Received September 14, 2019, accepted October 3, 2019, date of publication October 8, 2019, date of current version October 22, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2946168

Why Customers Don't Revisit in Tourism and Hospitality Industry?

JING-RONG CHANG[1], MU-YEN CHEN[2], LONG-SHENG CHEN[1], AND SHU-CIH TSENG[1]
[1]Department of Information Management, Chaoyang University of Technology, Taichung 41349, Taiwan
[2]Department of Information Management, National Taichung University of Science and Technology, Taichung 40401, Taiwan

Corresponding author: Long-Sheng Chen (lschen@cyut.edu.tw)

This work was supported in part by the Ministry of Science and Technology under Grant MOST 108-2410-H-324-009.

ABSTRACT The development of social media has changed the way that travelers visit sightseeing spots. The social Internet of Things (IoT) allows products to automatically generate posts, share content and location information, and help build an online community of users based on their company's products, so that marketing personnel can also get useful feedback and understand the user's opinions. In tourism and hospitality industry, to enhance the revisit intention of passengers is an important issue for the purpose of increasing margin. In recent years, related researches had focused on the customers' revisit behaviors and factors. However, few studies have investigated the related issues that travelers do not want to visit again. Failure to revisit may bring a great damage to the company's revenue in the future. To avoid this situation, a text mining based approach will be proposed to identify non-revisit factors from online textual reviews in social media. Because it is impossible to determine whether a passenger has intention to revisit, this study proposed a text mining based approach which uses sentiment of text reviews to identify the passenger's motivations (negative for revisit and non-negative for revisit). Then, feature selection methods, decision tree (DT), Least Absolute Shrinkage and Selection Operator (LASSO), and Support Vector Machines Recursive Feature Elimination (SVM-RFE) will be utilized to discover the important factors of non-revisit factor set. Back-propagation Neural Networks (BPN) and Support Vector Machines (SVM) will be employed to evaluate the effectiveness of selected feature sets. Finally, experimental results could be provided to travel service providers to improve service quality and effectively avoid non-revisit behaviors in the future.

INDEX TERMS Text mining, non-revisit, feature selection, tourism, social media.

I. INTRODUCTION

With the development of social media, the number of tourists has grown rapidly. The social Internet of Things (IoT) allows products/devices to automatically generate posts, share content and location information, and help build an online community of users based on their company's products [126], [127], so that marketing personnel can also get useful feedback and understand the user's opinions. In 2016, the scale of global tourism industry income has accounted for 10% of global GDP [1]. According to the 2017 report of the World Tourism Organization, the number of international tourists has reached to 1.235 billion in 2016. The tourism revenue has remarkably increased from 49.5 billion U.S. dollars in 2000 to 1,220 billion U.S. dollars in 2016. It is also predicted that the number of international tourists will reach 1.8 billion in 2030, and its economic growth rate averages 2.2% per year [2]. Consequently, tourism has made a significant contribution to the growth of the international economy and has become one of the important industries. Moreover, the Statista [3] company reported that the tourism economic contribution in 2016 has reached 7.6 trillion US dollars, of which accommodation, transportation, entertainment, attractions, services, restaurants, retail transactions are the main source of revenue [4]. The growth of tourism market and the development of low-cost operations have made the competition more intense.

In recent years, travel platform websites that have emerged around the world, such as Booking.com, TripAdvisor, Trivago, and Hotels.com, provide B2C (Business to Consumer) marketing channels and social community functions for tourists, hotels, travel agencies, restaurants, and other related businesses. These websites can give kinds of services, such as information sharing, reservations, travel guides,

The associate editor coordinating the review of this manuscript and approving it for publication was Patrick Hung.

price comparisons, etc. Raguseo *et al.* [5] shown that travel websites greatly influence tourism. Most reviews represent consumer decisions. These reviews are also main source of valuable information to consumers. In addition, due to the popularity of mobile devices and the evolution of application services, travelers can connect to the Internet anytime and anywhere. These habits have also changed tourism ecology [6]. Travelers use mobile services to plan travel and record footprint, and share experiences in social media [7].

The visitors who have already visited are different from the first-visit visitors. They have familiarity with and understanding of the places where they visited. They are more willing to pay for souvenir, native products and services in the expected travel process [8], [9]. They are also more willing to support the development of local tourism [10]. These travelers are willing to provide richer consumer experiences to share with others when their expectations are satisfied during the travel process [11]. Sharing a message is a source of decision making and future motivation for each reviewer [12], [13]. They may have different emotions including happiness, surprise, trust, joy, anger, disgust, sadness, fear and others when sharing comments [14]. Under different emotions they will write different comments for travel processes. For example, they will interact with friends on social media in order to increase personal well-being, to describe moments of travel in a pleasant mood story, and to promote the possibility of other people visiting their destinations [15]. When the process of tourism is frightening, they may stop other people from accessing or even sharing the travel experience [14].

Passengers' revisit intention has been recognized as one of the important factors for the survival and development of the tourism industry [16]. Revisit intention contains the future behaviors of passengers, including decision-making, destination selection, and post-trip assessment [17], and it is also one of the main factors that enhance the growth of tourism marketing operations [18]. The analysis of users' influence on the behavior of others in social media is increasingly important [19]. Ye *et al.* [20] showed that searching for travel related information is one of the most popular online activities. Passengers may influence the decision of future travel goals when reading the contents of the consumer reviews. The first-visit travelers will refer to these reviews to assess whether the tourist attractions are worth visiting, and the visitors who intend to revisit will seek suitable travel listings from the comments [21]. From a consumer's point of view, they trust travelers rather than official tourist messages [22].

Some studies have shown that if customers have behaviors that they will not visit in the future, they may lead to sharing negative experiences and may even defame service providers such as hotels and restaurants [23]. These behaviors will strongly affect future marketing of enterprises. Consequently, determining why the customer is no longer visiting can improve the corresponding services. Companies can also implement remedial actions based on this and transform a dissatisfied, complaining, or angry customer to a loyal customer [24].

Retaining customers and encouraging them to revisit is key to increase revenue. Most importantly, the cost of retaining old customers is only one-fifth of the cost of searching for new customers [25]. On the other hand, the company can improve all aspects of internal services, in addition to avoiding new passengers may produce a bad first impression, it may also reduce the motivation for customers to publish negative reviews online.

In the past years, many scholars have discussed issues of revisit intention. For examples, Som *et al.* [16] believed that relaxing, promoting relationships, improving social interactions, and improving reputation are factors of revisit intentions. Kim *et al.* [26] found that the overall quality of life is a factor that influences revisiting of older people. Hsu *et al.* [27] indicated that the perception of service by low-cost operation will strongly influence passenger revisits. Loi *et al.* [28] thought the quality of the tourism shuttle affects the revisit intentions due to destination image and destination satisfaction, and there are lots of studies that explored the relationship between trust, satisfaction, loyalty, and revisit intentions [29]–[31]. However, from available literature, relatively few studies have been conducted to discuss issues that passengers do not visit (non-revisit intention) anymore. Therefore, this study will study non-revisit intention, which means travelers will no longer re-visit.

In addition, online reviews have the characteristics of communication and global influence [32], and text reviews reflect the aggressive evaluation of customer satisfaction or dissatisfaction. Other consumers' decision-making behaviors are influenced by these text reviews [33]. Textual comments in social media can be considered one of source of voice of customers. However, the questionnaire survey cannot be immediately responded to by customers, and needs to consider costs of manpower and time, and there may be disadvantages such as misunderstanding of question items and sampling bias. Compared to traditional questionnaire surveys, the text in the commentary is more active in showing consumers' future intention [34]. Moreover, the emotions expressed in the comments have different influences and their importance has been recognized in previous studies [35]. In order to explain these emotions, scholars have conducted research through sentiment analysis in text mining [36]–[38]. Therefore, the study will use passenger reviews instead of questionnaires to conduct research.

To sum up, this study proposed a text mining based approach to identify non-revisit factors from online textual reviews in social media. We firstly define potential factors of revisit intention. This study will use related literature about tourism service, including restaurants, restaurants, and destinations, to identify potential factors of revisit intention. Secondly, a famous website, TripAdvisor, will be employed as our experimental case for collecting data. Because it is impossible to determine whether a passenger has intention to revisit, this study uses sentiment of text reviews to

TABLE 1. Related works in revisit intensions.

Authors	Research objects	Discussed Factors	Survey method
[45]	Tourist attractions, hotels	Trust	Questionnaire
[46]	Climbing attractions	Satisfaction	Questionnaire
[29]	Restaurants		Questionnaire
[47]	Destination image	Loyalty	Questionnaire
[44]	Comments	e-WOM	Literature review
[18]	E-commerce website	Revisit intention	Questionnaire
[48]	Aviation industry		Questionnaire
[49]	Destination image		Questionnaire

identify the passenger's motivations (negative for revisit and non-negative for revisit). Then, feature selection methods, decision tree (DT), Least Absolute Shrinkage and Selection Operator (LASSO), and Support Vector Machines Recursive Feature Elimination (SVM-RFE) will be utilized to discover the important non-revisit factors. Finally, through the analysis of the results, some suggestions will be provided for tourism industry to enhance their service quality, improve passengers' experiences after playing, and improve internal service management. Based on discovered results, tourism industry can enhance the passengers' revisit intention.

II. RELATED WORKS

A. REVISIT INTENTIONS

Kozak [39] thought revisit intention is the intention of the consumers to travel to a destination or other tourist attraction again. When the consumer is satisfied with the experience of the tour, or even exceeds his expectation, they will come back again or recommend it to others [40]. Engel *et al.* [41] pointed out that revisit intentions are mainly the understanding of tourism, the emotions generated by tourism, and the future behavior of tourism. To develop the revisit intentions of consumers, it is necessary to grasp the promotion at the first visit, including the overall experience regarding price, facilities, product core value, employee interaction, etc. Existing researches show that there has been a positive relationship between past travel experiences and future travel [42].

Phillips *et al.* [43] stated that to create revisit motivations is the first and crucial step for the growth of tourism. Researchers found that the trust, satisfaction, and loyalty of consumers are the main antecedents which influence revisit intentions [40], [44]. In addition, lots of researchers paid much attention on this issue. Related studies were conducted in different directions, as shown in Table 1. For examples, Artigas *et al.* [45] suggested that in order to create relationships with tourists, it is necessary to grasp the tourist's perception of the destination and make them feel affectionate to the local area. Taher *et al.* [46] indicated that travelers' satisfaction with the overall experience is the antecedents of driving visitors to revisit. The antecedents include landscape features, accessibility, organization and perceived risks. And attractive local features will strongly affect the overall satisfaction of tourists. Kim *et al.* [29] investigated if the provided meals are good for health, it will affect passengers' willingness to revisit in the future. Zhang *et al.* [47] stated that factors including enjoyment, refreshment, knowledge provision, participation, novelty and others will make them willing to revisit. Cantallops and Salvi [44] proposed to understand consumers' motivation for writing reviews and the effects of electronic word of mouth. Che *et al.* [18] indicated that the hospitality industry needs to master the personalized characteristics of products to satisfy customers and stimulate their willingness to revisit. Liu and Lee [48] thought that companies should give appropriate prices to customers to enhance electronic word-of-mouth marketing strategy. Stylos *et al.* [49] considered that customers should be given a good destination image, which will encourage tourists to recommend or revisit in the future.

According to literature review from available published works, the used survey or data collection methods for revisit intention are almost conducted in the form of questionnaires. For questionnaire survey, data collection probably consumes a lot of costs and time. In addition, the questions items might have potential problems which may be unable to be fully understood by respondents when filling in the questionnaires. Therefore, this study will attempt to use online textual reviews as research data and utilize text mining approach for further analysis.

Moreover, regarding potential factors of revisit intentions, we also surveyed some works. For instances, Knutson *et al.* [50] proposed four dimensions that affect the action experience in 2009. They are environment, accessibility, benefit, and incentive. Ren *et al.* [51] also presented three dimensions of sensory experience, staff performance, tangible-sensorial experience, and aesthetic perception. They believe these experience processes could motivate passengers' future visits. Tan [52] stated that there are three psychological factors for travelers to reduce their willingness to revisit, which include (1) internal constraints: inner fear in the travel experience; (2) interpersonal constraints: from friends and family members, and others' social interactions; (3) structural constraints: the impact of the external environment.

