KAMAYANI KATIYAR

NOVEL APPROACHES IN TOURISM MODELLING

THE INFO LIBRARY

Novel Approaches in Tourism Modelling

Novel Approaches in Tourism Modelling

Kamayani Katiyar

THE INFO
LIBRARY

Published by The InfoLibrary,
4/21B, First Floor, E-Block,
Model Town-II,
New Delhi-110009, India

Novel Approaches in Tourism Modelling
Kamayani Katiyar
ISBN: 978-93-5590-698-4

Table of Contents

Ranking of Performance Assessment Measures at Tehran Hotel by Combining DEMATEL, ANP, and SERVQUAL Models under Fuzzy Condition

Mehrdad Kargari

Faculty of Industries and Systems Engineering, Tarbiat Modares University of Tehran, Tehran, Iran

Correspondence should be addressed to Mehrdad Kargari; m_kargari@modares.ac.ir

Academic Editor: Giovanni Garcea

An effective hybrid model has been proposed by combining ANP, SERVQUAL, and DEMATEL techniques. This model aims to meet different purposes of the hotels and diverse needs of customers at different stages, that is, reservation, reception, accommodation, catering, and check-out. High quality services are ensured when customer expectations have been provided at the expectation level of the customers or beyond that. SERVQUAL model is used to assess the performance of the organizations in terms of five dimensions: responsiveness, empathy, reliability, assurance, and tangibles. Super matrix calculations and pair comparisons required in ANP model have been carried out using DEMATEL model in order to measure the influence of performance assessment measures on each other. In this paper, SERVQUAL model parameters have been considered as the expectations of the hotel clients. Then these customer expectations have been analyzed using DEMATEL model and finally have been ranked using ANP model. Parameters of SERVQUAL model are comprised of verbal and vague criteria in terms of the responses provided by the organizations and customers. This has led to fuzzy conditions in this research. The hybrid model provided better results compared with each individual model, in terms of meeting customer satisfaction and the organization's objectives.

1. Introduction

Quality is one of the most important and attractive concepts in management and industrial engineering disciplines that not only has maintained its value but also has gained much more significance due to the increasing competition in the global markets. Due to this fact, nowadays quality is considered as a competitive advantage and also a tool in order to maintain and attract customer satisfaction and profitability [1]. Global competitiveness recently has become the biggest concern for many companies, namely, the companies with continuous improvement to achieve rapid development and to innovate and meet the needs of their clients. Superior performance in services strengthens competitiveness and establishes a relationship with customer, consolidating the brand and communication with the market [2]. Today, business decisions at many companies are directed toward products and services provision to achieve customer satisfaction associated with high levels of profitability [3]. In addition to this, the role is also significant in the development or liquidation of service organizations and, in other words, it is nonnegligible. Customer satisfaction is a vital necessity for the survival of service-providing organizations. Quality enhancement in the hospitality industry can lead to satisfied customers, increase the number of visitors, and positively affect the GDP of countries [4]. In the hospitality industry, the lack of accurate and internationally defined standards has made it difficult to define quality standards for services in this industry, because the quality of service is different significantly from country to country, from city to city, from one hotel to another hotel, from one client to another client, and even from day to day [5]. One of the well-established models for the measurement of service quality and customer satisfaction in the hotel industry is SERVQUAL model. SERVQUAL model has 5 main factors and 22 pairs of questions. To determine needs of customers in this paper SERVQUAL model questionnaire has been used. The importance of each factor has been measured as a fuzzy five-point Likert scale from customers of three hotels "Parsiani," "Mehr," and "Tehran." To extract the final significance of customer needs and initiatives, the ANP model has been used. To investigate the relationship between these needs and demands with the initiatives adopted by the hotel, QFD model is used. This also has been performed in a five-point Likert scale. To determine the relationship between customer demands and initiatives of the QFD, the team used the average opinion. Next, measures to increase customer satisfaction and improve the organization's objectives have been ranked and prioritized by ANP weights and QFD matrix. In summary, this study has the following three steps.

Step 1. Determine the relevant importance of individual customer needs using Delphi method and SERVQUAL model questions with five fuzzy terms.

Step 2. Rank each and every requirement of hotel customers in order to enhance customer satisfaction and hotel objectives using ANP model and pairwise comparison with a five-point fuzzy range.

Step 3. Determine the relationship between customer requirements and hotel service quality improvement initiatives using Quality Function Deployment and five-point fuzzy scale and ranking them based on the weights obtained from ANP and QFD matrix.

The paper is structured as follows. In Section 2, literature review for five models has been presented: SERVQUAL, ANP, QDF, Fuzzy Logic, and Hybrid models. Identifying and categorizing customer needs (requirements) are performed in Section 3. Section 4 deals with prioritizing the customer needs (requirement) using ANP model. In Section 5 the case study is presented and in Section 6 interrelationships between customer requirements and improvement actions have been determined using QFD model. Section 7 is devoted to service quality improvement actions in order to meet prioritized customer needs and Section 8 is the results and future research.

2. Literature Review

Nowadays, customer satisfaction is known as one of the most common terms in business contexts but, no doubt, creating customer satisfaction and getting them excited by quality goods and services requires, in the first place, identifying their needs and requirements and then transforming these needs into technical specifications. Due to increasing complexity of economic, social, and cultural systems, these will not happen by themselves and methods and systematic events are required in order to transform these concepts to organizational processes. Under these circumstances, on one hand, customer satisfaction is essential in all aspects of the services and, on the other hand, service-providing organizations must obtain the defined goals while ensuring employees support and meeting the established constraints. In this paper, literature review is composed of five sections:

(1) SERVQUAL model (recognizing and classification of customer needs)

(2) ANP model (prioritizing customer needs in order to enhance customer satisfaction)

(3) QFD model (determining the relationships between customer needs and organizational initiatives using ANP weights)

(4) Fuzzy logic (converting customers imprecise language to fuzzy numbers for the calculations)

(5) Hybrid models (examining various combinations of the above-mentioned models in different industry and service sectors).

In studies conducted so far, a combination of ANP and SERVQUAL models (deterministic and fuzzy) and a combination of QFD, ANP, and SERVQUAL models have not been used simultaneously for the improvement of organizational initiatives in order to obtain organizational objectives and customer satisfaction simultaneously.

2.1. SERVQUAL Model. Study of the recent developments represents a vast expansion of services and suggests the ever increasing speed of this trend in the years to come. In other words, industry has undergone downsizing and service sector is growing [6]. Quality of services is critical to the success of any service-providing organization [7]. Customers, whether consumer or service provider, have a huge impact on different aspects of their organizations. In the literature, proper measurement of service quality is guided toward the definitions of quality, training more [8], the service quality dimensions [9, 10], significance of service quality [11], and customer satisfaction [5]. SERVQUAL model introduced in 1985 is one of the most well-known methods for measuring the quality of services [12]. Parasuraman et al. in their study concluded that by comparing customer service performance they were able to understand what customers think about needs and requirements and it allows them to assess the quality of service. If you lower the level of customer service performance, expectations gap will occur. SERVQUAL model is used to measure customer satisfaction and customer satisfaction calculation of the gap (the gap between customer expectations and organizational performance). SERVQUAL model initially had 10 questions and 97 pairs after the couple has been questioned. In order to alleviate this model, supplementary investigation was carried out after the final model in question is summarized in five and 22 pairs [13]. Five main aspects that have an impact on customer evaluation of service quality are defined as follows [14]:

(1) *Tangibles* include physical facilities, tools, and equipment and other tangible models of communication used to provide the service.

(2) *Credibility* includes factors such as trustworthiness, believability, and honesty in providing service for customers.

(3) *Responsiveness* is the readiness and willingness of employees to help customers by providing prompt timely services.

(4) *Courtesy* is the consideration for the customer's property and a clean and neat appearance of contact personnel, manifesting as politeness, respect, and friendliness.

(5) *Knowing the customer* means making an effort to understand the customer's individual needs, providing individualized attention, recognizing the customer when they arrive, and so on.

These aspects have remained unchanged up to this day. This reference framework has been used in almost every research with the objective of measuring and managing service quality by deploying a questionnaire that measures both the customer expectations of service quality in terms of these five dimensions and their perceptions of the service they receive. Parasuraman et al. have asked customers several times to score the organizations in order to determine the validity of the questionnaire. For the internal reliability test, correlation analysis using Alpha Cronbach has been performed. The obtained results are desirable and led to popularity of the model [15].

Five dimensions of service quality are examined with 22-pair questions in SERVQUAL questionnaire. The first question deals with customer waiting and the second question deals with customer perception based on five-point fuzzy Likert [16]. According to the collected data obtained from questionnaires and their analysis, the resulting gap analysis, resulting in a gap between what customers want and what they offer, will be calculated. In order to examine the specific characteristics of different industries and services, this model can be extended to other sectors also added to the questionnaire [17]. To calculate the deficit or gap between each dimension of service quality and customer expectations the following formula is used:

$$O_j = \frac{\sum_{i=1}^{n} \left(P_{ij} - E_{ij}\right)}{n_j}. \tag{1}$$

j is SERVQUAL model dimensions, n_j is number of questions in j dimension, P_{ij} is average of perceptions, E_{ij} is average of expectations, and O_j is gap between every dimension.

If O_j is positive, the level of service quality is higher in relation to customer expectations in terms of j dimension; otherwise, the level of service quality is lower in relation to customer expectations in terms of j dimension. In short, however negative the value is; it suggests that the higher the improvement priority must be assigned for the corresponding dimension. In this paper, SERVQUAL model was used to identify and classify the needs and demands of customers.

2.2. The ANP Model. The world around us is full of multicriteria problems and people mainly make decisions under such conditions. Some of these decisions are so important that any error may impose irreparable losses on us. It is therefore necessary to develop appropriate techniques for selecting optimal techniques and correct decisions. Analysis Network Process (ANP) is one of the most complete of these techniques, first developed by Saaty [18] in 1980. This process is one of the most comprehensive systems designed for multiple attribute decision-making, because this technique provides the possibility of formulating the question in the form of a network and also taking into account qualitative and quantitative criteria at issue. This process is involved in decision-making options and the criteria and subcriteria sensitivity analysis when possible. In addition, pair-wise comparison is based on the judgment and facilitates calculations. It shows the compatibility or incompatibility of the decision. In addition, it has a strong theoretical basis and is based on the principles of being self-evident. Analytic Network Process is a graphical representation of a real complex problem and headed by general purpose next level criteria, subcriteria, and their options. In this article, access to quality services at the top of the charts as the ideal objective is examined. In the next levels five dimensions of SERVQUAL model as service quality improvement criteria and SERVQUAL questions as alternatives have been used. Network analysis process diagram is illustrated in Figure 3.

2.3. QFD Model. To meet the demands of customers in each of the services provided to them, technical and practical measures turned out to be one of the most powerful tools to transform customer needs into technical specifications and practical measures; the QFD method is a method in which organizations have to focus more on your various units in order to achieve the characteristics of the customer of the goods or services utilize [19]. The philosophy of this approach is applied and the quality demands of customers in different development stages of a product service or project are defined. This technique as one of the modern methods of management and engineering quality, market study, and identifying customers began to work and the process of investigation and analysis in addition to identifying the needs and requirements of customers tries to incorporate these expectations at all stages, including design, production, and support [20]. QFD model determines the relationship between parameters such as consumer needs and requirements of engineering a comparative analysis from customer conception of other products, services, and similar projects to establish [21]. QFD model is based on identifying implicit and explicit needs of customers and translating them into the service specifications and reflecting them in all units of the organization. Also, the QFD tool that enables the product development team to identify potential conflicts in the process of identifying and translating the demands of customers into technical requirements, reduced as much as possible. Some of these conflicts include nonconformity of technical product requirements with the demands and needs of our customers, as well as nonconformity of the final product with technical product requirements mentioned. To reduce these conflicts, technical product requirements have to consider the customer needs and demands [3]. The QFD model is applied in various services and industries including education, healthcare, and service sector and helps to understand customer needs and requirements and to turn them into technical specifications, but prioritizing the demands of customers always requiring technical

$$\mu_A(X) = \begin{cases} \dfrac{X - a_1}{a_2 - a_1}, & a_1 \le X \le a \\ \dfrac{X - a_3}{a_2 - a_3}, & a_2 \le X \le a \\ 0 & \text{O.W} \end{cases}$$

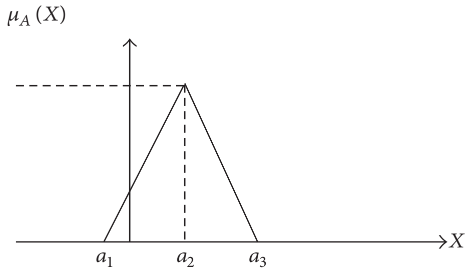


Figure 1: Set of fuzzy numbers and terms.

assistance in this model has been questioned [22], which was used in this article to answer the question of ANP model.

2.4. Fuzzy Logic. The history of criticism of the thought of the world being zero and one, black and white, and right and wrong dates back to more than fifty years ago. Although the exact sciences have been able to correctly explain many phenomena and classical logic has brought the right conclusions, they have not been able to model and explain everything that is all around us. Fuzzy Logic was presented by Dr. Lotfi Zadeh in 1965 at the University of California, Berkeley. He presented human logic into mathematics. If we consider black and white to be corresponding to zero and one in mathematics, the mathematical logic does not recognize the color spectrum existing between these two colors. But in fuzzy sets, there is a spectrum of grey colors and that is where human and machine intersect. The main advantage of fuzzy logic lies in the expression of vague or partially true variables or values. The application of fuzzy logic is widely diverse due to multiple characteristics of it. Also this concept plays an important role in decision-making sciences. Generally, characteristics of fuzzy logic have turned it into a much more efficient tool in quality subjects compared to other conventional mathematic tools. Prioritizing customer needs and quality improvement initiatives are also considered as quality and human fields of science in which various qualities and mathematical and engineering tools are used. So, employing fuzzy logic makes these tools more intelligent and brings them closer to human reasoning [23]. Linguistic variables related to the degree of satisfaction and importance are often ambiguous and vague in nature. For example, expression of satisfaction in terms of "Very Satisfactory," "Somewhat Satisfactory," and "Not Satisfactory" is often seen as a natural perception of customer priorities and judgments. The applicability of fuzzy theory in decision-making sciences lies in the ambiguity and vagueness of expression. When available information is subjective and inaccurate, fuzzy modeling is an effective way to formulate decision-making problems [24]. Herreva and Viedma emphasized that, in situations with subjective decision-making and uncertainty, linguistic variables are an effective tool for better and easier understanding and evaluation of the service quality performance [25].

Fuzzy number is an effective method in order to sufficiently entail subjective and objective knowledge and understanding. Lotfi Zadeh in 1975 developed fuzzy logic, introduced approximate reasoning concept, and showed that logical and vague statements comprise algorithms that help us to use ambiguous data to achieve optimal results [26]. In this study, the triangular fuzzy numbers have been used according to Figure 1.

A crisp of a triangular fuzzy number is obtained from the following equation:

$$A_a = [a_1^a, a_2^a] = [(a_2 - a_1)\,a - (a_3 - a_2)\,a + a_3] \qquad 0 \le a \le 1. \tag{2}$$

Every fuzzy term is represented by a triangular number in the range $(1, 0)$. The fuzzy numbers used in this research are provided in Table 3. In every management science research, respondents may have a different understanding and perception of the linguistic terms. Fortunately, these errors are of minor importance, because the default values (Table 3) have been used as representative values to reflect the priorities and the membership function of Figure 1; the formula explains the asymmetry of the fuzzy numbers by the asymmetry of the linguistic terms.

2.5. Hybrid Models Literature Review. Although SERVQUAL, QFD, and Fuzzy ANP techniques have been applied individually in service and industrial research settings, hybrid application of these three techniques under fuzzy conditions has not been seen so far. Some of the recent researches using SERVQUAL, QFD, and Fuzzy ANP techniques with fuzzy approach single or double or triple models are as follows.

Aysun and Masoudi proposed a hybrid QFD and SERVQUAL conceptual framework to measure the hotel service quality using the SERVQUAL model as a starting point and then identified service design and hotel guests' requirements using a QFD approach. In this research a six-step model has been proposed to improve the service quality in hotel industry [27].

Benítez et al. measured the service quality in hospitality industry using fuzzy numbers. In their research, they determined the priority of improvement initiatives using TOPSIS method [28].

Kazemi & Almardani developed a SERVQUAL based model prioritizing quality improvement measures in Chalous Province Power Distribution Office. They proposed effective initiatives for the improvement of each of the five service quality areas using statistical studies and analysis [29].

In the study, published in 2005 in the International Journal of Intelligent and Fuzzy Systems, which was titled "Implementation of QFD Based on ANP Process of Linguistic Data: An Application in the Automotive Industry," researchers intended to use QFD techniques in a fuzzy environment with the demands and needs of our customers, to prioritize design requirements by examining the correlation between customer needs, design requirements, and interdependence among them; the ANP have used these techniques [1].

In a research, titled "A Fuzzy Optimization Model for QFD Planning Process Using ANP Method" whose results were published in the International Journal of Operational Research in 2004, using quality function deployment (QFD) the authors tried to reduce the conflicts due to lack of conformity between customers' requirements and product characteristics and also lack of conformity between technical specifications and final product and, then, the correlations between customer requirements and technical specifications have been examined using AHP method and, due to imprecise and ambiguous data, linguistic parameters have been addressed in ANP. Due to imprecise and ambiguous data and lack of quantitative tools, a mixed integer linear programming model has been developed in order to meet technical specifications of the product and, finally, technical indices have been prioritized using a fuzzy ranking method [3].

In the research titled "Using QFD and ANP to Analyze the Environmental Production Requirements in Linguistic Preferences" published in the International Journal of Expert Systems with Applications in 2010, authors, at first, tried to facilitate the main issue of the QFD problem; however, the "What" questions of EPRs and "How" problems of the SPIs have to be made, which are two major components and they should be emphasized on the house of quality matrices. In conjunction with fuzzy sets theory and analytical network process, the systematic analytical procedures are proposed. Subsequently, the systematic network processes have been proposed using a combination of fuzzy sets and ANP method and, ultimately, the precise awareness of the SPI and focus on using ERP effects for the case company have been suggested [30].

In a paper titled "Rapid Tooling Route Selection for Metal Casting Using QFD–ANP Methodology," published in International Journal of Computer Integrated Manufacturing in 2007, the author has proposed an integrated methodology using QFD and ANP methods to determine and prioritize the technical and engineering requirements of casting parts, based on the customer needs, for selection and evaluation of an appropriate rapid prototyping- (RP-) based route for tooling fabrication. Ultimately, the author developed a planning matrix using a robust evaluation method based on ANP in order to translate customer needs to technical requirements [22].

In an article titled "A Model with a Customer-Manufacturer-Competitor Orientations for Life Cycle Analysis of Products Based on QFD, AHP/ANP, and TRIZ" published in 2007 in the Journal of Design Engineering, the researchers proposed a customer-manufacturer-competitor model to help manufacturers to analyze customers, suppliers, and competitors orientations and the issues related to PLC. This model is composed of three evaluation processes: (1) customer-oriented evaluation, (2) manufacturer-oriented evaluation, and (3) ideal formulation [31].

In an article titled "An Effective Decision Making Approach Using a Combination of QFD and ANP" published in 2008 in the WSEAS Transactions on Business and Economics Journal, researchers developed an effective decision-making approach based on QFD and ANP approaches in order to help decision-making in planning or evaluation problems using a practical example [32].

Atashsooz in 1383 in his master thesis in the field of industrial management in Tehran University titled as "Designing a Product Planning Model Using QFD, ANP and Goal Programming" suggests that in order to develop the houses of quality first it is required to examine the relationships between customer needs and product specifications regarding the interrelationships between customer needs and product specifications in order to be able to prioritize product specifications in house of quality. The research has exploited the analytical network process (ANP) for this purpose. Moreover, the proposed decision-making algorithm considers the multiobjective nature of the problem and includes other objectives related to the product design and development including the constraints related to human resource and product specification scalability and design. The researcher proposes a Zero-One Goal Programming model that incorporates the relative importance of the product technical specifications determined using ANP approach and, according to the obtained results, suggestions for design and development of the software in order to increase the customer satisfaction have been provided [33].

Zaheri, in a master's thesis in the field of industrial engineering at the University of Yazd titled "Prioritizing Strategic Actions at Strategic-Oriented Organizations by Combining AHP and QFD with Fuzzy Variable," in this line of research by identifying customer demands at Telecom Fars Province, tried to prioritize these demands using the AHP, TOPSIS, and AHP-LP techniques, determining strategic actions that eventually led to the realization of customers' demands, transforming the relationship between customer demands and strategic actions into the form of fuzzy numbers at fuzzy QFD house of quality. Ultimately, strategic actions have been carried out. The results of prioritizing strategic actions have been compared with the results of TOPSIS and AHP-LP [34].

3. Identifying and Classifying the Customer Needs

For information gathering in order to identify and classify customer needs and demands in the hospitality industry,

Table 1: Fuzzy representation of numbers and scales.

Symbol	Scale	Weight	Verbal expression	Verbal expression
V	Excellent	(0.75, 1, 1)	Extremely important	Very high
H	Good	(0.5, 0.75, 1)	Very important	High
M	Fair	(0.25, 0.5, 0.75)	Important	Average
L	Poor	(0, 0.25, 0.5)	Relatively important	Low
P	Very poor	(0, 0, 0.25)	Unimportant	Very low

Table 2: Service quality improvement indices.

Index code	Factors	Class
T_1	(i) The attractiveness of the location, facade design, and outdoor surrounding	Tangibles
T_2	(ii) Appealing and diversity of interior decoration	
T_3	(iii) Cleanliness, uniformity, and appearance of staff	
T_4	(iv) Modern entertainment and recreation facilities	
T_5	(v) The hotel's interior and exterior are clean	
L_1	(i) Orders done by staff properly and on time	Reliability
L_2	(ii) Cleanness and quality of rooms	
L_3	(iii) Facilities of rooms (equipment worked properly)	
S_1	(i) Proper welcoming of customers by employees at the front desk and reception	Responsiveness
S_2	(ii) Employees responded promptly to my requests	
S_3	(iii) Employees responded quickly to my problems (speed of service)	
S_4	(iv) Reservation and service cover area	
A_1	(i) Staff experience and professionalism in providing error-free services	Assurance
A_2	(ii) Staff politeness	
A_3	(iii) I got what I paid for	
A_4	(iv) Effort done by staff for security and comfort	
A_5	(v) Hotel atmosphere (calm and quiet)	
E_1	(i) Hotel prevision for customer necessities	Empathy
E_2	(ii) Staff availability	
E_3	(iii) I received undivided attention at the front desk	
E_4	(iv) Staff flexibility in order to receive and respond guest requests	
E_5	(v) Provision of special services for children, people with disabilities, and the elderly	

Table 3: Priority of SERVQUAL model index.

Dimension (1)	Dimension (2)	Dimension (3)	Dimension (4)	Dimension (5)
(0.2, 0.3, 0.35)	(0.7, 0.75, 0.8)	(0.3, 0.35, 0.45)	(0.5, 0.6, 0.7)	(0.4, 0.45, 0.5)
(0.4, 0.5, 0.67)	(0.8, 0.95, 1)	(0.25, 0.35, 0.4)	(0.4, 0.5, 0.55)	(0.15, 0.2, 0.25)
(0.7, 0.75, 0.95)	(0.65, 0.75, 0.85)	(0.5, 0.65, 0.7)	(0.75, 0.85, 0.95)	(0.2, 0.25, 0.35)
(0.4, 0.45, 0.5)		(0.6, 0.7, 0.8)	(0.7, 0.8, 0.9)	(0.35, 0.5, 0.74)
(0.75, 0.8, 0.85)			(0.7, 0.8, 0.85)	(0.6, 0.8, 0.9)

in field studies, the model SERVQUAL has been the most widely used. Accordingly, in this study, five dimensions and 22 pairs of questions in SERVQUAL model have been used. Based on interviews with experts and executives in the hospitality industry, 22 coupled questions in five dimensions of service quality, SERVQUAL model, have been selected and included in the questionnaire. To get customer feedback on the effectiveness of each of the major and minor criteria in assessing the quality of hotel services, a questionnaire was designed. The questionnaire includes a table comparing the main criteria and subcriteria of the five dimensions of SERVQUAL model. The criteria and measures are provided in Table 2. In these tables, the importance of each criterion from the customer perspective, with fuzzy scales of very low, low, medium, high, and very high, has been questioned. Out of 125 questionnaires distributed, 107 questionnaires have been completed. The corresponding fuzzy numbers are provided in Table 1 and Figure 2 [35]. Average importance of the survey is shown in Table 2.

4. Prioritizing Customer Needs Using ANP

In order to calculate the importance of each customer's needs in this study, questionnaire approach and fuzzy ANP have

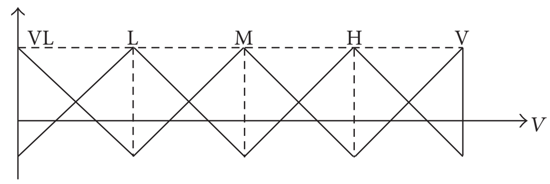


FIGURE 2: View fuzzy numbers and options.

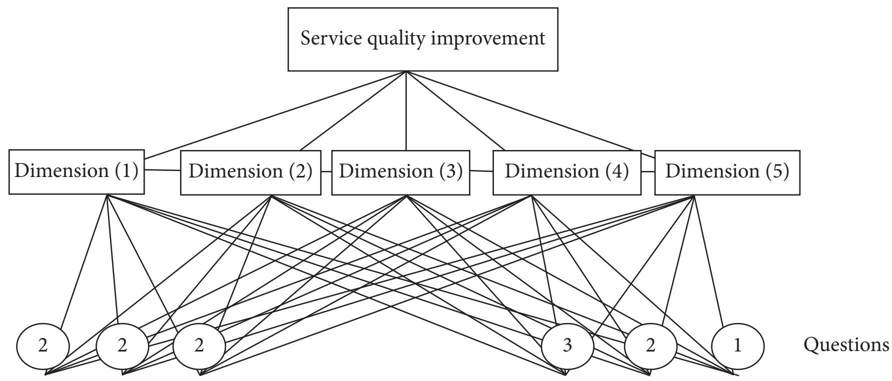


FIGURE 3: ANP diagram.

been used in combination. Also in order to convert the weights obtained from every customer questionnaire to pair-wise comparison of each individual customer questionnaire, every individual item in the questionnaire is divided by each other (W_j/W_i) and are compared to each other at every row and, finally, pair-wise comparison matrix is obtained based on the preferences. If the matrix of the weight of main dimensions of service quality is W_1 and the matrix of correlation among main options is W_3, then prioritizing of the main options of service quality regarding the internal correlation among these main service quality dimensions is the multiplication of two matrices $W_F = W_3 * W_1$. Also, whether the matrix of the weighted vectors of dimensions required strengthening the main dimensions of service quality is represented as W_2 and the internal correlation between dimensions is represented as W_4; then, priority of the quality improvement options is the multiplication of $W_P = W_F * W_R$. The final weight of each of the dimensions and options to improve are presented in Table 3. ANP model diagram is shown in Figure 3.

In this model, each dimension of service quality as compared to the overall goal of a whole-choice test (equal importance, a little more, more, much more, and more great) was assessed. As well as any of the SEVQUAL model questions, the five aspects of model-choice test have been questioned. In this model, internal cohesion and solidarity between the qualities of internal questions of SERVQUAL model for paired comparisons have been conducted in the same range of five options.

5. Identification and Classification of Quality Improvement Measures

After the identification, classification, and prioritization of needs and demands of customers QFD team has to identify measures to improve service quality and increase customer satisfaction. The number of these measures in reviewed literature [23, 27, 29] and others has been collected and categorized by QFD team. In short, this procedure is categorized in Table 4.

6. Determining the Relationship between Customer Demands and Measures to Help Improve the Quality of QFD

In this step, the intensity and quality of relationship between each customer need and requirements and each hotel service quality improvement initiative are determined. The relationship between customer demands and improvement initiatives adopted by hotel could be represented by numbers or contract signs. To do so, in this paper fuzzy terms of "*very high,*" "*high,*" "*medium,*" "*little,*" and "*very little*" are used. QFD team has used brainstorming method to identify this relationship. In this paper, every relationship is individually discussed by QFD team and then one of the fuzzy terms is chosen for that relationship. The relationships between customer needs and requirements and hotel initiatives at QFD house of quality matrix are shown in Table 5.

Table 4: Service quality improvement initiatives.

Action code	Initiatives	Class
P_1	(i) Automation	Acceptance and settlement system P
P_2	(ii) Reservation radius	
P_3	(iii) Room delivery & settlement	
K_1	(i) Proper and on-time order	Cleaning and service K
K_2	(ii) Help to carry luggage	
K_3	(iii) Room and hotel environment cleanliness	
F_1	(i) Quality, quantity, and diversity	Food and beverage F
F_2	(ii) Sanitation and health	
F_3	(iii) Service waiting time and cost	
U_1	(i) Belongings and documents	Hotel security U
U_2	(ii) Goods and items	
U_3	(iii) Spouse and children	
O_1	(i) Facilities and equipment	Hotel facilities O
O_2	(ii) Room size and hotel places	
O_3	(iii) Silence	
O_4	(iv) Service waiting time and cost	
O_5	(v) Hotel view and lighting	
M_1	(i) Staff politeness and courtesy	Hotel staff M
M_2	(ii) Staff education and training	
M_3	(iii) Team work	
M_4	(iv) Hotel staff uniformity and cleanness	
Q_1	(i) Diversity and attractiveness	Hotel recreational facilities Q
Q_2	(ii) Special services	
Q_3	(iii) Modernity	

7. Prioritize Actions to Help Improve the Quality of QFD

In this section, using QFD rules and the obtained results, quality improvement initiatives have been ranked. For this purpose, consider a column to the right of the QFD table. This column contains the priority of customer needs and is represented as a fuzzy number. If fuzzy numbers of this column are multiplied by the fuzzy numbers column p_i and then added together and divided by 22, the weight is achieved. This operation is similarly performed for other columns. The results of this operation are shown in the weight column of Table 6.

7.1. Ranking Quality Improvement Initiatives with Fuzzy Logic. For triangular fuzzy numbers $A_i = (L_i, M_i, U_i)$ ranking is based on the following three criteria [23, 26, 36].

(1) The calculation of the enclosed surface area: $S_i = (L_i + 2M_i + U_i)/4$

(2) Mode: $\text{Mode}(A_i) = M_i$

(3) The scope of fuzzy number: $R(A_i) = U_i - L_i$.

The three criteria are applied in a sequential order. This means that if, after calculating the first criterion, the number of conditions are equal, then the second initiative is applied and then again in the third criterion it is used equally. The results of the ranking are provided in Table 6. For ranking of these measures, only the first criterion is used and there is no need to apply the rest of the criteria.

This ranking reflects the results below.

The first priority of the hotels in order to meet customer satisfaction is the cost and time of stay and hospitality. Room cleanness and hotel environment come second. The third priority is the quality, quantity, and variety of food and beverages. The fourth priority is the courtesy and politeness of the hotel staff. The fifth priority is special customer service of the hotel. The remaining priorities for the satisfaction of hotel customers are provided in Table 6.

8. Discussion

Today, good service is one of the basic concerns of all service organizations, including the hotel industry, and many hotels have been carefully monitored by these indicators to increase customer satisfaction, especially in this type of industry. They use personalization as an attempt to increase customer satisfaction, in order to increase profits in raising quality assurance. In fact, you need to be sure that they provide the right product or service to the right person. In this research, the effectiveness of the combined model and its impact are investigated through the combination of the stated techniques, to meet the multiple objectives of the hotel and the various needs of customers in different stages. Customer loyalty is guaranteed when the customer's expectations of the service are provided and, on the other

Table 5: QFD house of quality.

Customer requirements	Hotel initiatives																							
		P			K			F			U				O				M				Q	
	P_1	P_2	P_3	K_1	K_2	K_3	F_1	F_2	F_3	U_1	U_2	U_3	O_1	O_2	O_3	O_4	O_5	M_1	M_2	M_3	M_4	M_1	M_2	M_3
T																								
T_1	P	P	P	P	P	P	P	P	M	P	P	M	M	M	M	M	H	P	P	P	H	H	H	M
T_2	M	M	P	M	P	H	M	M	H	M	M	P	H	M	H	H	M	L	P	P	M	V	H	H
T_3	P	P	L	H	M	H	V	V	V	M	L	M	H	P	M	V	M	M	M	M	V	H	M	M
T_4	H	H	M	M	M	M	M	M	M	M	M	M	H	P	M	H	M	P	P	M	P	H	H	H
T_5	P	P	P	M	M	H	H	H	V	M	V	H	M	M	P	M	H	P	P	M	V	H	M	H
L																								
L_1	H	H	V	H	H	M	H	H	V	H	H	H	M	M	L	H	P	M	M	H	L	P	H	M
L_2	P	P	P	P	P	V	H	H	V	L	L	L	L	P	P	H	L	M	M	M	V	P	L	L
L_3	L	P	P	H	H	M	M	M	H	H	H	H	V	L	L	V	H	P	P	L	M	H	H	V
S																								
S_1	P	P	M	M	M	M	M	M	M	L	L	L	L	L	L	M	P	H	H	P	M	P	M	P
S_2	P	P	M	M	M	H	M	H	H	L	L	P	M	P	P	H	L	H	M	L	L	P	M	P
S_3	V	V	V	H	M	P	M	P	V	H	H	H	H	P	P	V	L	L	P	P	L	M	V	V
S_4	V	V	V	H	M	H	H	M	H	H	H	M	M	P	P	M	P	M	P	M	L	P	M	M
A																								
A_1	H	H	V	H	M	M	V	V	V	M	M	M	M	L	M	H	L	H	M	M	H	M	M	M
A_2	P	P	M	M	H	H	H	H	V	M	M	M	P	P	P	M	P	V	H	M	H	M	M	M
A_3	M	H	H	V	H	H	V	H	V	M	M	M	M	V	M	V	M	M	M	M	H	H	V	H
A_4	P	M	L	M	H	V	H	H	V	H	H	H	L	P	V	H	P	H	M	L	H	P	M	L
A_5	P	P	P	P	P	P	P	P	H	V	V	V	L	P	V	H	L	P	P	P	L	P	M	M
E																								
E_1	L	M	H	H	H	H	H	H	H	L	L	L	M	P	P	L	P	M	L	L	M	P	M	L
E_2	H	L	H	V	H	M	L	L	H	P	P	L	L	L	P	M	P	M	M	L	P	P	M	L
E_3	H	L	M	M	L	L	L	L	L	P	P	P	L	P	P	M	P	M	L	L	L	P	H	H
E_4	L	P	M	L	M	M	M	M	H	M	M	M	L	P	P	M	P	M	M	P	L	M	M	L
E_5	M	V	M	H	H	H	H	H	V	H	H	H	H	M	L	V	L	M	M	M	H	V	V	V

hand, through the quality of service. SERVQUAL model for assessing hotel performance and comparing the ANP model needed by DEMATEL model is used to measure the effect of performance evaluation indicators on each other in a hybrid model and then ranked by the ANP model. The results of the implementation of this hybrid model indicate better solutions to the implementation of each of the above models in providing customer satisfaction and organizational goals simultaneously.

