common and contrast dining behavior of location residents in their home countries.

REFERENCES

Abbasi, A., A. Hassan, and M. Dhar 2014. “Benchmarking Twitter Sentiment Analysis Tools”. In *Proceeding of the 9th International Conference on Language Resources and Evaluation,* 823–29. Reykjavik, Iceland.

Abdelhamied, H. H. S. 2011. “Customers' Perceptions of Floating Restaurants in Egypt”. *Anatolia: An International Journal of Tourism and Hospitality Research* 22:1-15.

Alonso, A. D., and J. Northcote. 2010. “The Development of Olive Tourism in Western Australia: A Case Study of an Emerging Tourism Industry”. *International Journal of Tourism Research* 12 (6):696-708.

Anderson, A., and L. Law. 2012. “An Advanced Framework for Food Trail Performance”. *Journal of Vacation Marketing* 18 (4):275-86.

Australia Bureau of Statistics. 2015. “3218.0–Regional Population Growth, Australia, 2013-14”. Technical report, Australian Bureau of Statistics, Canberra. http://www.abs.gov.au/ausstats/abs@.nsf/mf/3218.0 (accessed May 25, 2016).

Australia Bureau of Statistics. 2016. “3412.0–Migration, Australia, 2014-15”. Australia Bureau of Statistics, Canberra. http://www.abs.gov.au/ausstats/abs@.nsf/mf/3412.0/ (accessed May 20, 2016).

Bruwer, J., I. Lesschaeve, and B. L. Campbell. 2012. “Consumption Dynamics and Demographics of Canadian Wine Consumers: Retailing Insights from the Tasting Room Channel”. *Journal of Retailing and Consumer Services* 19 (1):45-58.

Chang, R. C. Y., and A. H. N. Mak. 2010. “Food Preferences of Chinese Tourists”. *Annals of Tourism Research* 37 (4):989-1011.

Chang, R. C. Y., and A. H. N. Mak. 2011. “Attributes that Influence the Evaluation of Travel Dining Experience: When East Meets West”. *Tourism Management* 32:307-16.

Chaves, M. S., R. Gomes, and C. Pedron. 2012. “Analysing Reviews in the Web 2.0: Small and Medium Hotels in Portugal”. *Tourism Management* 33 (5):1286-87.

Choi, J., and J. Zhao. 2010. “Factors Influencing Restaurant Selection in South Florida: Is Health Issue One of the Factors Influencing Consumers' Behavior when Selecting a Restaurant?” *Journal of Foodservice Business Research* 13:237-51.

Cohen, E., and N. Avieli. 2004. “Food in Tourism: Attraction and Impediment”. *Annals of Tourism Research* 31 (4):755-78.

Correia, A., M. Moital, C. F. D. Costa, and R. Peres. 2008. “The Determinants of Gastronomic Tourists’ Satisfaction: A Second-order Factor Analysis”. *Journal of Food Service* 19 (3):164-76.

Cox, R. J., C. M. Cunial, and S. Winter. 2016. “Eliminating Cross Cultural Variations in Consumer Perceptions of Cooked Beef Steaks”. *Journal of Food Service Business Research* 8 (3):81-94.

Crotts, J. C., P. R. Peyton, and B. Davis. 2009. “Measuring Guests Satisfaction and Competitive Position in the Hospitality and Tourism Industry. *Journal of Travel Research* 48 (2):139-51.

Eric, F., A. Ramos, and F. E Amuquandoh. 2014. “Traditional Food Preferences of Tourists in Ghana”. *British Food Journal* 115 (7):987-1002.

Gan, Q., B. H. Ferns, Y. Yu, and L. Jin. 2016. “A Text Mining and Multidimensional Sentiment Analysis of Online Restaurant Reviews”. *Journal of Quality Assurance in Hospitality & Tourism* DOI:10.1080/1528008X.2016.1250243

Getz, D., and G. Brown. 2006. “Critical Success Factors for Wine Tourism Regions: A Demand Analysis. *Tourism Management* 27 (1):146-58.

Getz, D., and R. N. S. Robinson. 2014. ““Foodies” and Their Travel preferences”. *Tourism Analysis* 19:659-72.

Hall, C., and R. Mitchell. 2002. “Tourism as a Force for Gastronomic Globalisation and Localization”. In *Tourism and Gastronomy*, edited by A.-M. Hjalager, and G. Richards, 71-90. London, UK: Routledge.