Because there is no single one work can provide all factors of revisit intentions, this study attempts to find out the potential factors in the related works about hotels, restaurants, and destinations. Table 2 provides a summary of these literatures. From available works, there are some potential factors mentioned in related works. For examples, Han and Kim [53] found that location accessibility is a major factor affecting tourists' visits to green restaurants, where distance and ride difficulty are the main factors. Han and Hyun [54] investigated luxury hotels. They found that location accessibility, amenity, and food as intermediary variables. And these factors will increase the satisfaction of the travelers

TABLE 2. A brief summary of literature review in revisit intention.

Potential factors \ Literature types	Hotels	Restaurants	Destinations	Others
Location accessibility	[53], [54], [55]	[57]	[58]	
Employee services	[55]			
Amenity	[54]			
Food / drink	[54]		[58]	
Price	[56]	[59]		
Recommendation				[60], [61]
Trust				[62]
Cleanliness	[56]			
Comfort	[56], [51]			
Emotion word				[35]
Environment facilities	[54]	[59]		
Room facilities	[55]			
Shopping			[33]	

in their travels, and when the customers use hotel provided facilities will increase their impressions. Peng *et al.* [55] believed that it is particularly important for hotel management to enhance the customer experience through employee services and room facilities. Good service quality will impress customers and enhance the possibility of future customer visits. Gu and Ryan [56] thought that the price, cleanliness, and comfort will affect the customer's overall satisfaction. And they also are core services to improve hotel performance.

Besides, Mattila [57]considered that location accessibility greatly influences customer loyalty, which is one of the important factors for their willingness to revisit. Um *et al.* [58] also stated that the number of revisits at the location is an important indicator for the revisiting of a passenger. The quality of food/drink is also a reason for visitors to visit. Ryu and Han [59] conducted a survey between customer satisfaction responses and behavioral intent. The results showed that the price is a factor that influences the customer's travel satisfaction, and also will be used as the determinant of revisit. Bigne *et al.* [60] stated that the recommendation of tourism image is a reference for intentional revisit, and Hui *et al.* [61] also indicated that when passengers have recommendation behaviors, they intend to revisit. Garbarino and Johnson [62] suggested that establishing trust with customers will increase the possibility of revisiting. Kabadayıand Alan [35] found that taking into account the emotions of passengers, their emotional state directly affects the intention of revisit.

B. ONLINE TRAVEL REVIEWS

Amin *et al.* [126] defined social Internet of Things (IoT) which can be viewed as devices can create connections with each other to independently. They also indicated the importance of providing reliable data analyses by using trust and friendliness based properties. Therefore, Amin *et al.* [127] aims to solve link selection problems in social IoT. Moreover, in the work of Amin *et al.* [126], they aimed to explain social IoT, including the basic concept of trust, the properties of trust, and so on. Based on a survey over past studies, they also classified friendliness and trust of social IoT. Since social IoT allows products to automatically generate posts, share content and location information, and help build an online community, we need to discuss the influence of social media.

Due to the booming social media, its influence on the tourism industry is continuously increasing [63]. Consumers often take pictures on the trip and upload them to social media sites for comments. Their behaviors indirectly influence others' awareness of and access to tourist attractions [6]. Online reviews as a traveler to share their travel experiences, especially products and services that recommend or complain about travel [64] to other consumers. These comments reflect the true feelings of consumers, and also refer to the satisfaction of customers, and have important implications for online marketing [65]. Although the online rating system provides consumers with intuitive feedback on products and services, such as star rating, sharing, etc., in contrast, the content of reviews has more valuable information than star rating or sharing [66], [67]. When consumers write negative reviews, they implicitly refuse to visit their destination again [68]. Vásquez [69] pointed out that negative reviews include some of the consumer's speech behaviors toward the industry. Negative reviews include positive discussions, expectations, suggestions, content for improvements, and warnings.

Since travelers often read other travelers' comments, they are more likely to view other visitors' comments as a source of providing more reliable and relevant information, and these are more likely to be highly influenced by these comments [70]. Consumers' positive comments motivate other consumers to increase their purchases or increase their willingness to visit [71], and they have always been actively pursued by manufacturers. The main reason is the communication of electronic word-of-mouth (Electronic Word-of-Mouth, e- WON) will affect other consumers' decisions [72], and finally influence market competition trends [73]. In recent years, scholars have presented different opinions on online reviews. Cenni and Goethals [68] stated that although positive reviews influence other consumer decisions, negative reviews have more reference value for improving products and services. When the comments contain negative sentiment, they will have a negative impact on the revisits of other travelers after reading, and may eventually lead them to no longer visit. Sánchez-García and Currás-Pérez [74] stated that dissatisfied consumption experience leads to negative e-WOM, and that regrettable consumers are more likely to spread negative e-WOMs, thereby greatly reducing the likelihood that other consumers will be willing to make a first visit or revisit to the destinations.

The TripAdvisor website provides a feedback mechanism after the consumption experience, and the website does not influence the user to book the hotel by manipulating the comments. Recently, it has also been a research target for many scholars on tourism related issues [75], [76], [77], [78]. To sum up, the study will use the consumer reviews of the TripAdvisor website as research data and then confirm the relationship between these comments and revisits. The impact of sentiment on revisiting will be further analyzed for discovering the potential relationship of non-revisit.

C. TEXT MINING

Sullivan [79] defined text mining as a process of editing, organizing, and analyzing a large number of documents to provide specific user-specific information and to find certain features related to documents. Different from traditional data mining methods, the main task of text mining is to convert texts into numerical data through language analysis and natural language processing. The main purpose of text mining is to identify the important information of a document, to discover potentially useful information from the features in the document, and to further analyze the information. Due to the rapid development of social media and the explosion of internet traffic, these unstructured or semi-structured texts need to be processed through text mining techniques to explore the underlying structure and rules [80].

In recent years, tourism-related researches have also used text mining for data analysis. For examples, Godnov and Redek [81] conducted a sentiment analysis for reviews of 87 hotels in Croatia, to confirm consumer appeals. Wong and Qi [82] implemented an analysis on comments of the TripAdvisor's travelers from 2005 to 2013 to investigate the evolution of Macau hotels over the years. Kim *et al.* [34] classified the comments for discovering the visitors' views on the destination services. Hu *et al.* [83] clustered hotel reviews to find out more comprehensive information for hotels to improve internal management. In addition, Schuckert *et al.* [84] explained that since traditional questionnaires have experimental effects and online reviews have more objective, large-scale, and sample-free biases features, the information provided by written materials has more information than the questionnaires. So, text mining approaches will be employed as data analysis tools for collected textual comments.

1) WORD SEGMENTATION

In most natural language processing, words are the basic unit of processing in texts. Word segmentation is the process of making texts in a sequence of N-grams (N consecutive words). These segmented consecutive words can be used for text indexing (keyword search) or feature classification [85]. The common N-gram units are uni-gram (one word) and bi-gram (two consecutive words). When N is larger, the more accurate the model is, the more complicated it is to calculate.

Word segmentation has also been adopted in various fields, such as sentiment classification of online restaurant reviews [86], comparison of online consumer reviews between Chinese and English [87], extracting emotional characteristics by using intelligent text processing and computer linguistics from multilingual texts [88], and utilizing feature selection methods to optimize the system for network intrusion detection [89]. Consequently, this study will use the word segmentation as the basis for processing textual comments. We'll segment irregular, complex comments, and extract these segmented words.

2) SENTIMENT ANALYSIS

Sentiment analysis aims to judge or assess people's emotional state, attitude and opinion [90]. In the process of extracting valuable information from public tweeting, natural language is used to process specific topics or comments [91]. The process of calculating, identifying and classifying the opinions expressed in the text is used to determine the author's attitude toward products is positive, negative or neutral [92]. And sentiment analysis is mainly used to understand the user's intention and purpose, and then infer future behavior [19].

From available literature, Ferreira *et al.* [93] assessed sentence similarity through lexical, syntactic and sentiment analysis. Gaspar *et al.* [94] analyzed the emotional classification of social media users to external potential stress events. Sentiment classification merely judges the sentiment of a review is positive or negative. There are two common sentiment classification methods, including semantic orientation and machine learning. And they are also usually used hybrid analysis methods in recent years [36], [37]. Therefore, this study will use these two methods of analysis to conduct experiments.

Semantic orientation focuses on determining the polarity of text, sentences, or features (positive or negative) and measuring the degree of polarity in the text [95]. In other words, its main purpose is to compile or customize words as positive or negative thesaurus, and calculate the vocabulary relationship score based on the text's positive and negative, and finally judge its semantic orientation by its score [96]. At present, there are also related works that used semantic orientation to classify sentences and sentiments. For examples, Dun and Guo [97] classify text based on the multiple definitions of HowNet lexicon, and proposed a new method of semantic orientation measurement. Garcia-Moya *et al.* [98] retrieved product features and opinions from customer reviews. Chenlo and Losada [99] subjectively categorized sentences from online products, movies, reviews, and news articles.

Machine learning is a technology that uses collected data to train classifiers. Because it needs to train a large amount of data, its calculation and complexity are very difficult with artificial statistical techniques. Therefore, machine learning is needed to train the classification model offline, and then used the trained model online. For classifying sentiment of reviews, machine learning methods can achieve better accuracy, but they need lot of learning time. Semantic orientation can classify sentiment in a review very quick, but it cannot have a good accuracy. To reduce learning costs and time, machine learning techniques are often combined with semantic orientation. When the frequency of texts in various languages is counted, a term-document matrix (TDM) could be established. We can use semantic orientation method to label examples in TDM. And finally, a classification model is established by machine learning methods. This combination could achieve the purpose of reducing learning time and recognizing sentiment accurately [100]. Besides, the lexicon of semantic orientation methods should keep updated. And classification models built by machine learning methods should be repeatedly trained to achieve better performances when the text data changes [101]. Therefore, this study will label collected data by semantic orientation methods, and then establishes the sentiment classification model by machine learning.

D. FEATURE SELECTION

Feature selection has been widely used in data mining and machine learning. Its purpose is not only to significantly reduce the feature space, but also to improve the prediction accuracy of the classifier by eliminating redundant or irrelevant features [102]. This study will use feature selection methods to extract the crucial factors that influence the visitor's re-visit intention.

1) DECISION TREE

A decision tree (DT) is a method for establishing a classification model. Its purpose is to induce examples to generate a tree structure model [103]. Decision trees are one of very popular classification methods. Lots of successful applications have been presented. For examples, Han *et al.* [104] analyzed customers' value by using a decision tree model to segment telecommunications customers. Chen *et al.* [105] used Bayesian networks or principal component analysis, with back-propagation neural network or decision tree (C5.0) to predict the effectiveness of earnings management. They indicated that Bayesian networks with decision trees can have better results. Moro *et al.* [106] used Support Vector Machines (SVM) combining with decision trees for sensitivity analysis to predict social media performance, and to assess impacts of brand building. Perez-Alonso *et al.* [107] used a decision tree to predict the environmental and agronomic effects of sewage and sludge mixtures. To sum up, this study will use decision trees as one of the feature selection methods.

2) LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR

Least absolute shrinkage and selection operator (LASSO) is a compression coefficient and regression variable selection method proposed by Tibshirani [108]. It constructs a refined model and passes the characteristic coefficients. The square sum of the least-squares method reduces the sum of the absolute values of the coefficients to less than the constant 1 [108], [109]. In the feature selection, if the resulting feature coefficient is set to 0, it means that this feature is not recommended in the mode. LASSO has the advantage of retaining the subset contraction [109], [109].

In recent years, lots of applications have successfully used LASSO to conduct feature selection studies. For examples, Kamkar *et al.* [111] used decision trees in combination with LASSO for clinical prediction of concurrent risk of cancer and acute myocardial infarction to obtain better prediction accuracy and stability. Wang *et al.* [109] applied LASSO and regression analysis to predict the fuel consumption of the ship under different navigation environments. Gauthier *et al.* [112] used the LASSO regression model to predict the sound quality. Therefore, this study will use LASSO as one of the feature selection methods.

3) SUPPORT VECTOR MACHINE RECURSIVE FEATURE ELIMINATION

Support vector machine recursive feature elimination (SVM-RFE) is one of widely used feature selection approaches. It is mainly based on SVM algorithm to train classifiers. The input factors could be ranked by the weight vector w in descending order. The ranking of a factor represents its importance [113].