9. Conclusions and Future Research

In this study, a hybrid model SERVQUAL + QFD + ANP has been proposed and used to prioritize quality improvement initiatives at international hotels in fuzzy terms to satisfy the customer's needs and requirements.

SERVQUAL model has been applied to identify and classify the needs of customers, ANP model for careful prioritization of needs and demands of customers, and QFD house of quality in order to ensure the relationship between customer needs and demands and hotel service quality initiatives and prioritizing the service quality improvement initiatives. This hybrid model allows us to prioritize and rank the quality improvement initiatives focused and oriented toward performance enhancement and increasing the customer satisfaction. The extended form of this model could be used as a powerful tool that enables the hotel managers not only to increase customer satisfaction but also to reduce hotel costs in the long term. The results of the hybrid model while confirming the results of the previous studies and in a joint meeting with hotel managers and QFD team provide us with better results compared to individual and separate application of ANP, QFD, and SERVQUAL models which results in greater customer satisfaction and more effective organizational performance. The results of the SERVQUAL model could be used as input for QFD and prioritized needs and demands of customers using ANP model could be used as supplement for QFD model.

Table 6: Ranking of service quality improvement initiatives.

Initiative	Weight of initiative	Enclosed surface area	First rank
P			
P_1	(0.101705, 0.20625, 0.391364)	0.226392	19
P_2	(0.128409, 0.232955, 0.406818)	0.250284	18
P_3	(0.136364, 0.264205, 0.440227)	0.27625	15
K			
K_1	(0.182955, 0.338636, 0.537386)	0.349403	7
K_2	(0.155682, 0.302273, 0.507273)	0.316875	12
K_3	(0.196023, 0.358523, 0.560795)	0.468466	2
F			
F_1	(0.211364, 0.381818, 0.57875)	0.388438	3
F_2	(0.193182, 0.353409, 0.558295)	0.354574	7
F_3	(0.307955, 0.511364, 0.659545)	0.497557	1
U			
U_1	(0.157386, 0.324432, 0.529318)	0.333892	9
U_2	(0.140909, 0.306818, 0.508864)	0.315852	13
U_3	(0.156818, 0.319318, 0.520909)	0.329091	10
O			
O_1	(0.128409, 0.294318, 0.499545)	0.304148	14
O_2	(0.055682, 0.131818, 0.309659)	0.157244	24
O_3	(0.085795, 0.186932, 0.365568)	0.206307	22
O_4	(0.24375, 0.436932, 0.621591)	0.434801	4
O_5	(0.061932, 0.168182, 0.361932)	0.190057	23
M			
M_1	(0.113068, 0.248295, 0.452955)	0.375653	5
M_2	(0.078409, 0.185795, 0.386818)	0.209205	21
M_3	(0.078409, 0.202841, 0.398409)	0.220625	20
M_4	(0.168182, 0.336932, 0.523182)	0.341307	8
Q			
Q_1	(0.132386, 0.243182, 0.436932)	0.26392	16
Q_2	(0.186932, 0.372159, 0.569318)	0.365142	6
Q_3	(0.157386, 0.323295, 0.5075)	0.327869	11

References

[1] T. Ertay, G. Büyüközkan, C. Kahraman, and D. Ruan, "Quality function deployment implementation based on analytic network process with linguistic data: An application in automotive industry," *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology*, vol. 16, no. 3, pp. 221–232, 2005.

[2] N. M. Stefano, N. Casarotto Filho, R. Barichello, and A. P. Sohn, "A fuzzy SERVQUAL based method for evaluated of service quality in the hotel industry," in *Proceedings of the 7th CIRP Industrial Product-Service Systems Conference, IPSS 2015*, pp. 433–438, France, May 2015.

[3] C. Kahraman, T. Ertay, and G. Büyüközkan, "A fuzzy optimization model for QFD planning process using analytic network approach," *European Journal of Operational Research*, vol. 171, no. 2, pp. 390–411, 2006.

[4] M. A. Beheshtinia and M. Farzaneh Azad, "A fuzzy QFD approach using SERVQUAL and Kano models under budget constraint for hotel services," *Journal of Total Quality Management & Business Excellence, Taylor & Francis Online*, pp. 1–23, 2017.

[5] E. M. O'Brien and K. R. Deans, "Educational supply chain: A tool for strategic planning in tertiary education?" *Marketing Intelligence & Planning*, vol. 14, no. 2, pp. 33–40, 1996.

[6] Y. Horovitz, *The Seven Secret of Service Strategy*, Financial Times Prentice Hall, Harlow, UK, 1st edition, 2000.

[7] J. Kandampully, "The impact of demand fluctuation on the quality of service: A tourism industry example," *Managing Service Quality: An International Journal*, vol. 10, no. 1, pp. 10–19, 2000.

[8] M. Joseph and B. Joseph, "Service quality in education: A student perspective," *Quality Assurance in Education*, vol. 5, no. 1, pp. 15–21, 1997.

[9] M. S. Owlia and E. M. Aspinwall, "A framework for the dimensions of quality in higher education," *Quality Assurance in Education*, vol. 4, no. 2, pp. 12–20, 1996.

[10] J. B. Ford, M. Joseph, and B. Joseph, "Importance-performance analysis as a strategic tool for service marketers: The case of service quality perceptions of business students in New Zealand and the USA," *Journal of Services Marketing*, vol. 13, no. 2, pp. 171–186, 1999.

[11] J. Rowley, "Beyond service quality dimensions in higher education and towards a service contract," *Quality Assurance in Education*, vol. 5, no. 1, pp. 7–14, 1997.

[12] A. Parasuraman, A. V. Zeithaml, and L. L. Berry, "A conceptual model of service quality and its implication for future research," *Journal of Marketing*, vol. 49, no. 4, pp. 41–50, 1985.

[13] A. Parasuraman, L. L. Berry, and V. A. Zeithaml, "Perceived service quality as a customer-based performance measure: An empirical examination of organizational barriers using an extended service quality model," *Human Resource Management*, vol. 30, no. 3, pp. 335–364, 1991.

[14] J. Van Iwaarden, J. Wide. Vander, L. Bell, and R. Miller, "Applying SERVQUAL to websites: an exploratory study," *International Journal of Quality Management*, vol. 20, no. 8, pp. 919–935, 2003.

[15] L. J. Cronbach and L. Furby, "How we should measure "change": Or should we?" *Psychological Bulletin*, vol. 74, no. 1, pp. 68–80, 1970.

[16] G. Albaum, "The Likert scale revisited: An alternate version," *International Journal of Market Research*, vol. 39, no. 2, pp. 331–348, 1997.

[17] N. M. Kassim and J. Bojei, "Service quality: Gaps in the Malaysian telemarketing industry," *Journal of Business Research*, vol. 55, no. 10, pp. 845–852, 2002.

[18] T. L. Saaty, "Highlights and critical points in the theory and application of the analytic hierarchy process," *European Journal of Operational Research*, vol. 74, no. 3, pp. 426–447, 1994.

[19] k. Bhagirathi, *Determining customer needs based on service quality dimensions through Quality Function Deployment (QFD)*, 2009.

[20] E. Bottani and A. Rizzi, "Strategic management of logistics service: a fuzzy QFD approach," *International Journal of Production Economics*, vol. 103, no. 2, pp. 585–599, 2006.

[21] N. Kaneko, "QFD implementation in the service industry," in *Proceedings of the 45th Annual Quality Congress Transactions*, pp. 808–813, Milwaukee, WI, USA, May 1991.

[22] D. K. Pal, B. Ravi, and L. S. Bhargava, "Rapid tooling route selection for metal casting using QFD-ANP methodology," *International Journal of Computer Integrated Manufacturing*, vol. 20, no. 4, pp. 338–354, 2007.

[23] L. V. Vanegas and A. W. Labib, "A Fuzzy Quality Function Deployment (FQFD) model for deriving optimum targets," *International Journal of Production Research*, vol. 39, no. 1, pp. 99–120, 2001.

[24] H. Zimmermann, *Fuzzy Set Theory—and Its Applications*, Kluwer Academic, Dordrecht, The Netherlands, 2nd edition, 1992.

[25] F. Herrera and E. Herrera-Viedma, "Linguistic decision analysis: steps for solving decision problems under linguistic information," *Fuzzy Sets and Systems*, vol. 115, no. 1, pp. 67–82, 2000.

[26] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning-I," *Information Sciences*, vol. 8, no. 3, pp. 199–249, 1975.

[27] I. K. Aysun and A. Masoudi, "A QFD and SERVQUAL approach to hotel service design," *İşletme Fakültesi Dergisi*, vol. 9, no. 1, pp. 17–31, 2008.

[28] J. M. Benítez, J. C. Martín, and C. Román, "Using fuzzy number for measuring quality of service in the hotel industry," *Tourism Management*, vol. 28, no. 2, pp. 544–555, 2007.

[29] A. Kazemi and S. Almardani, "Development of the SERVQUAL model for prioritizing measures for quality of service Amelioration case study in the electricity industries," *Journal of science and technology Sharif*, vol. 49, pp. 139–150, 2009.

[30] Y. H. Lin, H.-P. Cheng, M.-L. Tseng, and J. C. C. Tsai, "Using QFD and ANP to analyze the environmental production requirements in linguistic preferences," *Expert Systems with Applications*, vol. 37, no. 3, pp. 2186–2196, 2010.

[31] J. Hou and D. Su, "A customer manufacturer competitor orientation model for product life cycle analysis based on QFD, AHP/ANP and TRIZ," *International Journal of Design Engineering*, vol. 1, no. 1, pp. 104–124, 2007.

[32] L. Yu-Ting, W. Wei-Wen, and T. Gwo-Hshiung, "An effective decision-making method using a combined QFD and ANP approach," *WSEAS Transactions on Business and Economics*, 2008.

[33] A. Atashsooz, *Designing of a product planning model using QFD, ANP and Goal programming, [Master, thesis]*, University of Tehran, 2004.

[34] M. Zaheri, *Prioritizing the Strategic Initiatives in Customer-oriented Organizations with integrating AHP, QFD and Fuzzy variables, [Master, thesis]*, Yazd University, 2007.

[35] L. A. Zadeh, "Is there a need for fuzzy logic?" *Information Sciences*, vol. 178, no. 13, pp. 2751–2779, 2008.

[36] Y. Taho and H. Chih-ching, "Multiple-attribute decision making methods for planet layout decision problem," *Robotic and Computer-Integrated Manufacturing*, vol. 23, pp. 126–137, 2007.

Large-Scale Image Retrieval of Tourist Attractions Based on Multiple Linear Regression Equations

Yinping Song

Xi'an International University, Xi'an, 710077 Shaanxi Province, China

Correspondence should be addressed to Yinping Song; xaiu00010@xaiu.edu.cn

Academic Editor: Miaochao Chen

This paper presents an in-depth study and analysis of large-scale tourist attraction image retrieval using multiple linear regression equation approaches. This feature extraction method often relies on the partitioning of the grid and is only effective when the overall similarity of different images is high. The BOF model is borrowed from the method for text retrieval, which generally extracts the local features of an image by the scale-invariant feature transform algorithm and clusters them using k-means to obtain a low-dimensional visual dictionary and characterizes the image features with a histogram vector based on the visual dictionary. However, when there are many kinds of images, the dimensionality of the visual dictionary will be large and it is not convenient to construct the BOF model. The last fully connected layer is taken as the image feature, and it is dimensionalized by the principal component analysis method, and then, the low-dimensional feature index structure is constructed using the locality-sensitive hashing- (LSH-) based approximate nearest neighbor algorithm. The accuracy of our graph retrieval has increased by 8%. The advantages of feature extraction by a convolutional neural network and the high efficiency of a hash index structure in retrieval are used to solve the shortcomings of traditional methods in terms of accuracy and other aspects in image retrieval. The results show that compared with the above two algorithms, for most of the attractions, the method has a relatively obvious advantage in the accuracy of retrieval, and when there are few similar images of a particular attraction in the attraction image library, the accuracy of the query results is not much different from the first two methods.

1. Introduction

With the rapid development of computer technology, various computer-related disciplines have emerged. Computer vision, as an important discipline, involves a wide range and extensive research. Image feature extraction is an important concept in computer vision, and during years of research, many scholars have also identified several of the most basic image features for images. These features are widely used because they are easy to extract, are widely applied, and can accurately describe images, among other properties. Image feature extraction is the key technology for image processing, and the extracted image features are used to perform digital image matching and image retrieval based on the matching results [1]. The existing image retrieval technology provides technical support for the research and implementation of the tour guide system. With the advent of the era of big data, how to collect, store, and process big data has gradually become an urgent problem in all walks of life, so in the face of such a large-scale image of tourist attractions, how to quickly and accurately retrieve the relevant information of the corresponding attractions is of great significance, which will help to improve the operation of China's tourism industry, optimize the industrial structure, and improve the competitiveness of China's tourism industry in the international arena [2]. In conclusion, with the rapid development of the tourism industry and the arrival of the era of big data, it is of great practical value to realize the fast and accurate retrieval of large-scale images of tourist attractions, which will certainly promote the further development of China's tourism industry.

This paper applies the recommendation system to travel websites, which not only can improve the user's satisfaction with the recommendation but also has important application value for the improvement of related travel websites. From the theoretical point of view, this paper combines the latent semantic space analysis model with visual Bayesian personalized ranking (VBPR) for the first time and proposes a new hybrid recommendation model multimodal visual Bayesian personalized ranking, which is a contribution to the theoretical aspect of the recommendation system [3] (sparsity problem) and improves user interaction experience, thus contributing to better development of tourism industry. The analysis of an ideal intelligent scenic spot construction system needs to be done from a tourist perspective. Tourists occupy the main position in the tourism industry, and solving the contradiction between tourism supply and demand is the communication bridge between the two [4]. In the scenic spot, intelligent tourism should be built with tourists as the center, and every time tourists visit a scenic spot, it is like putting themselves on the Internet of Things, using cloud computing and other information technology to make big data statistics and analysis of tourists' needs, and then starting from precise data, providing tourists with intelligent scenic spots' starting points, destinations, plans, and itinerary. Before visitors leave, they can also have a virtual experience of the attraction, which means that they can view the attraction and have a comprehensive understanding of it on the Internet, producing ways to attract visitors, such as tourist cartoons and caricatures; visitors to scenic spots can scan QR code e-tickets by using mobile devices, which will automatically save the visitor's information and ensure that the tickets will not be lost.

If we can use the cell phone, which is so intelligent now, to design and realize a powerful and convenient tour guide system software, it will be able to provide a great convenience for the travelers as follows: reduce the cost of travel, make the most suitable for their travel plans, and significantly improve the quality of tourism and the service level of cell phone information technology. The research work of this paper consists of two main parts: the mobile application is the mobile program of the intelligent tourism guide system developed by using Android native development technology, which is mainly responsible for realizing the functions of image acquisition and preprocessing, including compact feature extraction of CDVS of scenic spot images and display of retrieval results and push information from the server-side application; the server-side application is developed based on C++ technology, and the server-side application is mainly responsible for receiving requests from the mobile terminal for attraction retrieval, completing accurate attraction retrieval based on the image features and GPS information sent from the mobile terminal, and pushing the retrieval results and related attraction information to the mobile terminal for display. By introducing the visual retrieval technology based on CDVS features into the intelligent tourist guide system, this paper achieves precise location and accurate identification of attractions greatly improves the accuracy of attraction location and identification, provides technical guarantee for the subsequent accurate pushing of tourist attraction information and surrounding commercial information, and thus improves the user experience of the tourist guide system. Thus, the work of this paper has greater theoretical research value and practical application value.

2. Current Status of Research

For the currently used content-based large-scale image retrieval, most of the images are characterized by extracting the underlying features of the images, which are used as the basis for image retrieval, and can be divided into the global feature and local feature retrieval methods according to the underlying features [5]. The global features of the image are to treat the image as a whole and extract its global feature information directly by some global feature description method, which makes feature extraction convenient and facilitates retrieval, but the features extracted by this method only focus on the image as a whole and ignore the image detail information, which often has high requirements on the image quality [6]. This feature extraction method has achieved good research, generally by using some feature extraction method to get multiple local features of the image, which were then integrated to describe the whole image; using this method can get the details of the image features but ignore the overall information of the image, and often, the extracted local features are more complex, in large-scale image retrieval [7]. Multithreading plays a very important role in this part. The use of multithreaded programming allows the program to perform multiple tasks at the same time. The interface can be switched when extracting features, and the results of extracting features can be displayed at the same time when sending a request. This multistep simultaneous method can ensure the improvement of the overall operating efficiency of the system. It has a certain impact on the retrieval efficiency in large-scale image retrieval. It is very simple to use: you only need to input the destination information or its orientation information, its surrounding food and beverage information, gas stations, and tourist areas, and a series of information can be very clearly displayed, so this electronic tour guide can be accepted by many tourists and can be widely put into use [8]. However, for a picture, which presents the scenic spots that tourists are interested in and want to reach, it is difficult to get the corresponding tourist information and is even unable to determine the destination and make a travel plan [9]. There exist many tour guide tools as well as tour guide systems that are applied in the travel process. There is a lack of a system that can be used to determine travel locations and facilitate travel planning for tourists.

In large-scale image retrieval, usually, a low-dimensional indexing method needs to be constructed for the feature vector after extracting the global features of the image, and most of the indexing methods constructed based on the description of global features are using the nearest neighbor or approximate nearest neighbor retrieval methods [10]. One is the classical K-D (k-dimensional) tree method, which is a classical nearest neighbor retrieval method, mainly based on the original feature data to build a hierarchical index structure, and we then search from the root node of the tree to find the child node like the query image; this method is better in the retrieval efficiency when the dimensionality is relatively low than when the data dimensionality is high [11]. The retrieval complexity is relatively high. The other one is to perform cluster quantization analysis on feature vectors. Fujita used$k$-means to perform cluster analysis on the established tree structure, which results in lower dimensionality [12]. Based on this, Keane et al. proposed a hierarchical k-means tree index structure to compare the similarity between leaf nodes, which is suitable for large-scale image retrieval [13]. Another method is the locality-sensitive hashing- (LSH-) based method proposed in recent years to use hash learning in image retrieval, which is an approximate nearest neighbor search method [14] and improves the user's interactive experience, thereby helping to better develop the tourism industry. The analysis of the ideal intelligent scenic spot construction system needs to start from the perspective of tourists. The main method is to map the original space to the Hamming space and construct multiple hash functions under the corresponding conditions, i.e., the probability that two vectors that are adjacent or similar in the original space are similar enough after they are mapped to the Hamming space by hashing. The hash function can encode the high-dimensional vectors into binary form, which has obvious advantages over the K-D tree method in large-scale image retrieval, and the LSH method occupies less storage space and has more obvious advantages when the data dimension is larger [15].

Firstly, we introduce the background and significance of the current research on the mobile tour guide, analyze the current situation of the system at home and abroad, and introduce the research content of this paper given its problems; we design and implement an intelligent tour guide system. The overall design framework is outlined; then, the design ideas of each submodule in the framework are expanded and introduced, specifically including image acquisition and visual feature extraction on the mobile side, global feature generation on the mobile side, distributed inverted index construction on the server-side, attraction location and information push based on real-time image retrieval, etc.; finally, the release and operation of the designed system are introduced. Based on the interest preferences of the target user's most similar neighbors, the target user's preference level towards the recommended object is predicted, and the system makes recommendations to the target user according to the preference level. The biggest advantage of this recommendation system is that there is no special requirement for the recommended objects, and it can recommend objects that are difficult to represent in text structure. However, since the user rating data for the recommended objects are small compared to the total number of items, this type of system has the problem of data sparsity.

3. Multiple Linear Regression Equation for Large-Scale Tourist Attraction Image Retrieval Design

3.1. Multiple Linear Regression Equation Algorithm Analysis. Nonnegative matrix factorization (NMF) is the decomposition of a given original matrix into the product of two matrices, both of which are nonnegative, and the result of multiplying the decomposed two matrices together is approximately equal to the original matrix.

$$R \approx PQ^2. \quad (1)$$

Suppose R is an $m \times n$ matrix, and P and Q are $m \times k$ -dimensional and $k \times n$-dimensional nonnegative matrices, respectively. k generally satisfies $(n \times m)k \leq mn$. Because the decomposed matrix adds the nonnegative constraint, the product of the two decomposed matrices is hardly equal to the original matrix, and the decomposed matrix can only be made equal to the predecomposition matrix as much as possible. Therefore, the nonnegative matrix decomposition is transformed into the following optimization problem [16]:

$$\begin{cases} \phi_{\min}(R, P, Q), \\ P \leq 0, \\ Q \leq 0, \end{cases} \quad (2)$$

$$Q = \begin{cases} y_i = b_0 - \sum_{j=1}^{9} b_j x_{ij}, \\ \xi_i^{iid} = N\left(0, \sigma^2\right). \end{cases} \quad (3)$$

The Laplacian of Gaussian- (LoG-) based interest point detection algorithm used by CDVS uses a polynomial to approximate the LoG filtering effect, which is called the low-order polynomial (ALP).

$$p(x, y, \sigma) = \sum_{k=0}^{k-1} a_k L_k(x, y)\sigma^3 - \sum_{k=0}^{k-1} a_k L_k(x, y)\sigma^2. \quad (4)$$

The four-octave image is filtered by Laplacian of Gaussian filtering normalized to the scale of the original image. The ALP algorithm uses the point with the first-order derivative of the polynomial at scale 0 as the point of interest by computing the polynomial and later compares the polar values with the eight neighboring pixel points in the plane around the point of interest, as shown in Figure 1.

The stratified sampling statistical model classifies the samples in strata according to the characteristic distribution of the overall units to reduce the differences within each stratum while increasing the differences between strata. On this basis, a certain number of samples are drawn from each stratum separately to portray the distribution of the stratum and constitute the overall sample. Since the samples are reasonably stratified, the stratified sampling statistical model can better capture the travel preferences of users and lay the foundation for a more accurate generation of recommendation lists [17]. The server-side application is developed based on C++ technology. The server-side program is mainly responsible for receiving the scenic spot retrieval request from the mobile terminal and completes the accurate retrieval of the scenic spot according to the image characteristics and GPS information sent by the mobile terminal.

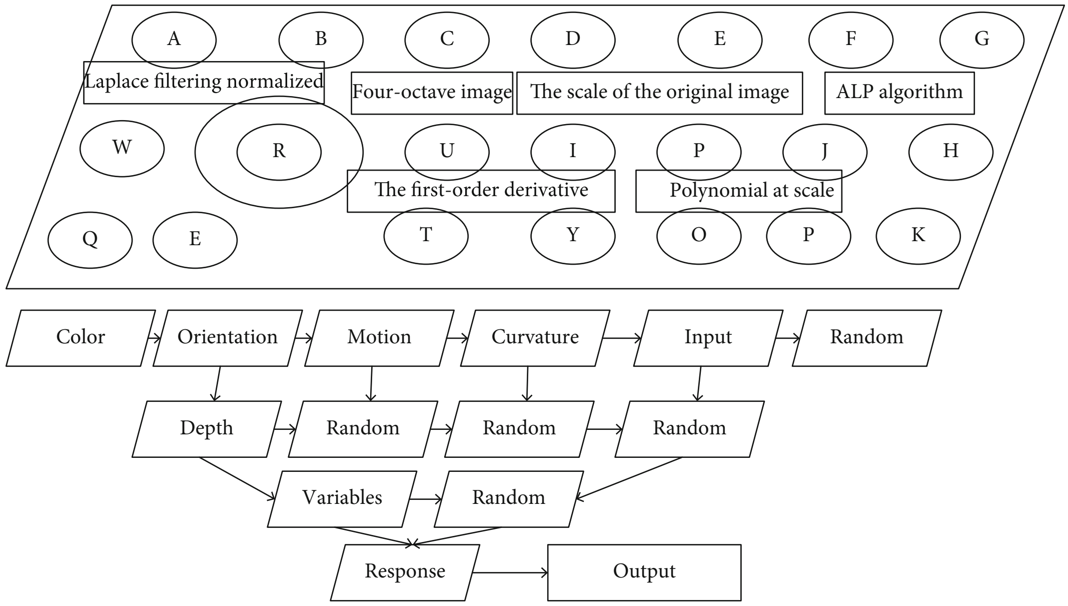


Figure 1: Algorithmic framework of multiple linear regression equations.

Setting the weights of sampling statistics based on the subjective assignment evaluation method, hierarchical analysis is applied to adjust the weights of different user attributes; i.e., the relative importance of the same level is compared to establish a new discriminant matrix, and the weights of each user attribute are determined according to the discriminant matrix.

$$G = \begin{bmatrix} G_1 & G_2 & G_3 \\ 1 & \frac{1}{2} & 2 \\ 2 & 1 & \frac{1}{2} \end{bmatrix}. \tag{5}$$

An image $g(x, y)$ of size is filtered with a two-dimensional Gabor filter set of scale p and direction q. This step is essentially a convolution operation with the image using this filter set, respectively, where the mathematical expression of the two-dimensional filter can be expressed as follows:

$$g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp \cos\left(2\pi f_0 - \varphi\right). \tag{6}$$

Texture features are another important global descriptive feature. Texture features describe the local spatial distribution in an image, including information about local light intensity, and are often used to distinguish between images that are rich in information, have similar colors, or are not easy to segment. This feature can describe the overall texture of the image, based on the relationship of the grayscale image response to the full image region values, will not be affected by local extremes and cause mismatching, and has rotational invariance and is not significantly affected by noise, and its research results have been applied to many important fields. Due to the above properties of texture features, texture features can play a good role in distinguishing landscapes and buildings in different countries and regions. Texture features of images are usually reflected in the form of a cogeneration matrix, and the grayscale cogeneration matrix reflects the grayscale of graphics in the direction, the correlation between adjacent pixels, and the transformation amplitude value, which is a more common method to analyze the distribution of local texture of images.

$$P(i, j) = f(x_1, y_1). \tag{7}$$

The BOF model is a feature that is widely used in the field of image processing. Drawing on a document representation method previously used in the retrieval of information such as text, the BOF model replaces text in the text retrieval model with images and uses the same idea to classify or retrieve images. In the bag-of-words model for text information retrieval, for a document, regardless of the word order and sentence syntax in the content of that document, it is treated as a combination of many individual words only, and the occurrence of each word in the document is random; i.e., the occurrence of any word in the document is independent of other words, and the content in the document is disordered.

$$F_{S_i} = \frac{s^2}{mn} \sum_{(x,y)\in S_i} F_{S_i}(x, y). \tag{8}$$

s represents the area of the graph. If we compare an image to a document, such that the set of images is equivalent to the set of documents and the features of the images are equivalent to the words in the documents, i.e., the image can be understood here as an ensemble of many "visual words," and there is no order among these visual words.

Then, we can apply the methods for text retrieval to image retrieval, for example, in the field of large-scale image retrieval, and use the efficiency in text processing to improve the speed in large-scale image retrieval.

For images, the "words" in images are not like the words that usually appear in the text, but images are usually multidimensional datasets. Therefore, the first thing we need to do is to extract independent "visual words" from the images, which constitute a visual vocabulary [18]. The commonly used method for extracting features is SIFT-based feature extraction, which is suitable for feature extraction to form a vocabulary vector because of the good uniqueness of the features extracted by the SIFT algorithm and the richness of the image information it contains. Using the SIFT algorithm, visual vocabulary is extracted from each type of image set, and all the visual vocabulary is formed into a visual vocabulary vector, as shown in Figure 2.

Due to the natural properties of the image, the statistical properties in one part of the image are the same as in other regions of the image. So, we can use the features learned in one part of the image in other regions, which enables parameter sharing. For example, when we select an arbitrary area from an image and get a small sample, we learn a set of features from this small sample, and then, this set of features is applied to any other part of the image, and the features are used as a detector to convolve with the original image, and we can get different feature values at different locations of the image. In a convolutional neural network, the convolution kernel of each convolutional layer does the convolution operation on the image to get the features of a certain aspect of the whole image. Each convolution kernel will share the same parameters so that when feature extraction is performed on the image, it is not necessary to know the location of the local features of the image, and it also reduces the parameters of the network convolutional neural network.

$$f(z) = \frac{1}{1 - e^{-z}}, \tag{9}$$

$$z_j^{l+1} = \lim_{S_l \longrightarrow \infty} \sum_{i=1}^{S_l} a_i^l W_{ij}^{l+1}, \tag{10}$$

$$a_j^{l-1} = f\left(z_j^{l-1}\right). \tag{11}$$

After the convolutional features are obtained by the above calculation, they can theoretically be used directly as the feature basis for training the classifier, but they are very computationally intensive and prone to overfitting. Considering the inherent properties of natural images, the features of one region are also applicable to another region; therefore, the features at different locations can be aggregated statistically, and the average or maximum value of features in a local region can be calculated to characterize the specialization of the whole region. The edge extraction algorithm is generally adopted in the implementation process, and the edges are connected to extract the shape of the object. Natural landscape images generally do not have regular shapes, and the shape features extracted from landscape images that exist in cities are not representative and do not distinguish well between different regions. This feature extraction method does not have application value compared with other features in the study of this paper.

$$r_{u,i} = \alpha - \beta_u - \beta_i, \tag{12}$$

$$D_S = \{(u, i, j) | u \in U\}. \tag{13}$$

The server-side module is the key to the operation of the whole system, which first accepts the images sent from the mobile client and extracts features from the images, extracting RGB color features, texture features, and GIST features of the images, respectively, and using the perceptual hash algorithm to rank the features of the images and perform image matching according to the feature values. By combining various basic image features and filtering them at different levels, the content-based image retrieval is completed, and a more desirable retrieval result can be obtained. In the second stage of "fine sorting," the RANSAC algorithm described in the previous section is generally used, but this algorithm is computationally intensive and takes a long time to compute.

$$S_{\mathrm{LDR}} = \left\{ \log \left(\frac{\mathrm{dict}\left(l_{q,i}, l_{q,m}\right)}{\mathrm{dict}\left(l_{q,i}, l_{q,m}\right)} \right) \middle| (i.j) \in M \right\}. \tag{14}$$

These scores are then quantified into a histogram, and later, the largest value in the histogram is taken as a measure of the score between the two images, with higher scores indicating greater similarity, and then, the candidate images are reordered to obtain the result.

$$C_{\mathrm{LDR}}(\alpha) = \sum_{z \in S_{\mathrm{LDR}}} I\left(\frac{\alpha}{c} \le z \le \frac{\alpha - 1}{c - 1}\right). \tag{15}$$

To ensure the security of information, first, the user logs in, and the server-side makes a judgment based on the username and password entered by the user, whether the login is successful or not. After successful login, users can select images for query and retrieval: images can come from local albums, or users can take pictures of images they are interested in through cell phone cameras, and after selecting the image to be retrieved, send the image to the server, and check the results returned from the server check.

3.2. Experimental Design for Large-Scale Tourist Attraction Image Retrieval. In the process of guiding, the user is the main one, and the design of the tourist guide system must meet the needs of the user and provide maximum satisfaction to the user. The designed tourist guide system should meet the following design principles. The user is mobile, and he can stay at any location in any attraction. When using a cell phone, wireless communication is needed between the user side and the central processing site, which allows the user to better experience the convenience [19]. The aggregation process of calculating the feature average or maximum value of a certain local area to characterize

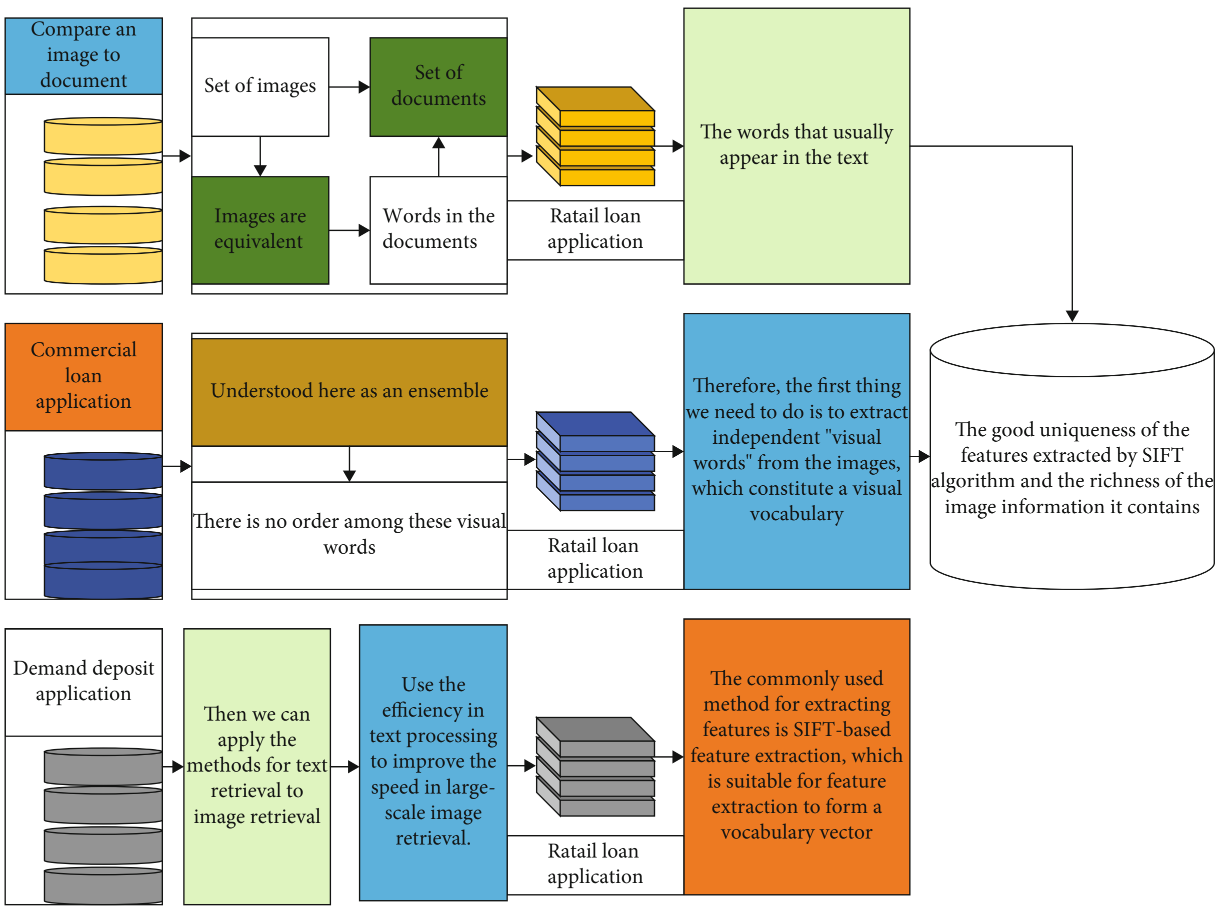


FIGURE 2: Schematic diagram of local connection.

the specialization of the entire area is called pooling or downsampling. Real time is reflected in two aspects. One is the rapid identification of the user's location because the user is mobile; if the user's location is not quickly identified, it may appear that the user has moved to the next attraction; the system only identified the last attraction, which is not feasible. Another is to respond promptly to changes in attraction infrastructure maintenance, environmental changes, and changes in surrounding stores. The user community includes people of all ages and knowledge levels, which means that the user-side interface should be user-friendly and able to be operated effectively by the elderly and even older children.