Harrington, R. J., M. C. Ottenbacher, and K. A. Way. 2016. “QSR Choice: Key Restaurant Attributes and the Roles of Gender, Age and Dining Frequency”. *Journal of Quality Assurance in Hospitality and Tourism* 14 (1):81-100.

Johnson, B. 2016. “The Great British Pub”. *Historic UK: The History and Heritage Accommodation Guide*. http://www.historic-uk.com/CultureUK/The-Great-British-Pub (accessed May 20, 2016).

Joo, N., S. Park, B. Lee, and J. Yoon. 2015. “Importance-Satisfaction Analysis of Street Food Sanitation and Choice Factor in Korea and Taiwan”. *Nutrition Research* 9 (1):296-303.

Kanji, G. K. 2006. “100 Statistical Tests” (3rd ed.). London, UK: SAGE Publication.

Kim, Y. G., A. Eves, and C. Scarles. 2009. “Building a Model of Local Food Consumption on Trips and Holidays: A Grounded Theory Approach”. *International Journal of Hospitality Management* 28 (3):423-31.

Kim, Y. S., C. Bergman, and C. Raab. 2016. “Factors that Impact Mature Customer Dining Choices in Las Vegas”. *Journal of Foodservice Business Research* 13 (3):178-92.

Kivela, J., and J. Crotts. 2006. “Tourism and Gastronomy: Gastronomy’s Influence on How Tourists Experience a Destination”. *Journal of Hospitality and Tourism Research* 30 (3): 354-77.

Kotiloglu, S., T. Lappas, K. Pelechrinis, and P. P., Repoussis. 2017. “Personalized multi-period tour recommendations”. *Tourism Management* 62:76-88.

Laesser, C., and G. I. Crouch. 2006. “Segmenting Markets by Travel Expenditure Patterns: The Case of International Visitors to Australia”. *Journal of Travel Research* 44 (4):397-406.

Law, R., J. Rong, H. Q. Vu, G. Li, and H. A. Lee. 2011. “Identifying Changes and Trends in Hong Kong Outbound Tourism”. *Tourism Management* 32 (5):1106-14.

Lee, A., C. U. Lambert, and R. Law. 2016. “Customer Preferences for Social Value Over Economic Value in Restaurants”. *Asia Pacific Journal of Tourism Research*, 17 (5):473-88.

Li, G., R. Law, H. Q. Vu, and J. Rong. 2013. "Discovering the Hotel Selection Preferences of Hong Kong Inbound Travelers Using the Choquet Integral". *Tourism Management* 36:321-30.

Li, G., R. Law, H. Q. Vu, J. Rong, and X. Zhao. 2015. "Identifying Emerging Hotel Preferences Using Emerging Pattern Mining Technique". *Tourism Management* 46:311-21.

Liu, C. H. S., C. S. Su, B. Gan, and S. F. Chou. 2014. "Effective Restaurant Rating Scale Development and a Mystery Shopper Evaluation Approach". *International Journal of Hospitality Management* 43:53-64.

Min, K. H., and T. J. Lee. 2014. "Customer Satisfaction with Korean Restaurants in Australia and Their Role as Ambassadors for Tourism Marketing". *Journal of Travel and Tourism Marketing* 31 (4):493-506.

Mynttinen, S., J. Logren, M. Sarkka-Tirkkonen, and T. Rautiainen. 2015. "Perceptions of Food and its Locality among Russian Tourists in the South Savo Region of Finland". *Tourism Management* 48:455-66.

Statistics New Zealand. 2014. 2013 Census QuickStats about culture and identity. New Zealand Government http://www.stats.govt.nz/~/media/Statistics/Census/2013%20Census/profile-and-summary-reports/quickstats-culture-identity/quickstats-culture-identity.pdf (accessed September 22, 2017)

Pang, B., and L. Lee. 2008. "Opinion mining and sentiment analysis". *Foundation and Trends in Information Retrieval* 2 (1-2):1-135.

Pizam, A., G. H. Jeong, A. Reichel, H. V. Boemmel, J. M. Lusson, L. Steynberg, O. State-Costache, S. Volo, C. Kroesbacher, J. Kucerova, and N. Montmany. 2004. "The Relationship between Risk-Taking, Sensation Seeking, and the Tourist Behavior of Young Adults: A Cross-Cultural Study". *Journal of Travel Research* 42 (3):251-60.