Therefore, lots of successful SVM-RFE applications in real world. For example, SVM-RFE has been employed in medical area, such as drug classification [114], cancer identification [115] and so on to diagnose the disease in a short period of time. Other applications include feature selection model for testing brain wave of emotional responses [116]. It has also been applied to electricity market analysis to predict fluctuations in electricity prices [117]. Therefore, this study will use SVM-RFE adoption as one of the feature selection methods to identify factors that affect passengers no longer visit.

III. EMPLOYED METHODLOGY

There are six steps in this experiment. The detailed implemental procedures are used methods have been described as below.

STEP 1 DATA COLLECTION

This work collects English text comments published by the consumers from the travel online website, TripAdvisor, using the crawler tool. All the comments containing special characters and non-English words will be cleaned to avoid errors during pre-processing. Table 3 shows an example of data clean. In this table, only the English words in the comments will be retained.

STEP 2 DATA PRE-PROCESSING

The steps of pre-processing could be divided into three parts, including filtering revisit related comments (building revisit lexicon), defining revisit factors, and establishing term-document matrix (TDM). In the process of data preprocessing, in order to establish lexicons, word segmentation and word frequency computation will be implemented for all collected comments. This study uses the commonly used uni-gram method to segment words [118], [119]. Besides, in order to keep suitable size of feature set in dictionary, we merely keep those words whose frequency is more than 5 times [120]. Moreover, to avoid a large number of

TABLE 3. An example of text comments before and after data cleaning.

Before cleaning	"My wife, my one year old son and I stayed in the hotel for 9 days. It's is a new well designed and equippe 鍫?橑 d hotel, with very friendly staff who are very keen to help and always there to give a hand. The room was always clean, and the amenities provided of very?揮□?擗□?Savvy?□漁?□ □捐?□?愬 □□ 鉎 □?撣?隞砍蒂? ?□說 good quality. The breakfast was great, with healthy selection of very tasty food, and again great service from the restaurant staff. The hotel location is also very convenient, a qui 曇鍫?橑抬?橑 ck walk to Nathan road and Jordan Road and Jordan MTR station. We definitely got our money's worth and will hope to stay there in our next visit to Hong Kong. ?□??輸□霈曇鍫?橑抬?橑□??□??圈?揮□?□?瘣餅??胯 ?Very modern"
After cleaning	"My wife, my one year old son and I stayed in the hotel for 9 days. It's is a new well designed and equipped hotel, with very friendly staff who are very keen to help and always there to give a hand. The room was always clean, and the amenities provided of very good quality. The breakfast was great, with healthy selection of very tasty food, and again great service from the restaurant staff. The hotel location is also very convenient, a quick walk to Nathan road and Jordan Road and Jordan MTR station. We definitely got our money's worth and will hope to stay there in our next visit to Hong Kong. Savvy Very modern"

TABLE 4. Experience dimension of hotels.

Type	Dimension	Definition	Supports
Hotel experiences	Benefit	There should be no surprises surrounding a product/service. All hotel products/services should be safe to use. Consistency in product/service performance makes me more confident.	[50]
	Conveniences	Hotels must be laid out so that I can find what I want. Product/service information should be readily available to me. Products/services must always be readily available. Hotel products (web-based or otherwise) must be clutter free. The process of buying and using the hotel's products/services should be simple. The products/services must be easy for me to acquire.	
	Incentive	Incentives increase the chance that I will purchase the featured hotel product/service. I am more likely to buy a hotel's product/service if incentives are offered. Price promotions that accompany a product/service are a bonus.	
	Environment	The surroundings should be entertaining to me. Music enhances my interaction with the hotel's products/services. The hotel's environment should provide sensory stimulation. Stimulating product/service environments make me more likely to buy. The product/service environment should motivate me. The hotel environment should be fun.	
	staff relational/ interactional experience	Service staff attitude and feelings of interaction	
Sensorial	Tangible-sensorial experience	The overall sensory experience of the tangible environment, such as the atmosphere in the bedroom.	[51]
	Aesthetic perception	The sense of design of the interview location, such as the decoration of the hotel.	
	Location	Convenient location and nearby facilities.	

redundant words, we also remove stop words, such as "a", "and", "the", "with", "or", etc. Detailed sub-steps of data pre-processing have been described as follows.

STEP 2.1 BUILD HOTEL REVISIT LEXICON AND REMOVING UNRELATED COMMENTS

This step aims to remove comments that are irrelevant to the revisit. To achieve this task, we should build revisit lexicon first. Hu and Liu [121] indicated that nouns can be regarded as a feature used to assess the importance of sentences. This study uses the related nouns and their relevance to the description of the experience dimension shown in Table 4. For every single dimension, we collect synonyms, similar words, related words, and antisense words from dictionaries and sample documents to build "hotel revisit" lexicon. When the comments mentioned one of hotel experience dimensions proposed by Knutson *et al.* [50] and also mentioned one of the sensorial experience dimensions proposed by Ren *et al.* [51], they will be considered as revisit related comments and kept for further analysis. In other words, when the description in one review is similar or close to these dimensions, the comment will be considered as relevant to the revisit, and it will be used as experimental data. If the comments are not similar, the comments will be discarded. In addition, this study will perform part-of-speech (POS) to tag words in sentences. The process of providing appropriate part-of-speech symbols and tags to reduce the possibility of misreading when reading reviews.

STEP 2.2 DEFINE FACTORS

This step aims to define candidate factors, extract semantic words, and build sentiment lexicon.

STEP 2.2.1 DEFINE CANDIDATE FACTORS AND CLASS LABELS

According to available literatures, we focus on revisit intentions issues in hotels, restaurants, tourist attractions and destinations, to identify the factors that might potentially affect visitors' re-visit intentions, and then build the feature vocabulary according to the definition of the factors. These features words will be extracted from reviewing the literature and from collected samples of text reviews.

STEP 2.2.2 EXTRACT FEATURE SEMANTIC WORD

Turney [122] indicated that adjectives are important indicators of emotion. In addition, Singh *et al.* [123] pointed out that verbs, adjectives or adverbs can reflect all behaviors, emotions, and opinions. Therefore, we will extract verbs, adjectives and adverbs to be our semantic words.

The feature word extraction process in this work will use two ways. In the first way, we select the relevant words presented in the past literatures. The second way uses the one-fifth of the total number of collected comments after implementing step 2.1. And then, we segment words using uni-gram. At the same time, in order to avoid too many redundant words and slogans, stop words will be removed. In order to extract more words for building semantic lexicon, this work will keep words whose frequency are more than 5 times [120], containing verbs, adjectives, and adverbs.

STEP 2.2.3 BUILD SENTIMENT LEXICON

Next, we will evaluate the relation between selected words and defined factors in previously relevant literatures. And the appropriate words will be classified into suitable factor. Other extracted feature semantic words which are not relative to defined factors will be ignored. Finally, we use selected words as our base, and extend lexicon by adding their synonyms, synonyms, and antonyms from English dictionaries.

STEP 2.3 BUILD TDM

In this step, we focus on revisit related comments. Then, we will perform word segmentation. The segmented words will be compared to lexicon and the frequency of segmented words will be calculated. Next, the sum of the frequency of each factor will be totaled as the attribute value of a factor.

In addition to attributes, class labels should be defined. Since it is impossible to confirm from the comments whether the passengers revisit or not, this word will use the sentiment of text reviews to determine if one review is positive to revisit or negative to revisit.

We implement two experiments in this study. In experiment #1, we follow the work of Hu and Liu [121] to compute the score of sentiment. If the sentiment score is large than zero, the class label will be determine as ''positive for revisit''. On the other hand, the class label will be considered as ''negative for revisit''.

Because the class imbalance problem which means the classifier has an extremely low ability of identifying the minority examples in experiment #1. On the other hand, Singh *et al.* [123] believed that the companies should investigate the negative emotions of the passengers to alleviate the pressure on both sides. Vásquez [125] said that when the negative words appeared in the comments, the passengers were provided with negative information about the experience process; Cenni and Goethals [68] also indicated that the more bad reviews or negative semantics in the comments will have a negative impact on the revisit, causing them to stop visiting. To sum up, the study will consider the negative meaning of the comments as negative comments on the passengers' revisiting, and adopt the lexicon established by Hu and Liu [121] to determine class labels. Therefore, in experiment #2, the class labels will be defined as ''not negative for revisit'' and ''negative for revisit''. Finally, based on defined class labels and attributes, we can build TDM for further analysis.

STEP 3 IMPLEMENT FEATURE SLELECTION

Before learning, we first normalize the collected data into the interval $\{0, 1\}$ by using equation (1).

$$X_{n,m} = \frac{X'_{n,m} - X'_{n,min}}{X'_{n,max} - X'_{n,min}}, \tag{1}$$

where $X^{'}_{n,m}$ is the value of factor X_n in document d_m, $X^{'}_{n,min}$ is the minimum of all values in factor X_n, $X^{'}_{n,max}$ is the maximum of all values in factor X_n. Then, we implement 5-fold cross-validation experiments which divide the collected data into 5 equal parts. One part of them will be taken as test set, and the other 4 parts are used as training set in turn. Therefore, we'll have 5 different training/test sets.

STEP 3.1: DT

This study will use the decision tree (C5.0 algorithm) as one of our feature selection methods. Before building a decision tree, the conclusions and residuals of the previous tree will be set as the learning coefficient. In the subsequent training, the prediction error is reduced to improve the accuracy of the model, and the system resources and memory are less. Therefore, the speed of the calculation process is increased, and the training time is reduced. The training steps are as follows:

STEP 3.1.1

Construct training and test data sets.

STEP 3.1.2

Utilize C5.0 algorithm to establish a decision tree.

STEP 3.1.3

Create an initial rule tree.

STEP 3.1.4

Prune this rule tree to make it more readable.

STEP 3.1.5

Select the best performing rule tree.

In decision trees, the factors remaining in the tree will be considered as important. Since we use 5-fold cross-validation experiment, we will build 5 trees. One factor is selected more than 3 times will be kept as our candidate feature subset.

STEP 3.2 LASSO

LASSO is one of our feature selection methods. Since the selection criteria, defined in equation (2), must be considered, the evaluation values of the appropriate parameters can be used.

$$min \sum_{t=1}^{T} (y_t - \beta_0 - \beta_1 x_{1,t} - \cdots - \beta_k x_{k,t})^2, \ s.t. \sum_{j=1}^{k} |\beta_j| \leq \lambda \tag{2}$$

As shown in equation (2), since the regression parameter value βi is based on a certain penalty function selection criterion, an appropriate change is selected. Where λ is the adjustment system, T is the number of data, and x is the explanatory variable. In the case of given k explanatory variables, the selection of the λ value affects the parameter estimation value, and when the estimated value is set to 0, it indicates that the variable is not suitable for selection into a mode. So, the estimate can be selected to fit the appropriate value. Variables are used as guidelines. This study considers factors with estimates that are not zero as factors that have potential impact in the subset of features that are no longer being accessed.

STEP 3.3 SVM-RFE

SVM-RFE is based on the SVM classifier. The obtained weight vector w is ranked in order. The minimum weight of the feature vector is removed each time, and the remaining training set is new. The feature combination will reclassify the sorted feature values by repeating the above steps. The detailed SVM-RFE algorithm has been shown as bellow.

INPUTS

Training examples $X_0 = [x_1, x_2, ...x_k, ...x_l]^T$, class labels $y = [y_1, y_2, ...y_k, ...y_l]^T$

INITIALIZE

Subset of surviving features $s = [1, 2, ...n]$ feature rank list $r = []$ repeat until $s = []$ restrict training examples to good feature indices $X = X_0(:, s)$ train the classifier $\alpha = SVM - train(X, y)$ compute the weight vector of dimension length(s) $w = \sum_k \alpha_k y_k x_k$ compute the ranking criteria $c_i = (w_i)^2$, for all i find the feature with smallest ranking criterion $f = \arg min(c)$ update feature rank list $r = [s(f), r]$ eliminate the feature with smallest ranking criterion $s = s(1 : f - 1, f + 1 : length(s))$.

OUTPUT

Feature ranked list r.

STEP 4 BUILD THE CLASSIFICATION MODELS

In this work, all feature subsets and original feature sets have been used to build BPN and SVM classifiers, respectively. Based on classification performances, we can evaluate the performances of selected feature subsets.

STEP 4.1 TRAIN SVM CLASSIFIERS

The selected feature subsets and original feature sets are firstly trained and classified by SVM algorithm, and the classification results are verified. The training steps are as follows:

STEP 4.1.1

Transform data format.