Once the tour guide system based on mobile visual search technology is put into operation, the number of users will proliferate, and many unavoidable MIS operations will occur during the use process according to the different cognitive levels of different users. A reliable and stable tour guide system based on mobile visual search technology must solve the MIS operations of users in time to ensure the stable and normal use of the system. It is worth noting that real time and accuracy are trade-offs. This is because to improve accuracy, a large amount of computational processing is required, which increases processing time and decreases real-time performance. The reverse is also true; if real time is to be improved, fast recognition is required, which will reduce accuracy, as shown in Figure 3.

As an innovative concept, smart tourism has been accepted by increased tourism enterprises. In addition to the ticketing system, the scenic guide system is also a breakthrough in smart tourism, which gives tourists a better touring experience while also realizing the transparent display of scenic information. The scenic guide system needs to combine artificial intelligence, mobile Internet, Internet of Things, and other technologies so that tourists can get one-to-one in-depth guide service through cell phones, to meet the demand of tourists to find information about scenic spots; to help scenic spots to achieve panoramic display, voice explanation, route planning, information transfer, and other integrated guide services; and to improve the quality of service in scenic spots and improve the tour experience of tourists [20]. Scenic areas cannot implement accurate marketing, high conversion costs, and low tourist independent experience. Scenic areas have a high cost of the manual guide, low efficiency, and no intelligent guide service for foreign tourists. Using audio, video, pictures, text, etc., as the main presentation methods, the scenic spot information is displayed to tourists, and the problems of passenger flow guidance, information lag, and play guides are solved. The intelligent guide system helps scenic spots to provide intelligent self-service for tourists. Through the network control system formed by the electronic guide hardware equipment and the central database in the background, the information

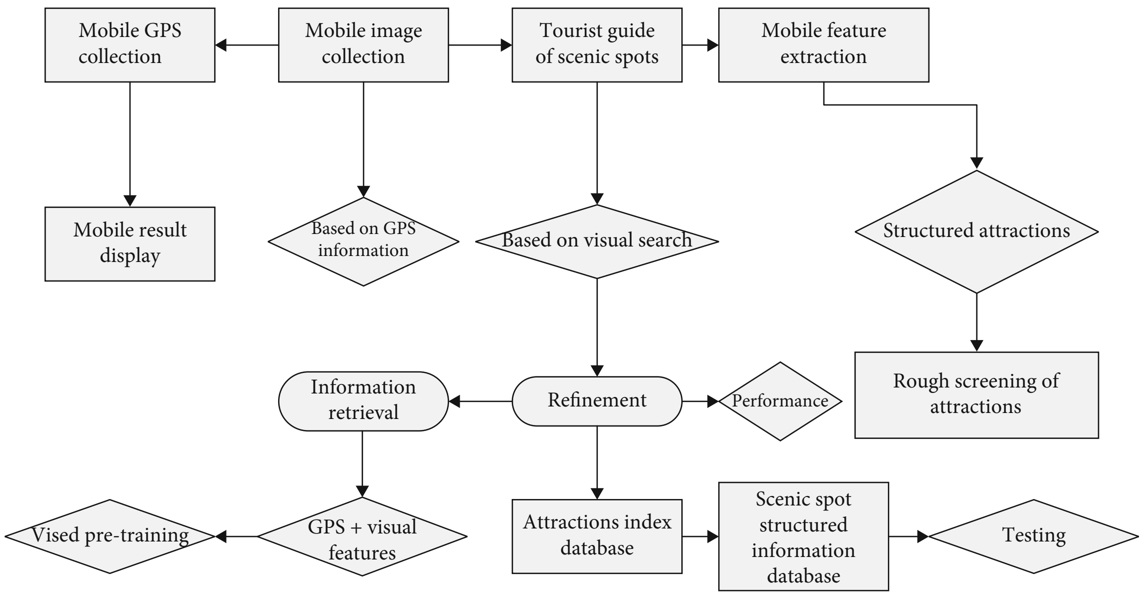


Figure 3: Overall framework of the tour guide system.

of scenic spots is displayed to tourists with audio, video, pictures, and text as the main presentation methods, solving the problems of passenger flow guidance, information lag, and tour guides. From the whole situation, it helps the scenic spot to guide the flow intelligently; prevent the tourist attractions, line congestion, and other problems; and improve the tourists' touring experience.

The information processing on the server-side consists of two main steps: the first one is to retrieve attractions, and the second one is to add attraction-related information. Based on these two steps, the database is divided into an attraction index database and an attraction structured information database. Using the attraction index database, the attraction retrieval step is performed by the GPS information-based attraction cluster filtering module and the visual search-based refinement module. Based on the retrieved attraction information and the attraction structured information database, the attraction structured information retrieval module adds attraction-related information and downlinks this information to the mobile terminal. The information display on the mobile terminal is done by the mobile terminal result display module and the attraction tour guide module. The mobile terminal result display module gives basic information about the attraction, and the attraction tour guide module introduces further information about the attraction, such as the humanities and the surrounding business information related to the attraction. In this paper, a grayscale co-occurrence matrix is selected and applied to the system. This method counts the relationship between the grayscale in each direction of the image, and between different pixels, and reflects the characteristics of the spatial distribution of the image color and light intensity, as shown in Figure 4.

In this paper, a rapid and accurate method is designed using the acceleration based on a hash algorithm for various descriptors, different types of classifiers based on a human visual perspective based on the characteristics of different image features, and the calculation of image similarity using scientific methods, which improves the accuracy of image retrieval and makes this system more application-worthy. Image retrieval is performed using annotated images downloaded on Flickr, and the data given in the experiments are tested on 2000 randomly selected images [21]. All the images are divided into 16 image collections according to their features to speed up their retrieval by classifying the images. Based on human vision, this paper classifies the images using the grayscale correlation of the images and the entropy value of the images, i.e., the first classification of the images using the texture distribution of the images and the complexity of the images.

4. Analysis of Results

4.1. Analysis of the Multiple Linear Regression Algorithm. Based on human vision, this paper uses the grayscale correlation of images and the entropy value of images to classify images, i.e., the first classification of images using the texture distribution of images and the complexity of images. When using correlation for classification, this paper divides the original dataset into 4 subsets, and after many data tests, the boundary values that can divide the subsets are obtained. After evaluating all the image correlation values accordingly, a range of approximate values is obtained, and 3 values should be finally determined to classify the subsets. To facilitate the statistics of the data, a value greater than a certain value is used for comparison, and the number of images such as 500, 1000, and 1500 is obtained, and the equal cases are divided into the next subset for statistics. Based on this, a second division is performed, the same as the previous method, and the final range is determined by continuous experiments with different values. The results of the test on the image set were also extremely accurate, and according to the correlation shown in Figure 5, 1001 images met the

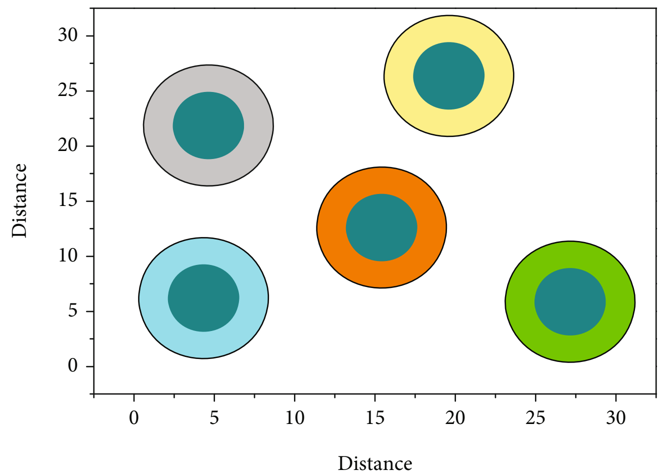


FIGURE 4: Feature map.

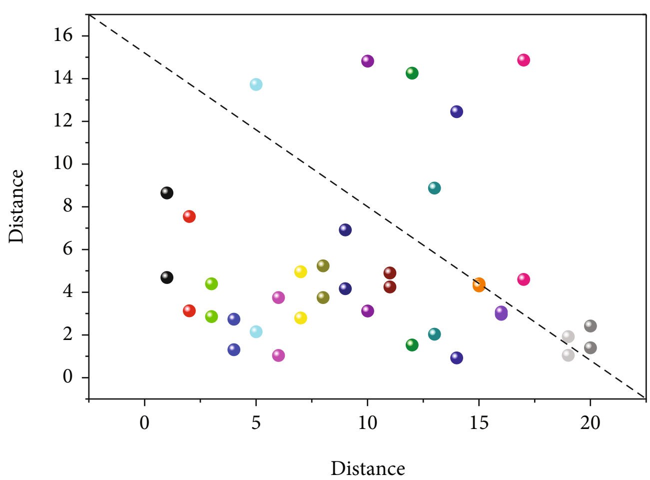


FIGURE 5: Multiple linear regression performance.

correlation with a value greater than 0.156 so that the division of the first two groups was completed.

For values greater than 0.076, 1501 images are obtained. Finally, four subsets containing 500 images each can be obtained, and experiments on larger datasets can also yield better results. Based on the above data, the values determined in this paper can satisfy the average distribution of the image sets. In the next step, the entropy value is used for further classification, and the subset obtained in the previous step is again divided into four subsets. Both the average precision rate and the average recall rate are decreasing, mainly since the amount of data in the image database of different scenic spots is different and the number of similar pictures in the image database is different from that of the query image. The accuracy rate is slightly reduced. The entropy value can effectively describe the complexity of the images, so it can play a good role in classifying the natural environment and urban buildings, people, animals, etc. However, if a lot of human factors are added when the images are taken, it will affect the classification effect of this layer, and this effect will be reduced in the case of very large image sets. Based on the normalization of the color features, the similarity calculation is performed simultaneously in groups of 8 numbers. In the process of calculation, if the specified degree of similarity is not satisfied, the image calculated now is discarded and the color information of the next image is calculated so that the efficiency and quality of the returned information can be guaranteed at the same time. A similarity level of 45 does result in fewer results, but more than half of the images are not like the original. For example, if the main scene depicted in the original image is a mountain peak, then more than half of the images have a mountain peak. In this step of the retrieval process, the similarity is also used to determine the number of images, which is passed into the third feature of the retrieval.

In large-scale tourist attraction image retrieval, we need to query the accuracy and completeness of different attractions to calculate the average accuracy and average completeness of multiple attractions, which is used to characterize the overall performance of query results. For a total of 400,000 images of 1740 attractions online, we return 30 similar images for each attraction in order of similarity; query 100, 200, 500, 700, and 1000 attractions, respectively; and calculate the average accuracy and average completeness of the returned results belonging to the queried attractions, as shown in Figure 6.

From Figure 6, we can find that as the number of query sites increases, the average search accuracy and average search completeness decrease, mainly due to the different data volume of different sites in the image database and the different number of images like the query images in the image database, which leads to a slight decrease in accuracy. However, we can also find that the overall decrease of the average accuracy and average completeness is not significant, so we can consider that the method used in this paper is stable for large-scale tourist attraction image retrieval.

4.2. Graphical Retrieval Results of Large-Scale Tourist Attractions. In Figure 7, the recall rate of the improved

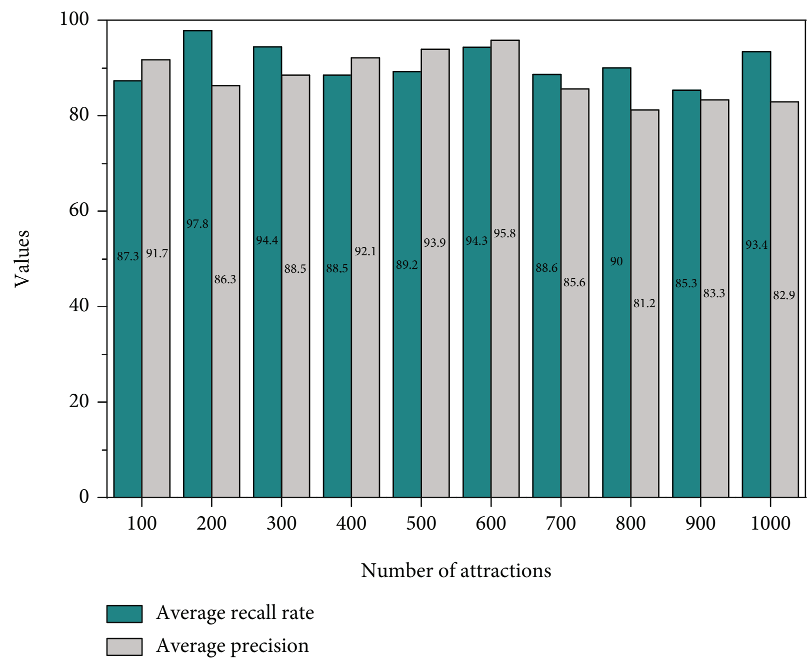


Figure 6: Average accuracy rate and average completeness rate.

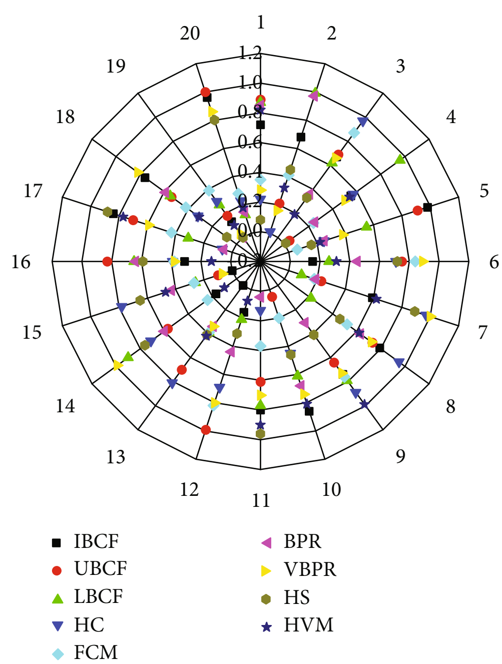


Figure 7: Comparison of recall values for recommended systems.

VBPR model is significantly better than the comparison baseline for the same reason as above. As the number of recommended attractions increases, its recall rate steadily increases and is less volatile compared to NMF and KNN. This indicates that the improved VBPR model is better at recommending multiple attractions. The hybrid recommendation (HVM) makes full use of the user travel preference information and the recommendation results of the improved VBPR model, and its recommendation performance is optimal; i.e., the user travel preferences better smoothen the final recommendation results. The VBPR model plays the main role, while the user travel preference information based on the stratified sampling statistical model plays a secondary role.

The recommendation system of tourist attractions based on stratified sampling statistics and an improved VBPR model can effectively improve the recommendation performance and meet the user requirements to a greater extent, thus alleviating the data sparsity problem. And the stability of recommendations is better. Although the improved VBPR model can complete the recommendation well, the image features used in this chapter are all independent features, and the multimodal semantic correlation between different image features is not fully explored, so the recommendation performance still needs to be improved.

When the feature extraction is finished, APP will generate a string with the extracted features and send it to the server. Since the retrieval result returned by the server is the URL of the image, when the APP receives the returned result, it sends a request to the server again to get the image and then displays it. Each network request here requires a multithreaded operation to ensure operational efficiency and user experience at the same time. Multithreading plays a very important role in this part. Using multithreaded programming allows the program to perform multiple tasks at the same time, converting the interface when extracting features and displaying the results of extracted features when sending a request at the same time. Compared with other methods, our method improves the efficiency and accuracy by about 8%. This multistep simultaneous approach ensures that the overall system runs more efficiently and that the program does not get stuck due to certain operations that take a long time to compute.

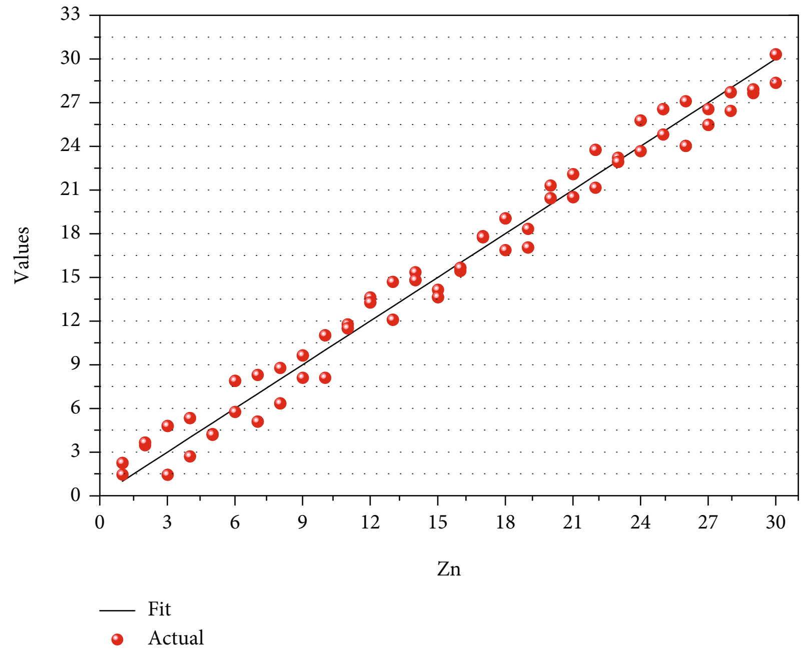


FIGURE 8: Search accuracy results.

This multistep simultaneous method can ensure the improvement of the overall operating efficiency of the system and objectively describe the rich semantic information contained in the scenic spot images. Therefore, it is necessary to deeply explore the multimodal correlation between different image features to better portray the visual content of attraction images from the perspective of multimodal feature fusion. The DCA-VBPR model plays this role well, as it utilizes label information and suppresses correlations between classes in the feature mapping process, so different image features are fully fused, feature discriminability is continuously improved, content analysis of objects is very successful, and recommendation performance is improved. The DCA model can generate high-quality recommendation results compared with scheme 1 and scheme 2. In the image upload interface, we select an image to upload for retrieval; at this time, we select the familiar Badaling Great Wall attraction image; we click the upload button; the image is uploaded to the server; the server will, according to the received image, carry out the corresponding feature extraction and use the extracted features to carry out the corresponding matching; and the initial matching result will be filtered to get the retrieval result, and the detection result will be returned to the cell phone. The detection results are viewed in the View Results option, and we click the View Results button, and there will be the results of the images we have queried, as shown in Figure 8.

Visitors take pictures of the attractions by using the mobile client, use CDVS descriptors to extract the corresponding texture features from the pictures, and upload the feature information and GPS information to the server. Based on the received information, the server performs image matching using CDVS feature-based visual retrieval and rearrangement techniques, performs image retrieval based on the matched results, filters the matched results, and returns them to the mobile client. A front-to-back separation method is used in the server module, which improves retrieval efficiency. This paper's travel guide system, based on mobile visual search technology, facilitates the travel needs of tourists, eliminates unnecessary expenses, makes travel easier and faster, and improves the quality of tourism in general.

5. Conclusion

This paper applies the retrieval of large-scale tourist attraction images. The WAN method is a more classical algorithm; of course, these two algorithms have certain shortcomings; for example, the fully connected layer in the convolutional network structure, due to too many parameters in the fully connected layer, often leads to overfitting; in the locality-sensitive hashing, we hope that the adjacent data can be integrated into the same bucket after hashing, while the nonadjacent data fall into different buckets. However, in practice, the adjacent data will fall into different buckets and the nonadjacent data will fall into adjacent buckets, which we do not want to see. Therefore, we can improve the accuracy of retrieval and reduce the error rate by improving the structure of optimized convolutional networks (e.g., using more network layers and reducing the number of fully connected layers) and the structure of locality-sensitive hashes (e.g., using more hash pylons within a hash table or building more hash tables). In summary, on the one hand, the accuracy of image retrieval can be improved by building a larger and better image database of tourist attractions, and on the other hand, the error rate of image retrieval can be reduced by continuously optimizing the feature extraction algorithm and index structure, to optimize the image retrieval of large-scale tourist attractions.

References

[1] M. Ren, H. Q. Vu, G. Li, and R. Law, "Large-scale comparative analyses of hotel photo content posted by managers and customers to review platforms based on deep learning: implications for hospitality marketers," *Journal of Hospitality Marketing & Management*, vol. 30, no. 1, pp. 96–119, 2021.

[2] D. Li, L. Deng, and Z. Cai, "Statistical analysis of tourist flow in tourist spots based on big data platform and DA-HKRVM algorithms," *Personal and Ubiquitous Computing*, vol. 24, no. 1, pp. 87–101, 2020.

[3] Y. Liang, S. Gao, Y. Cai, N. Z. Foutz, and L. Wu, "Calibrating the dynamic Huff model for business analysis using location big data," *Transactions in GIS*, vol. 24, no. 3, pp. 681–703, 2020.

[4] C. Tosun, B. B. Dedeoğlu, C. Çalışkan, and Y. Karakuş, "Role of place image in support for tourism development: the mediating role of multi-dimensional impacts," *International Journal of Tourism Research*, vol. 23, no. 3, pp. 268–286, 2021.

[5] S. du, S. du, B. Liu, X. Zhang, and Z. Zheng, "Large-scale urban functional zone mapping by integrating remote sensing images and open social data," *GIScience & Remote Sensing*, vol. 57, no. 3, pp. 411–430, 2020.

[6] Q. Wang, J. Gao, W. Lin, and X. Li, "NWPU-Crowd: a large-scale benchmark for crowd counting and localization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 6, pp. 2141–2149, 2021.

[7] S. Giglio, F. Bertacchini, E. Bilotta, and P. Pantano, "Machine learning and points of interest: typical tourist Italian cities," *Current Issues in Tourism*, vol. 23, no. 13, pp. 1646–1658, 2020.

[8] J. Kang, R. Fernandez-Beltran, Z. Ye, X. Tong, P. Ghamisi, and A. Plaza, "Deep metric learning based on scalable neighborhood components for remote sensing scene characterization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 12, pp. 8905–8918, 2020.

[9] K. L. van Leeuwen, R. A. Hill, and A. H. Korstjens, "Classifying chimpanzee (Pan troglodytes) landscapes across large-scale environmental gradients in Africa," *International Journal of Primatology*, vol. 41, no. 6, pp. 800–821, 2020.

[10] N. Tadi Bani and S. Fekri-Ershad, "Content-based image retrieval based on combination of texture and colour information extracted in spatial and frequency domains," *The Electronic Library*, vol. 37, no. 4, pp. 650–666, 2019.

[11] X. Li, "Space-time distribution model of visitor flow in tourism culture construction via back propagation neural network model," *Personal and Ubiquitous Computing*, vol. 24, no. 2, pp. 223–235, 2020.

[12] H. Fujita, "AI-based computer-aided diagnosis (AI-CAD): the latest review to read first," *Radiological Physics and Technology*, vol. 13, no. 1, pp. 6–19, 2020.

[13] A. Keane, J. F. Lund, J. Bluwstein, N. D. Burgess, M. R. Nielsen, and K. Homewood, "Impact of Tanzania's Wildlife Management Areas on household wealth," *Nature Sustainability*, vol. 3, no. 3, pp. 226–233, 2020.

[14] F. Z. Benkhattab, M. Hakkou, I. Bagdanavičiūtė et al., "Spatial–temporal analysis of the shoreline change rate using automatic computation and geospatial tools along the Tetouan coast in Morocco," *Natural Hazards*, vol. 104, no. 1, pp. 519–536, 2020.

[15] M. Ligero, O. Jordi-Ollero, K. Bernatowicz et al., "Minimizing acquisition-related radiomics variability by image resampling and batch effect correction to allow for large-scale data analysis," *European Radiology*, vol. 31, no. 3, pp. 1460–1470, 2021.

[16] S. E. Rigby, T. J. Lodge, S. Alotaibi et al., "Preliminary yield estimation of the 2020 Beirut explosion using video footage from social media," *Shock Waves*, vol. 30, no. 6, pp. 671–675, 2020.

[17] N. Hor and S. Fekri-Ershad, "Image retrieval approach based on local texture information derived from predefined patterns and spatial domain information.," vol. 8, no. 6, pp. 246–254, 2019, https://arxiv.org/abs/1912.12978.

[18] K. A. Shinde, "Religious theme parks as tourist attraction systems," *Journal of Heritage Tourism*, vol. 16, no. 3, pp. 281–299, 2021.

[19] S. O. Petrovan, C. G. Vale, and N. Sillero, "Using citizen science in road surveys for large-scale amphibian monitoring: are biased data representative for species distribution?," *Biodiversity and Conservation*, vol. 29, no. 6, pp. 1767–1781, 2020.

[20] M. T. Falk and E. Hagsten, "Visitor flows to World Heritage Sites in the era of Instagram," *Journal of Sustainable Tourism*, vol. 29, no. 10, pp. 1547–1564, 2021.

[21] K. E. Callahan, C. J. Clark, A. F. Edwards et al., "Automated frailty screening at-scale for pre-operative risk stratification using the electronic frailty index," *Journal of the American Geriatrics Society*, vol. 69, no. 5, pp. 1357–1362, 2021.

Game Models on Optimal Strategies in a Tourism Dual-Channel Supply Chain

Lei Yang,[1] Jingna Ji,[1] and Kebing Chen[2]

[1]*School of Economics and Commerce, South China University of Technology, Guangzhou, Guangdong 510006, China*
[2]*Department of Mathematics, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu 210016, China*

Correspondence should be addressed to Lei Yang; yang@scut.edu.cn

Academic Editor: Luca Guerrini

This paper explores a two-echelon tourism supply chain consisting of a hotel and an online travel agency. The upside hotel rooms can be sold through the downside hotel alliance and online travel agency. The hotel alliance, selling rooms at a lower price, is a direct sale platform with a negligible entry fee. Notwithstanding, the online travel agency sells the room at a higher price with related personalized service. Customers will be refunded partially in case of their cancellation or no-show. An integrated model and two decentralized models based on Bertrand and Stackelberg games are developed, respectively. The results show that when the wholesale price is lower than a certain value, both the hotel and the online travel agency can gain more profit from the Stackelberg game than that from the Bertrand game. In the case that the hotel allows overbooking, the optimal overbooking quantity is obtained. If the overbooking proportion is too high, overbooking is profitable for the hotel only when the overbooking cost is lower than a certain value. At the end of the study, some experiments are conducted to analyze the sensitivity of the optimal prices and profits in the light of certain parameters.

1. Introduction

The improvement of transport infrastructure has significantly reduced the trip cost and promoted the development of tourism [1]. In recent years, the tourism industry has evolved considerably and plays an essential role in local economic development. In the supply chain management, specifically, tourism has never been so much concerned by academic researchers and industry practitioners since it provides a new perspective for the tourism firms to improve their competitiveness. As we all know, supply chain management is a kind of process to increase the efficiency of a supply chain network via conducting and integrating fund, information, and goods flows from suppliers to end customers [2]. The tourism supply chain is defined as a network of tourism organizations engaged in different activities ranging from the supply of tourism products or services such as flights and accommodation to the distribution and marketing of the final tourism product at a specific tourism destination [3]. In China, for example, the hotel industry is one of the fastest growing worldwide, with revenue increasing 1.8 times during the period from 2000 to 2009 [4]. Therefore, we study the tourism supply chain involving hotels.

The development of Internet has provided a strong incentive for manufacturers to engage in direct sales. The providers of tourism products are also inclined to build dual-channels. By the end of December 2015, China has had 688 million Internet users, among which there are 260 million online travel product users (http://cnnic.cn/gywm/xwzx/rdxw/2015/201601/W020160122639198410766.pdf). On the other hand, in 2014, the online tourism market transactions totaled 307.79 billion yuan, which increases by 38.9% comparing with the same period in 2013 (http://report.iresearch.cn/report/201504/2341.shtml). The Internet has been widely used by travelers as a crucial channel for booking hotel rooms [5]. Hotels, therefore, usually provide two booking channels, the online travel agency (OTA) and the hotel alliance (HA).

OTA, as an intermediary between hotels and customers, plays an essential role in the tourism supply chain, such as

Ctrip.com, eLong.com, and Mangocity.com. However, a common phenomenon is that hotels must pay high commission for each sold room [6, 7]. In order to reduce commission fees and sell rooms to customers directly, hotels join a HA, where the HA can be regarded as a direct sales platform. For instance, WeChat public platform and alitrip.com offer the direct selling platforms for hotels. Compared with developing official websites, joining HA can help the hotels reduce cost. Hotels have not been benefited much from website development investment [8]. Take, for example, Four Seasons hotel; it has invested 18 million for its new website while online revenue only increases by 2% in five years [9]. On the contrary, hotels can pay 0 to 15000 yuan per year to join the alitrip.com, so as to deal with orders through this website directly. Generally speaking, the reservation price through the alliance is up to 20% to 70% of the rack rate. In the light of the uncertain factors such as time, city, room type, and room quantity, the discount is not fixed. In most cases, the price of HA is 10% to 30% lower than OTA, except for a plan of promotion, with the extent to which the price will be equal to that of OTA. Pricing competition exists between the two channels. Therefore, how to determine the pricing decisions has much practical significance for hotels.

Customer cancellations and no-show occur frequently in the yield management of hotel reservations [10]. Once customers cancel the reservation or fail to show up, the hotel will suffer loss for the rooms. That is because rooms are perishable products. In order to deal with this case, the hotel usually permits overbooking. However, overbooking will bring a risk that the actual booking quantity is greater than the number of hotel rooms. As a result, the overbooking costs must be paid. A question therefore arises as to how a hotel determines the overbooking quantity so as to keep a balance between the vacancy lost and the overbooking costs.

Since more and more hotels employ dual-channel structure, we develop a two-echelon tourism supply chain model involving one hotel and one OTA to study hotel room pricing strategies, with which OTA provides related service. Bertrand and Stackelberg game models are both analyzed. Furthermore, this paper considers customers' no-show phenomenon in a dual-channel setting under tourism supply chain. The results of the study will assist hotels to manage and optimize their pricing decisions.

The remainder of this paper is organized as follows: Section 2 provides a literature review and Section 3 presents notations. Section 4 analyzes the integrated tourism supply chain. Section 5 shows the Bertrand and Stackelberg game models, respectively. The overbooking situation is illustrated in Section 6. Section 7 conducts numerical experiments to provide more insights. Conclusions and outlooks are finally presented in Section 8. To make the paper more readable, all proofs are presented in Appendix.

2. Literature Review

This research is closely related to tourism supply chain, pricing strategies in dual-channel supply chain, and revenue management. To highlight our contributions, we review only the literature that is representative and particularly relevant to our study.

2.1. Tourism Supply Chain Management. Tourism supply chain management has been studied since the 1990s. For example, Yang et al. [11] utilized game theory to investigate the cooperation and competition between two tourism supply chains, which consist of service providers, a theme park operator, and accommodation providers. They concluded that, in order to optimize performance, the decision makers of the two tourism supply chains are expected to adopt appropriate product differentiation strategies. Differently, we highlight the comparison between the Bertrand and Stackelberg game in a two-echelon tourism supply chain. Huang et al. [12] did research on competition strategies in a tourism supply chain network consisting of theme parks, accommodation providers, and tour operators, which are involved in producing and providing package holidays. They analyzed the differences between the impacts of quantity and price competitions. Huang et al. [13] studied the impact of the involvement of tour operators in a tourism supply chain with multiple hotels and travel agencies and showed that when the market size of travel agencies is lower than a certain level, both travel agencies and the hotel can benefit more from the presence of a tour operator in a tourism supply chain. Similarly, we focus on the supply chain involving hotels and tour operators. However, our contributions are directed to formulating a dual-channel environment by considering the no-show phenomenon. Furthermore, Lee and Fernando [14] developed a model for the medical tourism supply chain and showed that supply chain coordination and information sharing have a direct effect on organizational performance. We aim to contribute to this stream of literature by considering hotels' pricing decisions under dual-channel structure.

2.2. Pricing Strategies in Dual-Channel Supply Chain. The research on dual-channel supply chain management has gained much attention among the marketing and supply chain management. In particular, the studies focusing on pricing strategies are closely related to our study. In this aspect, Chiang et al. [15] make a great contribution. They constructed a price-setting game between a manufacturer and its independent retailer. They pointed out that the direct sales channel can help the manufacturer improve overall profitability by reducing the degree of inefficient price double marginalization. Tsay and Agrawal [16] found that the manufacturer can coordinate the supply chain and achieve a win-win situation through adjusting the direct price. Q.-H. Li and B. Li [17] studied a dual-channel supply chain in which the retailer provides value-added services to products. Their results indicate that the entire supply chain cannot be coordinated with a constant wholesale price when the retailer provides value-added services and has fairness concerns. Similarly, we also assume that the OTA adds additional value to the tourism product. Panda et al. [18] explored pricing and replenishment policies for a high-tech product in a dual-channel supply chain that consists of a brick-and-mortar channel and an internet channel.

Liu et al. [19] investigated the pricing decisions under dual-channel structure. In their model, fairness and free-riding behavior are considered. They detected that manufacturer's equilibrium price based on channel fairness is below its equilibrium price ignoring fairness. More related articles on this issue have been presented by Chen et al. [20], Ding et al. [21], and Rodriguez and Aydin [22].