Pizam, A., and S. Sussmann. 1995. "Does Nationality Affect Tourist Behavior?" *Annals of Tourism Research* 22 (4):901-17.

Restaurant and Catering Australia. 2016. "Federal Budget 2016-2017: Submission to the Australian Government". Restaurant & Catering, New South Wales, Australia. https://www.treasury.gov.au/~/media/Treasury/Consultations%20and%20Reviews/Consultations/2015/2016%20Pre%20Budget%20submissions/Submissions/PDF/Restaurant%20and%20Catering%20Australia.ashx (accessed June 21, 2017)

Rhee, H. T., S. B. Yang, and K. Kim. 2016. "Exploring the Comparative Salience of Restaurant Attributes: A Conjoint Analysis Approach". *International Journal of Information Management* 36 (6):1360-70.

Robinson, R. N. S., and D. Getz. 2014. "Profiling Potential Food Tourists: an Australian Study". *British Food Journal* 116 (4):690-706.

Rong, J, H. Q. Vu, R. Law, and G. Li. 2012. "A Behavioral Analysis of Web Sharers and Browsers in Hong Kong Using Targeted Association Rule Mining". *Tourism Management* 33:731-40.

Stewart, J. W., L. Bramble, and D. Ziraldo. 2008. "Key Challenges in Wine and Culinary Tourism with Practical Recommendations. *International Journal of Contemporary Hospitality Management* 20 (3):303-12.

Tan, P.-N., M. Steinbach, and V. Kumar. 2006. “Introduction to Data Mining”, Boston, MA: Addison-Wesley Longman.

Thelwall, M. 2017. “The Heart and Soul of the Web? Sentiment Strength Detection in the Social Web with SentiStrength”. In *Cyberemotions: Collective Emotions in Cyberspace*, edited by J. A. Holyst, 119-34. Switzerland: Springer International Publishing.

Tourism Australia. 2016. “Restaurant Australia: Key Messages”. http://www.tourism.australia.com/documents/Campaigns/Restaurant_Australia_key_messages.pdf (accessed May 20, 2016).

Tourism Australia. 2010. “Restaurants and Dining: Key Facts”. http://www.tourism.australia.com/story-ideas/food-and-wine-1693.aspx (accessed May 20, 2016).

Tourism Research Australia. 2016. “International Visitors in Australia: December 2015 Quarterly Results of the International Visitor Survey”. Tourism Research Australia. http://www.tra.gov.au/research/latest-ivs-report.html (accessed May 20, 2016).

Torres, R. 2002. “Toward a Better Understanding of Tourism and Agriculture Linkages in the Yucatan: Tourist Food Consumption and Preferences”. *Tourism Geographies* 4:282-306.

Tse, P., and J. C. Crotts. 2005. “Antecedents of Novelty Seeking: International Visitors' Propensity to Experiment across Hong Kong's Culinary Traditions”. *Tourism Management* 26:965-68.

Updhyay, Y. 2014. “Culinary Preferences of Foreign Tourists in India”. *Journal of Vacation Marketing* 20 (1):29-39.

Verbeke, W., and G. P. López. 2005. “Ethnic Food Attitudes and Behaviour among Belgians and Hispanics Living in Belgium”. *British Food Journal* 107 (11):823-40.

Vu, H. Q., G. Li, R. Law, and B. H. Ye. 2015. “Exploring the Travel Behaviors of Inbound Tourists to Hong Kong using Geotagged Photos”. *Tourism Management* 46:222-32.

Vu, H. Q., G., Li, R. Law, and Y. Zhang. 2017. “Tourist activity analysis by leveraging mobile social media data”. *Journal of Travel Research*, In Press. DOI: 10.1177/0047287517722232

Zhang, H.-Y., P. Ji, J.-Q. Wang, and X.-H. Cheng. 2017. “A Novel Decision Support Model for Satisfactory Restaurants Utilizing Social Information: A Case Study of TripAdvisor.com”. *Tourism Management* 59:281-97.

Zhang, Z., Z. Zhang, and R. Law 2014. “Positive and Negative Word of Mouth about Restaurants: Exploring the Asymmetric Impact of the Performance of Attributes”. *Asia Pacific Journal of Tourism Research* 19 (2):162-80.