STEP 4.1.2

Select RBF kernel function.

$$K(x, y) = e^{-r||\text{x-y}||^2} \tag{3}$$

STEP 4.1.3

Find optimal settings of parameters C and γ.

STEP 4.1.4

Build SVM model using selected parameter settings.

STEP 4.1.5

Test the built SVM classifier.

STEP 4.2 TRAIN BPN CLASSIFIERS

BPN is also used to evaluate the results of feature selection. The implementing steps are as follows:

STEP 4.2.1

Determine the number of network layers and the number of nodes in each layer.

STEP 4.2.2

Setup the network initial weights and initial intercept

STEP 4.2.3

Input training sample and target output value.

STEP 4.2.4

Calculate the network output value.

STEP 4.2.5

Calculate differences between the output layer and the hidden layer.

STEP 4.2.6

Compute the adjustments of weights and intercept of each layer.

STEP 4.2.7

Update weights and intercept of each layer.

STEP 4.2.8

Repeat step 4 to step 7 until the network converges or the number of training reaches the upper limit.

STEP 4.2.9

Test built BPN model.

STEP 5 EVALUATE PERFORMANCES

In this study, we use several metrics to evaluate performances of classifiers. These metrics will be computed using the

TABLE 5. Confusion matrix.

Actual \ Predicted	Positive examples	Negative examples
Positive examples	True Positive, TP	False Negative, FN
Negative examples	False Positive, FP	True Negative, TN

confusion matrix shown in Table 5. In this table, TP and TN represent positive and negative samples that are classified as right class. FP and FN mean positive and negative samples that are classified into wrong class.

Next, we introduce the used metrics in this study. Positive Accuracy (PA) and Negative Accuracy (NA) defined in equations (4) and (5) represent the ability of classifying positive and negative examples, respectively.

$$PA = \frac{TP}{TP + FN} \tag{4}$$

$$NA = \frac{TP}{FP + TN} \tag{5}$$

The third used metric is the geometric mean (GM) of PA and NA. The last two used indicators are overall accuracy (OA) and F1-Measure (F1). These metrics have been defined in equations (6)~(8).

$$GM = \sqrt{PA \times NA} \tag{6}$$

$$OA = \frac{TP + TN}{TP + FP + FN + TN} \tag{7}$$

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{8}$$

STEP 6 RESULTS AND DISCUSSION

Finally, the effectiveness of selected feature sets will be evaluated by SVM and BPN classifiers. Based on classification results, we can determine the critical factors that affect tourists' non-revisit intentions. Moreover, the differences of factors between luxury and economic hotels will be discussed. Finally, we can provide some suggestions provided to travel service providers to improve service quality and effectively avoid non-revisit behaviors in the future.

IV. EXPERIMENTAL RESULTS

A. DATA COLLECTION AND PRE-PROCESS

In this study, the crawler tool "Kimono" (https://www.kimonolabs.com/) was used to extract online text reviews. We focus on reviews of Hong Kong restaurants and hotels in TripAdvisor (https://www.tripadvisor.com.tw/), a world-renowned tourism review site. Figure 1 provides one example of collected reviews. We collected top five hotels' reviews according to the ranking of the site. Moreover, in order to enhance the effectiveness of this study, we will collect different types of restaurants and hotels, including reviews of consumers in both luxury and economic hotels. The dates of collected reviews are from January 1, 2015 to October 30, 2017. The summary of collected data has been

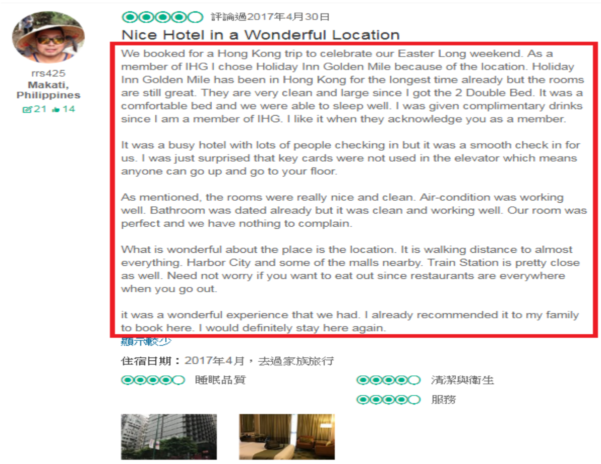

FIGURE 1. **An example of collected reviews in TripAdvisor.**

TABLE 6. Summary of collected data.

Type	Number of hotels	Rank	Hotel name	Number of reviews	Size of valid reviews
Luxury hotels	49	No.1	The Upper House	938	6402
		No.2	The Ritz-Carlton, Hong Kong	1655 (not valid 1)	
		No.3	The Landmark Mandarin Oriental Hong Kong	440	
		No.4	Four Seasons	1305	
		No.5	Hotel ICON Hong Kong	2064	
Economic hotels	354	No.1	Cordis Hotels and Resorts	566	2069
		No.2	The T Hotel	175	
		No.3	The Jervois Hong Kong	279	
		No.4	Mira Moon Hotel	822	
		No.5	Hotel Stage	227	

shown in Table 6. Finally, there are 6402 valid reviews from luxury hotels, and 2069 from economic hotels.

Next, text mining tool, QDA Miner, was used for processing collected text data. Those nouns whose frequency is greater than 10 will be kept as our input features. Moreover, we used the natural language tagging tool, Stanford Part-Of-Speech Tagger (https://nlp.stanford.edu/software/tagger.shtml#About).

Finally, a review related to the dimensions of the mobile experience [50] and the dimensions of the sensory experience [51] will be considered as revisit related reviews. After screening, 4476 reviews of luxury hotels and 1569 comments of economic models were used for further analysis.

Then, we will label the class attribute, and build TDM. In this study, we have two experiments. In experiment #1,

TABLE 7. Employed revisit related reviews and their distribution.

Hotel type	Revisit related reviews	Data size	Distribution of class	
			Experiment #1	Experiment #2
Luxury hotels	427	4476	Positive to revisit: 98.3% Negative to revisit: 1.7%	Not negative to revisit 45.4% Negative to revisit: 54.6%
	1010			
	262			
	725			
	2052			
Economical hotels	455	1569	Positive to revisit: 97.8% Negative to revisit: 2.2%	Not negative to revisit: 46.5% Negative to revisit: 53.5%
	132			
	226			
	497			
	259			

the class is determined by scores of sentiment. The labels include "positive" and "negative". In luxury hotels, the ratio of positive and negative comments is 98.3%: 1.7%. In economic hotels, the ratio of positive and negative comments is 97.8%: 2.2%. Since it might cause class imbalance problems in experiment #1, we implement experiment #2 by adjusting class distribution. In experiment #2, for balancing class distribution, we re-define class labels as "not negative for revisit" and "negative for revisit". The ratios of "not negative for revisit" and "negative for revisit" are 45.4%:54.6% and 46.5%:53.5%, respectively, for luxury and economic hotels data sets. The detailed information could be found in Table 7.

In order to reduce the error of the experimental results, the input data was normalized before the experiment, and the values were compressed between 0 and 1, and for each classifier. Moreover, to improve the accuracy of experiments, we use the 5-fold cross validation experiment. For implementing feature selection methods, the tools "See5" and "Weka 3.8" have been employed for implementing decision trees and SVM-RFE algorithms, respectively. And LASSO algorithm is programmed in the environment of Matlab.

B. DEFINED FACTORS

This work collected relevant literature on revisit intentions in recent years. Then, we defined 13 potential factors that affected consumer revisit intentions. The defined attributes and their definitions have been summarized in Table 8. Next, for each defined factor, we build a lexicon to determine the value of this factor. In order to compensate for the shortcomings of the adopted words, the thesaurus of the synonyms, synonyms, and antonyms will be found from the dictionary to be our lexicons.

C. RESULTS

Firstly, in experiment #1, this study uses feature selection methods to identify the factors that potentially affect passengers' no-visit, and establishes feature subsets. Then, BPN and

TABLE 8. The defined candidate factors for feature selection.

Notation	Factors	Definitions	Supports
A	Amenity	Service personnel's attitudes, expressions, behaviors, etc. affect the feelings of consumers.	[54]
C	Comfort	Have a feeling of relaxation in the tour.	[56]
CL	Cleanliness	The cleanliness of the facility and attractions.	[56]
EF	Environment facilities	Leisure and relaxation facilities provided by the hotel.	[54], [58]
EM	Emotional words	Words expressed in comments that can reflect consumers' psychological feelings during the vacation experiences.	[35]
ES	Employee services	The service of the hotel employees makes the passengers have a mood.	[55]
FD	Food and drink	Travelers expect the food and drink provided by the attraction is healthy and delicious.	[54], [58]
LA	Location and accessibility	The ease with which passengers arrive at a tourist destination	[53], [54], [55], [57], [58],
P	Price	Price, discount, and promotions.	[51], [56], [59]
R	Recommendation	Travelers share or recommend attractions and hotels to others.	[60], [61]
RF	Room facilities	Availability of related facilities and equipment in the hotels.	[55]
S	Shopping	Consumers can purchase souvenirs and commemorative goods	[33]
T	Trust	Tourist destinations make visitors feel at ease	[62]

SVM classifiers are used to validate the candidate feature subsets. Finally, the feature subsets and the original feature sets will be compared to determine the important non-revisit factors.

1) RESULTS OF EXPERIMENT #1

In results of the DT feature selection for luxury hotels, we can build a feature subset based on occurrence frequency. In table 9, only one factor, "S", is contained in this subset. In LASSO, we merely find one factor "CL" in 2 experiments. The occurrence frequency is less than the average of 5 experiments. Therefore, we cannot find build feature subset in LASSO. Next, we implement SVM-RFE. According to rankings of factors, we can extract feature subsets. For luxury hotels, we extracted 3 feature subsets, including SVM-RFE #1 (3 factors: EM, S, FD), SVM-RFE #2 (6 factors: EM, S, FD, T, LA, R) and SVM-RFE #3 (9 factors: EM, S, FD, T, LA, R, ES, CL, EF).

Then, we are going to evaluate the effectiveness of the extracted feature subsets by using BPN and SVM. Tables 10∼11 show the results of evaluations, compared to

TABLE 9. The extracted feature subsets in DT and LASSO (Luxury hotels).

Fold Method/Factor		1	2	3	4	5	Frequency
DT	X		X	X	X	4	4
LASSO	CL	-1.02E-14	0	0	-1.16E-14	0	2

TABLE 10. Results of BPN evaluation (Luxury hotels).

Feature sets	Original set	DT #1	SVM-RFE #1	SVM-RFE #2	SVM-RFE #3
Factors	13 factors	S	EM,S,FD	ES,S,FD,T,LA,R	ES,S,FD,T,LA,R,ES,CL,EF
Metrics	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
PA (%)	99.28 (0.38)	99.17 (0.30)	99.20 (0.33)	99.19 (0.33)	99.21 (0.29)
NA (%)	13.33 (14.49)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	5.00 (11.18)
GM (%)	98.64 (0.71)	99.17 (0.30)	99.17 (0.30)	98.64 (0.38)	98.66 (0.31)
OA (%)	27.56 (26.59)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	9.94 (22.22)
F1 (%)	99.31 (0.36)	99.58 (0.15)	99.58 (0.15)	99.31 (0.19)	99.32 (0.16)
Time (hh/mm/ss)	00:05:30	00:00:36	00:03:21	00:06:21	00:07:46

TABLE 11. Results of SVM evaluation (Luxury hotels).

Feature sets	Original set	DT #1	SVM-RFE #1	SVM-RFE #2	SVM-RFE #3
Factors	13 factors	S	EM,S,FD	ES,S,FD,T,LA,R	ES,S,FD,T,LA,R,ES,CL,EF
Metrics	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
PA (%)	99.15 (0.34)	99.13 (0.27)	99.20 (0.33)	99.20 (0.33)	99.20 (0.33)
NA (%)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
GM (%)	99.15 (0.34)	99.13 (0.27)	99.20 (0.33)	99.20 (0.33)	99.20 (0.33)
OA (%)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
F1 (%)	99.57 (0.17)	99.56 (0.13)	99.60 (0.17)	99.60 (0.17)	99.60 (0.17)
Time (hh/mm/ss)	00:09:47	00:01:21	00:03:51	00:05:14	00:06:05

original feature set. As shown in these tables, it can be found that whether the results of BPN or SVM have great deviations in the results NA. It means the classifiers cannot detect the minority (negative) examples. We can make a concluding remark that there is a serious class imbalance problem due to unbalanced distribution of class. Under such situation, the extracted factors are useless for luxury hotels dataset.