2.3. Revenue Management. Revenue management, which was originated in the airline industry, is also widely used in the hotel industry. That is because airline and hotel have similar characteristics (e.g., perishable service or product, fixed capacity, distinct customer segmentation, and price differentiation). In this aspect, the room pricing decision has been studied extensively since price has an important influence on customer's accommodation selection decisions [23]. It is also relevant to our study.

Chen et al. [24] examined the impact of hotel pricing on guest satisfaction. Their results imply that, at the low price level, room price and food and beverage price lead to an increase in guest satisfaction, whereas the high price level could have just the opposite effect. Ling et al. [6] considered the case that customers can make reservations directly through the distribution channel of the hotel or indirectly through the OTA. Although the background of the study is similar to that in our model, they aim to find a method for hotels to manage their room availability for cooperative OTAs. Espinet et al. [25] analyzed how hotel characteristics affect seasonality in prices and find that more hotel services and higher star ratings are associated with fewer seasonal variations in hotel prices. Juaneda et al. [26] investigated the price component of physical characteristics and the location of apartments and hotels and compare their effect on the final price of both types of accommodation. Abrate et al. [27] collected data from almost 1000 hotels in Europe to analyze the dynamic pricing strategies of hotels; they showed that the intertemporal pricing structure primarily depends on the type of customer, the star rating, and the number of suppliers with available rooms.

The extant literature about revenue management is mostly from the perspective of the hotel itself. However, nowadays, the competition among tourism enterprises turns to be gradually and distinctly embodied in the competition among their tourism supply chains. In contrast to the above papers, our model is more practical by considering pricing strategies from the perspective of tourism dual-channel supply chain.

This paper makes the following contributions. Firstly, this paper introduces the tourism products into the dual-channel supply chain framework while the aforementioned literature ignores this point. Secondly, the rooms in our research are classified into economic and luxury type, which can be better to reflect the reality. Thirdly, this paper aims to illustrate the pricing decisions under several common cases, which can provide a valuable reference for hotels and OTAs.

3. Model Setup and Notations

We consider a tourism supply chain model involving one hotel and one OTA (Figure 1). The hotel and OTA are both

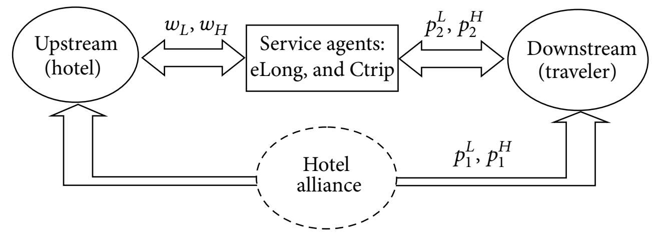


Figure 1: Tourism dual-channel supply chain model diagram.

risk neutral and completely rational. The hotel is required to sign a cooperation agreement to join a HA. However, HA charges the hotel trivial commissions. Thus, HA can be treated as a hotel's direct sales platform and HA's price is decided by the hotel directly. Therefore, it is negligible to consider the contractual relationship between the hotel and HA.

For the hotel, customers are usually divided into two types: business travelers and leisure travelers. Leisure travelers are more sensitive to room price than business ones so that they often make a reservation before a long time. Accordingly, hotels prefer to sell luxury rooms to business travelers. Correspondingly, the hotel rooms are usually classified according to different customers' need, such as luxury rooms, economy rooms, standard rooms, and business rooms. Through setting different rates for different rooms, hotels are able to sell different types of rooms to specific target customers. Based on the above background, in our model, the hotel rooms are classified into economic and luxury type. In this paper, hotel sells room means that hotel sells the right for guest to use the room; that is, the hotel rents the room to guest for a period of time. The hotel sells the economy (luxury) rooms with the price of p_1^L (p_1^H) to the customer directly, and it sells rooms to OTA with the wholesale prices of w_L (w_H), respectively. Providing combination of products and additional services, OTA sells rooms to customers with the prices of p_2^L and p_2^H.

According to international practice, in order to reduce the phenomenon of no-show, customers must provide hotel reservation deposit with credit card or prepay the fee when they make reservation. If customers cancel the booking or choose no-show, the hotel can deduct all or part of the room charge. Take, for example, Ctrip.com; the orders can be divided into two types: orders that cannot be cancelled and time-limited orders. When customers choose the orders that cannot be cancelled, they cannot cancel the order after they conform the booking; if not, the reservation deposit or prepaid fee will not be returned. This kind of situation usually occurs when guest room resources are intense. The time-limited order means that customers can cancel the order for free before the deadline (usually 1 to 3 days before check-in). If customers cancel the booking after the deadline or choose no-show, Ctrip.com will charge the customers for the first day's room fee as a punishment. This also means that when the customer reserves a room for more than one day, he will receive partial refund. Such methods are also adopted by Booking.com and eLong.com. Based on reality, we assume that customers must provide full guarantee with credit card when they make reservation, but if the customers cancel the booking or choose no-show finally, the hotel would only return certain expense rate of α ($0 \leq \alpha \leq 1$). If a customer is booking through OTA, then OTA would return to the customer with the proportion of the retail prices. At the same time, the hotel would return to OTA with the proportion of the wholesale prices. We assume that the proportion that hotel returns to the customer, the proportion that hotel returns to OTA, and the proportion that OTA returns to the customer are the same.

We list the following notations, where i ($i = L, H$) in subscripts or superscripts are for economic (low price) and luxury (high price) rooms, respectively.

Q: the total number of rooms provided for renting to guests.

Q_i: the total number of economic/luxury rooms.

D_1^i: the reservation amount of economic/luxury room booking through HA.

D_2^i: the reservation amount of economic/luxury room booking through OTA.

a_i: the market base demand of economic/luxury room.

a: the total market base demand of rooms ($a = a_L + a_H$).

s: the market share of HA in the tourism dual-channel supply chain.

p_1^i: HA's price of economic/luxury rooms, that is, the direct selling prices.

p_2^i: OTA's price of economic/luxury rooms, that is, the retail prices.

w_i: the wholesale price of economic/luxury rooms.

b_i: the price sensitivity coefficient of i price rooms ($b_L > b_H$).

θ: the diffusion intensity, which describes the shift between two channels with regard to the price and value.

λ: the proportion of customer cancellations and no-show.

α: the return expense proportion of customer cancellations and no-show ($0 \leq \alpha \leq 1$).

β: the services elasticity of demand.

v: the value added by OTA's services.

η: the coefficient of service cost.

$c(v)$: the function of service cost, $c(v) = \eta v^2/2$.

π_1: the profit of the hotel.

π_2: the profit of the OTA.

In the model, we do not consider the marginal sales cost of the hotel since it is very low. In reality, OTA will firstly introduce to customers the hotels' facilities and quota and afterwards confirm the booking and finally book rooms from the hotel for customers. Therefore, the rooms sold by HA and OTA are homogeneous. In addition to providing hotel reservations, OTA also provides other related services as well, such as offering plane ticket booking, holiday booking, and travel information. OTA aims to provide a full range of travel program for travelers. In order to provide such services, OTA needs to pay for additional operating costs. The customers who want to get more tourist information and services will prefer to reserve rooms through OTA. By contrast, HA only provides hotel room information for customers. For instance, customers cannot book the suitable hotel and air ticket at the same time through the HA platform. Consequently, in order to distinguish the influence of different information services to customer's channel choice, we suppose that OTA provides additional service and the value added from the services is v. Further, the cost for providing service is $c(v) = \eta v^2/2$, where the parameter η measures the cost effectiveness of the service. Such cost function is commonly adopted in previous literature and can well describe the relationship between additional services and cost [28, 29]. When OTA does not provide the service, the cost should be zero.

According to the process of demand function in Tsay and Agrawal [30] and Choi [31], the demands of economic/luxury rooms through each channel are expressed as follows:

$$\begin{aligned} D_1^i &= \left(sa_i - b_i p_1^i\right) + \theta\left(p_2^i - v - p_1^i\right), \\ D_2^i &= \left[(1-s)\, a_i - b_i p_2^i\right] + \beta v + \theta\left(p_1^i + v - p_2^i\right). \end{aligned} \tag{1}$$

Such linear demand functions are used because they can achieve approximate and satisfactory fit to certain types of demand, and they are tractable and widely used in marketing and supply chain management literature [32]. In our model, the parameter b represents the price sensitivity coefficient. It describes the marginal channel demand per respective channel price. The own price elasticity is normally negative [33]. And the parameter β measures the marginal demand of OTA per service value added. In addition, the parameter θ is understood as the diffusion intensity (cross-price sensitivity) and describes the shift between two channels with regard to the price and the value. Such modeling method can guarantee that the self-price sensitivity, in absolute value, is stronger than the cross-price sensitivity, which is proved by Hanssens et al. [34] and used by Yao and Liu [28], Choi [31], and Kurata et al. [35].

Now consider the situation without overbooking; that is, $D_1^L + D_2^L \leq Q_L$ and $D_1^H + D_2^H \leq Q_H$. Thus, the reservation amount can be regarded as the final demand.

The total profit for the hotel, OTA, and the whole supply chain can be expressed as follows:

$$\pi_1 = (1-\lambda\alpha)\left(p_1^L D_1^L + p_1^H D_1^H + w_L D_2^L + w_H D_2^H\right), \tag{2}$$

$$\begin{aligned} \pi_2 = &\left[(1-\lambda\alpha)\left(p_2^L - w_L\right) - c(v)\right] D_2^L \\ &+ \left[(1-\lambda\alpha)\left(p_2^H - w_H\right) - c(v)\right] D_2^H, \end{aligned} \tag{3}$$

$$\pi = \pi_1 + \pi_2. \tag{4}$$

4. Integrated System

As a benchmark, we first consider an integrated system where the common goal of supply chain upstream and downstream members is to maximize the overall profit of the supply chain. This is an ideal model that the members in the supply chain operate as an entirety. In such an integrated model of tourism supply chain, the hotel and OTA cooperate to determine the pricing strategies.

Proposition 1. *In the integrated system, the optimal pricing decisions of HA and OTA are as follows:*

$$p_1^{i*} = \frac{a_i (b_i s + \theta) + \theta v (\beta - b_i)}{2 (b_i)^2 + 4 b_i \theta},$$
$$p_2^{i*} = \frac{(b_i + \theta - b_i s) a_i + b_i \beta v + \theta v (\beta + b_i)}{2 (b_i)^2 + 4 b_i \theta} + \frac{c(v)}{2(1 - \lambda\alpha)}, \quad i = L, H. \tag{5}$$

From Proposition 1, the optimal prices shown above can explain that in the integrated system the optimal prices of HA will not be affected by the proportion of customer cancellations and no-show (λ) or the return expense proportion (α). On the other hand, with the increase of λ or α, the optimal prices for OTA will increase. This is reasonable because when OTA faces unstable demand or it should return more to customers' cancellation or no-show, it would like to increase the price to guarantee its profit.

Corollary 2. *p_2^{L*} and p_2^{H*} are increasing in v. Moreover,*

(1) *if $\beta > b_L > b_H$, then $\partial p_1^{L*}/\partial v > 0$, $\partial p_1^{H*}/\partial v > 0$,*

(2) *if $b_L > \beta > b_H$, then $\partial p_1^{L*}/\partial v < 0$, $\partial p_1^{H*}/\partial v > 0$,*

(3) *if $b_L > b_H > \beta$, then $\partial p_1^{L*}/\partial v < 0$, $\partial p_1^{H*}/\partial v < 0$.*

From Corollary 2, we can know that when OTA offers more services, OTA's optimal price will be increased accordingly. It is reasonable that if v increases, the cost of OTA will also definitely increase. So there is always a positive correlation between OTA's optimal prices and the added value.

If $\beta > b_L > b_H$, the market of economy rooms and that of luxury rooms are both service-oriented rather than price-oriented. The implication is that the effective service will increase the optimal price of HA. When OTA offers more effective service, its price will increase. But some customers want to enjoy the additional services, so they are willing to make a booking through OTA despite spending more money. Therefore, the demand in OTA will increase while the demand in HA will decrease.

If $b_L > \beta > b_H$, the market of economy rooms is service-oriented, while the market of luxury rooms is price-oriented. The HA's optimal price of economy rooms will decrease with respect to v. Conversely, the HA's optimal price of luxury rooms will increase with respect to v. For examples, the change of price has a great impact on leisure travelers' decisions, but for business travelers, the price is not the most concern. If the price of OTA increases, OTA's demand of economy rooms will significantly reduce and the HA's demand will increase, thus the HA's optimal price will decline. On the other hand, for luxury rooms, if OTA provides more effective service, the customer would be willing to pay higher prices for value-added service. This will bring the increase in OTA's demand and the reduction of HA's demand, thus the optimal price of HA will increase.

If $b_L > b_H > \beta$, the market of economy rooms and that of luxury rooms are both price-oriented than service-oriented. The HA's optimal price of both economy rooms and luxury rooms will decrease with the increase of v. When the market is price-oriented, the promotion of price in OTA will lead to a reduction in OTA's demand. Thereby, the demand in HA will increase and optimal price in HA will decline.

5. Two Game Models

In the tourism dual-channel supply chain, the relationship between hotel and OTA is cooperative as well as competitive. In the decentralized decision-making model, they are more competitive than cooperative. Their purposes are to maximize their own profits. In reality, hotel can choose whether to inform the price of HA or not, for HA is competing with OTA. Uninformed of the prices in HA, OTA can predict the range of price on HA by the star-rated and room type of the hotel but it cannot know the exact price. Therefore, this section examines the pricing equilibrium between the hotel and OTA under two types of games, namely, the Bertrand and Stackelberg games. For the Bertrand game, both of the participants regard price as a decision variable; the hotel and OTA are uncooperative. For the Stackelberg game, the participants are divided into two roles, the leader and the follower. As a leader in tourism supply chain, the hotel will provide the policy to maximize its interests. As a follower, OTA accepts if profitable. By comparing the two models, we can understand how the behavior of hotel influences the pricing decisions.

5.1. Bertrand Game. In a Bertrand game, the procedure is as follows: keeping the wholesale price unchanged, the hotel determines HA's prices p_1^L and p_1^H, so as to maximize profit π_1. Uninformed of the prices of HA, OTA decides prices p_2^L and p_2^H, so as to maximize his profit π_2. Let $(p_1^i)^B$, $(p_2^i)^B$ ($i = L, H$) denote HA's and OTA's prices under a Bertrand model, respectively.

In the following, L_n ($n = 1, 2, \ldots, 5$), s_0, and G_m ($m = 1, 2$) are the threshold values and are given in Appendix G.

Proposition 3. *There exist Bertrand equilibrium prices $(p_1^i)^{B*}$, $(p_2^i)^{B*}$ ($i = L, H$), where*

$$\left(p_1^i\right)^{B*} = \frac{(1 - \lambda\alpha) L_1 + c\theta (b_i + \theta)}{(1 - \lambda\alpha)\left[4 (b_i + \theta)^2 - \theta^2\right]},$$
$$\left(p_2^i\right)^{B*} = \frac{(1 - \lambda\alpha) L_2 + 2c (b_i + \theta)^2}{(1 - \lambda\alpha)\left[4 (b_i + \theta)^2 - \theta^2\right]}. \tag{6}$$

From Proposition 3, in the Bertrand model, the equilibrium prices for hotel and OTA are both affected by α and λ. With the increase of α or λ, both $(p_1^i)^{B*}$ and $(p_2^i)^{B*}$ will be higher. In other words, with higher uncertainty of room occupancy, the hotel and OTA tend to increase prices to protect their own profits.

Denote $c_1 = (1-\lambda\alpha)(2b_L+\theta-\beta)/(b_L+\theta)$ and $c_2 = (1-\lambda\alpha)(2b_H+\theta-\beta)/(b_H+\theta)$; then $c_1 > c_2$.

Corollary 4. *$(p_2^L)^{B*}$ and $(p_2^H)^{B*}$ are increasing in v. $(p_1^L)^{B*}$ and $(p_1^H)^{B*}$ have the following relationships with respect to OTA's service:*

(1) *if $c'(v) > c_1$, then $\partial(p_1^L)^{B*}/\partial v > 0$, $\partial(p_1^H)^{B*}/\partial v > 0$,*

(2) *if $c_1 > c'(v) > c_2$, then $\partial(p_1^L)^{B*}/\partial v < 0$, $\partial(p_1^H)^{B*}/\partial v > 0$,*

(3) *if $c'(v) < c_2$, then $\partial(p_1^L)^{B*}/\partial v < 0$, $\partial(p_1^H)^{B*}/\partial v < 0$.*

From Corollary 4, when OTA offers more effective services, its optimal prices will increase due to the increment in the cost. Marginal cost measures the input costs and the output value per unit. If the marginal cost is greater than c_1, then the costs of OTA to provide additional services are higher. So the equilibrium prices of HA will increase as the value added by OTA's services increases. When the marginal cost is between c_1 and c_2, HA's equilibrium price of economy rooms and the value added by OTA's service have a negative correlation, but HA's equilibrium price of luxury rooms would increase with the increment of v. In addition, when the marginal cost is less than c_2, the costs of OTA to provide additional service are lower. And the equilibrium prices of HA will decrease when v increases.

Corollary 5. *There exist threshold unit wholesale prices w_i^0, w_i^1 $(i = L, H)$, such that*

(1) *$(p_1^i)^* < (p_1^i)^{B*}$, if $w_i > w_i^0$; $(p_1^i)^* = (p_1^i)^{B*}$, if $w_i = w_i^0$; $(p_1^i)^* > (p_1^i)^{B*}$, if $w_i < w_i^0$.*

(2) *$(p_2^i)^* < (p_2^i)^{B*}$, if $w_i > w_i^1$; $(p_2^i)^* = (p_2^i)^{B*}$, if $w_i = w_i^1$; $(p_2^i)^* > (p_2^i)^{B*}$, if $w_i < w_i^1$,*

where

$$\begin{aligned} w_i^0 &= \frac{L_3}{6b_i(b_i+\theta)(b_i+2\theta)} - \frac{c}{3(1-\lambda\alpha)}, \\ w_i^1 &= \frac{\theta\left[(1-\lambda\alpha)L_4 - cb_i\theta(b_i+2\theta)\right]}{2b_i(1-\lambda\alpha)(b_i+2\theta)(3\theta^2+2b_i^2+2b_i\theta)}. \end{aligned} \tag{7}$$

From Proposition 1, it is easy to know that the equilibrium prices in the integrated system are not affected by the wholesale prices. However, from Proposition 3, in the Bertrand model, there is a positive correlation between the equilibrium prices and the wholesale prices. Corollary 5 shows that the equilibrium price of HA in the Bertrand game is higher than that in the integrated system if the wholesale price is greater than a certain critical point (w_i^0). Additionally, if the wholesale price is greater than w_i^1, then the equilibrium price of OTA in Bertrand model will be higher than that in the integrated model.

Corollary 6. *There exists a threshold unit wholesale price w_i^2 $(i = L, H)$, such that if $w_i > w_i^2$, then $D_i^* > D_i^{B*}$ $(i = L, H)$, where*

$$w_i^2 = \frac{\theta\left[(1-\lambda\alpha)(v\beta+a_i) - b_ic\right]}{2b_i(1-\lambda\alpha)(b_i+2\theta)}. \tag{8}$$

Corollary 6 shows that when the wholesale price is greater than w_i^2, the high price will lead to a lower demand in the Bertrand game. That is, the reservation amount of Bertrand model is less than that in the integrated model. The higher sale price in OTA results from the higher wholesale price. Thus, the above result is reasonable because travelers are price sensitive; especially for leisure travelers, the increase of price will bring the decline of reservation amount. When the relationship between the hotel and OTA is modeled as a Stackelberg game, the hotel can motivate customers to reserve rooms through a reasonable wholesale price. If the wholesale price is greater than the threshold, the sales volume in the centralized case is lower than that in the Bertrand case.

5.2. Stackelberg Game. In a Stackelberg game, the procedure is as follows: keeping the wholesale price unchanged, the hotel (as the leader) announces HA's prices p_1^L and p_1^H of two types of rooms, respectively, in order to maximize its profit π_1. In response to HA's prices, OTA (as the follower) determines prices p_2^L and p_2^H to maximize its profit π_2. Let $(p_1^i)^S$, $(p_2^i)^S$ $(i = L, H)$ denote, respectively, HA's and OTA's prices under a Stackelberg game.

Proposition 7. *The optimal prices are $(p_1^i)^{S*}$, $(p_2^i)^{S*}$ $(i = L, H)$, and there exists a threshold unit wholesale price w_i^3 $(i = L, H)$, such that if $w_i > w_i^3$, then $(p_1^i)^* < (p_1^i)^{S*}$, where*

$$\begin{aligned} \left(p_1^i\right)^{S*} &= \frac{(1-\lambda\alpha)\left(L_1 - w_i\theta^2 - w_ib_i\theta\right) + c\theta(b_i+\theta)}{(1-\lambda\alpha)\left[4(b_i+\theta)^2 - 2\theta^2\right]}, \\ \left(p_2^i\right)^{S*} &= \frac{\theta\left(p_1^i\right)^{S*} + (1-s)a_i + v(\theta+\beta) + w_i(b_i+\theta)}{2(b_i+\theta)} + \frac{c}{2(1-\lambda\alpha)}, \\ w_i^3 &= \frac{b_iv\beta + ab_i + \theta vb_i + a_i\theta + \theta v\beta - a_ib_is}{2b_i(b_i+2\theta)} - \frac{c}{2(1-\lambda\alpha)}. \end{aligned} \tag{9}$$

Similar to the Bertrand competition, the equilibrium prices in Stackelberg game will increase with the increase of the return expense proportion (α) or the cancellations and no-show proportion (λ).

From Proposition 7, we can see that the Stackelberg model is similar to the Bertrand model that the equilibrium prices would increase with the increase of wholesale prices. Furthermore, when the wholesale price is greater than w_i^3, the equilibrium price of HA in Stackelberg model will be higher than that in the integrated model.

In the Stackelberg game, if OTA offers more effective services, the equilibrium prices of OTA would increase accordingly. The relationship between the Stackelberg equilibrium prices and v is similar to that in the Bertrand model.

5.3. Contrastive Analysis of Bertrand and Stackelberg Games. Based on the equilibrium prices obtained in the previous section, we in this section compare the equilibrium prices and optimal profit under Bertrand and Stackelberg games.

Corollary 8. *There exists a threshold unit wholesale price w_i^4 $(i = L, H)$, such that*

(1) $(p_1^i)^{B*} > (p_1^i)^{S*}$ *and* $(p_2^i)^{B*} > (p_2^i)^{S*}$ *if* $w_i > w_i^4$,

(2) $(p_1^i)^{B*} = (p_1^i)^{S*}$ *and* $(p_2^i)^{B*} = (p_2^i)^{S*}$ *if* $w_i = w_i^4$,

(3) $(p_1^i)^{B*} < (p_1^i)^{S*}$ *and* $(p_2^i)^{B*} < (p_2^i)^{S*}$ *if* $w_i < w_i^4$,

where

$$w_i^4 = \frac{\theta\left[(1-\lambda\alpha)\,L_5 + c\theta\left(b_i+\theta\right)\right]}{4b_i\,(1-\lambda\alpha)\left(b_i+\theta\right)\left(b_i+2\theta\right)}. \tag{10}$$

Corollary 8 compares the equilibrium prices under the Bertrand and Stackelberg game with respect to the wholesale price. It shows that if the wholesale price is greater than w_i^4, the pricing equilibrium under the Bertrand is higher than that under the Stackelberg game. Conversely, the pricing equilibrium is lower if the wholesale price is less than w_i^4.

In the decentralized case, both the manufacturer and the retailer want to maximize their profit. A high wholesale price could be passed on to customers. Therefore, the above results exist. In addition, how to share the profit between the manufacturer and the retailer depends on their bargaining powers.

Corollary 9. *If $w_i \neq w_i^4$ $(i = L, H)$, then $\pi_1^{S*} > \pi_1^{B*}$. OTA's profit has the following relationship with respect to w_i^4:*

(1) $\pi_2^{B*} > \pi_2^{S*}$ *if* $w_L > w_L^4$ *and* $w_H > w_H^4$,

(2) $\pi_2^{B*} = \pi_2^{S*}$ *if* $w_L = w_L^4$ *and* $w_H = w_H^4$,

(3) $\pi_2^{B*} < \pi_2^{S*}$ *if* $w_L < w_L^4$ *and* $w_H < w_H^4$.

The hotel is the leader in Stackelberg game and can always benefit from it except for the critical point w_i^4. Compared with the Bertrand game, the hotel can get higher profits in Stackelberg game when the wholesale price is not equal to w_i^4. Therefore, in decentralized decision-making model, the hotel will be more inclined to be the price leader in order to ensure that it can achieve higher profits.

Corollary 9 indicates that the Stackelberg model is superior to the Bertrand model when the wholesale price is less than the critical point w_i^4. Not only the hotel but also OTA can gain more profit from the Stackelberg game. Though OTA is a follower in Stackelberg game, the profit will not be reduced because the lower wholesale prices can compensate for the loss of profit caused by the opening of HA channel. Consequently, the hotel should play the role of the price leader in the supply chain and develop reasonable wholesale prices to achieve optimization of the whole supply chain profit while optimizing its own profit.

Corollary 10. *Under either Bertrand or Stackelberg equilibrium channel prices, there exists optimal wholesale price w_i^* $(i = L, H)$ that maximizes the total equilibrium channel profit, where $w_i^* = G_1/G_2$.*

According to Corollary 10, there exists optimal wholesale price w_i^* $(i = L, H)$ that maximizes the total equilibrium channel profit under either Bertrand or Stackelberg game. The optimal wholesale price will optimize the decentralized decision-making supply chain. Meanwhile, both the hotel and the OTA can benefit from the optimal wholesale price. Therefore, the reasonable wholesale price can help promote mutual cooperation and communication between the supply chain members.

5.4. Contrastive Analysis of Two Channels in the Supply Chain

Corollary 11. *There exists a threshold unit market share of HA in the tourism dual-channel supply chain (s_0), such that when $s > s_0$, the hotel's profit from HA is higher than that from OTA.*

Keeping the wholesale prices and sales prices unchanged, when the market share of HA in the tourism dual-channel supply chain exceeds a certain value s_0, the hotel's profit from HA is higher than that from OTA. The result is suitable for the integrated system, Bertrand and Stackelberg games. From the perspective of hotel's revenue, the hotel should explore the channel for direct selling actively. Also, the hotel can strengthen the promotion of HA channel to improve the customer's understanding and preference for the HA channel. But the premise is that the profit growth from the promotion can compensate for the cost. According to the statistical data from the hotel price competitiveness analysis report in the first half of 2012, we know that the booking cost from direct selling channel is about 1 dollar per night, but that from OTA is about 10 dollars per night. Booking cost can be reduced by 46% if the direct selling level is doubled. Then improving the market share of direct selling will greatly reduce the cost. More comparisons of different channel's profits in each model can be found in part of numerical analysis.

6. Overbooking

This section extends to consider the overbooking case. Regarding overbooking, in the hospitality industry, the common practice is that the hotel upgrades the room type or transfers customers to other hotels nearby that have the same level and same type. In airline industry, if customers are denied boarding after arriving at the airport, the airline will provide a free upgrade, endorse the ticket, or provide compensation for the customers according to specific situation. Compared to the hotels, the airlines cannot transfer customers to other companies. The air ticket alteration will usually disrupt customers' traveling schedule. Though the overbooking by airlines has been well studied in the literature, our research is new in the hospitality industry by considering the overbooking with/without room transfer, with the hope that certain implications can be enlightened to the hotel managers.

The hotel wants to increase the occupancy rates to increase the profits through overbooking. However, overbooking will lead to the risk that the actual occupancy amount is greater than the number of hotel rooms. Therefore, the hotel has to provide compensation to travelers who cannot check in, that is, overbooking cost. The greater the numbers of overbooking, the less the probability of room vacancy, but at the same time travelers are more likely to be refused by the hotel. Therefore, the hotel must balance between the vacancy lost and the overbooking costs. We develop a model to analyze the overbooking choice.

In reality, the OTA is always a crucial partner of hotels for it can attract a large number of customers for hotels and improve their occupancy rates [36]. Spontaneously, the OTA plays an increasingly significant role in the distribution systems of hotels [5, 7]. By contrast, the HA mainly provides a platform for individual guest. Compared with individual customers, hotel pays more attention to the long-term cooperation relationship with the OTA. Therefore, in our model, we assume that when the market demand is greater than the supply and the hotel does not carry out overbooking, the hotel would give priority to customers who are booking from OTA.

6.1. Without Room Transfer. Here, we consider the case with no room transfer. When the actual occupancy amount exceeds the total quantity of rooms, the hotel does not arrange other alternative accommodation for the customers who cannot check in. But the hotel must pay the loss of goodwill cost. Suppose that the overbooking quantity of economy and luxury rooms is ΔC_L and ΔC_H and the overbooking cost is c_L and c_H, respectively.

We discuss economy rooms' overbooking. The actual occupancy can be either of the following two cases: (1) when $Q_L < D_1^L + D_2^L \le Q_L + \Delta C_L$, the reservation amount of rooms is $D_1^L + D_2^L$, so the actual occupancy amount is $(D_1^L + D_2^L)(1-\lambda)$; (2) when $D_1^L + D_2^L > Q_L + \Delta C_L$, the reservation amount is $Q_L + \Delta C_L$; then the actual occupancy amount is $(Q_L + \Delta C_L)(1-\lambda)$. Therefore, if $(Q_L + \Delta C_L)(1-\lambda) \le Q_L$, that is, $\Delta C_L \le \lambda Q_L/(1-\lambda)$, the profit of hotel is

$$\pi_{11}^L = (1-\lambda\alpha)\left[p_1^L\left(Q_L + \Delta C_L - D_2^L\right) + w_L D_2^L\right]. \quad (11)$$

If $(Q_L + \Delta C_L)(1-\lambda) \ge Q_L$, that is, $\Delta C_L \ge \lambda Q_L/(1-\lambda)$, the economy rooms' actual occupancy amount exceeds the economy rooms' number in the hotel, so the hotel must pay the overbooking cost; then the profit of hotel is

$$\begin{aligned}\pi_{12}^L &= (1-\lambda\alpha)\left[p_1^L\left(Q_L + \Delta C_L - D_2^L\right) + w_L D_2^L\right] \\ &\quad - \left(p_1^L + c_L\right)\left[\left(Q_L + \Delta C_L\right)(1-\lambda) - Q_L\right].\end{aligned} \quad (12)$$

Proposition 12. *Without considering rooms' transfer, the optimal overbooking quantity of economy rooms and luxury rooms is*

$$\Delta C_i^* = \frac{\lambda}{1-\lambda} Q_i, \quad i = L, H. \quad (13)$$

From Proposition 12, we know that the higher the proportion of customer cancellations and no-show is, the larger the optimal overbooking quantity of the hotel will be. With the increase of α, the penalty that customers need to pay for their cancellation or no-show decreases. Two following extreme situations are considered. When $\alpha = 1$, the hotel will return all the guarantee fee in case of customers' cancellation or no-show. That is the same as the hotel does not charge customers any guarantee fee when customers book the rooms. Under such circumstances, customers can book rooms for several days firstly; after they identify a specific itinerary, they can cancel the extra rooms for they do not suffer any loss. It is no doubt that such regulation will increase the cancellation and no-show rate. On the other hand, if $\alpha = 1$, all the guarantee fee will not be returned after cancellation or no-show. In this case, customers need to reserve a room cautiously. If it is not necessary, customers will not choose cancellation or no-show because the penalty cost is too high. Such policy will reduce the cancellation and no-show rate. Spontaneously, to some extent, α and λ are positive correlation. When the hotel increases proportion α, it would like to increase the overbooking quantity, so as to increase the actual occupancy rate. Nowadays, the time-limited orders are widely used among hotels. Such regulation cannot only reduce customers' loss for the change of schedule but also help the hotel confirm the orders so as to handle the overbooking phenomenon in advance.

In addition, the risk of overbooking the economy rooms is relatively lower than the luxury rooms. Compared to economic rooms, it is more difficult for the hotel to find alternative luxury rooms to satisfy the customers. Naturally, overbooking cost of luxury room is higher. For instance, if economy rooms are overbooked, the hotel can transfer the customers to luxury rooms and the customers are willing to accept. By contrast, if luxury rooms are overbooked, customers are not willing to accept the adjustment from luxury rooms to economic rooms. This implies that the hotel has to pay a higher default cost. Therefore, the hotel should be more careful to choose the overbooking quantity of luxury rooms.

6.2. With Room Transfer. The following discusses the case with room transfer. Firstly, when the economy rooms are insufficient and luxury rooms have a surplus, customers can check in the hotel's luxury rooms with the price of economy rooms. Customers spend the same money and can enjoy better service, so they are willing to accept this adjustment. Secondly, when the luxury rooms' actual occupancy amount exceeds the number of luxury rooms in the hotel, the hotel usually transfers customers to other hotels nearby that have the same level and same type because the customers who book luxury room have a high requirement. Customers can enjoy the same service, so they are willing to accept this transfer. However, the hotel's credibility in customers' mind will be damaged and it must pay the additional costs. Thirdly, if both the economy rooms and luxury rooms are unable to meet the demand, the hotel usually transfers the customers who cannot check in to other hotels. Note that when $D_1^i + D_2^i > Q_i$, $i = L, H$, if the hotel does not overbook, its profit is

$$\pi_{111} = (1 - \lambda\alpha)\left[p_1^L\left(Q_L - D_2^L\right) + p_1^H\left(Q_H - D_2^H\right) + w_L D_2^L + w_H D_2^H\right]. \tag{14}$$

Proposition 13. *With room transfer, there exists threshold unit overbooking cost c_{i0}, such that if $c_i < c_{i0}$, then overbooking is profitable for the hotel, where*

$$c_{i0} = \frac{(1 - \lambda\alpha)\, p_1^i}{1 - \lambda}, \quad i = L, H. \tag{15}$$

From Propositions 12 and 13, we conclude that the hotel should control the overbooking quantity in a certain range to prevent lacking of rooms. If the overbooking proportion is too large, then the hotel may face a loss. Only when the overbooking cost is low, the hotel can obtain additional profit. Otherwise, if the overbooking cost exceeds a certain value, the hotel's profit would be lower when it carries out overbooking. As the market demand increases, the supply is close to saturation, and then the overbooking costs will be higher. Therefore, when the market demand is excessive, the hotel should strictly control the proportion of overbooking. In addition, if the number of hotels that have the same level and type in the same region is little, that is, the possibility to transfer customer is small, then the hotel should lower the proportion of overbooking.