Next, we do the same procedure for economic hotels dataset. Evaluation results could be summarized in Tables 12∼13. From these two tables, we also found class imbalance problems (very low ability to detect negative examples). Therefore, we redefine class labels to balance class distribution, and implement experiment #2.

TABLE 12. Results of BPN evaluation (Economic hotels).

Feature sets	Original set	SVM-RFE #1	SVM-RFE #2	SVM-RFE #3
Factors	13 factors	EM,S,FD	ES,S,FD,T,LA,R	ES,S,FD,T,LA,R,ES,CL,EF
Metrics	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
PA (%)	97.86 (0.71)	97.83 (0.73)	97.82 (0.72)	97.81 (0.72)
NA (%)	3.33 (7.45)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
GM (%)	95.98 (0.95)	97.83 (0.73)	97.07 (0.76)	96.62 (1.16)
OA (%)	8.07 (18.05)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
F1 (%)	97.95 (0.49)	98.90 (0.37)	98.51 (0.39)	98.28 (0.60)
Time (hh/mm/ss)	00:05:02	00:01:22	00:01:45	00:01:56

TABLE 13. Results of SVM evaluation (Economic hotels).

Feature sets	Original set	SVM-RFE #1	SVM-RFE #2	SVM-RFE #3
Factors	13 factors	EM,S,FD	ES,S,FD,T,LA,R	ES,S,FD,T,LA,R,ES,CL,EF
Metrics	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
PA (%)	97.83 (0.73)	97.83 (0.73)	97.83 (0.73)	97.83 (0.73)
NA (%)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
GM (%)	97.83 (0.73)	97.83 (0.73)	97.83 (0.73)	97.83 (0.73)
OA (%)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
F1 (%)	98.90 (0.37)	98.90 (0.37)	98.90 (0.37)	98.90 (0.37)
Time (hh/mm/ss)	00:03:21	00:00:52	00:01:27	00:02:45
OA (%)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
F1 (%)	98.90 (0.37)	98.90 (0.37)	98.90 (0.37)	98.90 (0.37)
Time (hh/mm/ss)	00:03:21	00:00:52	00:01:27	00:02:45

2) RESULTS OF EXPERIMENT #2

In experiment #2, to avoid class imbalance problems, we redefine our class labels from "Positive, Negative for revisit" to "Not negative to revisit & Negative to revisit". So, we can focus on extracting important factors of non-revisit.

In DT, the number of feature subsets is determined by the fact that the frequency of the factors in the five sub-sets is greater than three. Therefore, from Table 14, three feature

TABLE 14. The extracted feature subsets in DT (Luxury hotels).

Fold / Factors	1	2	3	4	5	Frequency
C	✓	✓	✓	✓	✓	5
EM	✓	✓	✓	✓	✓	5
ES	✓	✓	✓	✓	✓	5
LA	✓	✓	✓	✓	✓	5
P	✓	✓	✓	✓	✓	5
RF	✓	✓	✓	✓	✓	5
T	✓	✓	✓	✓	✓	5
FD		✓	✓	✓	✓	4
CL			✓	✓	✓	3
S	✓				✓	2
EF			✓			1
R					✓	1

TABLE 15. The extracted feature subsets in LASSO (Luxury hotels).

Fold / Factors	1	2	3	4	5	Frequency
CL	0.0783	0.0741	0.2182	0.0888	0.0267	5
EM	0.4777	0.3934	0.1849	0.2248	0.4016	5
ES	0.2383	0.2837	0.2647	0.2737	0.4929	5
FD	0.3808	0.3145	0.2072	0.1320	0.1562	5
LA	1.7585	1.4935	1.5904	1.5732	1.4538	5
P	0.3215	0.3226	0.3031	0.2370	0.2532	5
RF	0.5279	0.8480	0.7619	0.7819	0.9767	5
T	0.3515	0.2817	0.2151	0.2708	0.3840	5
C	0.2493	0.2984	0.1976	0.2533	0	4
EF	0	0	0	0	0	0
R	0	0	0	0	0	0
S	0	0	0	0	0	0
A	0	0	0	0	0	0

subsets can be established. They are DT #1 (7 factors: C, EM, ES, LA, P, RF, T), DT #2 (8 factors: C, EM, ES, LA, P, RF, T, FD) and DT #3 (9 factors: C, EM, ES, LA, P, RF, T, FD, CL).

According to the LASSO compression picking results, the estimated coefficients for each factor under the five subsets are extracted as if the coefficient is not zero. Therefore, as shown in Table 15, we found two LASSO feature subsets, namely LASSO #1 (8 factors: CL, EM, ES, FD, LA, P, RF, T), LASSO #2 (10 factors: CL, EM, ES, FD, LA, P, RF, T, C). As implementing the same procedure of extracting important factors, we can build three feature subsets, namely SVM-RFE #1 (3 factors: LA, T, RF) and SVM-RFE #2 (6 factors: LA, T, RF, P, CL, EM) and SVM-RFE # 3 (10 factors: LA, T, RF, P, CL, EM, ES, A, C, R).

Next, all extracted feature subsets will be evaluated by BPN and SVM. The performance results of DT subset classification are shown in Table 16. It can be found that DT #3 with SVM classifier can outperform other classification results (including BPN and SVM). As a result, the DT #3 is determined to be our best feature set of DT feature selection.

The evaluation results of the LASSO subsets are shown in Table 17. From this table, it can be found that LASSO #2 with BPN classifier has better performance than other feature subsets, when considering OA, GM, and F1. In the part of learning time, we can see SVM is much more time-consuming than BPN in this case. Therefore, this study will select LASSO #2 to be our best feature set.

TABLE 16. Results of BPN and SVM evaluation for DT feature selection (Luxury hotels).

Feature sets	DT #1		DT #2		DT #3	
Factors	C,EM,ES,LA,P,RF,T		C,EM,ES,LA,P,RF,T,FD		C,EM,ES,LA,P,RF,T,FD,CL	
Methods	SVM	BPN	SVM	BPN	SVM	BPN
Metrics	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
PA (%)	62.53 (1.58)	72.19 (2.05)	62.88 (1.30)	72.09 (2.53)	63.93 (2.15)	69.35 (7.91)
NA (%)	72.07 (1.76)	64.28 (1.48)	73.18 (2.76)	63.95 (0.94)	72.98 (1.99)	61.12 (6.42)
GM (%)	67.14 (1.43)	68.36 (1.55)	67.83 (0.92)	68.13 (1.56)	**68.45** (1.67)	65.37 (7.10)
OA (%)	67.12 (1.36)	68.12 (1.52)	67.82 (1.19)	67.89 (1.58)	**68.30** (1.66)	65.11 (7.11)
F1 (%)	67.92 (2.22)	70.12 (1.82)	68.86 (0.53)	69.93 (1.52)	69.89 (1.95)	67.43 (6.67)
Time (hh/mm/ss)	03:26:21	01:20:02	03:48:31	01:25:23	03:43:01	01:28:02

TABLE 17. Results of BPN and SVM evaluation for LASSO feature selection (Luxury hotels).

Feature sets	LASSO #1		LASSO #2	
Factors	CL,EM,ES,FD,LA,P,RF,T		CL,EM,ES,FD,LA,P,RF,T,C	
Methods	SVM	BPN	SVM	BPN
Metrics	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
PA (%)	63.00 (1.51)	71.99 (2.47)	63.93 (2.15)	72.54 (2.70)
NA (%)	73.11 (2.20)	64.04 (1.69)	72.98 (1.99)	64.88 (1.72)
GM (%)	67.92 (1.09)	68.16 (1.85)	68.45 (1.67)	**68.88** (2.00)
OA (%)	67.86 (1.24)	67.89 (1.85)	68.30 (1.66)	**68.60** (2.04)
F1 (%)	69.01 (0.86)	70.08 (1.87)	69.89 (1.95)	**70.85** (1.94)
Time (hh/mm/ss)	04:04:21	01:24:24	04:09:24	02:04:54

Table 18 summarizes evaluation results of SVM-RFE feature subsets. Taking GM, OA, and F1, it can be seen that the SVM-RFE #2 with SVM classifier has better performances than other feature sets. Consequently, this study will use SVM-RFE #2 to be our best LASSO feature set.

Table 19 lists all the comparison between/among extracted feature subsets after doing feature selection in Experiment #2. GM, OA, and F1 of SVM-RFE #3 are slightly inferior to others, however, when considering learning time and the number of utilized features, we can conclude it's our best choice. Based on the results, 6 factors that influence passengers to stop visiting are found. They are LA (location accessibility), T (trust), RF (room facilities), P (price), CL (cleanliness), and EM (emotional words).

TABLE 18. **Results of BPN and SVM evaluation for SVM-RFE feature selection (Luxury hotels).**

Feature sets	SVM-RFE #1		SVM-RFE #2	
Factors	LA,T,RF		LA,T,RF,P,CL,EM	
Methods	SVM	BPN	SVM	BPN
Metrics	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
PA (%)	63.21 (1.91)	68.24 (1.87)	63.96 (2.12)	72.04 (2.07)
NA (%)	71.92 (3.13)	66.56 (2.61)	72.97 (2.38)	64.32 (2.35)
GM (%)	67.61 (2.02)	67.45 (2.09)	**68.50** (2.00)	68.34 (1.88)
OA (%)	67.41 (2.07)	67.39 (2.12)	**68.31** (2.02)	68.07 (1.91)
F1 (%)	69.29 (1.83)	68.69 (1.84)	**69.98** (1.96)	70.33 (1.77)
Time (hh/mm/ss)	00:09:47	00:01:24	00:16:32	00:01:51

TABLE 19. **Comparisons between original feature sets and extracted feature sets (Luxury hotels).**

Feature sets	Original set		DT #3	LASSO #2	SVM-RFE #2
Factors	13 factors		C,EM,ES,LA,P,RF,T,FD,CL	CL,EM,ES,FD,LA,P,RF,T,C	LA,T,RF,P,CL,EM
Methods	SVM	BPN	SVM	BPN	SVM
Metrics	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
PA (%)	72.15 (2.23)	63.47 (1.83)	63.93 (2.15)	72.54 (2.70)	63.96 (2.12)
NA (%)	63.25 (1.45)	74.07 (2.69)	72.98 (1.99)	64.88 (1.72)	72.97 (2.38)
GM (%)	67.74 (1.31)	68.54 (1.42)	68.45 (1.67)	68.88 (2.00)	**68.50** (2.00)
OA (%)	67.54 (1.46)	68.55 (1.48)	68.30 (1.66)	68.60 (2.04)	**68.31** (2.02)
F1 (%)	69.34 (0.99)	69.45 (1.58)	69.89 (1.95)	70.85 (1.94)	**69.98** (1.96)
Time (hh/mm/ss)	00:30:24	05:24:11	03:43:01	02:04:54	**00:16:32**

Following the same procedure for luxury hotels data, Table 15 shows the evaluation results of selected feature sets for economic hotels. As shown in the table, it can be found that the LASSO #2 with SVM classifier has the best classification performance in all candidate feature sets. F1 and learning time are also ranked as the 2nd place. Therefore, in economical hotels, LASSO #2 is our best feature set. Consequently, 8 important factors can be determined based on the results. Factors affecting travelers no longer visiting economic hotels are C (comfort), EM (emotional words), ES (employee service), FD (diet), LA (positional accessibility), P (price), S (Shopping), and CL (Cleanliness).