The proportion of overbooking depends on the proportion of customer cancellations and no-show as well as the total number of hotel rooms, so the hotel can analyze the historical data to develop an appropriate overbooking proportion, so as to maximize the hotel's expected profit.

Proposition 14. *When the economy rooms and luxury rooms are both overbooked, the optimal centralized prices are as follows: $i = L, H$,*

$$p_{14}^{i*} = p_1^{i*} + \frac{c_u^i (1 - \lambda)}{2(1 - \lambda\alpha)},$$
$$p_{24}^{i*} = p_2^{i*} + \frac{c_u^i (1 - \lambda)}{2(1 - \lambda\alpha)}. \tag{16}$$

Bertrand and Stackelberg prices are as follows:

$$\left(p_{14}^i\right)^{B*} = \left(p_1^i\right)^{B*} + \frac{2c_u^i b_i (b_i + \theta)(1 - \lambda)}{(1 - \lambda\alpha)\left[4(b_i + \theta)^2 - \theta^2\right]},$$

$$\left(p_{24}^i\right)^{B*} = \left(p_2^i\right)^{B*} + \frac{\theta c_u^i b_i (1 - \lambda)}{(1 - \lambda\alpha)\left[4(b_i + \theta)^2 - \theta^2\right]},$$

$$\left(p_{14}^i\right)^{S*} = \left(p_1^i\right)^{S*} + \frac{c_u^i b_i (2b_i + 3\theta)(1 - \lambda)}{(1 - \lambda\alpha)\left[4(b_i + \theta)^2 - 2\theta^2\right]},$$

$$\left(p_{24}^i\right)^{S*} = \frac{\theta\left(p_{14}^i\right)^{S*} + (1 - s) a_i + v(\theta + \beta) + w_i(b_i + \theta)}{2(b_i + \theta)} + \frac{c}{2(1 - \lambda\alpha)}. \tag{17}$$

For the hotel, the premise to overbook rooms is that the market demand is greater than the supply. When the hotel faces the risk of overbooking, the equilibrium prices in both centralized and decentralized decision-making supply chain will improve. In the busy season, in order to reduce the risk of overbooking, the hotel would increase the prices and the increment will increase as the transfer cost increases. From (16) and (17) we can know that the increment in centralized decision-making model is the greatest and that in Bertrand game is the smallest.

7. Numerical Analysis

In this section, we do some numerical experiment to gain more insights. According to the China Hotel Sales Channel Report by I-Research (http://report.iresearch.cn/report/201303/1892.shtml), there are 38.1% of the hotels that are suitable for high-end travelers and 61.9% of the hotels that are suitable for low-end travelers. Therefore, this part assumes that the proportion of the economical rooms and the luxury rooms is 4 : 6; that is, $a_L = 0.6a$, $a_H = 0.4a$. Other parameter settings are made as follows: $a = 600$, $s = 0.5$, and $b_L = 0.6$; $b_H = 0.2$; $\beta = \theta = 0.5$; $\eta = 0.04$; $\alpha = 0.8$; $\lambda = 0.1$; $w_i = 0.5p_1^i$ $(i = L, H)$. In the following we investigate the impact of v on the optimal prices and profits.

When $v = 0$, that is, the OTA does not offer any service, then the optimal prices of HA are equal to the prices of OTA. Since $b_L > \beta > b_H$, then the market of economy rooms is service-oriented, while the market of luxury rooms is price-oriented. From Figure 2, we can know that when OTA offers more services, OTA's optimal price will be increased accordingly. It is reasonable that if v increases, the cost of OTA will also definitely increase. With the increase of v, p_1^{L*} will decline slowly, while p_1^{H*} will increase.

The OTA's profit status becomes better with the improvement of its price. But as the v increases, the cost of OTA will also definitely increase. When v is equal to 27, then the profit of OTA is the highest. If v is larger, then the high costs will reduce the profits of OTA. When v is equal to 34, the profit of the whole supply chain is the highest. So, from the perspective of the supply chain, v is expected to be 34 rather than 27. Similarly for the integrated system, the profit of the OTA also rises first and then declines with the increase of v. After calculation we can get that when $v = 27$, both π_2^{B*} and π_2^{S*} will be the highest. From Figures 3 and 4, we know that the OTA would tend to offer service that valued 27, so as to maximize its own profit.

After calculation we can conclude that when $v = 34$, then π^{B*} is highest; when $v = 35$, then π^{S*} is highest (Figure 5).

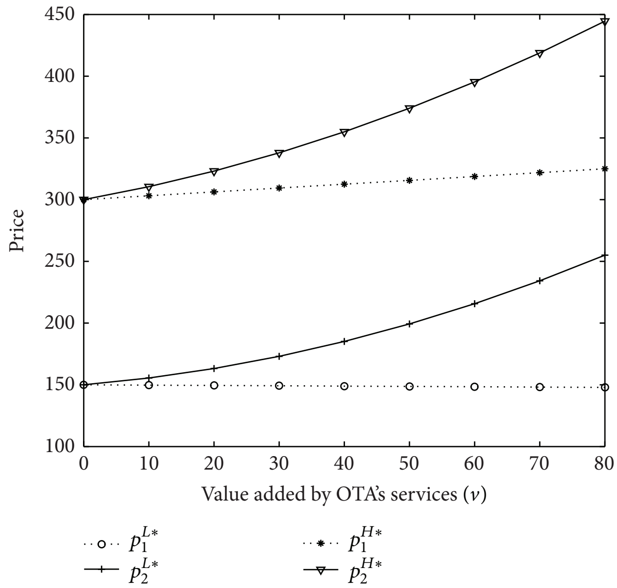


Figure 2: Equilibrium prices in the integrated system.

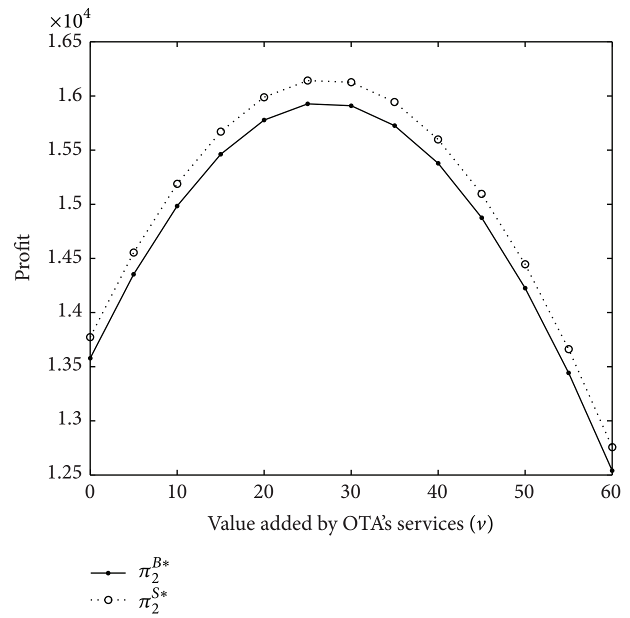


Figure 4: OTA's profits under two games.

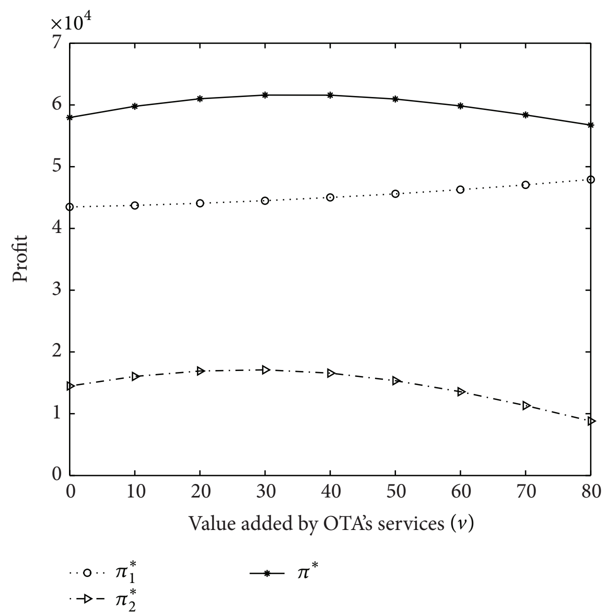


Figure 3: Profits in the integrated system.

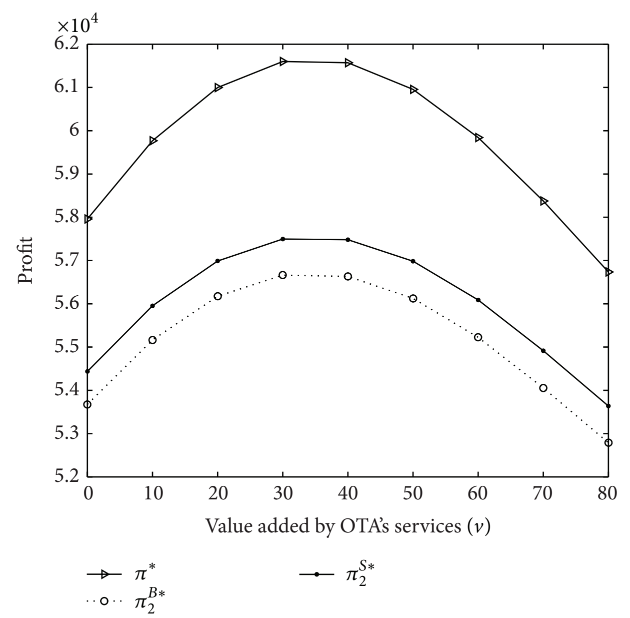


Figure 5: The profit comparison.

From the perspective of the supply chain, the OTA is expected to offer more services. Despite the rising cost of OTA will lower its profit, the hotel can lower the wholesale prices to coordinate the distribution of profits, so as to achieve a win-win situation.

In the following, we focus on the comparison of different channel's profit in each model. Figure 6 depicts the curves of the total profits of the HA/OTA channels in the integrated system with respect to v. Figure 7 draws the curves of the different channels' profits under both Bertrand and Stackelberg game. The profit of the HA channel refers to the profit that the hotel derived from the HA channel. The profit of the OTA channel involves the OTA's profit and the hotel's profit gained from the OTA channel.

As shown in Figures 6 and 7, with the increase of v, the profit of the OTA channel will increase first and will then decrease. This indicates that a high level of service by OTA will be detrimental to the channel profit. That is because a high level of service will bring a high cost undoubtedly. For the integrated model and Bertrand game model, this disadvantage is more evident. The OTA's service can motivate the actual purchase behavior of customers in the OTA channel. When v is relatively lower, the additional profits from OTA's

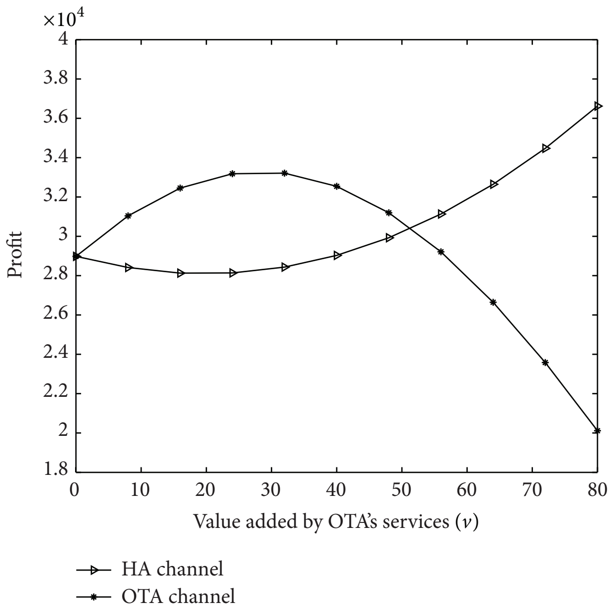


Figure 6: Channel profits in the integrated system.

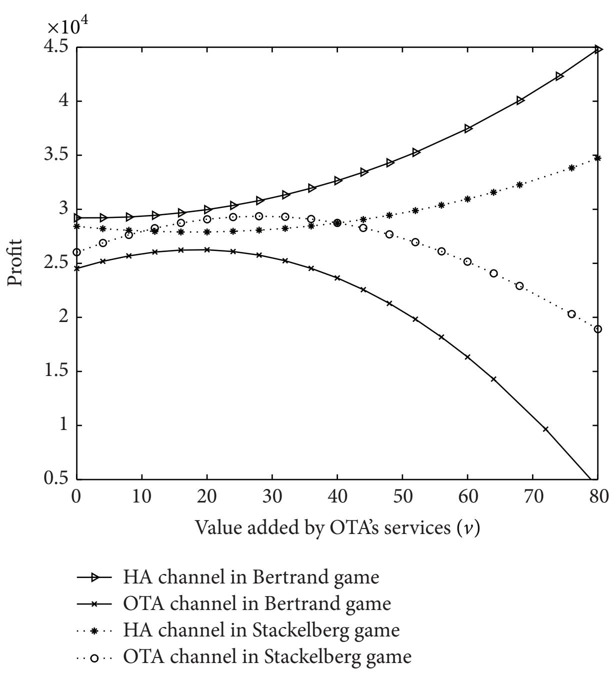


Figure 7: Channel profits in two games.

service are enough to compensate for the cost. In this case, the hotel and the OTA can get more profit from providing the service. However, when v is high, the additional profits from OTA's service cannot make up for the high cost. In such phenomenon, the OTA should not increase investment in service.

For both integrated model and Bertrand game model, with the increase of v, the profit of the HA channel will decrease first and will then increase, while in the Stackelberg game model, the profit of the HA channel will always increase with v. With Figure 2 we can know that when OTA provides more additional services, it tends to increase its price. Some of the customers will turn to reserve rooms from the HA because of the excessively high price in OTA. As a result, the hotel can gain more profits from the HA channel.

8. Conclusion

With the development of e-commerce, the providers of tourism products are more inclined to build dual-channels. In this paper, we consider the direct selling channel of HA and the traditional retail channel of OTA from the perspective of tourism supply chain. We develop a two-echelon tourism supply chain model involving one hotel and one OTA where the OTA provides related service. Customers will be refunded partially in case of cancelation or no-show. We investigate the optimal pricing strategies under several common cases. The integrated operation and the decentralized operation based on Bertrand and Stackelberg games are studied and corresponding equilibrium pricing policies are obtained.

Our findings demonstrate that the hotel plays a crucial role in optimizing the benefits of supply chain. Considering the overall profit of the supply chain, the hotel should act as a leader in dual-channel supply chain and formulate reasonable wholesale prices to improve the cooperative enthusiasm of OTA, so as to achieve optimization of the whole supply chain. A lower wholesale price can help reduce OTA's loss of profit because of the opening of HA channel. In addition, we consider the overbooking of the hotel and obtain the optimal overbooking quantity. We conclude that the hotel can always obtain higher profits when the overbooking proportion is controlled within a certain range. Otherwise, if the overbooking proportion is too high, the hotel can get extra profits only when the overbooking cost is lower than a certain value.

The contribution of this paper is that we extend tourism supply chain to the dual-channel structure in which the hotel opens its direct channel to sell tourism products. And the extant literature only refers to the single channel. In addition, we realize that customer cancellations and no-show occur frequently in hotel industry. Thus, we develop the game models based on such phenomenon. At last, we derive the conditions under which the hotel dominant case is profitable for both the supply chain members. Our results can provide hotels and OTAs with some new managerial insights under the dual-channel environment.

This research can be extended in several directions in future work. For example, in this paper we assume that all information is known to the hotel and OTA. However, some information could be asymmetric (e.g., see Chen [37]). Thus, we can explore the Bayesian equilibrium under asymmetric information settings. Secondly, the assumption of linear demand is a limitation of this paper. We can further study the dynamic pricing in tourism dual-channel supply chain.

Appendix

A. Proof of Corollary 8

First we set $(p_1^L)^{B*} > (p_1^L)^{S*}$, which can be expressed as

$$\frac{(1-\lambda\alpha)\left(\theta a_L+\theta s a_L+2b_L s a_L+\theta\nu\beta+3w_L\theta^2+3w_L b_L\theta-\theta^2\nu-2\theta b_L\nu\right)+c\theta\left(b_L+\theta\right)}{(1-\lambda\alpha)\left[4\left(b_L+\theta\right)^2-\theta^2\right]} > \frac{(1-\lambda\alpha)\left(\theta a_L+\theta s a_L+2b_L s a_L+\theta\nu\beta+2w_L\theta^2+2w_L b_L\theta-\theta^2\nu-2\theta b_L\nu\right)+c\theta\left(b_L+\theta\right)}{(1-\lambda\alpha)\left[4\left(b_L+\theta\right)^2-2\theta^2\right]}. \tag{A.1}$$

The above equation can be simplified as

$$w_L > \frac{\theta\left[(1-\lambda\alpha)\left(\theta a_L+\theta s a_L+2b_L s a_L+\theta\nu\beta-\theta^2\nu-2\theta b_L\nu\right)+c\theta\left(b_L+\theta\right)\right]}{4b_L(1-\lambda\alpha)\left(b_L+\theta\right)\left(b_L+2\theta\right)}. \tag{A.2}$$

So we can get Corollary 8.

B. Proof of Corollary 9

If $(p_1^i)^{S*} > (p_1^i)^{B*}$, then

$$\begin{aligned}&(1-\lambda\alpha)\left(\theta a_i+\theta s a_i+2b_i s a_i+\theta\nu\beta+2w_i\theta^2+2w_i b_i\theta-\theta^2\nu-2\theta b_i\nu\right)+c\theta\left(b_i+\theta\right)\\&\quad-(1-\lambda\alpha)\left(\left(p_1^i\right)^{S*}+\left(p_1^i\right)^{B*}\right)\left[2\left(b_i+\theta\right)^2-\theta^2\right]>0,\\&\pi_1^{S*}-\pi_1^{B*}=\frac{(1-\lambda\alpha)\left(\theta a_L+\theta s a_L+2b_L s a_L+\theta\nu\beta+2w_L\theta^2+2w_L b_L\theta-\theta^2\nu-2\theta b_L\nu\right)+c\theta\left(b_L+\theta\right)}{2\left(b_L+\theta\right)}\left(\left(p_1^L\right)^{S*}-\left(p_1^L\right)^{B*}\right)\\&\quad+\frac{(1-\lambda\alpha)\left(\theta a_H+\theta s a_H+2b_H s a_H+\theta\nu\beta+2w_H\theta^2+2w_H b_H\theta-\theta^2\nu-2\theta b_H\nu\right)+c\theta\left(b_H+\theta\right)}{2\left(b_H+\theta\right)}\left(\left(p_1^H\right)^{S*}-\left(p_1^H\right)^{B*}\right)\\&\quad-(1-\lambda\alpha)\frac{2\left(b_L+\theta\right)^2-\theta^2}{2\left(b_L+\theta\right)}\left[\left(\left(p_1^L\right)^{S*}\right)^2-\left(\left(p_1^L\right)^{B*}\right)^2\right]-(1-\lambda\alpha)\frac{2\left(b_H+\theta\right)^2-\theta^2}{2\left(b_H+\theta\right)}\left[\left(\left(p_1^H\right)^{S*}\right)^2-\left(\left(p_1^H\right)^{B*}\right)^2\right]>0.\end{aligned} \tag{B.1}$$

If $(p_1^L)^{S*} < (p_1^L)^{B*}$ and $(p_1^H)^{S*} < (p_1^H)^{B*}$, we can prove $\pi_1^{S*} > \pi_1^{B*}$ similarly.

From Propositions 3 and 7 with (3), we obtain that

$$\begin{aligned}\pi_2^{B*}-\pi_2^{S*}&=\frac{\theta}{4\left(b_L+\theta\right)}\Big\{2(1-\lambda\alpha)\\&\quad\cdot\left[(1-s)a_L+\nu(\beta+\theta)-w_L\left(b_L+\theta\right)\right]\\&\quad-2c\left(b_L+\theta\right)+\theta(1-\lambda\alpha)\left(\left(p_1^L\right)^{B*}+\left(p_1^L\right)^{S*}\right)\Big\}\\&\quad\cdot\left(\left(p_1^L\right)^{B*}-\left(p_1^L\right)^{S*}\right)+\frac{\theta}{4\left(b_H+\theta\right)}\Big\{2(1-\lambda\alpha)\\&\quad\cdot\left[(1-s)a_H+\nu(\beta+\theta)-w\left(b_H+\theta\right)\right]\\&\quad-2c\left(b_H+\theta\right)+\theta(1-\lambda\alpha)\left(\left(p_1^H\right)^{B*}+\left(p_1^H\right)^{S*}\right)\Big\}\\&\quad\cdot\left(\left(p_1^H\right)^{B*}-\left(p_1^H\right)^{S*}\right).\end{aligned} \tag{B.2}$$

Note that

$$D_2^i=\left[(1-s)a_i-b_i p_2^i\right]+\beta\nu+\theta\left(p_1^i+\nu-p_2^i\right)>0, \tag{B.3}$$

so we show that

$$\begin{aligned}&2(1-\lambda\alpha)\left[(1-s)a_i+\nu(\beta+\theta)-w_i\left(b_i+\theta\right)\right]\\&\quad-2c\left(b_i+\theta\right)+\theta(1-\lambda\alpha)\left(\left(p_1^i\right)^{B*}+\left(p_1^i\right)^{S*}\right)\\&\quad>0.\end{aligned} \tag{B.4}$$

Thus, we have the following:

(1) $\pi_2^{B*} > \pi_2^{S*}$, if $(p_1^L)^{B*} > (p_1^L)^{S*}$ and $(p_1^H)^{B*} > (p_1^H)^{S*}$.

(2) $\pi_2^{B*} = \pi_2^{S*}$, if $(p_1^L)^{B*} = (p_1^L)^{S*}$ and $(p_1^H)^{B*} = (p_1^H)^{S*}$.

(3) $\pi_2^{B*} < \pi_2^{S*}$, if $(p_1^L)^{B*} < (p_1^L)^{S*}$ and $(p_1^H)^{B*} < (p_1^H)^{S*}$.

Combined with Corollary 8, the proposition is proved.

C. Proof of Corollary 10

The profit of hotel from the economy rooms is as follows:

$$\begin{aligned}
\pi_L &= \left[(1-\lambda\alpha)\left(sa_L-\theta v\right)-c\theta\right] p_1^L \\
&\quad + \left[(1-\lambda\alpha)\left(a_L-sa_L+\beta v+\theta v\right)+c\left(b_L+\theta\right)\right] p_2^L \\
&\quad + 2\theta\left(1-\lambda\alpha\right) p_1^L p_2^L - \left(1-\lambda\alpha\right)\left(b_L+\theta\right)\left(p_1^L\right)^2 \\
&\quad - \left(1-\lambda\alpha\right)\left(b_L+\theta\right)\left(p_2^L\right)^2 \\
&\quad - \left(a_L-sa_L+\beta v+\theta v\right) c.
\end{aligned} \tag{C.1}$$

Define $(p_1^L)^{B(S)} = (A_1^L)^{B(S)} + (B_1^L)^{B(S)} w_L$, $(p_2^L)^{B(S)} = (A_2^L)^{B(S)} + (B_2^L)^{B(S)} w_L$, where

$$\begin{aligned}
\left(B_1^L\right)^B &= \frac{3\theta\left(b_L+\theta\right)}{4\left(b_L+\theta\right)^2-\theta^2}, \\
\left(B_1^L\right)^S &= \frac{\theta\left(b_L+\theta\right)}{2\left(b_L+\theta\right)^2-\theta^2}. \\
\left(B_2^L\right)^{B(S)} &= \frac{\theta}{2\left(b_L+\theta\right)}\left(B_1^L\right)^{B(S)}+\frac{1}{2}, \\
\left(A_1^L\right)^B &= \frac{(1-\lambda\alpha)\left(\theta a_L+\theta sa_L+2b_L sa_L+\theta v\beta-\theta^2 v-2\theta b_L v\right)+c\theta\left(b_L+\theta\right)}{(1-\lambda\alpha)\left[4\left(b_L+\theta\right)^2-\theta^2\right]}, \\
\left(A_1^L\right)^S &= \frac{(1-\lambda\alpha)\left(\theta a_L+\theta sa_L+2b_L sa_L+\theta v\beta-\theta^2 v-2\theta b_L v\right)+c\theta\left(b_L+\theta\right)}{(1-\lambda\alpha)\left[4\left(b_L+\theta\right)^2-2\theta^2\right]}, \\
\left(A_2^L\right)^{B(S)} &= \frac{\theta}{2\left(b_L+\theta\right)}\left(A_1^L\right)^{B(S)}+\frac{(1-\lambda\alpha)\left[(1-s)\,a_L+v\left(\theta+\beta\right)\right]+c\left(b_L+\theta\right)}{2\left(1-\lambda\alpha\right)\left(b_L+\theta\right)}.
\end{aligned} \tag{C.2}$$

Let $\partial\pi_L/\partial w = 0$, then we get w_L^*. Similarly we can get w_H^*.

D. Proof of Proposition 12

Let $\pi_d - \pi_r > 0$, then we can get s_0, where the profits of hotel from HA and OTA are as follows:

$$\begin{aligned}
\pi_d &= (1-\lambda\alpha)\left(p_1^L D_1^L + p_1^H D_1^H\right), \\
\pi_r &= (1-\lambda\alpha)\left(w_L D_2^L + w_H D_2^H\right).
\end{aligned} \tag{D.1}$$

E. Proof of Corollary 11

If $\Delta C_L \le \lambda/(1-\lambda)Q_L$, then the profit of hotel is π_{11}^L. The derivative of $\pi_{11}^L(\Delta C_L)$ is $\partial\pi_{11}^L/\partial\Delta C_L = (1-\lambda\alpha)p_1^L > 0$. So with the increase of ΔC_L, π_{11}^L will increase. When $\Delta C_L = \lambda/(1-\lambda)Q_L$, π_{11}^L is maximized. If $\Delta C_L \ge \lambda/(1-\lambda)Q_L$, then the profit of hotel is π_{12}^L. The derivative of $\pi_{12}^L(\Delta C_L)$ is $\partial\pi_{12}^L/\partial\Delta C_L = (1-\lambda\alpha)p_1^L - (1-\lambda)(p_1^L + c_L) < 0$. So with the increase of ΔC_L, π_{11}^L will decline. When $\Delta C_L = \lambda/(1-\lambda)Q_L$, π_{11}^L is maximized.

Similarly we can get ΔC_H^*.

F. Proof of Proposition 13

Assume that the overbooking proportion is γ_L and γ_H.

(1) Consider $(D_1^L + D_2^L + D_1^H + D_2^H)(1-\lambda) \le Q$ and $(D_1^L + D_2^L)(1-\lambda) > Q_L$, $(D_1^H + D_2^H)(1-\lambda) < Q_H$.

That is,

$$\begin{aligned}
\gamma_H &< \frac{\lambda}{1-\lambda} < \gamma_L \\
&\le \frac{\lambda Q_L + Q_H\left[1-(1-\lambda)\left(1+\gamma_H\right)\right]}{(1-\lambda)\,Q_L}.
\end{aligned} \tag{F.1}$$

Thus, the hotel would like to overbook to get more profit rather than room vacancy.

(2) Consider $(D_1^L+D_2^L)(1-\lambda) < Q_L$ and $(D_1^H+D_2^H)(1-\lambda) > Q_H$.

That is, $\gamma_L < \lambda/(1-\lambda) < \gamma_H$. Then the profit of hotel is

$$\begin{aligned}
\pi_{112} &= (1-\lambda\alpha)\left(p_1^L D_1^L + p_1^H D_1^H + w_L D_2^L + w_H D_2^H\right) \\
&\quad - c_u^H\left[\left(D_1^H + D_2^H\right)(1-\lambda) - Q_H\right].
\end{aligned} \tag{F.2}$$

Then when $c_u^H < (1-\lambda\alpha)p_1^H/(1-\lambda)$, that is, $\pi_{112} - \pi_{111} > 0$, the hotel would carry out overbooking.

(3) Consider $(D_1^L + D_2^L + D_1^H + D_2^H)(1-\lambda) > Q$ and $(D_1^L + D_2^L)(1-\lambda) > Q_L$, $(D_1^H + D_2^H)(1-\lambda) < Q_H$.

That is, $\gamma_L > (\lambda Q_L + Q_H[1-(1-\lambda)(1+\gamma_H)])/(1-\lambda)Q_L$ and $\gamma_H < \lambda/(1-\lambda)$. Then the profit of hotel is

$$\pi_{113} = (1-\lambda\alpha)\left(p_1^L D_1^L + p_1^H D_1^H + w_L D_2^L + w_H D_2^H\right) - c_u^L\left[\left(D_1^L + D_2^L + D_1^H + D_2^H\right)(1-\lambda) - Q\right]. \quad \text{(F.3)}$$

Then, when $c_u^L < (1-\lambda\alpha)p_1^L/(1-\lambda)$, $\pi_{113} - \pi_{111} > 0$, thus the hotel would carry out overbooking.

(4) Consider $(D_1^L + D_2^L + D_1^H + D_2^H)(1-\lambda) > Q$ and $(D_1^L + D_2^L)(1-\lambda) > Q_L$, $(D_1^H + D_2^H)(1-\lambda) > Q_H$.

That is, $\gamma_L > (\lambda Q_L + Q_H[1-(1-\lambda)(1+\gamma_H)])/(1-\lambda)Q_L$ and $\gamma_H > (\lambda Q_H + Q_L[1-(1-\lambda)(1+\gamma_L)])/(1-\lambda)Q_H$.

Then, the profit of hotel is as follows:

$$\pi_{114} = (1-\lambda\alpha)\left(p_1^L D_1^L + p_1^H D_1^H + w_L D_2^L + w_H D_2^H\right) - c_u^L\left[\left(D_1^L + D_2^L\right)(1-\lambda) - Q_L\right] - c_u^H\left[\left(D_1^H + D_2^H\right)(1-\lambda) - Q_H\right]. \quad \text{(F.4)}$$

Then, when $c_u^L < (1-\lambda\alpha)p_1^L/(1-\lambda)$ and $c_u^H < (1-\lambda\alpha)p_1^H/(1-\lambda)$, $\pi_{114} - \pi_{111} > 0$, overbooking is preferred.

G. Threshold Values

Threshold values are as follows:

$$L_1 = \theta a_i + \theta s a_i + 2b_i s a_i + \theta v\beta + 3w_i\theta^2 + 3w_i b_i\theta - \theta^2 v - 2\theta b_i v,$$

$$L_2 = 2a_i(b_i+\theta) - s a_i(2b_i+\theta) + \theta^2 w_i + 2(\theta+\beta)(b_i + \theta)v + 2w_i(b_i+\theta)^2 - \theta^2 v,$$

$$L_3 = 4\theta v b_i\beta - \theta s a_i b_i + 2v\beta b_i^2 + 3a_i\theta^2 + 4a_i b_i\theta - 2a_i b_i^2 s + 3\theta^2 v\beta + \theta^2 v b_i + 2\theta v b_i^2 + 2a_i b_i^2,$$

$$L_4 = 3a_i b_i\theta + 3\theta^2 v\beta - 2\theta v b_i^2 - \theta^2 v b_i + 3a_i\theta^2 + 2a_i b_i^2 s + a b_i s\theta + 3b_i\beta v\theta,$$

$$L_5 = \theta a_i + \theta s a_i + 2b_i s a_i + \theta v\beta - \theta^2 v - 2\theta b_i v,$$

$$G_1 = \left[(1-\lambda\alpha)(s a_i - \theta v) - c\theta\right]\left(B_1^i\right)^{B(S)} + \left[(1-\lambda\alpha)(a_i - s a_i + \beta v + \theta v) + c(b_i+\theta)\right] \cdot \left(B_1^i\right)^{B(S)} + 2\theta(1-\lambda\alpha)\left[\left(A_1^i\right)^{B(S)}\left(B_2^i\right)^{B(S)} + \left(B_1^i\right)^{B(S)}\left(A_2^i\right)^{B(S)}\right] - 2(1-\lambda\alpha)(b_i+\theta) \cdot \left[\left(A_1^i\right)^{B(S)}\left(B_1^i\right)^{B(S)} + \left(A_2^i\right)^{B(S)}\left(B_2^i\right)^{B(S)}\right],$$

$$G_2 = 2(1-\lambda\alpha) \cdot \left\{(b_i+\theta)\left[\left(\left(B_1^i\right)^{B(S)}\right)^2 + \left(\left(B_2^i\right)^{B(S)}\right)^2\right] - 2\theta\left(B_1^i\right)^{B(S)}\left(B_2^i\right)^{B(S)}\right\},$$

$$s_0 = \left((b_L+\theta)\left(p_1^L\right)^2 + (b_H+\theta)\left(p_1^H\right)^2 + \theta w_L\left(p_1^L - p_2^L\right) + \theta w_H\left(p_1^H - p_2^H\right) + \theta v\left(p_1^L + p_1^H\right) + (w_L + w_H)(\beta+\theta)v + w_L\left(a_L - b_L p_2^L\right) + w_H\left(a_H - b_H p_2^H\right) - \theta\left(p_1^L p_2^L + p_1^H p_2^H\right)\right)\left(a_L\left(p_1^L + w_L\right) + a_H\left(p_1^H + w_H\right)\right)^{-1}. \quad \text{(G.1)}$$

Acknowledgments

The authors thank associate editor and anonymous referees for their numerous constructive comments and encouragement that have improved their paper greatly. The work was partly supported by (i) the National Natural Science Foundation of China (71201083, 71572058, and 71571100); (ii) the Fundamental Research Funds for the Central Universities, SCUT (2015ZZ057); (iii) Jiangsu Province Science Foundation for Youths (BK2012379); and (iv) the Fundamental Research Funds for the Central Universities (NS2015076).

References

[1] J. Li, W. N. Zhang, H. Xu, and J. Jiang, "Dynamic competition and cooperation of road infrastructure investment of multiple tourism destinations: a case Study of Xidi and Hongcun World Cultural Heritage," *Discrete Dynamics in Nature and Society*, vol. 2015, Article ID 962028, 10 pages, 2015.

[2] T. Paksoy, E. Ozceylan, and G.-W. Weber, "Profit oriented supply chain network optimization," *Central European Journal of Operations Research*, vol. 21, no. 2, pp. 455–478, 2013.

[3] X. Zhang, H. Song, and G. Q. Huang, "Tourism supply chain management: a new research agenda," *Tourism Management*, vol. 30, no. 3, pp. 345–358, 2009.

[4] Z. Yang and J. Cai, "Do regional factors matter? Determinants of hotel industry performance in China," *Tourism Management*, vol. 52, pp. 242–253, 2016.

[5] B. Pan, L. Zhang, and R. Law, "The complex matter of online hotel choice," *Cornell Hospitality Quarterly*, vol. 54, no. 1, pp. 74–83, 2013.

[6] L. Ling, X. Guo, and C. Yang, "Opening the online marketplace: an examination of hotel pricing and travel agency on-line

distribution of rooms," *Tourism Management*, vol. 45, pp. 234–243, 2014.

[7] L. Ling, Y. Dong, X. Guo, and L. Liang, "Availability management of hotel rooms under cooperation with online travel agencies," *International Journal of Hospitality Management*, vol. 50, pp. 145–152, 2015.

[8] J. N. K. Liu and E. Y. Zhang, "An investigation of factors affecting customer selection of online hotel booking channels," *International Journal of Hospitality Management*, vol. 39, pp. 71–83, 2014.

[9] L. Fox, "Four Seasons Unveils $18 Million Website as Luxury Travel Grows," 2012, http://www.tnooz.com/article/four-seasons-unveils-18-million-dollar-website-as-luxury-travel-grows.

[10] C.-M. Chen and Y.-C. Lin, "The influence of uncertain demand on hotel capacity," *International Journal of Hospitality Management*, vol. 34, no. 1, pp. 462–465, 2013.

[11] S. Yang, G. Q. Huang, H. Song, and L. Liang, "Game-theoretic approach to competition dynamics in tourism supply chains," *Journal of Travel Research*, vol. 47, no. 4, pp. 425–439, 2009.

[12] G. Q. Huang, H. Song, and X. Zhang, "A comparative analysis of quantity and price competitions in tourism supply chain networks for package holidays," *Service Industries Journal*, vol. 30, no. 10, pp. 1593–1606, 2010.

[13] Y. Huang, H. Song, G. Q. Huang, and J. Lou, "A comparative study of tourism supply chains with quantity competition," *Journal of Travel Research*, vol. 51, no. 6, pp. 717–729, 2012.

[14] H. K. Lee and Y. Fernando, "The antecedents and outcomes of the medical tourism supply chain," *Tourism Management*, vol. 46, pp. 148–157, 2015.

[15] W.-Y. K. Chiang, D. Chhajed, and J. D. Hess, "Direct marketing, indirect profits: a strategic analysis of dual-channel supply-chain design," *Management Science*, vol. 49, no. 1, pp. 1–20, 2003.

[16] A. A. Tsay and N. Agrawal, "Channel conflict and coordination in the E-commerce age," *Production and Operations Management*, vol. 13, no. 1, pp. 93–110, 2004.

[17] Q.-H. Li and B. Li, "Dual-channel supply chain equilibrium problems regarding retail services and fairness concerns," *Applied Mathematical Modelling*, 2016.

[18] S. Panda, N. M. Modak, S. S. Sana, and M. Basu, "Pricing and replenishment policies in dual-channel supply chain under continuous unit cost decrease," *Applied Mathematics and Computation*, vol. 256, pp. 913–929, 2015.

[19] Y. Liu, C. Ding, C. Fan, and X. Chen, "Pricing decision under dual-channel structure considering fairness and free-riding behavior," *Discrete Dynamics in Nature and Society*, vol. 2014, Article ID 536576, 10 pages, 2014.

[20] J. Chen, H. Zhang, and Y. Sun, "Implementing coordination contracts in a manufacturer Stackelberg dual-channel supply chain," *Omega*, vol. 40, no. 5, pp. 571–583, 2012.

[21] Q. Ding, C. Dong, and Z. Pan, "A hierarchical pricing decision process on a dual-channel problem with one manufacturer and one retailer," *International Journal of Production Economics*, vol. 175, pp. 197–212, 2016.

[22] B. Rodriguez and G. Aydin, "Pricing and assortment decisions for a manufacturer selling through dual channels," *European Journal of Operational Research*, vol. 242, no. 3, pp. 901–909, 2015.

[23] T. Lockyer, "The perceived importance of price as one hotel selection dimension," *Tourism Management*, vol. 26, no. 4, pp. 529–537, 2005.

[24] C.-M. Chen, H.-W. Yang, E. Y. Li, and C.-C. Liu, "How does hotel pricing influence guest satisfaction by the moderating influence of room occupancy?" *International Journal of Hospitality Management*, vol. 49, pp. 136–138, 2015.

[25] J.-M. Espinet, M. Fluviá, R. Rigall-I-Torrent, and A. Saló, "Hotel characteristics and seasonality in prices: an analysis using Spanish tour operators' brochures," *Tourism Economics*, vol. 18, no. 4, pp. 749–767, 2012.

[26] C. Juaneda, J. M. Raya, and F. Sastre, "Pricing the time and location of a stay at a hotel or apartment," *Tourism Economics*, vol. 17, no. 2, pp. 321–338, 2011.

[27] G. Abrate, G. Fraquelli, and G. Viglia, "Dynamic pricing strategies: evidence from European hotels," *International Journal of Hospitality Management*, vol. 31, no. 1, pp. 160–168, 2012.

[28] D. Q. Yao and J. J. Liu, "Competitive pricing of mixed retail and e-tail distribution channels," *Omega*, vol. 33, no. 3, pp. 235–247, 2005.

[29] J. Zhao and L. Wang, "Pricing and retail service decisions in fuzzy uncertainty environments," *Applied Mathematics and Computation*, vol. 250, pp. 580–592, 2015.

[30] A. A. Tsay and N. Agrawal, "Channel dynamics under price and service competition," *Manufacturing & Service Operations Management*, vol. 2, no. 4, pp. 372–391, 2000.

[31] S. C. Choi, "Price competition in a duopoly common retailer channel," *Journal of Retailing*, vol. 72, no. 2, pp. 117–134, 1996.

[32] Y. Bakos and E. Brynjolfsson, "Bundling and competition on the internet," *Marketing Science*, vol. 19, no. 1, pp. 63–82, 2000.

[33] H. Song, J. H. Kim, and S. Yang, "Confidence intervals for tourism demand elasticity," *Annals of Tourism Research*, vol. 37, no. 2, pp. 377–396, 2010.

[34] D. M. Hanssens, L. J. Parsons, and R. L. Schultz, *Market Response Models: Econometric and Time Series Analysis*, Kluwer Academic Publishers, Boston, Mass, USA, 2nd edition, 2001.

[35] H. Kurata, D.-Q. Yao, and J. J. Liu, "Pricing policies under direct vs. indirect channel competition and national vs. store brand competition," *European Journal of Operational Research*, vol. 180, no. 1, pp. 262–281, 2007.

[36] R. Law and C. Cheung, "A study of the perceived importance of the overall website quality of different classes of hotels," *International Journal of Hospitality Management*, vol. 25, no. 3, pp. 525–531, 2006.

[37] K. B. Chen, "Procurement strategies and coordination mechanism of the supply chain with one manufacturer and multiple suppliers," *International Journal of Production Economics*, vol. 138, no. 1, pp. 125–135, 2012.

A Novel Travel Group Recommendation Model Based on User Trust and Social Influence

Zhiyun Xu,[1] Xiaoyao Zheng,[2] Haiyan Zhang,[2] and Yonglong Luo[1,2]

[1]*School of Geography and Tourism, Anhui Normal University, Wuhu 241002, China*
[2]*School of Computer and Information, Anhui Normal University, Wuhu 241002, China*

Correspondence should be addressed to Yonglong Luo; ylluo@ustc.edu.cn

Academic Editor: Nicola Bicocchi

The interactions between group members often have a significant impact on the results of group recommendations. The traditional group recommendation algorithm does not consider the trust and social influence among users. It involves a low utilization rate of social relationship information, which leads to a low accuracy and satisfaction of group recommendations. Considering these issues, in this study, we propose a travel group recommendation model based on user trust and social influence. Based on the user trust relationship, this model defines the user direct and indirect trust and calculates the user global trust by combining the two trusts. Subsequently, the PageRank algorithm is used to calculate the social influence of users based on their interaction relationship history. Thereafter, a consensus model integrating the intra- and intergroup prediction scores is designed by integrating users' global trust and social influence to realize group recommendations for tourist attractions. Comparison experiments with several well-known group recommendation models for datasets of different scenic spots in Beijing demonstrate that the proposed model provides a better recommendation performance.

1. Introduction

Online searches have become the main method for tourists to obtain information before traveling. However, with the rise of social networks and travel websites, tourists are often exposed to a large quantity of information and product selections; therefore, a travel recommender system is an effective method for overcoming the issue of information overload [1]. By learning user history, such recommendation systems establish a description of user interest preferences and recommend items or sets of items that the user could be potentially interested in, thereby providing a personalized service. Currently, such recommendation technology has been widely used by major e-commerce sites and has promoted the sales, as well as improved user satisfaction and loyalty [2–4].

Massive data present in social networks are rich in information regarding the users. Mining interaction rules among users can effectively improve the effectiveness of group recommendations. Tourism is an activity with rich and varied contextual information, including food, accommodation, transportation, travel, shopping, and entertainment information, and each aspect has its own attributes. Compared with other items, tourism products comprise increasingly complex information. The choice of a tourist destination is a complex decision-making process, which often requires a collective decision, and the entirety of the decision-making process requires the support of relevant tourism information [5]. However, current research focuses on the tourism project recommendation of a single user, which relies on user and project information and does not sufficiently consider decision-making within a group [6]. Therefore, it is necessary to expand the traditional single-user recommendation method. According to online evaluation information of tourists, the opinions and social relationships of users within a group are utilized to form a common decision-making mechanism and improve the accuracy and satisfaction of a tourism recommendation. This is of great importance to the research of tourism recommendations. Christensen et al. [7] proposed a society-based approach to travel recommendation systems, which constructs a group strategy model by analyzing users' preferences and the social relationships between group members. Guo et al. [8] attempted to use tag information to determine if users had similar neighbors and extracted tag information from the photos of neighbors to design a group recommendation scheme based on similar neighbors. Currently, the application of matrix factorization method to recommendation has significantly improved the accuracy of recommendation [9–11]. However, this method is not effective for data sparsity, and it cannot solve the problem of new user and new item recommendation. Mining user behavior [12] and extracting item features [13,14] can effectively improve the accuracy of recommendation. These methods are all applied to single-user recommendation, and the current research on group users is still insufficient. The key issue of group recommendations is the preference fusion strategy. At present, the fusion strategy mainly considers the introduction of contextual information, neighbor interactions of user preference, and geographical, spatial, and temporal locations. However, the current group recommendation system is entirely based on the static relationships between users, and the social influence of users within and between groups is not adequately considered. Moreover, the preferences of group users change as social interactions change [15].

It is easier for high-trust group users to reach consensus and achieve accurate item recommendation [16]. How to accurately measure the trust value between users is one of the key steps of using trust to improve the group recommendation efficiency. In addition, due to the complexity of attributes such as context, item, and user in travel recommendation, this paper employs users' social influence to improve the weight of each user in the group user's recommendation to improve the accuracy of recommendation. To overcome this issue, this study proposes a tourist group recommendation model based on a social influence and user trust recommendation model, namely, TSTGR. This model considers the social influence of social networks and trust between the users in the group integration strategy, optimizes the differences within the group consensus, and realizes tourism destination recommendations. Finally, the validity of the proposed group recommendation algorithm is verified using a real dataset from Beijing.

2. Related Works

2.1. Group Recommendations. The group recommender system expands the recommended service objects from one user to multiple users. Compared with the traditional recommender system for a single user, the group recommender system better considers the interactions between group members and social factors [17,18]. Many differences exist between the group and traditional recommendation systems. First, the recommendation service object is changed from a single user to a group. The group recommender system must consider user interaction behaviors and preference fusion, and the degree of association and interaction between group members will also affect the recommendation results. Second, the recommendation results of the group recommendation system are shared by all members of the group and provide a reference for group decision-making, while the recommendation results of the traditional recommendation system are unique to a user. Therefore, the difficulty in researching group recommendation systems stems from the reasonable coordination of the different preferences of group members such that the recommendation results can meet the preferences of the group members to the greatest extent. There are many different preference fusion strategies for group recommendation systems. Previous literature [19] has elaborated on common fusion strategies. Generally, they can be categorized as single-preference [20] and mixed [21] fusion strategies. Tang et al. [15] proposed a preference fusion strategy based on user interaction behaviors, but this strategy did not consider the influence of the consumed items in the group on the recommendation results. Hong et al. [22] implemented group recommendations based on social affinity and credibility according to users' historical records and evaluation content characteristics but did not consider users' ratings. Additionally, their fusion strategy was too simple, leading to a low recommendation accuracy. Wang et al. [23] proposed a bidirectional tensor decomposition model for group recommendations and used the Bayesian personalized ordering technique to learn the parameters in the proposed BTF-GR model. Zhang et al. [24] provided a recommendation list by determining the most similar group to which the target customer belongs by combining a personalized recommendation method for group relevance and customer preference. Xiao et al. [25] used adaptive weighting to aggregate group members' preferences to determine the group's decision for a certain service, which was based on the opinions of familiar members and group influence. Therefore, their method can implement group recommendations. Many useful explorations of group recommendations have been conducted. However, the existing preference fusion strategies for group recommendations are relatively simple while the divergence degree between group members is high, and the accuracy of recommendations must be improved. Moreover, the results of existing research are all based on static relationships, which ignore the influence of changes in the social influence of users on the recommendation performance. Therefore, this study conducts further research on the preference fusion of group recommendation systems.

2.2. Tourism Recommendations. Online searches have become the main method for tourists to obtain information before traveling. However, with the increasing prevalence of social networks and travel websites, tourists are often exposed to a large amount of information and product options. A travel recommendation system is an effective means of overcoming the issue of information overload. Kofler et al. [26] collected tourism photos shared on the Flickr platform and obtained the metadata and user-generated data of reach photo, among other information. After the user selects the tourist destination, their method recommends photo sets of tourist attractions related to the destination. Moreno et al. [27] designed a personalized tourism destination recommendation system. They considered the motivation of tourists, as well as user access, evaluation, and personal history records, and then used collaborative filtering technology to recommend scenic spots similar to those visited by the users. Loh et al. [28] established a tourism ontology database and used the text mining method to mine users' preferences for tourism destinations or scenic spots and then query destinations or scenic spots with high similarity to the users' preferences from the tourism ontology knowledge base and recommend these destinations to users. Levi et al. [29] extracted the characteristic values of hotels by analyzing the evaluation records of users on TripAdvisor and Venere using a clustering algorithm. Then, according to the motivation and preferences of the target users, the content-based recommendation technology was adopted to recommend hotels of potential interest to the users, and the satisfaction of the users was verified using a survey. The impact of a single-user travel recommendation is relatively small and the recommendation accuracy is high for this method. Tourism is a complex human activity, and single-user travel recommendations cannot consider all the travel needs of tourists. In the above studies, little attention has been paid to group tourism, the description of tourist user groups has not been adequately precise or detailed, and the granularity of all methods is too coarse. As a result, the recommended tourism items cannot meet the personalized needs of group users. Therefore, with the help of group recommendation technology, this study proposes a group travel recommendation model based on user trust and social influence.

3. Group Recommendation Model Incorporating User Trust and Social Influence

3.1. Recommendation Framework. Group recommendation generally consists of three components: group discovery, preference fusion, and prediction recommendation. According to the social relationships of travel users, this study proposes a group recommendation framework that integrates user trust and social influence, as shown in Figure 1. This framework is mainly composed of three parts: a data acquisition module, preference modeling module, and group recommendation algorithm design module. The data collection module is mainly responsible for the collection and sorting of data for tourist attractions, as well as the collection and processing of social relationships between users. The preference modeling module mainly involves group discovery, trust modeling, and social influence modeling. It is responsible for the division of travel user groups, the quantitative analysis of user influence, and the establishment of trust for users within the group. According to the divided group and preference model, the group recommendation algorithm design module completes the group's rating predictions and recommendations to the tourist.

3.2. User Trust Modeling. Most of the existing group recommendation studies have only considered whether a trust relationship exists between users. Typically, a trusted relationship is denoted as 1 while an untrusted relationship is denoted as 0. This measurement method is relatively simple, and it does not specifically consider the level of trust between users. However, the degree of trust between users will have different effects on the final decision.

This study divides the trust relationship between users into direct and indirect trust. Direct trust is defined when there are common rating items among users and their ratings are consistent. Rating consistency means that the rating is divided into two parts according to the rating level. If a user's rating is greater than the median rating, the user's rating is positive; otherwise, the user's rating is negative. For example, the ratings of user u_i and u_j on item S_k are 4 and 5, and the median rating is 3, indicating that the ratings of user u_i and u_j are both positive, so the ratings of user u_i and u_j are consistent. Indirect trust is obtained by the weighted transfer of direct trust between users. The specific calculation method is shown in Figure 2. Therefore, the trust between u_i and u_j can be obtained by the weighted sum of direct trust and indirect trust, namely, the global trust defined in this paper. Meanwhile, the consensus function among group users can be calculated using the global trust. For users with direct trust relationships, the specific definitions are shown as follows.

Definition 1 (direct trust). Assuming there are N user reviews in the dataset, the direct trust between users is defined by the following equation:

$$D_{ij} = \frac{\sum_{N_{ij}} f\left(u_i, u_j\right)}{N}. \tag{1}$$

Here, N is the total number of evaluations in the user dataset, N_{ij} represents the number of scenic spots jointly evaluated by users u_i and u_j, and $f(u_i, u_j)$ represents the common evaluation function of scenic spots evaluated by users u_i and u_j. If they are consistent, this equation is set to 1; otherwise, it is set to 0. Assume that the rating level is $r = \{1, 2, 3, 4, 5\}$. For example, the rating by user u_i of scenic spot s is 5 and the rating by user u_j of scenic spot s is 4. Both users' ratings are greater than or equal to the median of the rating scale. We then set $f(u_i, u_j) = 1$, which is otherwise set to 0.

Definition 2 (indirect trust). Indirect trust between users is defined using the following equation:

$$I_{ij} = \sum_{k=1, k \neq i}^{N} w_k \times D_{kj}. \tag{2}$$

Here, D_{kj} is the direct trust between users u_k and u_j. The indirect trust between users u_i and u_j is the weighted sum of the direct trust between users u_j and u_k that does not include the direct trust between them, and w_k is the weight.

Based on the direct and indirect trust between users, this study provides a definition of user global trust as follows:

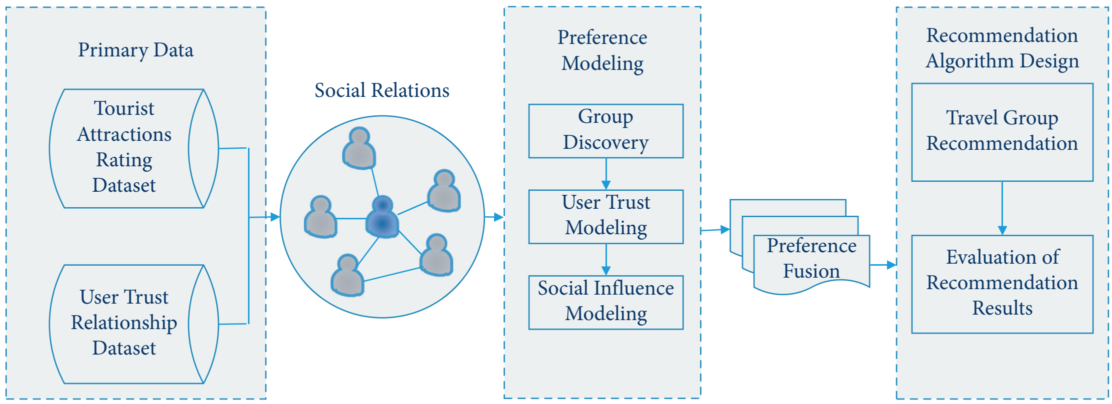


FIGURE 1: Tourism group recommendation framework integrating user trust and social influence.

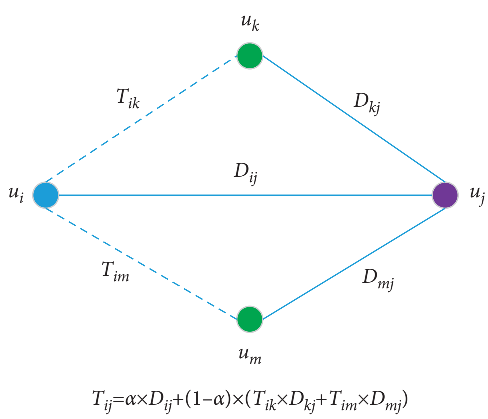


FIGURE 2: Schematic diagram of the global trust.

$$T_{ij} = \alpha D_{ij} + (1-\alpha) I_{ij}. \quad (3)$$

Equation (3) can also be written in the following specific form:

$$T_{ij} = \alpha \frac{\sum_{N_{ij}} f\left(u_i, u_j\right)}{N} + (1-\alpha) \sum_{k=1,k\neq i}^{N} w_k \times \frac{\sum_{N_{ik}} f\left(u_i, u_k\right)}{N}. \quad (4)$$

The global trust of users u_i and u_j is shown in Figure 2, where α is the regulating coefficient of direct trust and indirect trust in the global trust, and $\alpha \in [0, 1]$. When $\alpha = 1$, the global trust of users u_i and u_j is completely determined by the direct trust. When $\alpha = 0$, the global trust of users u_i and u_j is completely determined by the indirect trust. The weight is set to $w_k = T_{ik}$, which represents the global trust of users u_i and u_j.

The collection of trust data is to use the user's rating dataset, and set the users with common ratings and rating consistency to direct trust. Then, the indirect trust is calculated through the direct trust iteration, and finally the trust relationship and trust value between the whole users are obtained.

3.3. Social Influence. Social networks contain a wealth of information, and each user's influence in the network is different. The social network graph of the whole user set can be obtained by abstractions of the connections among users in the social network. After the users submit ratings and text reviews of tourist attractions online, they usually interact with other online users. If there is a common review between user u_i and u_j, then they are called to have social influence, and edge (u_i, u_j) is established. By establishing the connection between users, the user set can be abstract into a graph with users as nodes and their connections as edges, and the improved PageRank algorithm proposed in this paper can be used to calculate the influence of each user in the social network. Therefore, based on the online interaction information of users, we apply the improved PageRank algorithm proposed in this paper to describe the social network influence of users to provide a decision basis for group recommendations.

In this study, the social rating information of users is described as a social network G and users are set as nodes, represented by V. If there is a common rating among users, it is considered that there is a connection between users, which is denoted as an edge set E. If G has n nodes, then the node set is $V = \{v_1, v_2, \ldots, v_n\}$. If there is a connection between users u_i and u_j, denoted by v_{ij}, then the edge set is $E = \left\{v_{ij} \middle| i, j \in n\right\}$. According to the above definition, all users and their relationships are represented as $G = \{V, E\}$ in this study. To obtain the social influence of each user node, we improved the traditional PageRank algorithm and weighted the trust function with the damping coefficient to obtain a new damping coefficient, $\beta = a \cdot T_{ij} + b$, where a and b are linear weights used to adjust the damping coefficient. We set the default values for a and b to 0.5. The setting of this damping coefficient can dynamically adjust the transmission of influence between users. Therefore, the greater the trust between users, the greater the transmission of influence between them, and vice versa. Through the above settings, the PageRank algorithm is used to obtain the social influence of each user. The specific steps are as follows:

Step 1: traverse each node v_i in the node set V and randomly initialize the PageRank value of node v_i to obtain $PR(v_i)$

Step 2: calculate the degree N_i of v_i

Step 3: double traverse the user node and calculate $PR(v_i) = (1-\beta)PR(v_i) + \beta \times (PR(v_i)/N_j)$

Step 4: repeat Step 3 until the entire PageRank influence converges

3.4. Implementation of the Tourism Group Recommendation Algorithm. To improve the accuracy of group recommendations and reduce the degree of disagreement among group members, this study first proposes a new group consensus model that integrates group social influence and trust among users. In this study, $F(G_i, S_j)$ is used to represent the predicted rating of group G_i for a tourist attraction S_j, which is defined as follows:

$$F\left(G_i, S_j\right) = \gamma\left(\sum_{u_i \in G_i} PR\left(u_i\right) \cdot R_i + \sum_{u_i \in G_i} R_{\mathrm{CF}}\left(u_i, S_j\right) + \sum_{u_i \in G_i} R_{\mathrm{TF}}\left(u_i, S_j\right)\right) + (1-\gamma)\left(\sum_{G_t \in \mathrm{Neighor}\left(G_i\right)} \mathrm{Sim}\left(G_t, G_i\right) \cdot R_{G_t}\right). \quad (5)$$

Here, γ is the weighting parameter, $PR(u_i)$ represents the social influence of user u_i, and R_i represents the average rating of the scenic spots reviewed by user u_i. $R_{\mathrm{CF}}(u_i, S_j)$ represents the predicted rating of a single user in the group for tourist attraction S_j based on collaborative filtering technology, wherein $R_{\mathrm{CF}}(u_i, S_j) = (\sum_{u_t \in \mathrm{Neighor}(u_i)} \mathrm{Sim}(u_i, u_t) \cdot r_{tj}) / (\sum_{u_t \in \mathrm{Neighor}(u_i)} \mathrm{Sim}(u_i, u_t))$, and $\mathrm{Neighor}(u_i)$ represents similar neighbors of user u_i. $R_{\mathrm{TF}}(u_i, S_j)$ represents the predicted rating of tourist attraction S_j by a single user in the group based on the global trust among users; i.e., $R_{\mathrm{TF}}(u_i, S_j) = (\sum_{u_t \in \mathrm{Neighor}(u_i)} T_{it} \cdot r_{tj}) / (\sum_{u_t \in \mathrm{Neighor}(u_i)} T_{it})$. $\mathrm{Sim}(G_t, G_i)$ represents the similarity between groups G_i and G_t, and R_{G_t} represents the average rating given by users in group G_t to tourist attraction S_j. The consensus model is composed of two parts, namely, intra- and intergroup prediction ratings, and the weights of these two parts are adjusted using parameter γ. The first part is composed of the social influence score, collaborative filtering score, and trust user score. The second part is composed of the cooperative prediction scores between groups. The specific implementation steps are shown in Algorithm 1.

This algorithm is mainly composed of three parts: the PageRank algorithm used to calculate social influence, traversing user sets for calculating the global trust among users, and the Top-K group rating prediction for scenic spots. The complexity of the PageRank algorithm in the first part is $O(kn^2)$, where k is the number of iterations and n is the number of users. The second part involves the time complexity of calculating the degree of global trust, which is mainly composed of the degrees of direct and indirect trust. Its complexity is $O(C_1(n^2+n))$, where C_1 is a constant. The third part involves the calculation of the consensus function to achieve the Top-K group recommendation, and its complexity is $O(Kmn + n^2 + C_2 n)$, where m is the number of scenic spots in the dataset, K is the number of scenic spots recommended by the Top-K group, and C_2 is a constant. Therefore, the time complexity of the entire algorithm is $O(Kmn + C(n^2+n))$, where C is a constant.

4. Evaluation

4.1. Experimental Dataset and Environment. To verify the effectiveness of the group tourism recommendation scheme proposed in this study, we collected the evaluation data [30] of 200 scenic spots from 37000 tourists extending from July 1, 2014, to June 30, 2017, including the user ID, scenic spot, ticket prices, scores, text evaluation, evaluation time, travel types, etc. A total of 472,710 comment data were collected. The distribution of the scenic spots is shown in Figure 3. In addition, we added a Yelp Restaurant dataset to further verify the effectiveness of the method proposed in this paper. In this dataset, we extracted data from the US state of Arizona, including 622,446 reviews of 9,427 restaurants (https://www.yelp.com/dataset/).

The experimental environment in this study is a 64-bit operating system on the Windows 10 platform, the CPU is an Intel(R) Core(TM) i7-8750H, the main frequency of the processor is 2.20 GHz, and the physical memory is 16.0 GB. The algorithm proposed in this study is implemented by Microsoft Visual C++. To evaluate the recommendation results and algorithms effectively, 80% of the dataset is randomly selected as the training set for training the algorithm and the remaining 20% is used as the test set. The recommendation results are then verified using the test data, and the experimental results are compared and analyzed.

4.2. Evaluating Indicator. In this study, the accuracy rate, recall rate, and normalized loss cumulative gain are used as performance indicators to measure the performance of the group recommendation model such that the performance of several comparison models may be measured more accurately.

(1) Accuracy is defined as follows:

$$\text{precision} = \frac{N_{\mathrm{RL}}}{N_R}, \quad (6)$$

where N_R denotes the total number of items recommended by the recommendation system to users, N_{RL} denotes the number of favorite items in the recommended item set, and N_L denotes the number of items that users like in the whole dataset.

(2) Normalized discounted cumulative gain (nDCG) is an evaluation method based on ranking, which is an

Input: tourism user evaluation dataset **R**, tourism user dataset U, scenic spot dataset S, and parameters $\{\alpha, \gamma\}$
Output: **Top-K** prediction rating of the group for the tourist attraction
divide Groups
for each u **in U**
calculate $PR(u)$ and $\mathrm{Sim}(G_t, G_i)$
end for
for each u_i **in U**
for each u_j **in U**
calculate T_{ij}
end for
end for
for $k \longleftarrow 1$ **to** K
for each G in Group
$F_{\mathrm{IntraGroup}} \longleftarrow \sum_{u_i \in G_i} PR(u_i) \cdot R_i + \sum_{u_i \in G_i} R_{\mathrm{CF}}(u_i, S_j) + \sum_{u_i \in G_i} R_{\mathrm{TF}}(u_i, S_j)$
$F_{\mathrm{InterGroup}} \longleftarrow \sum_{G_i \in \mathrm{Neighbor}(G_i)} \mathrm{Sim}(G_t, G_i) \cdot R_{G_t}$
$F \longleftarrow \gamma F_{\mathrm{IntraGroup}} + (1-\gamma) F_{\mathrm{InterGroup}}$
end for
Update recommendation scenic spot list
end for
return Top K scenic spot

Algorithm 1: Group recommendation algorithm based on social influence and user trust recommendation.

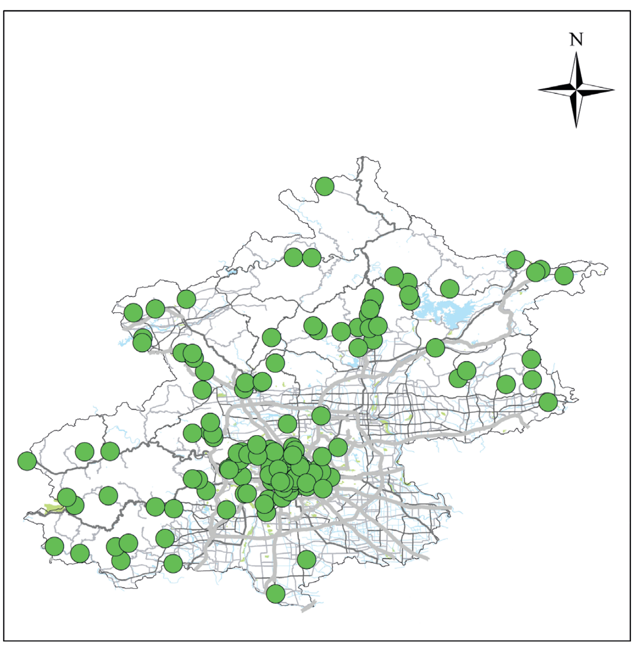


Figure 3: Distribution map of Beijing attractions.

important indicator of the recommendation accuracy of the evaluation group. It is defined as follows:

$$\mathrm{nDCG@}k = \frac{\sum_{g \in G} \mathrm{DCG}_g @k}{\mathrm{IDCG}_g}, \tag{7}$$

where $\mathrm{DCG}_g@k$ denotes the cumulative discount revenue of the recommendation algorithm to the group g recommendation list and IDCG_g denotes a list of the best recommended results for the group g.

4.3. Contrast Model. Herein, four classical group recommendation models are selected and compared with the TSTGR model proposed in this study. The details are as follows:

(1) GRSAT model [22]: this model realizes group recommendations based on social affinity and credibility according to user history and evaluation content characteristics, without considering the user's score.
(2) PLTSGR model [24]: this model adopts a personalized recommendation method combining group relevance and customer preference. An unsupervised method and PLTS are used to determine group association between a customer group and a restaurant group, and a recommendation list is provided by finding the most similar group to which target customers belong.
(3) PFGR model [25]: based on the opinions of familiar members and group influence, this scheme uses adaptive weighting to aggregate the preferences of group members to determine the group's decision for a certain service. A strategy based on the alliance game is used to realize the group recommendation.
(4) SIGR model [18]: this group recommendation model uses the attention mechanism to learn the social impact of each user and adapt their social impacts to different groups. Group information fusion is realized by using and integrating the global and local social network structure information of users.
(5) TSTGR model: in this study, considering the difficulty of merging the preferences of group members for group recommendations, a global trust model is

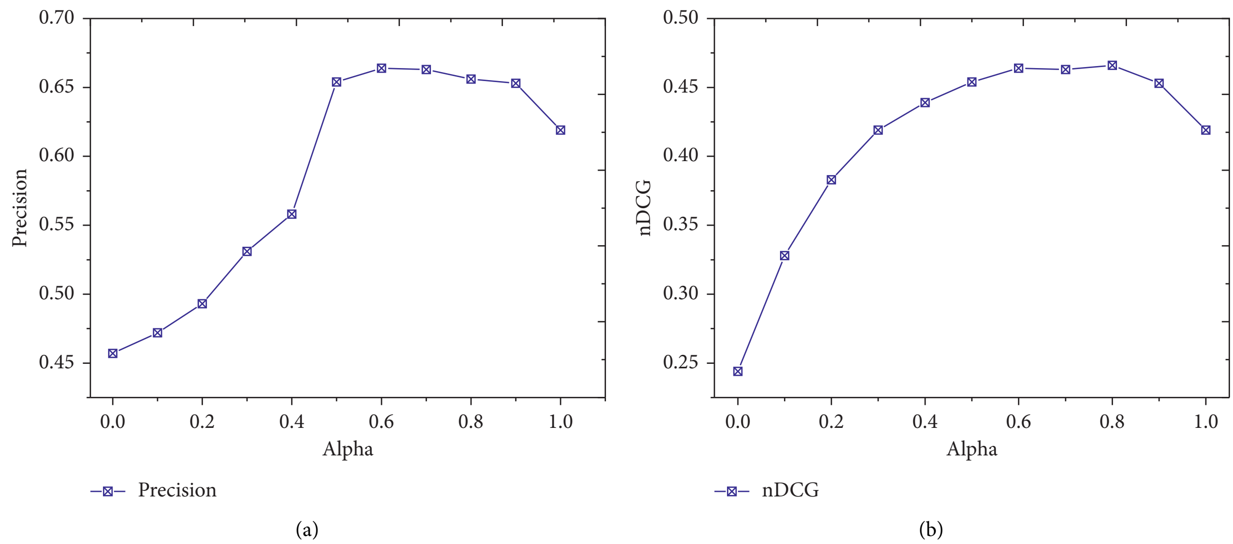


Figure 4: Influence of parameter α on the performance of the TSTGR model. (a) Influence of parameter α on the precision index. (b) Influence of parameter α on the nDCG index.