D. DISCUSSIONS

Based on results of experiment #2, we can identify 6 and 8 important factors that will affect passengers no longer visit for luxury hotels, and economic hotels, respectively. Table 21 provides the comparison of selected

TABLE 20. **Comparisons between original feature sets and extracted feature sets (Economic hotels).**

Feature sets	Original set		DT #3	LASSO #2	SVM-RFE #3
Factors	13 factors		EM,C,ES,LA,FD	C,EM,ES,FD,LA,P,S,CL	LA,P,T,A,EF,FD,EM,ES,RF,S
Methods	SVM	BPN	BPN	SVM	SVM
Metrics	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
PA (%)	66.35 (3.58)	63.64 (2.98)	71.12 (4.52)	63.49 (2.60)	62.33 (1.10)
NA (%)	59.47 (4.65)	73.93 (2.42)	63.10 (1.62)	75.49 (1.97)	73.55 (4.43)
GM (%)	62.78 (3.83)	68.20 (1.55)	66.86 (1.92)	**68.64** (1.21)	67.24 (1.77)
OA (%)	62.79 (3.72)	68.55 (1.26)	66.95 (2.02)	**69.20** (1.02)	67.68 (1.99)
F1 (%)	63.77 (4.44)	67.74 (2.25)	67.65 (1.69)	67.72 (1.76)	66.53 (1.89)
Time (hh/mm/ss)	00:03:14	00:35:02	00:01:25	00:23:52	00:27:11

TABLE 21. **Selected important factors affecting visitors' non-revisit intentions.**

Luxury hotels		Economic hotels	
Notation	Important factors	Notation	Important factors
LA	Location and accessibility	LA	Location and accessibility
P	Price	ES	Employee services
T	Trust	FD	Food and drink
CL	Cleanliness	P	Price
EM	Emotional words	CL	Cleanliness
RF	Room facilities	C	Comfort
		EM	Emotion words
		S	Shopping

factors in luxury and economic hotels. 4 factors, including "Location and accessibility", "Price", "Cleanliness" and "Emotional words" are important no matter in luxury or economic hotels. These 4 factors can be considered as basic features those influencing tourists' non-revisit behaviors. In addition, "Trust" and "Room facilities" are considered as crucial factors for luxury hotels. Customers who live in luxury hotels pay much more expense than economic hotels. So, they will care about related facilities and equipment, such as SPA, swimming pool, gym and so on in hotels. Besides, they might pay the higher fee on internet. So, they should trust how much they paid can transform into what they receive. Once the trust factor doesn't exist, they won't book or revisit a specific luxury hotel again.

Moreover, "Employee services", "Food and drink", "Comfort", and "Shopping" have been viewed as important for economic hotels. Customers who choose economic hotels might think cost-price ratio is very important for them. They want to receive more than how much they paid. Therefore, they will evaluate one hotel by reviewing their employee services, provided food and drink, feelings of comfort, and shopping environment. These factors usually are guaranteed to be an acceptable level in luxury hotels.

V. CONCLUSION

Emotions are important for users' decision making in social media. Due to its objective expression, the visitors' comments have the subjective sense that affects other visitors who no longer visit. It is very difficult to identify important factors that reduce passengers' non-revisit intention in huge amount of textual reviews. Therefore, the purpose of this study is to discover the factors that passengers do not visit again from textual comments in social media. First, according to the tourism industry related literatures, 13 potential factors regrading textual comments have been defined. Next, opinion lexicon has been built based on the definition of the factors, and then we can use text mining technique to process the collected comments in social media. Three feature selection methods including Decision Tree (DT), Least Absolute Compression Sampling (LASSO), and Support Vector Machine Sequential Feature Removal (SVM-RFE) have been employed to select the important factors.

The experimental results of this study could be summarized as follows. Firstly, it can be found that "location and accessibility", "price", "cleanliness", and "emotion word" affect the traveler's future non-revisit intention no matter for luxury or economic restaurants. Secondly, for different types of hotels, "trusts" and "room facilities" are crucial factors that influence travelers not to visit luxury hotels. But, on the other hand, "employee services", "food and drink", "comfort", "shopping", will affect the travelers' non-revisit behaviors for economic hotels. Finally, when using the same data for two classifiers (BPN and SVM) training, we found that the results of the two classifications are almost the same, but the SVM classifier has a better performance in the negative examples (negative for revisit), and the BPN classifier is in the positive examples (positive for revisit). And, the time required for BPN to perform training is much less than SVM.

In addition, for different types of hotel administrators in Hong Kong, the experimental results of this study can be used as a reference basis to adjust the internal management of the hotel and the improvement of service quality. Administrators of different types of hotels should in particular strengthen different factors to make improvements to increase market share. According to different factors, this study provides relevant suggestions to reduce the non-revisit intention in the future. The recommendations have been made in this study are summarized in Table 22.

Finally, in order to highlight the contribution of our proposed text mining based method, we made a comparison table as below. In addition, the contribution of this work can be explained by social, economic and academic aspects. In terms of social impact, this study used text comments in social media to identify important factors that affect tourism customers' revisit intentions. It can provide a broad reference for consumers. In terms of economic impact, this study allows hotels companies to quickly process large customer comments, reduce data collection and pro-

TABLE 22. Discovered important factors for non-revisit and suggestions of reducing non-revisit intention for hotels.

Type of hotels	Non-revisit factors	Suggestions for hotel managers
Luxury and economic hotels	Location and accessibility	Provide customers clear guidance of routes or arrange transportation vehicles to pick up passengers directly from stations, ports, and airports, including round trip plan.
	Price	Plan scheduled sales or promotions depending on seasons or festivals.
	Cleanliness	Employees should always pay attention to the facility environment. Hotels can hire professional cleaning personnel to clean the facilities.
	Emotion word	Hotels internet/website management personnel should be able to identify "negative emotional word" in customers' comments, and carefully respond to them timely, and maintain good interaction with social media users.
Luxury hotels	Trust	Hotels should build customers' trust from daily operation and long-term relation maintenance. Hotels also have good interactive relationship to find the needs of customers.
	Room facilities	In addition to the basic room facilities, some unexpected/attractive facilities and equipment such as massage chairs and game machines (VR/AR, Nintendo Switch, and so on) can be added.
Economic hotels	Employee services	Hotels employees should have good training in providing services, so they can anytime and anyway pay attention to travelers, and actively ask customer needs.
	Food and drink	Provide excellent food and drink to exceed customers' expectation. Hotel can inquire about customers' dietary preference before meals, and investigate after meals for customers to understand their needs.
	Comfort	Hotels can provide additional tools or services to increase the comfort of accommodation, such as SPA, free Wi-Fi, hair dryers, free laundry services, ATM, and so on.
	Shopping	Hotels can cooperate with local manufacturers/farmers to provide native products such as local gifts, agricultural products, specialty, and souvenirs for tourists.

TABLE 23. Comparison between the proposed text mining based method and traditional works in studying revisit intentions.

Comparisons	Conventional method	Our method
Employed methods	Multivariate statistics/ Structural Equation Modelling	Text mining/ data mining
Data collection tools	Questionnaires	Reviews/ comments
Data collection time length	Long	Short
Customer response update	Several months	Immediately

cessing time, and provide recommendations for improving service quality. In terms of academic development, this study proposed a text-mining based evaluation system that not only can instantly understand the factors that customers no longer visit, but also can improve the sampling bias generated by questionnaires.

Since this work is a pioneer work in related domain, our purpose to prove it's feasible and prove it could work in real world. We demonstrate the proposed method not only can achieve the same goal, but also have better performances, compared to conventional methods. Therefore, for possible directions of future works, readers could focus on presenting better models based our work. For examples, different language processing or feature selection methods could be employed to find the optimal performance.

REFERENCES

[1] (2016). *WTTC*. [Online]. Available: https://www.wttc.org/-/media/files/reports/economic%20impact%20research/regions%202016/world2016.pdf

[2] UNWTO. (2017). *UNWTO Tourism Highlights*. [Online]. Available: http://www.e-unwto.org/doi/pdf/10.18111/9789284419029

[3] (2017). *Statista*. [Online]. Available: https://www.statista.com/statistics/233223/travel-and-tourism–total-economic-contribution-worldwide/

[4] (2017). *Statista*. [Online]. Available: https://www.statista.com/topics/962/global-tourism/

[5] E. Raguseo, P. Neirotti, and E. Paolucci, "How small hotels can drive value their way in infomediation. The case of 'Italian hotels vs. OTAs and TripAdvisor,'" *Inf. Manage.*, vol. 54, no. 6, pp. 745–756, 2017.

[6] S. O. Lyu, "Travel selfies on social media as objectified self-presentation," *Tourism Manage.*, vol. 54, pp. 185–195, Jun. 2016.

[7] H. Daniszewski. (2012). *Smartphones Changing Tourism Business—Smartphones the Latest Travel Tool*. [Online]. Available: http://www.lfpress.com/2012/10/24/smartphones-changing-ontario-tourism-business

[8] M. Meleddu and M. Pulina, "Evaluation of individuals' intention to pay a premium price for ecotourism: An exploratory study," *J. Behav. Exp. Econ.*, vol. 65, pp. 67–78, Dec. 2016.

[9] W.-K. Tan and C.-E. Wu, "An investigation of the relationships among destination familiarity, destination image and future visit intention," *J. Destination Marketing Manage.*, vol. 5, no. 3, pp. 214–226, 2016.

[10] D. Stylidis, A. Shani, and Y. Belhassen, "Testing an integrated destination image model across residents and tourists," *Tourism Manage.*, vol. 58, pp. 184–195, Feb. 2017.

[11] M. Yadav and Z. Rahman, "Measuring consumer perception of social media marketing activities in e-commerce industry: Scale development & validation," *Telematics Inform.*, vol. 34, no. 7, pp. 1294–1307, 2017.

[12] K. Zhao, A. C. Stylianou, and Y. Zheng, "Sources and impacts of social influence from online anonymous user reviews," *Inf. Manage.*, vol. 55, no. 1, pp. 16–30, 2018.

[13] X. Shi and Z. Liao, "Online consumer review and group-buying participation: The mediating effects of consumer beliefs," *Telematics Inform.*, vol. 34, no. 5, pp. 605–617, 2017.

[14] D. A. Fennell, "Towards a model of travel fear," *Ann. Tourism Res.*, vol. 66, pp. 140–150, Sep. 2017.

[15] A. Kumar and T. Gilovich, "To do or to have, now or later? The preferred consumption profiles of material and experiential purchases," *J. Consum. Psychol.*, vol. 26, no. 2, pp. 169–178, 2016.

[16] A. P. M. Som, A. Marzuki, M. Yousefi, and A. N. AbuKhalifeh, "Factors influencing visitors' revisit behavioral intentions: A case study of Sabah, Malaysia," *Int. J. Marketing Stud.*, vol. 4, no. 4, pp. 39–50, 2012.

[17] L.-L. Chang and K. F. Backman. (2016). *An Investigation of Creative Tourists' Experience and Revisit Intention*. Travel and Tourism Research Association: Advancing Tourism Research Globally. 23. [Online]. Available: https://scholarworks.umass.edu/ttra/2012/Visual/23

[18] T. Che, Z. Peng, K. H. Lim, and Z. Hua, "Antecedents of consumers' intention to revisit an online group-buying Website: A transaction cost perspective," *Inf. Manage.*, vol. 52, no. 5, pp. 588–598, 2015.

[19] C. Liao, C. Feng, S. Yang, and H. Huang, "Topic-related Chinese message sentiment analysis," *Neurocomputing*, vol. 210, pp. 237–246, Oct. 2016.

[20] Q. Ye, R. Law, B. Gu, and W. Chen, "The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings," *Comput. Hum. Behav.*, vol. 27, no. 2, pp. 634–639, 2011.

[21] D. Schellong, J. Kemper, and M. Brettel, "Clickstream data as a source to uncover con-sumer shopping types in a large-scale online setting," in *Proc. ECIS*, 2016, pp. 1–15.

[22] S. A. Lee and H. Oh, "Sharing travel stories and behavioral outcomes: A case of travel," *Tourism Manage.*, vol. 62, pp. 147–158, Oct. 2017.

[23] B. A. Sparks, K. K. F. So, and G. L. Bradley, "Responding to negative online reviews: The effects of hotel responses on customer inferences of trust and concern," *Tourism Manage.*, vol. 53, pp. 74–85, Apr. 2016.

[24] M. A. Lapré, "Reducing customer dissatisfaction: How important is learning to reduce service failure?" *Prod. Oper. Manage.*, vol. 20, no. 4, pp. 491–507, 2011.

[25] C. W. Hart, J. L. Heskett, and W. E. Sasser, Jr., "The profitable art of service recovery," *Harvard Bus. Rev.*, vol. 68, no. 4, pp. 148–156, 1990.

[26] M. J. Kim, C.-K. Lee, and M. Bonn, "The effect of social capital and altruism on seniors' revisit intention to social network sites for tourism-related purposes," *Tourism Manage.*, vol. 53, pp. 96–107, 2016.

[27] C.-J. Hsu, J.-R. Yen, Y.-C. Chang, and H. K. Woon, "How do the services of low cost carriers affect passengers' behavioral intentions to revisit a destination?" *J. Air Transp. Manage.*, vol. 52, pp. 111–116, Apr. 2016.