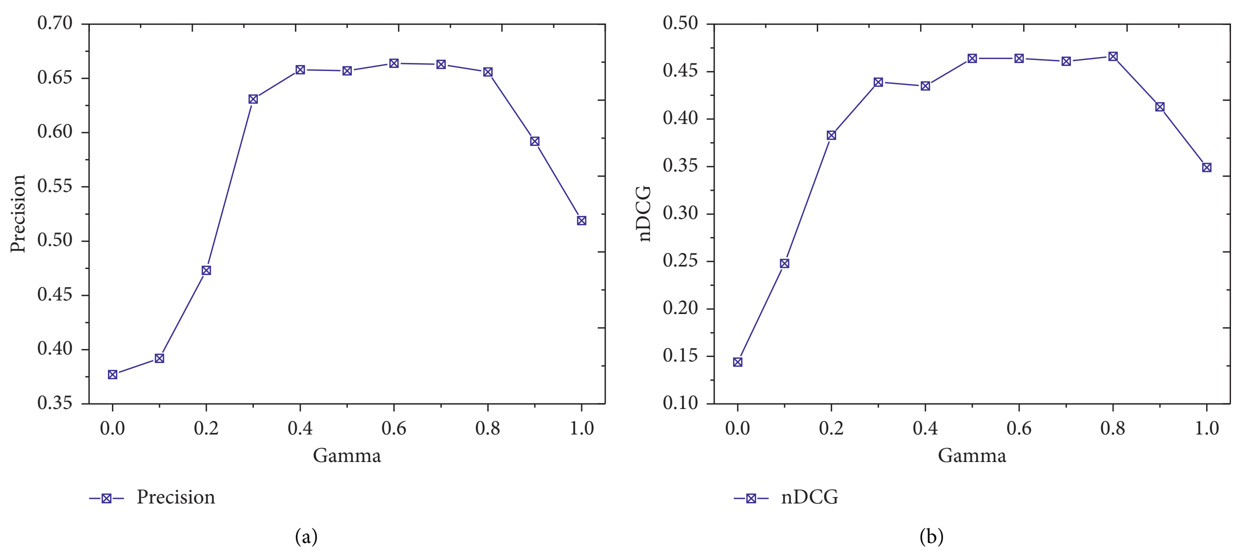


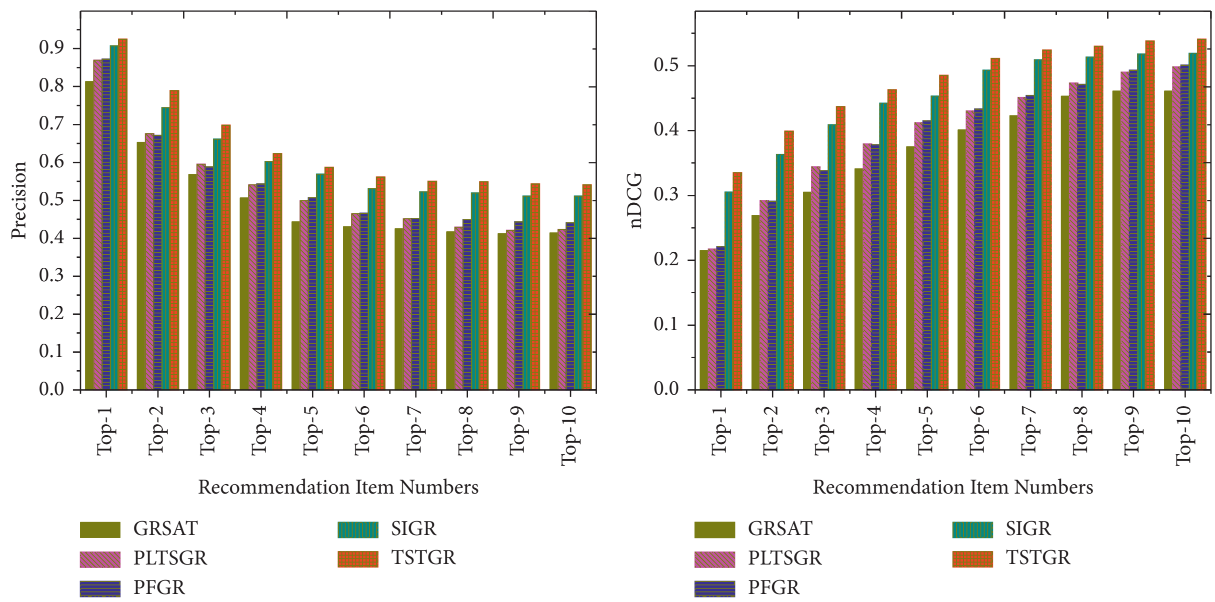
Figure 5: Influence of parameter γ on the performance of the TSTGR model. (a) Influence of parameter γ on the precision index. (b) Influence of parameter γ on the nDCG index.

constructed using direct and indirect trust between users, and the PageRank algorithm is used to measure user influence. The global trust and social influence of users are integrated into the group consensus model to realize the Top-K recommendation for tourist attractions.

4.4. Sensitivity Analysis.

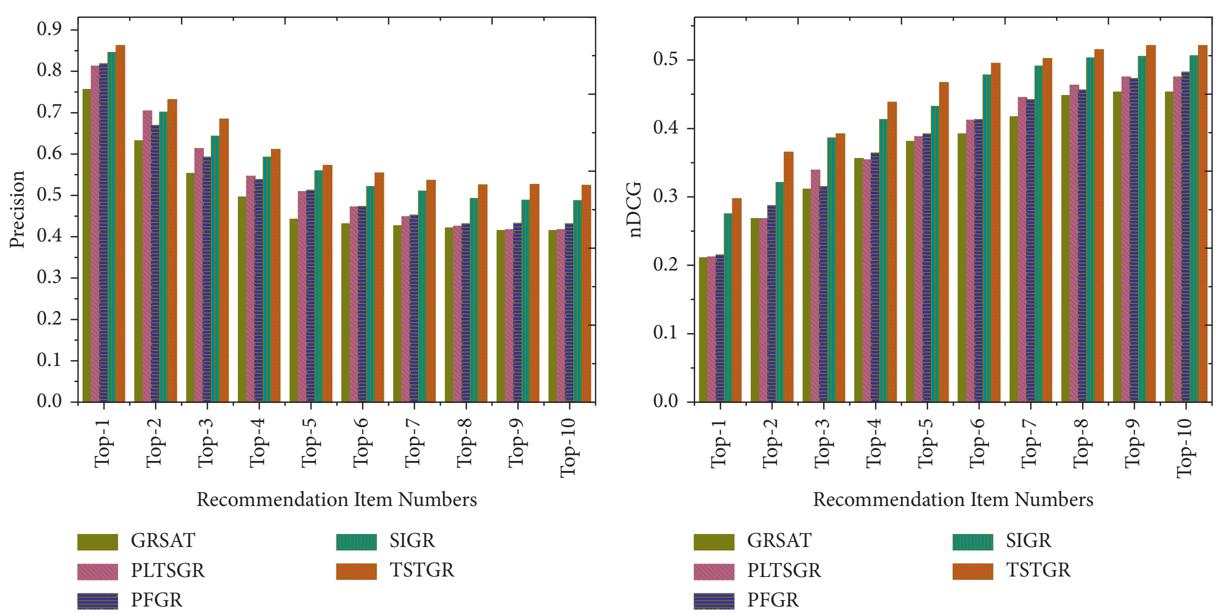
(1) The parameter α: this parameter is the weighting coefficient between direct and indirect trust, and its value range is [0, 1]. If the value of α is 0, the global trust is composed solely of indirect trust. If the value of α is 1, the global trust is only composed of direct trust. In this experiment, the value of α is selected between 0 and 1 to test the influence of this parameter on the performance of the group recommendation method proposed in this study. It can be seen from Figure 4 that when the parameter α is set within [0.5, 0.9], we can achieve a relatively good recommendation performance. Therefore, in the subsequent experiments in this study, the default value of α is set to 0.7.

(2) The parameter γ: the parameter γ is the weighting coefficient of the intra- and intergroup prediction scores in the parameter group consensus model, with a value range of [0, 1]. If the value of γ is 1, the prediction scores of the consensus model are determined by the scores in the group. If the value of γ is 0, the prediction score of the consensus model is



(a) (b)

Figure 6: Performance comparison of the five models on Beijing attractions. (a) Comparative experimental results for the precision index. (b) Comparative experimental results for the nDCG index.



(a) (b)

Figure 7: Performance comparison of the five models on Yelp restaurant. (a) Comparative experimental results for the precision index. (b) Comparative experimental results for the nDCG index.

determined by the scores between groups. In this experiment, the value of γ is selected between 0 and 1, and the influence of this parameter on the performance of the TSGGR model proposed in this study is evaluated. It can be concluded from Figure 5 that the TSTGR model proposed in this study can achieve the best recommendation accuracy and performance when the value of γ is selected within [0.3, 0.8]. In the subsequent comparative experiment in this study, we set $\gamma = 0.5$.

4.5. Contrast Experiment. To further verify the feasibility of the proposed method and evaluate the performance of the TSTGR method, the number of items recommended by a group was varied from 1 to 10 in this experiment to compare the performances of various comparison models considering the recommendation of Top-K. Figure 6 shows the comparison experiment on the dataset of Beijing attractions. Figure 6(a) demonstrates the recommendation precision of the five comparison models when the number of recommended items changes from 1 to 10, and Figure 6(b) shows the nDCG results of the five comparison models when the number of recommended items changes from 1 to 10. Figure 7 shows the comparative experiment results on the Yelp Restaurant dataset. It can be concluded from Figures 6 and 7 that the TSTGR model provides great advantages in terms of precision and nDCG. In addition, as the amount of recommended items is increased, the recommendation accuracy decreases and the nDCG index gradually increases. The GRSAT model only considers the trust and rating of uses, and its decision-making factors for group recommendations are inadequate. The PLTSGR and PFGR models consider the preference and group influence of users within a group and construct a consensus model to achieve group recommendations. They demonstrated performance improvements compared with the GRSAT model. The SIGR model uses an attention mechanism to learn the social impacts of each user, and it then integrates the global and local social network structure information of users to achieve group information fusion, resulting in a relatively high recommendation accuracy. The TSTGR model proposed in this study not only considers social influence, but also increases global user trust by combining direct and indirect trust. In addition, this model also integrates intragroup decision-making and intergroup collaborative recommendations, which further improves the performance of the resulting group recommendation. Therefore, compared with the four previously developed models, the model proposed in this study is more competitive.

5. Conclusion

With the rapid development of smart tourism, tourism group recommendations have become an important topic of research in the field of recommendation systems. In this study, a tourist attraction group recommendation model based on user trust and social influence was proposed. First, according to the dataset of a trust relationship between users, the model integrated the direct and indirect trust between users and calculated the degree of global trust between users. Second, according to the historical evaluation interaction records between users, the PageRank algorithm was used to determine the social influence of the users. Finally, a new consensus function was designed by combining the intra- and intergroup prediction scores to complete the evaluation score for tourist attractions and realize Top-K recommendations. Compared with many well-known group recommendation models, the proposed method demonstrated a good performance. The experimental results demonstrated that the integration of global trust and social influence can effectively improve the accuracy of the resulting group recommendation. In a future work, we will further explore the application of group recommendations in other tourism recommendation fields.

Acknowledgments

This research was supported by the Natural Science Foundation of China (no. 61972439), Natural Science Foundation of Anhui Province (no. 1808085MF172), and Key Program in the Youth Elite Support Plan in Universities of Anhui Province (gxyqZD2019010).

References

[1] J. Borràs, A. Moreno, and A. Valls, “Intelligent tourism recommender systems: a survey,” *Expert Systems with Applications*, vol. 41, no. 16, pp. 7370–7389, 2014.

[2] S. Zhang, L. Yao, A. Sun, and Y. Tay, “Deep learning based recommender system,” *ACM Computing Surveys*, vol. 52, no. 1, pp. 1–38, 2019.

[3] A. Ramlatchan, M. Yang, Q. Liu, M. Li, J. Wang, and Y. Li, “A survey of matrix completion methods for recommendation systems,” *Big Data Mining and Analytics*, vol. 1, no. 4, pp. 308–323, 2018.

[4] J. Han and H. Yamana, “A survey on recommendation methods beyond accuracy,” *IEICE-Transactions on Info and Systems*, vol. 100, no. 12, pp. 2931–2944, 2017.

[5] Y. Li, C. Hu, C. Huang, and L. Duan, “The concept of smart tourism in the context of tourism information services,” *Tourism Management*, vol. 58, pp. 293–300, 2017.

[6] M. Karl, “Risk and uncertainty in travel decision-making: tourist and destination perspective,” *Journal of Travel Research*, vol. 57, no. 1, pp. 129–146, 2018.

[7] I. Christensen, S. Schiaffino, and M. Armentano, “Social group recommendation in the tourism domain,” *Journal of Intelligent Information Systems*, vol. 47, no. 2, pp. 209–231, 2016.

[8] C. Guo, B. Li, and X. Tian, “Flickr group recommendation using rich social media information,” *Neurocomputing*, vol. 204, pp. 8–16, 2016.

[9] Y. Xu, Y. Wu, H. Gao, S. Song, Y. Yin, and X. Xiao, “Collaborative APIs recommendation for artificial intelligence of things with information fusion,” *Future Generation Computer Systems*, vol. 125, pp. 471–479, 2021.

[10] H. Gao, X. Qin, R. J. D. Barroso, W. Hussain, Y. Xu, and Y. Yin, “Collaborative learning-based industrial IoT API recommendation for software-defined devices: the implicit knowledge discovery perspective,” *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 99, pp. 1–11, 2020.

[11] X. Zheng, Y. Luo, L. Sun, J. Zhang, and F. Chen, “A tourism destination recommender system using users’ sentiment and temporal dynamics,” *Journal of Intelligent Information Systems*, vol. 51, no. 3, pp. 557–578, 2018.

[12] H. Gao, L. Kuang, Y. Yin, B. Guo, and K. Dou, "Mining consuming behaviors with temporal evolution for personalized recommendation in mobile marketing apps," *Mobile Networks and Applications*, vol. 25, no. 4, pp. 1233–1248, 2020.

[13] H. Gao and Y. Duan, "The cloud-edge-based dynamic reconfiguration to service workflow for mobile ecommerce environments," *ACM Transactions on Internet Technology*, vol. 21, no. 1, pp. 1–23, 2021.

[14] M. Cao, S. Zhou, and H. Gao, "A recommendation approach based on product attribute reviews: improved collaborative filtering considering the sentiment polarity," *Intelligent automation and soft computing*, vol. 25, no. 3, pp. 593–602, 2019.

[15] F. Tang, K. Liu, L. Feng, and J. Jingwei, "Research on the integration strategy of group recommendation based on user's interactive behaviors," in *Proceedings of the 2016 IEEE International Conference on Cloud Computing and Big Data Analysis*, pp. 367–372, IEEE, Changsha, China, December 2016.

[16] L. Ardissono and N. Mauro, "A compositional model of multi-faceted trust for personalized item recommendation," *Expert Systems with Applications*, vol. 140, Article ID 112880, 2020.

[17] S. Amer-Yahia, S. B. Roy, A. Chawlat, G. Das, and C. Yu, "Group recommendation," *Proceedings of the VLDB Endowment*, vol. 2, no. 1, pp. 754–765, 2009.

[18] H. Yin, Q. Wang, K. Zheng, Z. Li, J. Yang, and X. Zhou, "Social influence-based group representation learning for group recommendation," in *Proceedings of the 2019 IEEE 35th International Conference on Data Engineering (ICDE)*, pp. 566–577, IEEE, Macao, China, April 2019.

[19] D. Qin, X. Zhou, L. Chen, G. Huang, and Y. Zhang, "Dynamic connection-based social group recommendation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 3, pp. 453–467, 2018.

[20] Y.-L. Chen, L.-C. Cheng, and C.-N. Chuang, "A group recommendation system with consideration of interactions among group members," *Expert Systems with Applications*, vol. 34, no. 3, pp. 2082–2090, 2008.

[21] H. Yin, Q. Wang, K. Zheng, Z. Li, and X. Zhou, "Overcoming data sparsity in group recommendation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 1, 2020.

[22] M. Hong, J. J. Jung, and D. Camacho, "GRSAT: a novel method on group recommendation by social affinity and trustworthiness," *Cybernetics & Systems*, vol. 48, no. 3, pp. 140–161, 2017.

[23] J. Wang, Y. Jiang, J. Sun, Y. Liu, and X. Liu, "Group recommendation based on a bidirectional tensor factorization model," *World Wide Web*, vol. 21, no. 4, pp. 961–984, 2018.

[24] C. Zhang, H. Zhang, and J. Wang, "Personalized restaurant recommendation method combining group correlations and customer preferences," *Information Sciences*, vol. 454, pp. 128–143, 2018.

[25] Y. Xiao, Q. Pei, L. Yao, S. Yu, L. Bai, and X. Wang, "An enhanced probabilistic fairness-aware group recommendation by incorporating social activeness," *Journal of Network and Computer Applications*, vol. 156, Article ID 102579, 2020.

[26] C. Kofler, L. Caballero, M. Menendez, V. Occhialini, and M. Larson, "Near2me: an authentic and personalized social media-based recommender for travel destinations," in *Proceedings of the 3rd ACM SIGMM international workshop on Social media*, pp. 47–52, Scottsdale, AZ, USA, November 2011.

[27] A. Moreno, A. Valls, D. Isern, L. Marin, and J. Borràs, "SigTur/E-destination: ontology-based personalized recommendation of tourism and leisure activities," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 1, pp. 633–651, 2013.

[28] S. Loh, F. Lorenzi, R. Saldaña, and D. Licthnow, "A tourism recommender system based on collaboration and text analysis," *Information Technology & Tourism*, vol. 6, no. 3, pp. 157–165, 2003.

[29] A. Levi, O. Mokryn, C. Diot, and N. Taft, "Finding a needle in a haystack of reviews:cold start context-based hotel recommender system demo," in *Proceedings of the ACM Conference on Recommender Systems*, pp. 305-306, Dublin, Ireland, UK, September 2012.

[30] X. Zheng, *Beijing Poi Datasets with Geographical Coordinates and Ratings*, Anhui Normal University, Anhui, China, 2019.

Tourist Behavior Pattern Mining Model Based on Context

Dong-sheng Liu and Shu-jiang Fan

College of Computer Science & Information Engineering, Zhejiang Gongshang University, Xiasha University Town, Hangzhou 310018, China

Correspondence should be addressed to Dong-sheng Liu; lds1118@163.com

Academic Editor: Tinggui Chen

Personalized travel experience and service of tourist has been a hot topic research in the tourism service supply chain. In this paper, we take the context into consideration and propose an analyzed method to the tourist based on the context: firstly, we analyze the context which influences the tourist behavior patterns, select the main context factors, and construct the tourist behavior pattern model based on it; then, we calculate the interest degree of the tourist behavior pattern and mine out the rules with high interest degree with the association rule algorithm; we can make some recommendations to the tourist with better personalized travelling experience and services. At last, we make an experiment to show the feasibility and effectiveness of our method.

1. Introduction

With the development of economy and the improvement of people's living standard, more and more people pay more attention to the quality of personalized travelling experience and service. In recent years, there has emerged more and more personalized ways to travel in tourism, such as FIT travel and independent travel. The traditional mode of travel service limits the diversity of service options, and it cannot fully meet the personalized needs of tourists. How to find the laws and the features of the tourist behavior through mining tourist behavior patterns and offer them better services has been a problem in the tourism service supply chain.

There are many researches concentrating on the tourist behavior pattern. Qing analyzed the characteristics of tourism services and the structural properties, constituent elements, and operation mechanism of tourism service supply chain in the context of modern information technology, and he put forward a new tourism service supply chain conceptual model based on tourist personalized demand [1]. Farmaki took the Troodos (Cyprus) as a case to research on the tourist motivation [2]; Martin and Witt proposed tourism demand forecasting model to represent tourists' cost of living [3]; Smallman and Moore studied on the tourists' decision making [4]; Kim et al. has worked on the Japanese tourists' shopping preference with the decision tree analysis method [5].

These studies only analyzed the tourist from the view point of the psychology and behavioral science of the tourist and have not considered the context set which will influence the tourist behavior patterns. So, in this paper, we take the context into consideration and propose an analyzed method to the tourist based on context to find out the relationship between services in the travel and the context and analyse the important contexts which will influence the tourist behavior. To mine out rules with high interest degree with the association rule algorithm and do some recommendations to the tourist with better personalized travelling experience and services, we propose a method based on network diagram, and it can reflect the relationship of the contexts which influence the tourist behaviour clearly. Through this method, we can delete the low interest degree of tourist behavior patterns; then, we use the Apriori algorithm to mine the association rules of tourist behavior which have high interest degree. Finally, we take an experiment to show the feasibility and effectiveness of our method.

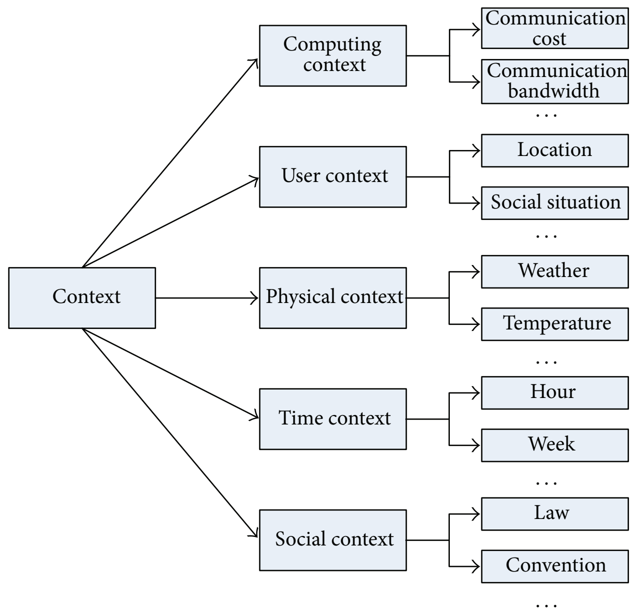


FIGURE 1: Context spectrum.

2. Related Works

2.1. Context. There are many definitions on the context and many researchers work on it. Schilit et al. defined the context as identifications and change of location, people, and objects around them [6]. Brown et al. thought that the context should be defined as the symbols around people or other objects such as location, time, season, temperature, and so on [7]. In paper [8], the definition of context would be extended to the feature information of some objects' situation, such as people, location, and so on. Snowdon and Grasso defined the context as the multilevel structure, mainly including the individual layer, the project layer, the group layer, and the organization layer [9]; Gu thought that the context would respond to the transformation based on the computers which are used as the centers to the people; in fact, he defined the context as a spectrum in his paper, as shown in Figure 1. He divided the context into computing context (such as communication bandwidth), user context (such as location), physical context (such as weather, temperature), time context (such as hour), and social context (such as law) [10].

In this paper, we think that the context is the influence factors of the tourist behavior pattern; different contexts will lead the tourist to different behavior patterns. We may take the following contexts into consideration: user, location, time, and device, and service type.

2.2. Association Rule and Apriori Algorithm. There are many association rule algorithms, and these algorithms can be divided into two classes: the first one is mainly focused on improving the analytical efficiency of the association rules; the other one pays more attention to the application of association rule algorithm and how to deal with value type variables and promotes the association of the single concept layer to multiple concept layers include and further reveals the inner structure of objects.

Apriori algorithm is one of the classical association rule algorithms; the earliest Apriori algorithm was proposed by Agrawal et al. [11]. The algorithm mainly including two parts: producing frequent item sets and producing association rules according to the frequent item sets. The algorithm scans data base, accumulates each item count, collects the items which meet the minimum support (min_sup), finds out the frequent 1-itemsets, and named it L_1. Then, the algorithm uses L_1 to find out the frequent 2-item sets L_2 and uses L_2 to find out the frequent 2-item sets L_3 and so on and keeps doing these until it cannot find out the frequent k-item sets. In these frequent item sets, it will be defined as a strong-association rule if it reaches the minimum confidence [12]. Since the association rule algorithm was proposed, it has been improved and applied in many fields. For example, Kang et al. applied the association rule algorithm in the Smart home [13], and Zhang et al. used the improved association rule algorithm in the university teaching managements [14].

3. Modeling and Mining Method for Tourist Behavior Pattern Based on Context

3.1. The Context Influence Factors Analysis of Tourist Behavior Pattern. We can consider a tourist as a mobile customer because the tourist moved anytime and anywhere. Presently, there are only a few researchers who work on the mobile customer behavior pattern. Tseng and Lin thought that the service and location are the influence factors of customer behavior in mobile service environment; they proposed a method named SMAP-Mine to mine customer behaviors [15]. Ma et al. took the time context into consideration and constructed a temporal sequence mobile access patterns mining model based on context awareness [16]. Chen et al. studied in the terms of the problem of mining matching mobile access patterns based on joining the following four kinds of characteristics: user, location, time, and service [17]. So in this paper, we think that the context influence factors of mobile customer behavior pattern includes mobile user, location, time, and service type.

At the same time, we take different capabilities of the mobile devices that the customer use, such as screen size, battery durability, and access bandwidth, into consideration. We consider that these capabilities will influence the mobile customer behavior pattern directly or indirectly. To prove that, we make an experiment as as follows. In the particular context, we observed behavior patterns of three customers who used different equipments and recorded the service types, the trajectory at which they moved, and time and type of service. Finally we got the customer movement trajectories as shown in Figure 2 and the service request information table as shown in Table 1. We can conclude from Figure 2 that customers have different behavior patterns when they use different mobile devices. For example, when the user u_1 used the device d_1, his movement trajectory was $l_2 \rightarrow l_6 \rightarrow l_8 \rightarrow l_9$; when he used the device d_2, his movement trajectory changed to $l_2 \rightarrow l_6 \rightarrow l_9$. Then, we can conclude form

Table 1: Customer service information table.

Time instances	Users and devices					
	(u_1, d_1)	(u_2,d_1)	(u_3, d_1)	(u_1, d_2)	(u_2, d_2)	(u_3, d_2)
TI_1	(l_2, t_1, s_1)	(l_9, t_5, s_1)	(l_5, t_{13}, s_1)	(l_2, t_1, s_1)	(l_9, t_1, s_1)	(l_5, t_1, s_1)
TI_2	(l_6,t_2, s_2)	(l_4, t_6, s_2)	(l_6, t_{14}, s_3)	(l_6, t_2, s_3)	(l_4, t_4, s_2)	(l_6, t_4, s_3)
TI_3	(l_8, t_3, s_3)	(l_1, t_8, s_2)	(l_4, t_{15}, s_2)	(l_9, t_3, s_4)	(l_7, t_5, s_4)	(l_3, t_5, s_2)
TI_4	(l_9, t_4, s_4)	(l_9, t_9, s_3)	(l_3, t_{16}, s_4)		(l_1, t_6, s_5)	(l_9, t_7, s_6)
TI_5		(l_7, t_{10}, s_4)	(l_9, t_{17}, s_6)		(l_3, t_8, s_6)	
TI_6		(l_1, t_{11}, s_5)				
TI_7		(l_3, t_{12}, s_6)				

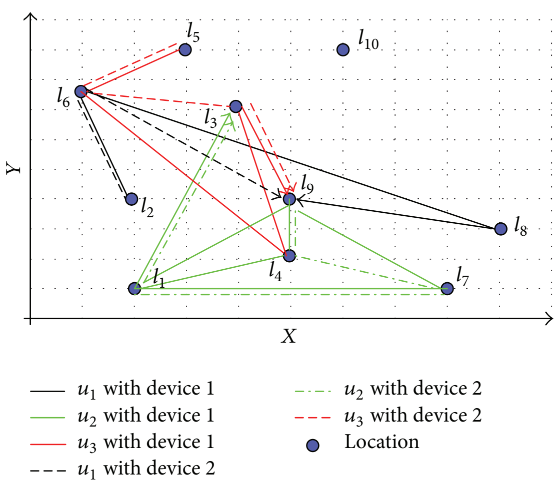


Figure 2: Movement trajectories of customers when they use different devices.

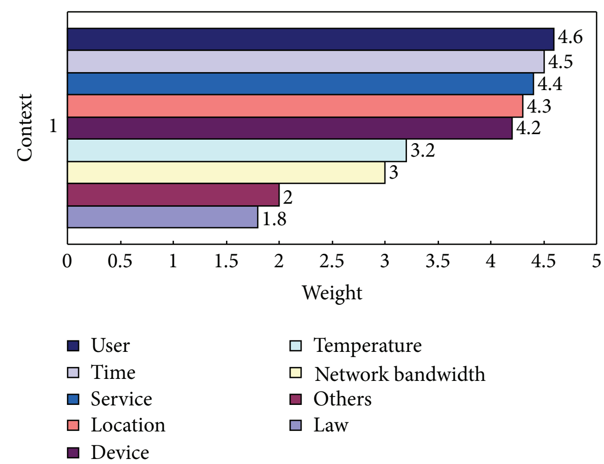


Figure 3: The results of the questionnaire.

Table 1 that the customer requested different services when he used different devices in the same time or requested the same service in different times. For example, when the user u_1 used device d_1 at time t_3; he requested the service s_3; when he used device d_2 at time t_3, he requested the service s_4; the user u_3 requested the service s_2 at the location l_4 when he used device d_1; he requested the service s_2 at the location l_3 when he used device d_2. Through these analyses, we can conclude that the mobile customer has different movement trajectories, request different services at the same times and requests the same service in different places when he or she uses different devices. So we take the mobile device as a context influence factor of mobile customer behavior pattern.

There are other context factors which influence the mobile customer behavior pattern, such as the physically environmental condition in which the customer stays, including weather, temperature, humidity, and so on; and the social situations in which the customer is involved (e.g., manners and customs and laws) will influence the mobile customer behavior pattern.

We use the form of the questionnaire to determine the main context factors. In this questionnaire, we design nine questions. Each of the nine questions involves a context factor which will influence the tourist behavior pattern. From these questions, we can study which contexts will influence the tourist behavior pattern most. A total of 102 individuals participate in the survey; they are all tourists. After stating these questionnaires, we use SPSS to analyze the results. We set that different option to different weight (1–5), and then statistically averaging, what are the context weights influence the behavior. We can get the results as shown in Figure 3. So in this paper, we choose the following five context factors as the main context factors: tourist (user), device, location, time, and service.

3.2. Modeling the Tourist Behavior Pattern Based on Context. The preceding part of this paper has a brief analysis on the context factors which influence the tourist behavior pattern, and then we will build a model based on these context factors. In the following part of this paper, we will give relational definitions about the tourist behavior patterns firstly and construct a model of the tourist behavior pattern based on context latterly.

Definition 1 (tourist user). $U = \{u_1, u_2, u_3, \ldots, u_i\}$ is the set of all the users; every user denotes a person who uses the mobile device to request mobile service messages from the mobile service supplier when he or she was travelling.

Definition 2 (devices of the tourist use). The device of the user use is a set of the devices of the user use to request mobile services and can be defined as $D = \{d_1, d_2, d_3, \ldots, d_h\}$.

Table 2: Timestamp table.

Timestamps	Time intervals
t_1	0:00–1:00
t_2	1:00–2:00
t_3	2:00–3:00
$\vdots$	$\vdots$
t_{22}	21:00–22:00
t_{23}	22:00–23:00
t_{24}	23:00–24:00

Definition 3 (location). Location denotes a set of places in which the tourist moves some times, and we use the set $L = \{l_1, l_2, l_3, \ldots, l_j\}$ to represent it.

Definition 4 (service). Service is a set of messages in which the tourist requests tourism services from the suppliers, and we use the set $S = \{s_1, s_2, s_3, \ldots, s_n\}$ to represent it.

Definition 5 (timestamp, sojourn time and service request time). To represent the time quantum of the forming of the tourist behavior pattern approximately, this paper divides a day's 24 hours into 24 time intervals simply, as shown in Table 2; every time interval denotes one hour, and the hour denotes one timestamp; sojourn time t_s denotes the time in which the user sojourns at somewhere; service request time t_r denotes the time in which the tourist requests some tourism services.

According to the previous definitions, this paper assumes $p = \{u, d, t, l, t_s, s, t_r\}$ as one tourist behavior, where u is an element of the tourist user set U, d is an element of the device of the user use set D, t is an element of the time set T, l is an element of the location set L, t_s is the time in which the tourist sojourns at location l, s denotes an element of service messages set S, and t_r denotes the time in which the tourist requests for tourism services.

In the graph theory, there is a structure called network whose structure is composed of nodes and edges. Every edge has its quantitative index related to the nodes or edges; this quantitative index is normally called weight which could denote distance, expenses, carrying capacity, and so on [18]. Namely, the structure of the network is composed of nodes and edges involving weight; taking this advantage of the network, this paper makes the context factors which influence the tourist behavior pattern as the nodes of the network, the connected relationship among the context factors as the edge of the network, and the connect coefficient among different context factors as the weight of the edge (the specific connect relationship and the connect coefficient will be demonstrated in detail in the following part of this paper). Like this, the behavior pattern of a tourist can be clearly portrayed in the network. Figure 4 illustrates the network structure of the behavior patterns of two different mobile users; we use $p_1 = \{u_1, d_2, t_1, l_2, 5, s_2, 2\}$ and $p_2 = \{u_2, d_1, t_2, l_1, 4, s_1, 3\}$ to represent their behavior patterns, respectively.

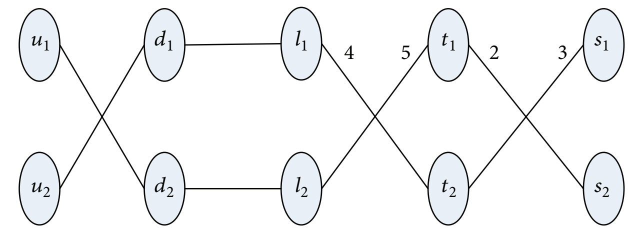


Figure 4: The network model of the tourist behavior pattern.