[28] L. T. I. Loi, A. S. I. So, I. S. Lo, and L. H. N. Fong, "Does the quality of tourist shuttles influence revisit intention through destination image and satisfaction? The case of Macao," *J. Hospitality Tourism Manage.*, vol. 32, pp. 115–123, Sep. 2017.

[29] H. J. Kim, J. Park, M.-J. Kim, and K. Ryu, "Does perceived restaurant food healthiness matter? Its influence on value, satisfaction and revisit intentions in restaurant operations in South Korea," *Int. J. Hospitality Manage.*, vol. 33, pp. 397–405, Jun. 2013.

[30] N. Y. Jung, S. Kim, and S. Kim, "Influence of consumer attitude toward online brand community on revisit intention and brand trust," *J. Retailing Consum. Services*, vol. 21, no. 4, pp. 581–589, 2014.

[31] M. Meleddu, R. Paci, and M. Pulina, "Repeated behaviour and destination loyalty," *Tourism Manage.*, vol. 50, pp. 159–171, Oct. 2015.

[32] Z. Liu and S. Park, "What makes a useful online review? Implication for travel product Websites," *Tourism Manage.*, vol. 47, pp. 140–151, Apr. 2015.

[33] X. Xu and Y. Li, "The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach," *Int. J. Hospitality Manage.*, vol. 55, pp. 57–69, May 2016.

[34] K. Kim, O.-J. Park, S. Yun, and H. Yun, "What makes tourists feel negatively about tourism destinations? Application of hybrid text mining methodology to smart destination management," *Technol. Forecasting Social Change*, vol. 123, pp. 362–369, Oct. 2017.

[35] E. T. Kabadayı and A. K. Alan, "Revisit intention of consumer electronics retailers: Effects of customers' emotion, technology orientation and WOM influence," *Procedia-Social Behav. Sci.*, vol. 41, pp. 65–73, Jan. 2012.

[36] M. D. Molina-González, E. Martínez-Cámara, M. T. Martín-Valdivia, and J. M. Perea-Ortega, "Semantic orientation for polarity classification in Spanish reviews," *Expert Syst. Appl.*, vol. 40, no. 18, pp. 7250–7257, 2013.

[37] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," *Knowl.-Based Syst.*, vol. 89, pp. 14–46, Nov. 2015.

[38] F. Kou, J. Du, Y. He, and L. Ye, "Social network search based on semantic analysis and learning," *CAAI Trans. Intell. Technol.*, vol. 1, no. 4, pp. 293–302, 2016.

[39] M. Kozak, "Repeaters' behavior at two distinct destinations," *Ann. Tourism Res.*, vol. 28, no. 3, pp. 784–807, 2001.

[40] T. Kim, W. G. Kim, and H.-B. Kim, "The effects of perceived justice on recovery satisfaction, trust, word-of-mouth, and revisit intention in upscale hotels," *Tourism Manage.*, vol. 30, no. 1, pp. 51–62, 2009.

[41] J. F. Engel, R. D. Blackwell, and P. W. Miniard, *Consumer Behavior*, 8th ed. Fort Worth, TX, USA: Dryden Press, 1995, p. 237.

[42] A. R. Julaimi, S. A. Talib, and M. Z. Suhaimi, "International tourists revisit intention: A case of the United Arab Emirates," *J. Tourism, Hospitality Culinary Arts*, vol. 8, no. 1, pp. 35–42, 2016.

[43] W. J. Phillips, K. Wolfe, N. Hodur, and F. L. Leistritz, "Tourist word of mouth and revisit intentions to rural tourism destinations: A case of North Dakota, USA," *Int. J. Tourism Res.*, vol. 15, no. 1, pp. 93–104, 2013.

[44] A. S. Cantallops and F. Salvi, "New consumer behavior: A review of research on eWOM and hotels," *Int. J. Hospitality Manage.*, vol. 36, pp. 41–51, Jan. 2014.

[45] E. M. Artigas, C. C. Yrigoyen, E. T. Moraga, and C. B. Villalón, "Determinants of trust towards tourist destinations," *J. Destination Marketing Manage.*, vol. 6, no. 4, pp. 327–334, 2017.

[46] S. H. M. Taher, S. A. Jamal, N. Sumarjan, and N. Aminudin, "Examining the structural relations among hikers' assessment of pull-factors, satisfaction and revisit intentions: The case of mountain tourism in Malaysia," *J. Outdoor Recreation Tourism*, vol. 12, pp. 82–88, Dec. 2015.

[47] H. Zhang, Y. Wu, and D. Buhalis, "A model of perceived image, memorable tourism experiences and revisit intention," *J. Destination Marketing Manage.*, vol. 8, pp. 326–336, Jun. 2018.

[48] C. H. S. Liu and T. Lee, "Service quality and price perception of service: Influence on word-of-mouth and revisit intention," *Journal Air Transp. Manage.*, vol. 52, pp. 42–54, Apr. 2016.

[49] N. Stylos, V. Bellou, A. Andronikidis, and C. A. Vassiliadis, "Linking the dots among destination images, place attachment, and revisit intentions: A study among British and Russian tourists," *Tourism Manage.*, vol. 60, pp. 15–29, Jun. 2017.

[50] B. J. Knutson, J. A. Beck, S. Kim, and J. Cha, "Identifying the dimensions of the guest's hotel experience," *Cornell Hospitality Quart.*, vol. 50, no. 1, pp. 44–55, 2009.

[51] L. Ren, H. Qiu, P. Wang, and P. M. C. Lin, "Exploring customer experience with budget hotels: Dimensionality and satisfaction," *Int. J. Hospitality Manage.*, vol. 52, pp. 13–23, Jan. 2016.

[52] W.-K. Tan, "Repeat visitation: A study from the perspective of leisure constraint, tourist experience, destination images, and experiential familiarity," *J. Destination Marketing Manage.*, vol. 6, no. 3, pp. 233–242, 2017.

[53] H. Han and Y. Kim, "An investigation of green hotel customers' decision formation: Developing an extended model of the theory of planned behavior," *Int. J. Hospitality Manage.*, vol. 29, no. 4, pp. 659–668, 2010.

[54] H. Han and S. S. Hyun, "Impact of hotel-restaurant image and quality of physical-environment, service, and food on satisfaction and intention," *Int. J. Hospitality Manage.*, vol. 63, pp. 82–92, May 2017.

[55] J. Peng, X. Zhao, and A. S. Mattila, "Improving service management in budget hotels," *Int. J. Hospitality Manage.*, vol. 49, pp. 139–148, Aug. 2015.

[56] H. Gu and C. Ryan, "Chinese clientele at Chinese hotels—Preferences and satisfaction," *Int. J. Hospitality Manage.*, vol. 27, no. 3, pp. 337–345, 2008.

[57] A. S. Mattila, "Emotional bonding and restaurant loyalty," *Cornell Hotel Restaurant Admin. Quart.*, vol. 42, no. 6, pp. 73–79, 2001.

[58] S. Um, K. Chon, and Y. Ro, "Antecedents of revisit intention," *Ann. Tourism Res.*, vol. 33, no. 4, pp. 1141–1158, 2006.

[59] K. Ryu and H. Han, "Influence of the quality of food, service, and physical environment on customer satisfaction and behavioral intention in quick-casual restaurants: Moderating role of perceived price," *J. Hospitality Tourism Res.*, vol. 34, no. 3, pp. 310–329, 2010.

[60] J. E. Bigné, M. I. Sánchez, and J. Sánchez, "Tourism image, evaluation variables and after purchase behaviour: Inter-relationship," *Tourism Manage.*, vol. 22, no. 6, pp. 607–616, 2001.

[61] T. K. Hui, D. Wan, and A. Ho, "Tourists' satisfaction, recommendation and revisiting Singapore," *Tourism Manage.*, vol. 28, no. 4, pp. 965–975, 2007.

[62] E. Garbarino and M. S. Johnson, "The different roles of satisfaction, trust, and commitment in customer relationships," *J. Marketing*, vol. 63, no. 2, pp. 70–87, 1999.

[63] T. H. Rashidi, A. Abbasi, M. Maghrebi, S. Hasan, and T. S. Waller, "Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges," *Transp. Res. C, Emerg. Technol.*, vol. 75, pp. 197–211, Feb. 2017.

[64] A. M. Munar and J. K. S. Jacobsen, "Motivations for sharing tourism experiences through social media," *Tourism Manage.*, vol. 43, pp. 46–54, Aug. 2014.

[65] L. Chen, T. Jiang, W. Li, S. Geng, and S. Hussain, "Who should pay for online reviews? Design of an online user feedback mechanism," *Electron. Commerce Res. Appl.*, vol. 23, pp. 38–44, May/Jun. 2017.

[66] S. S. Sohail, J. Siddiqui, and R. Ali, "Feature extraction and analysis of online reviews for the recommendation of books using opinion mining technique," *Perspect. Sci.*, vol. 8, pp. 754–756, Sep. 2016.

[67] B. Bhole and B. Hanna, "The effectiveness of online reviews in the presence of self-selection bias," *Simul. Model. Pract. Theory*, vol. 77, pp. 108–123, Sep. 2017.

[68] I. Cenni and P. Goethals, "Negative hotel reviews on TripAdvisor: A cross-linguistic analysis," *Discourse, Context Media*, vol. 16, pp. 22–30, Apr. 2017.

[69] C. Vásquez, "Complaints online: The case of TripAdvisor," *J. Pragmatics*, vol. 43, no. 6, pp. 1707–1717, 2011.

[70] S. Amaro, P. Duarte, and C. Henriques, "Travelers' use of social media: A clustering approach," *Ann. Tourism Res.*, vol. 59, pp. 1–15, Jul. 2016.

[71] A. M. Abubakar, M. Ilkan, R. M. Al-Tal, and K. K. Eluwole, "eWOM, revisit intention, destination trust and gender," *J. Hospitality Tourism Manage.*, vol. 31, pp. 220–227, Jun. 2017.

[72] F. Bronner and R. De Hoog, "Vacationers and eWOM: Who posts, and why, where, and what?" *J. Travel Res.*, vol. 50, no. 1, pp. 15–26, 2011.

[73] I. Erkan and C. Evans, "The influence of eWOM in social media on consumers' purchase intentions: An extended approach to information adoption," *Comput. Hum. Behav.*, vol. 61, pp. 47–55, Aug. 2016.

[74] I. Sánchez-García and R. Currás-Pérez, "Effects of dissatisfaction in tourist services: The role of anger and regret," *Tourism Manage.*, vol. 32, no. 6, pp. 1397–1406, 2011.

[75] S. Banerjee and A. Y. K. Chua, "In search of patterns among travellers' hotel ratings in TripAdvisor," *Tourism Manage.*, vol. 53, pp. 125–131, Apr. 2016.

[76] A. Ganzaroli, I. De Noni, and P. van Baalen, "Vicious advice: Analyzing the impact of TripAdvisor on the quality of restaurants as part of the cultural heritage of Venice," *Tourism Manage.*, vol. 61, pp. 501–510, Aug. 2017.

[77] G. Tontini, G. dos Santos Bento, T. C. Milbratz, B. K. Volles, and D. Ferrari, "Exploring the nonlinear impact of critical incidents on customers' general evaluation of hospitality services," *Int. J. Hospitality Manage.*, vol. 66, pp. 106–116, Sep. 2017.

[78] A. F. Y. Chao and H.-L. Yang, "Using Chinese radical parts for sentiment analysis and domain-dependent seed set extraction," *Comput. Speech Lang.*, vol. 47, pp. 194–213, Jan. 2018.

[79] D. Sullivan, *Document Warehousing and Text Mining: Techniques for Improving Business Operations, Marketing, and Sales*. Hoboken, NJ, USA: Wiley, 2001.

[80] G. M. Thomaz, A. A. Biz, E. M. Bettoni, L. Mendes-Filho, and D. Buhalis, "Content mining framework in social media: A FIFA world cup 2014 case analysis," *Inf. Manage.*, vol. 54, no. 6, pp. 786–801, 2017.

[81] U. Godnov and T. Redek, "Application of text mining in tourism: Case of Croatia," *Ann. Tourism Res.*, vol. 58, no. c, pp. 162–166, 2016.

[82] C. U. I. Wong and S. Qi, "Tracking the evolution of a destination's image by text-mining online reviews—The case of Macau," *Tourism Manage. Perspect.*, vol. 23, pp. 19–29, Jul. 2017.

[83] Y.-H. Hu, Y.-L. Chen, and H.-L. Chou, "Opinion mining from online hotel reviews—A text summarization approach," *Inf. Process. Manage.*, vol. 53, no. 2, pp. 436–449, 2017.

[84] M. Schuckert, X. Liu, and R. Law, "A segmentation of online reviews by language groups: How English and non-English speakers rate hotels differently," *Int. J. Hospitality Manage.*, vol. 48, pp. 143–149, Jul. 2015.

[85] X. Sun, Y. Zhang, T. Matsuzaki, Y. Tsuruoka, and J. I. Tsujii, "Probabilistic Chinese word segmentation with non-local information and stochastic training," *Inf. Process. Manage.*, vol. 49, no. 3, pp. 626–636, 2013.

[86] Z. Zhang, Q. Ye, Z. Zhang, and Y. Li, "Sentiment classification of Internet restaurant reviews written in Cantonese," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7674–7682, 2011.
[87] J. Huang, "The dining experience of Beijing roast duck: A comparative study of the Chinese and English online consumer reviews," *Int. J. Hospitality Manage.*, vol. 66, pp. 117–129, Sep. 2017.
[88] V. K. Jain, S. Kumar, and S. L. Fernandes, "Extraction of emotions from multilingual text using intelligent text processing and computational linguistics," *J. Comput. Sci.*, vol. 21, pp. 316–326, Jul. 2017.
[89] T. Hamed, R. Dara, and S. C. Kremer, "Network intrusion detection system based on recursive feature addition and bigram technique," *Comput. Secur.*, vol. 73, pp. 137–155, Mar. 2018.
[90] B. Liu, "Sentiment analysis and subjectivity," in *Handbook of Natural Language Processing*, vol. 2. London, U.K.: Chapman & Hall, 2010, pp. 627–666.
[91] F. Ali, D. Kwak, P. Khan, S. M. R. Islam, K. H. Kim, and K. S. Kwak, "Fuzzy ontology-based sentiment analysis of transportation and city feature reviews for safe traveling," *Transp. Res. C, Emerg. Technol.*, vol. 77, pp. 33–48, Apr. 2017.
[92] P. G. Preethi, V. Uma, and A. Kumar, "Temporal sentiment analysis and causal rules extraction from tweets for event prediction," *Procedia Comput. Sci.*, vol. 48, pp. 84–89, Jan. 2015.
[93] R. Ferreira, R. D. Lins, S. J. Simske, F. Freitas, and M. Riss, "Assessing sentence similarity through lexical, syntactic and semantic analysis," *Comput. Speech Lang.*, vol. 39, pp. 1–28, Sep. 2016.
[94] R. Gaspar, C. Pedro, P. Panagiotopoulos, and B. Seibt, "Beyond positive or negative: Qualitative sentiment analysis of social media reactions to unexpected stressful events," *Comput. Hum. Behav.*, vol. 56, pp. 179–191, Mar. 2016.
[95] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, nos. 1–2, pp. 1–135, 2008.
[96] M. D. Molina-González, E. Martínez-Cámara, M. T. Martín-Valdivia, and L. A. Ureña-López, "A Spanish semantic orientation approach to domain adaptation for polarity classification," *Inf. Process. Manage.*, vol. 51, no. 4, pp. 520–531, 2015.
[97] D. Li, M. A. Yong-Tao, and J.-L. Guo, "Words semantic orientation classification based on HowNet," *J. China Univ. Posts Telecommun.*, vol. 16, no. 1, pp. 106–110, 2009.
[98] L. Garcia-Moya, H. Anaya-Sanchez, and R. Berlanga-Llavori, "Retrieving product features and opinions from customer reviews," *IEEE Intell. Syst.*, vol. 28, no. 3, pp. 19–27, May 2013.
[99] J. M. Chenlo and D. E. Losada, "An empirical study of sentence features for subjectivity and polarity classification," *Inf. Sci.*, vol. 280, pp. 275–288, Oct. 2014.
[100] G. Vinodhini and R. M. Chandrasekaran, "Sentiment analysis and opinion mining: A survey," *Int. J.*, vol. 2, no. 6, pp. 282–292, 2012.
[101] R. Prabowo and M. Thelwall, "Sentiment analysis: A combined approach," *J. Informetrics*, vol. 3, no. 2, pp. 143–157, 2009.
[102] M. Dash and H. Liu, "Feature selection for classification," *Intell. Data Anal.*, vol. 1, nos. 1–4, pp. 131–156, 1997.
[103] A. J. Kim and K. P. Johnson, "Power of consumers using social media: Examining the influences of brand-related user-generated content on Facebook," *Comput. Hum. Behav.*, vol. 58, pp. 98–108, May 2016.
[104] S. H. Han, S. X. Lu, and S. C. H. Leung, "Segmentation of telecom customers based on customer value by decision tree model," *Expert Syst. Appl.*, vol. 39, no. 4, pp. 3964–3973, 2012.
[105] F.-H. Chen, D.-J. Chi, and Y.-C. Wang, "Detecting biotechnology industry's earnings management using Bayesian network, principal component analysis, back propagation neural network, and decision tree," *Econ. Model.*, vol. 46, pp. 1–10, Apr. 2015.
[106] S. Moro, P. Rita, and B. Vala, "Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach," *J. Bus. Res.*, vol. 69, no. 9, pp. 3341–3351, 2016.
[107] D. Perez-Alonso, S. Peña-Tejedor, M. Navarro, C. Rad, Á. Arnaiz-González and J. F. Díez-Pastor, "Decision trees for the prediction of environmental and agronomic effects of the use of compost of sewage slugde (CSS)," *Sustain. Prod. Consumption*, vol. 12, pp. 119–133, Oct. 2017.
[108] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc., B (Methodol.)*, vol. 58, no. 1, pp. 267–288, 1996.
[109] S. Wang, B. Ji, J. Zhao, W. Liu, and T. Xu, "Predicting ship fuel consumption based on LASSO regression," *Transp. Res. D, Transp. Environ.*, vol. 65, pp. 817–824, Dec. 2018.
[110] Z. B. Yan and Y. Yao, "Variable selection method for fault isolation using least absolute shrinkage and selection operator (LASSO)," *Chemometrics Intell. Lab. Syst.*, vol. 146, pp. 136–146, Aug. 2015.
[111] I. Kamkar, S. K. Gupta, D. Phung, and S. Venkatesh, "Stable feature selection for clinical prediction: Exploiting ICD tree structure using Tree-Lasso," *J. Biomed. Inform.*, vol. 53, pp. 277–290, Feb. 2015.
[112] P. A. Gauthier, W. Scullion, and A. Berry, "Sound quality prediction based on systematic metric selection and shrinkage: Comparison of stepwise, lasso, and elastic-net algorithms and clustering preprocessing," *J. Sound Vib.*, vol. 400, pp. 134–153, Jul. 2017.
[113] O. Frank, B. Brors, A. Fabarius, L. Li, M. Haak, S. Merk, U. Schwindel, C. Zheng, M. C. Müller, N. Gretz, R. Hehlmann, A. Hochhaus, and W. Seifarth, "Gene expression signature of primary imatinib-resistant chronic myeloid leukemia patients," *Leukemia*, vol. 20, no. 8, pp. 1400–1407, 2006.
[114] S. Korkmaz, G. Zararsiz, and D. Goksuluk, "Drug/nondrug classification using support vector machines with various feature selection strategies," *Comput. Methods Programs Biomed.*, vol. 117, no. 2, pp. 51–60, 2014.
[115] R. Stoean and C. Stoean, "Modeling medical decision making by support vector machines, explaining by rules of evolutionary algorithms with feature selection," *Expert Syst. Appl.*, vol. 40, no. 7, pp. 2677–2686, 2013.
[116] A. R. Hidalgo-Muñoz, M. M. López, I. M. Santos, A. T. Pereira, M. Vázquez-Marrufo, A. Galvao-Carmona, and A. M. Tomé, "Application of SVM-RFE on EEG signals for detecting the most relevant scalp regions linked to affective valence processing," *Expert Syst. Appl.*, vol. 40, no. 6, pp. 2102–2108, 2013.
[117] Z. Shao, S. Yang, F. Gao, K. Zhou, and P. Lin, "A new electricity price prediction strategy using mutual information-based SVM-RFE classification," *Renew. Sustain. Energy Rev.*, vol. 70, pp. 330–341, Apr. 2017.
[118] S. Li, Z. Zhao, T. Liu, R. Hu, and X. Du, "Initializing convolutional filters with semantic features for text classification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 1884–1889.
[119] M. Shirakawa, T. Hara, and T. Maekawa, "Never abandon minorities: Exhaustive extraction of bursty phrases on microblogs using set cover problem," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 2358–2367.
[120] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 3111–3119.
[121] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2004, pp. 168–177.
[122] P. D. Turney, "Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews," in *Proc. 40th Annu. Meeting Assoc. Comput. Linguistics. Assoc. Comput. Linguistics*, 2002, pp. 417–424.
[123] J. P. Singh, S. Irani, N. P. Rana, Y. K. Dwivedi, S. Saumya, and P. K. Roy, "Predicting the 'helpfulness' of online consumer reviews," *J. Bus. Res.*, vol. 70, pp. 346–355, 2017.
[124] W. Wei, L. Miao, and Z. J. Huang, "Customer engagement behaviors and hotel responses," *Int. J. Hospitality Manage.*, vol. 33, pp. 316–330, Jun. 2013.
[125] C. Vásquez, "'Usually not one to complain but...': Constructing identities in user-generated online reviews," in *The Language of Social Media*. Basingstoke, U.K.: Palgrave Macmillan, 2014, pp. 65–90.
[126] F. Amin, A. Ahmad, and G. S. Choi, "Towards trust and friendliness approaches in the social Internet of Things," *Appl. Sci.*, vol. 9, no. 1, p. 166, 2019.
[127] F. Amin, R. Abbasi, A. Rehman, and G. S. Choi, "An advanced algorithm for higher network navigation in social Internet of Things using small-world networks," *Sensors*, vol. 19, no. 9, p. 2007, 2019.

JING-RONG CHANG received the M.S. degree from the Department of Information Management, National Yunlin University of Science and Technology, Taiwan, in 2002, and the Ph.D. degree from the College of Management, National Yunlin University of Science and Technology, in 2006. He is currently an Assistant Professor with the Department of Information Management, Chaoyang University of Technology, Taiwan. His teaching and research interests include data mining, soft computing, decision analysis, and intelligent systems.

MU-YEN CHEN received the Ph.D. degree in information management from National Chiao-Tung University, Taiwan. He is currently a Professor of information management with the National Taichung University of Science and Technology, Taiwan. He has co-edited 15 special issues in international journals, such as *Computers in Human Behavior*, *Applied Soft Computing*, *Soft Computing*, the IEEE ACCESS, *Information Fusion*, *Neurocomputing*, the *Journal of Medical and Biological Engineering*, *The Electronic Library*, and *Library High Tech*. His current research interests include artificial intelligent, soft computing, bio-inspired computing, data mining, deep learning, context awareness, machine learning, and financial engineering, with more than 100 publications in these areas. He has served as the Editor-in-Chief and an Associate Editor for international journals, such as the *International Journal of Big Data and Analytics in Healthcare*, the IEEE ACCESS, *Applied Soft Computing*, *Human-centric Computing and Information Sciences*, the *Journal of Information Processing Systems*, and the *International Journal of Social and Humanistic Computing*, while he is an Editorial Board Member of several SCI journals.

LONG-SHENG CHEN received the B.S. and M.S. degrees in industrial management from National Cheng Kung University, Tainan, Taiwan, in 1998 and 2000, respectively, and the Ph.D. degree from the Department of Industrial Engineering and Management, National Chiao Tung University, Taiwan, in 2006. He is currently a Professor with the Department of Information Management and an Assistant Vice President for Academic Affairs at the Chaoyang University of Technology, Taiwan. His current research interests include data mining, text mining, social media, quality management, customer relationship management, context awareness, and machine learning, with more than 100 publications in these areas.

SHU-CIH TSENG received the M.S. degree from the Department of Information Management, Chaoyang University of Technology, Taiwan, in 2013, where he is currently a Graduate Student. His current research interests include data mining, text mining, feature selection, and social media.