3.3. Tourist Behavior Pattern Mining Method Based on the Network. The preceding part of this paper has a model analysis on the structure of the network of the tourist behavior; in the following part of this paper, we will give out the related definitions firstly and the specific procedures of the tourist behavior mining pattern based on the network latterly.

3.3.1. Basic Definitions. To explain the content of the mining method more clearly, we will give relational definitions firstly.

Definition 6 (connect coefficient). Connect coefficient denotes the connection relationship between two different attributes; the specific connect coefficients are $U \bowtie D$, $D \bowtie L$, $L \bowtie T$, and $T \bowtie S$. The connect coefficient of $U \bowtie D$ denotes the connection times between a mobile user u and a device d. The connect coefficient of $D \bowtie L$ is LC_{nu} which denotes the connection times of a device d with a location l. The connect coefficient of $L \bowtie T$ is t_s which denotes the time in which a user sojourns at location l. The connect coefficient of $T \bowtie S$ is t_r which denotes the time in which a mobile user requests for services.

Definition 7 (interesting locations and interesting services). When the length of time in which a tourist sojourns somewhere is larger than the threshold value we set, we think that the tourist is interested in this place. Similarly, when the length of time in which a mobile user requests for a service s_j is larger than the threshold value or the connection times is larger than a threshold value, we think that the mobile user is interested in this service. Usually the length of time will be set up to 30 minutes and the connection times will be set up to 10 times.

Definition 8 (repeated edge). For a tourist, he may have the same connection edge in two different behavior patterns; such edge will be called repeated edges in this paper. For example, in the following behavior patterns $p_1 = \{u_1, d_1, l_3, t_{17}, s_3\}$ and $p_2 = \{u_1, d_2, l_3, t_{17}, s_3\}$, they have two repeated edges, namely,$l_3 t_{17}$ and $t_{17} s_3$.

Definition 9 (connect edge value). Connect edge value is a standard value obtained with standardizing the connect coefficient (Definition 6) in the case where the different quantity levels of input variables affect the final mining result. In this paper we use "$\bowtie$" to present the connect relationship between different attributes, and specific weights are $U \bowtie D$,

$D \bowtie L$, $L \bowtie T$, and $T \bowtie S$; the computational formulas of every edge weight are as follows.

Connect Edge Value of $U \bowtie D$. The connect edge value of mobile user u_i and device d_j equals the ratio of the connect times between user u_i and device d_j to the sum times between user u_i and device set; the specific formula is

$$w_{ij} = \frac{\text{count}\left(u_i d_j\right)}{\sum_{j=1}^{n} \text{count}\left(u_i d_j\right)}, \quad (1)$$

where n denotes the amount of devices. Similarly, the connect edge value of $D \bowtie L$ is as follows:

$$w_{jk} = \frac{LC_{nu}\left(jk\right)}{\sum_{k=1}^{m} LC_{nu}\left(jk\right)}, \quad (2)$$

where m denotes the amount of locations and $LC_{nu}(jk)$ and $\sum_{k=1}^{m} LC_{nu}(jk)$ denote the connect times between devices and locations in the same behavior pattern of a mobile user.

The Connect Edge Value of $L \bowtie T$

$$w_{kh} = \frac{t_s\left(kh\right)}{\sum_{h=1}^{24} t_s\left(kh\right)}, \quad (3)$$

where $t_s(kh)$ and $\sum_{h=1}^{24} t_s(kh)$ denote the time in which a mobile user requests services at somewhere in his behavior pattern.

The Connect Edge Value of $T \bowtie S$

$$w_{hz} = \frac{t_r\left(hz\right)}{\sum_{z=1}^{n} t_r\left(hz\right)}, \quad (4)$$

where n denotes the amount of the connect service set and $t_r(hz)$ and $\sum_{z=1}^{n} t_r(hz)$ denote the time in which a mobile user requests for services in his behavior pattern. An edge will be deleted if its connection edge value is smaller than a threshold value. A behavior pattern will not be involved in the calculation of the connect edge value if it contains interesting locations or interesting services.

Definition 10 (connect edge coefficient e). When a repeated edge appears, this edge value constitutes of several behavior patterns; connection edge coefficient e denotes the incidence a behavior pattern has on this edge. Its value equals ratio of the connect edge coefficient of this behavior pattern to the sum of all the connect edge coefficients of the same mobile user at this edge.

Definition 11 (interesting degree id). Interesting degree id is an index to reflect the degrees of interests of the mobile user behavior pattern. Specifically, it equals the value that the sum of all the tuple (u, d, t, l, s) weight, the formula of interesting degree id $= e_{ij} * w_{ij} + e_{jk} * w_{jk} + e_{kh} * w_{kh} + e_{hz} * w_{hz}$. If the value of interesting degree id is smaller than a threshold value th_1, we will regard the degree of interests of this mobile user pattern as low interest level and delete this pattern from

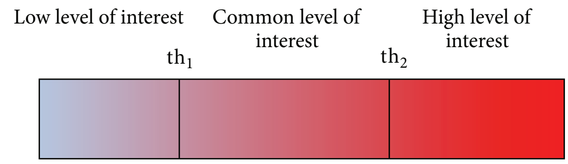


Figure 5: Two thresholds and three degrees.

Table 3: Tourist information.

U_i	d_h	L_j	LC_{nu}	T_m	T_s	S_n	T_r
u_1	d_1	l_1	1	t_{15}	8	s_4	6
u_1	d_2	l_1	1	t_{15}	8	s_2	2
u_2	d_1	l_4	2	t_{16}	9	s_3	9
u_2	d_2	l_3	2	t_{17}	9	s_3	9
u_3	d_1	l_5	6	t_{17}	5	s_5	5
u_3	d_2	l_5	6	t_{18}	12	s_5	8
u_4	d_3	l_7	5	t_{14}	8	s_6	7
u_5	d_2	l_6	3	t_{16}	10	s_8	9
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

the network. If the value of interesting degree id is larger than another threshold value th_2, we will regard the degree of interests of this mobile user pattern as high interest level. Like this, we divide mobile user behavior patterns into three parts, namely, low level of interest, common level of interest, and high level of interest. We can set them by our need; the larger the value is, the higher degree of interest the rules of the results will have. As is illustrated in Figure 5, we can use it as a behavior prediction model to predict the behavior pattern of a mobile user in the future. If a behavior pattern contains interesting locations or interesting services, we will regard it as the high interesting level behavior pattern without calculating the specific value of its interestingness.

3.3.2. Mining Steps

First Step (collecting data). To mine tourist behavior pattern, we must collect data about the tourist. We can get the information table as is shown in Table 3 through collecting user data, mainly including tourist information (U_i), mobile device (d_h), location (L_j), collecting times (LC_{nu}), time (T_m), time of the user stay the location (T_s), the service type the user request (S_n), and time of the user request the service (T_r).

Second Step. Extracting context attribute number of the context set which influences the mobile customer behavior pattern and design corresponding layers of the network diagram; in this paper, we should design a network diagram with five layers, each layer corresponds to all nodes of U_i, D_h, L_j, T_m, and S_n, respectively, and the number of the layer nodes corresponds to each attribute value number, as shown in Figure 6.

Third Step. Collecting the adjacent nodes, each connection coefficient should be marked as Definition 7; we need to add the connection coefficient of the side when it repeats several

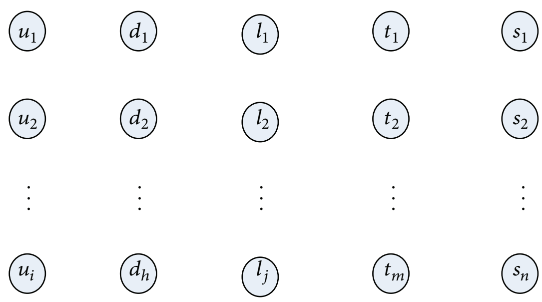


FIGURE 6: Network diagram model.

times. For example, there are two situations when the device d_1 collects the location l_3: one is 4 and the other is 2, so the connection coefficient of d_1l_3 equals 4 + 2 = 6.

Forth Step. Considering different customers have different behavior patterns, we classify each user into a group and calculate the collection weight according to Definition 8; when the collection weight is lesser than the threshold, the edge will be deleted.

Fifth Step. Calculating the remaining customer interest degree according to Definition 11 and set the low interest degree th_1 and the high interest degree th_2 value. When the customer interest degree is lesser than the low interest degree th_1, this customer behavior pattern will be unconcerned, and the general interest degree and the high interest degree pattern will be conducted in the next step.

Sixth Step. Using Apriori algorithm to mine the frequent pattern to the general interest degree and the high interest degree pattern, mine out the association rules with higher degree value on support and confidence; we can use these rules to forecast the customer' behaviors in future or recommend some services to mobile customers.

In order to show the availability of our method, we propose the concept of "coverage," which means the ratio of the number of the same rules that are produced by our model to the number of rules that produced directly. If the coverage is larger than a threshold, we say that the method we proposed is available. Generally, the larger the threshold is, the more the availability of the method is. In this paper, we set the threshold to be equal to 80%.

4. Experiment and Analysis

4.1. Example and Analysis. We take the West Lake of Hangzhou, for example, to illustrate the application of the model, via GPS and RFID provide personalized services to users combined with requirements and preference of the user. So we select part of the information data about tourist behavior from West Lake of Hangzhou Scenic Area Management Committee as is shown in Table 4.

To verify the effects of the proposed method, we use two standard metrics: interest degree and coverage.

TABLE 4: Tourist behavior information table.

P	U_i	d_h	L_j	LC_{nu}	T_m	T_s	S_n	T_r
p_1	u_1	d_1	l_1	1	t_{15}	8	s_4	6
p_2	u_1	d_2	l_1	1	t_{15}	8	s_2	2
p_3	u_1	d_1	l_1	1	t_{16}	11	s_2	3
p_4	u_1	d_2	l_1	1	t_{16}	11	s_2	4
p_5	u_1	d_1	l_3	4	t_{17}	21	s_3	17
p_6	u_1	d_2	l_3	4	t_{17}	4	s_3	3
p_7	u_1	d_1	l_2	11	t_{16}	35	s_4	30
p_8	u_2	d_1	l_4	2	t_{16}	9	s_3	9
p_9	u_2	d_2	l_3	2	t_{17}	9	s_3	9
p_{10}	u_2	d_1	l_3	2	t_{17}	2	s_3	1
p_{11}	u_2	d_2	l_2	2	t_{17}	28	s_5	15
p_{12}	u_2	d_2	l_2	8	t_{16}	38	s_4	31
p_{13}	u_3	d_1	l_5	6	t_{17}	5	s_5	5
p_{14}	u_3	d_2	l_5	6	t_{18}	12	s_5	8
p_{15}	u_3	d_1	l_5	6	t_{18}	12	s_4	4
p_{16}	u_3	d_1	l_5	6	t_{18}	18	s_3	13
p_{17}	u_3	d_2	l_5	6	t_{18}	18	s_3	5
p_{18}	u_3	d_2	l_5	10	t_{16}	32	s_4	30
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

We can conclude from Table 4 that the patterns p_7, p_{12}, and p_{18} are the patterns with interested locations or interested services, and we think that these patterns are the behavior patterns with high interest degree patterns. To show the processes of our method, we choose three users' patterns in this paper and design a network diagram with five layers and collect the adjacent layers and then calculate the connection coefficients according to Definition 7, as shown in Figure 7.

Then, we divide each user into a group and calculate the collection weight according to Definition 8; we set that the edges whose weight is lesser than 0.2 would be deleted, so the edges d_1l_1, d_2l_1, $t_{15}s_2$, $t_{15}s_2$, $t_{15}s_4$, l_4t_{15}, l_5t_{17}, $t_{17}s_5$, and $t_{18}s_4$ will be deleted and the patterns with edges d_1l_1, d_2l_1, $t_{15}s_2$, $t_{15}s_2$, $t_{15}s_4$, l_4t_{15}, l_5t_{17}, $t_{17}s_5$, and $t_{18}s_4$ will be deleted too, as shown in Figures 8, 9, 10.

So remain following patterns: $p_5 = \{u_1, d_1, l_3, t_{17}, s_3\}$, $p_6 = \{u_1, d_2, l_3, t_{17}, s_3\}$, $p_{10} = \{u_2, d_1, l_3, t_{17}, s_3\}$, $p_{11} = \{u_2, d_2, l_2, t_{17}, s_5\}$, $p_{14} = \{u_3, d_2, l_5, t_{18}, s_5\}$, $p_{16} = \{u_3, d_1, l_5, t_{18}, s_3\}$, $p_{17} = \{u_3, d_1, l_5, t_{18}, s_3\}$; then we calculate the interesting degree of these patterns according to Definition 11, as is shown in the following expressions:

$$\begin{aligned}
\text{id}_{p_5} &= e_{11} * w_{u_1d_1} + e_{13} * w_{d_1l_3} \\
&\quad + e_{3-17} * w_{l_1t_{17}} + e_{17-3} * w_{t_{17}s_3} \\
&= \frac{1}{3} * 0.5 + 1 * 0.333 + \frac{21}{25} * 0.397 \\
&\quad + \frac{17}{20} * 0.571 \approx 1.305; \\
\text{id}_{p_6} &= e_{12} * w_{u_1d_2} + e_{23} * w_{d_2l_3} \\
&\quad + e_{3-17} * w_{l_3t_{17}} + e_{17-3} * w_{t_{17}s_3}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{3} * 0.5 + 1 * 0.333 + \frac{4}{25} * 0.397 \\
&\quad + \frac{3}{20} * 0.571 \approx 0.647; \\
\mathrm{id}_{p_9} &= e_{22} * w_{u_2 d_2} + e_{23} * w_{d_2 l_3} \\
&\quad + e_{3-17} * w_{l_3 t_{17}} + e_{17-3} * w_{t_{17} s_3} \\
&= 0.5 * \frac{1}{2} + 1 * 0.25 + 0.229 * \frac{9}{11} \\
&\quad + 0.294 * \frac{9}{10} \approx 0.952; \\
\mathrm{id}_{p_{10}} &= e_{21} * w_{u_2 d_1} + e_{13} * w_{d_1 l_3} \\
&\quad + e_{3-17} * w_{l_3 t_{17}} + e_{17-3} * w_{t_{17} s_3} \\
&= 0.5 * \frac{1}{2} + 1 * 0.25 + 0.229 * \frac{2}{11} \\
&\quad + 0.294 * \frac{1}{10} \approx 0.571; \\
\mathrm{id}_{p_{11}} &= e_{22} * w_{u_2 d_2} + e_{22} * w_{d_2 l_2} \\
&\quad + e_{2-17} * w_{l_2 t_{17}} + e_{17-5} * w_{t_{17} s_5} \\
&= 0.5 * \frac{1}{2} + 1 * 0.25 + 1 * 0.583 \\
&\quad + 1 * 0.441 \approx 1.524; \\
\mathrm{id}_{p_{14}} &= e_{3-2} * w_{u_3 d_2} + e_{25} * w_{d_2 l_5} \\
&\quad + e_{5-18} * w_{l_5 t_{18}} + e_{18-5} * w_{t_{18} s_5} \\
&= 0.4 * \frac{1}{2} + 0.4 * \frac{1}{2} + 0.923 * \frac{12}{60} \\
&\quad + 1 * 0.229 \approx 0.815; \\
\mathrm{id}_{p_{16}} &= e_{31} * w_{u_3 d_1} + e_{15} * w_{d_1 l_5} \\
&\quad + e_{5-18} * w_{l_5 t_{18}} + e_{18-3} * w_{t_{18} s_3} \\
&= 0.6 * \frac{1}{3} + 0.6 * \frac{1}{3} + 0.923 * \frac{18}{60} \\
&\quad + 0.514 * \frac{13}{18} \approx 1.05; \\
\mathrm{id}_{p_{17}} &= e_{31} * w_{u_3 d_1} + e_{15} * w_{d_1 l_5} \\
&\quad + e_{5-18} * w_{l_5 t_{18}} + e_{18-3} * w_{t_{18} s_3} \\
&= 0.4 * \frac{1}{2} + 0.4 * \frac{1}{2} + 0.923 * \frac{18}{60} \\
&\quad + 0.514 * \frac{5}{18} \approx 0.820.
\end{aligned}
\tag{5}
$$

In this paper, we set the low interesting degree th_1 value to be equal to 0.8 and the high interesting degree th_2 value to be equal to 1. So the patterns whose interesting degrees are lesser than 0.8 interesting degree are the low interesting patterns, the patterns whose interesting degrees are higher than 1 interesting degree are the high interesting patterns, and the patterns whose interesting degree between 0.8 and 1 are the common patterns. So $p_6 = \{u_1, d_2, l_3, t_{17}, s_3\}$ and $p_{10} = \{u_2, d_1, l_3, t_{17}, s_3\}$ are the low interesting degree patterns, $p_9 = \{u_2, d_2, l_3, t_{17}, s_3\}$, $p_{14} = \{u_3, d_2, l_5, t_{18}, s_5\}$, and $p_{17} = \{u_3, d_1, l_5, t_{18}, s_3\}$ are the common interesting degree patterns, and $p_5 = \{u_1, d_1, l_3, t_{17}, s_3\}$, $p_{11} = \{u_2, d_2, l_2, t_{17}, s_5\}$, and $p_{16} = \{u_3, d_1, l_5, t_{18}, s_3\}$ are the high interesting degree patterns. We delete the low interesting degree patterns and get the patterns with high interesting degree as is shown in Table 5.

Table 5: Patterns with high interesting degree.

P_i	U_i	d_h	L_j	T_m	S_n
p_5	u_1	d_1	l_3	t_{17}	s_3
p_7	u_1	d_1	l_2	t_{16}	s_4
p_9	u_2	d_2	l_3	t_{17}	s_3
p_{11}	u_2	d_2	l_2	t_{17}	s_5
p_{12}	u_2	d_2	l_2	t_{16}	s_4
p_{14}	u_3	d_2	l_5	t_{18}	s_5
p_{16}	u_3	d_1	l_5	t_{18}	s_3
p_{17}	u_3	d_2	l_5	t_{18}	s_3
p_{18}	u_3	d_2	l_5	t_{16}	s_4

Then, we use the Apriori algorithm to mine rules on the high interesting degree patterns; we set the minimum support to 20% and the minimum confidence to 80%, then we can get the results as follows.

The lift denotes the ratio of the confidence to the support of the consequent item; the computational formula is followed: $L_{x \to y} = C_{x \to y}/S_y$. The lift reacts the influence degree of the antecedent item X to the consequent item Y appears. Generally, the lift value should be larger than 1, and it means that the antecedent item X has a positive influence on the consequent item Y appears. The larger the life value is, the better the rule is.

From Table 6, we can conclude that we can get 39 association rules when we use the method we proposed in this paper. These rules were obtained from the high interesting pattern; we thought that these rules were interesting rules. Then we observe the rule with the maximum lift, time $= t_{17}$, and service $= s_3 \to$ location $= l_3$. The value is 4.5. It means that this association rule has the highest realistic guidance. So this rule will be firstly considered when we use the rules of the result. We can use these association rules to recommend some services to tourist to offer them better services; for example, using the rule location $= l_2$ and time $= t_{16} \to$ service $= s_4$, we can recommend the s_4 to the tourist when the tourist stays in the context with location $= l_2$ and time $= t_{16}$. In this paper, the service s_4 is the tourism route guide, so we can send the tourism route guide to the tourist as is shown in Figure 11.

4.2. Comparison and Discussion. To verity the effects of the method we proposed in this paper, we use the Apriori

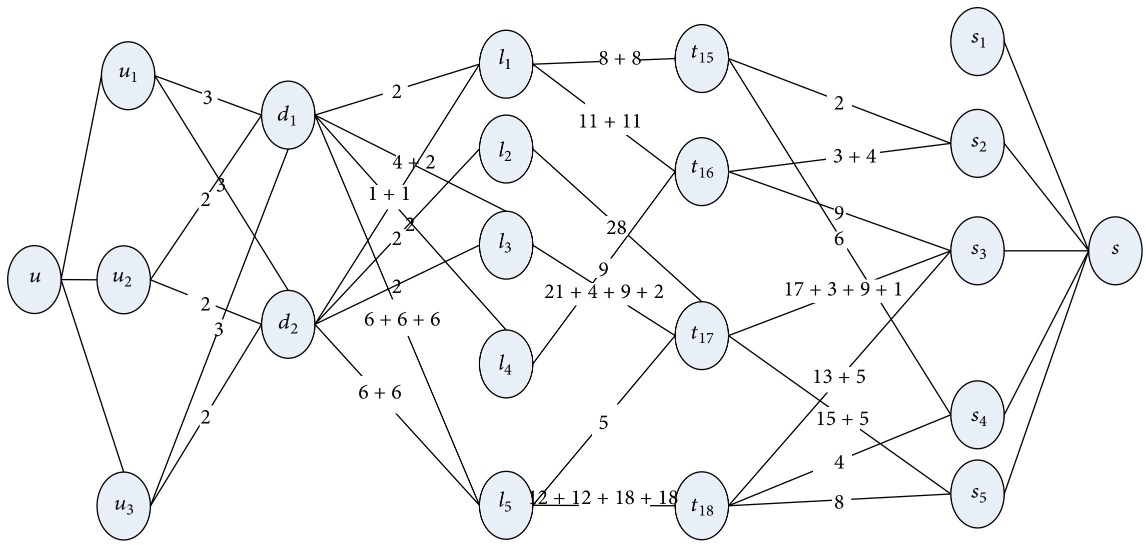


FIGURE 7: The network we constructed.

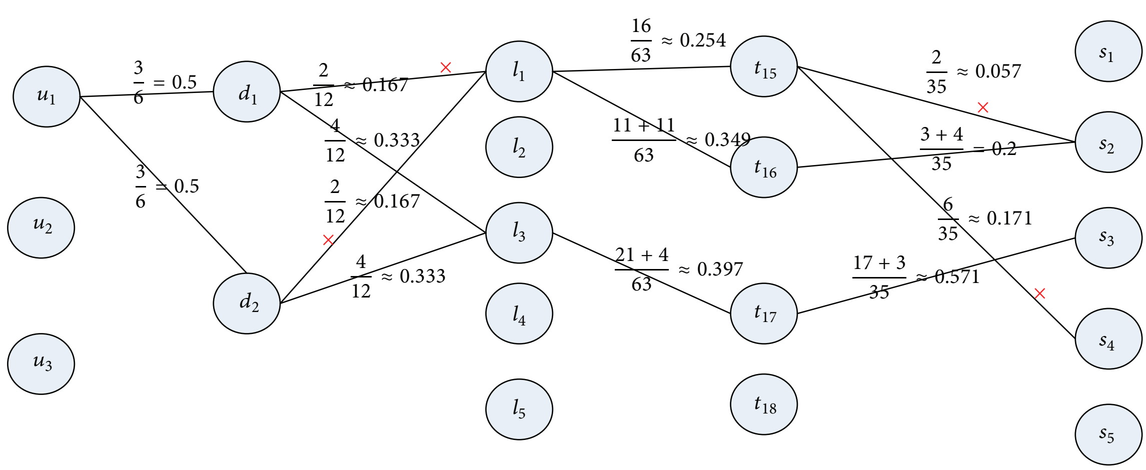


FIGURE 8: Calculate the tourist u_1 collection weight.

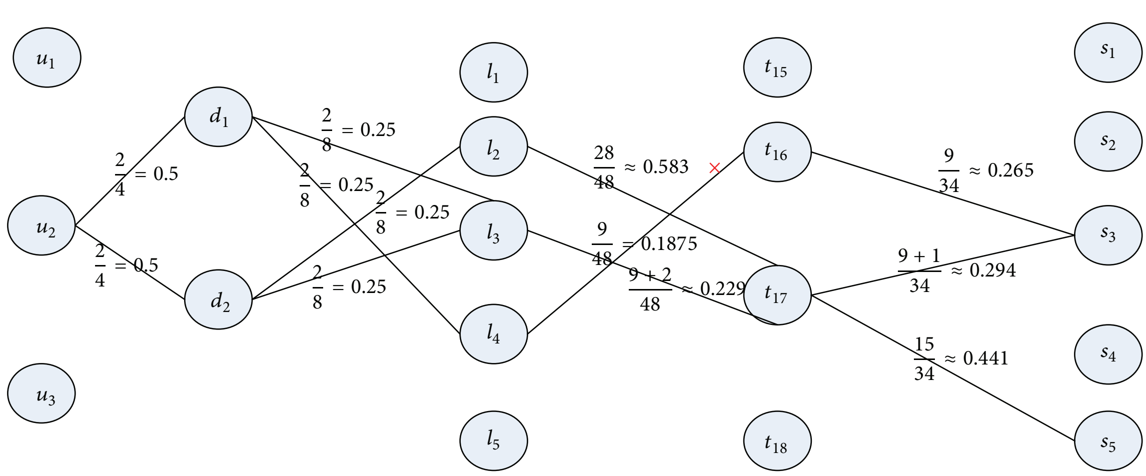


FIGURE 9: Calculate the tourist u_2 collection weight.

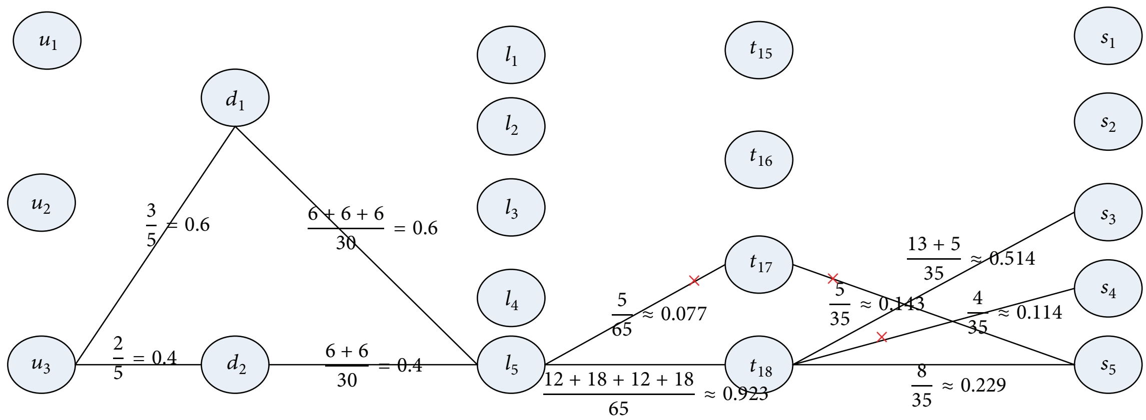


FIGURE 10: Calculate the tourist u_3 collection weight.

FIGURE 11: Tourism route guide.

algorithm, the GRI algorithm, the CARAMA algorithm and Predictive-Apriori algorithm on the original data (here we set the minimum support equals to 20% and the minimum confidence equals to 80%; too), and we get following rules as is shown in Tables 7, 8, 9, and 10.

Comparing Table 6 with Tables 7 and 8, there are 11 rules from Table 7 which have been emerged in Table 6 (the rules marked with yellow as is shown in Table 6), and all rules are in Table 8 have been emerged in Table 6. So we think that the method we proposed to mine the mobile customer behavior pattern has the merit of effectiveness; in this experiment the validity of the method is about 91.67% (11/12) to the Apriori algorithm and 100% (6/6) to the GRI algorithm, which means the coverage values are 91.67% and 100%, which are larger than the threshold we set before. It means that the method we proposed is feasibile and effective. Excluding the 11 rules in Table 7, Table 6 has other 28 rules and these rules have the feature of high interest, so they will provide more choices to the service provider and more services to the mobile customer. Then we observe the rule which has the maximum value of lift from Tables 6 and 7, the rule is time = t_{17} and service = s_3 $\rightarrow$ location = l_3; it means that the method we proposed is similar to the classical Apriori algorithm. At last, the rule whose ID = 1 in Table 7: location = l_1 $\rightarrow$ user = u_1,It is the only rule that is not included in Table 6, although this rule meets the minimum support and the minimum confidence; the pattern with $\{l_1, u_1\}$ is a low interesting pattern as we definite before, and the rule location = l_1 $\rightarrow$ user = u_1 is an uninteresting rule. In our method, we can reject uninteresting rules like this. Through the analysis, the method we proposed in this paper is more feasible and advanced when being compared with the Apriori algorithm.

Table 6: Mining results.

ID	Rules	Count	Lift
1	User = u_1 → device = d_1	2	3
2	Location = l_3 → time = t_{17}	2	3
3	Location = l_3 → service = s_3	2	2.25
4	Service = s_5 → device = d_2	2	1.5
5	User = u_2 → device = d_2	3	1.5
6	Time = t_{16} → service = s_4	3	3
7	Service = s_4 → time = t_{16}	3	3
8	Time = t_{18} → user = u_3	3	2.25
9	Time = t_{18} → location = l_5	3	2.25
10	User = u_3 → location = l_5	4	2.25
11	Location = l_5 → user = u_3	4	2.25
12	Location = l_3 and time = t_{17} → service = s_3	2	2.25
13	Location = l_3 and service = s_3 → time = t_{17}	2	3
14	Time = t_{17} and service = s_3 → location = l_3	2	4.5
15	User = u_2 and location = l_2 → device = d_2	2	1.5
16	Location = l_2 and device = d_2 → user = u_2	2	3
17	User = u_2 and time = t_{17} → device = d_2	2	1.5
18	Time = t_{17} and device = d_2 → user = u_2	2	3
19	Location = l_2 and time = t_{16} → service = s_4	2	3
20	Location = l_2 and service = s_4 → time = t_{16}	2	3
21	Time = t_{16} and device = d_2 → service = s_4	2	3
22	Service = s_4 and device = d_2 → time = t_{16}	2	3
23	Time = t_{18} and user = u_3 → location = l_5	3	2.25
24	Time = t_{18} and location = l_5 → user = u_3	3	2.25
25	Time = t_{18} and service = s_3 → user = u_3	2	2.25
26	User = u_3 and service = s_3 → time = t_{18}	2	3
27	Time = t_{18} and device = d_2 → user = u_3	2	2.25
28	Time = t_{18} and service = s_3 → location = l_5	2	2.25
29	Location = l_5 and service = s_3 → time = t_{18}	2	3
30	Time = t_{18} and device = d_2 → location = l_5	2	2.25
31	User = u_3 and service = s_3 → location = l_5	2	2.25
32	Location = l_5 and service = s_3 → user = u_3	2	2.25
33	User = u_3 and device = d_2 → location = l_5	3	2.25
34	Location = l_5 and device = d_2 → user = u_3	3	2.25
35	Time = t_{18}, user = u_3 and service = s_3 → location = l_5	2	2.25
36	Time = t_{18}, location = l_5 and service = s_3 → user = u_3	2	2.25
37	User = u_3, location = l_5 and service = s_3 → time = t_{18}	2	3
38	Time = t_{18}, user = u_3 and device = d_2 → location = l_5	2	2.25
39	Time = t_{18}, location = l_5 and device = d_2 → user = u_3	2	2.25

5. Conclusion

In this paper we considered the context factors which influence the tourist behavior pattern comprehensively, such as the device the tourist use, time, location, and service types, and got the context set which influences the tourist behavior pattern. Then we proposed a method to mine tourist behavior patterns based on the network diagram; this method constructed a network diagram firstly. Then, we got the behavior patterns with high interesting degree and did association rule mining in the patterns and got the rules; at last, we made an experiment to show the feasibility and effectiveness of our method. In our experiment, we set the low interest degree th_1 value to be equal to 0.8 and the high interest degree th_2 value to be equal to 1 and deleted the low interest pattern; then we did association mining with Apriori algorithm to the remainder of the patterns and got 39 rules; we can do some recommendations to the tourist with these high interest rules. Compared to the results which do not use this method, it has the following advantages: (1) it can keep the interest rules and delete the uninterested rules in the results; (2) it can produce many other interest rules, which

Table 7: The results based on Apriori algorithm.

ID	Rules	Count	Lift
1	Location = l_1 → user = u_1	4	2.57
2	Time = t_{18} → user = u_3	4	3.00
3	Time = t_{18} → location = l_5	4	3.00
4	Location = l_3 → time = t_{17}	4	3.00
5	Location = l_3 → service = s_3	4	2.57
6	User = u_3 → location = l_5	6	3.00
7	Location = l_5 → user = u_3	6	3.00
8	Time = t_{18} and user = u_3 → location = l_5	4	3.00
9	Time = t_{18} and location = l_5 → user = u_3	4	3.00
10	Location = l_3 and time = t_{17} → service = s_3	4	2.57
11	Location = l_3 and service = s_3 → time = t_{17}	4	3.00
12	Time = t_{17} and service = s_3 → location = l_3	4	4.50

Table 8: The results based on GRI algorithm.

ID	Rules	Count	Lift
1	Location = l_1 → user = u_1	4	2.5
2	Location = l_3 → service = s_3	4	2.5
3	User = u_3 → location = u_3	6	3
4	Location = l_5 → user = u_3	6	3
5	Location = l_3 → time = t_{17}	4	3
6	Time = t_{17} and service = s_3 → location = l_3	4	4.5

Table 9: The results based on CARMA algorithm.

ID	Rules	Count	Lift
1	User = u_3 → location = l_5	6	3
2	Location = l_5 → user = u_3	6	3
3	Location = l_1 → user = u_1	4	2.5
4	Location = l_3 → time = t_{17}	4	3
5	Location = l_3 → service = s_3	4	2.5
6	Location = l_3 → time = t_{17} and service = s_3	4	4.5
7	Time = t_{18} → user = u_3	4	3
8	Time = t_{18} → location = l_5	4	3
9	Location = l_3 and time = t_{17} → service = s_3	4	2.5
10	Location = l_3 and service = s_3 → time = t_{17}	4	3
11	Time = t_{17} and service = s_3 → location = l_3	4	4.5

we can use them to make more recommendations for the tourist; (3) it can produce the same rule which has the highest lift compared to the result that does not use this method. That is, the method we used in this paper is feasible and superior.

The future work will be further researching on the context factors which influence the tourist behavior pattern and expanding the context set; also we will analyze the performance of the method we proposed and optimize the method and so on.

Table 10: The results based on Predictive-Apriori algorithm.

ID	Rules
1	User = u_3 → location = l_5
2	Location = l_5 → user = u_3
3	Location = l_1 → user = u_1
4	Location = l_3 → time = t_{17} and service = s_3
5	Time = t_{18} → user = u_3 and location = l_5
6	Time = t_{17} and service = s_3 → location = l_3
7	Service = s_2 → user = u_1 and location = l_1
8	Time = t_{15} → user = u_1
9	User = u_1 and service = s_3 → location = l_3
10	User = u_2 and device = d_1 → service = s_3
11	User = u_3 and service = s_3 → location = l_5 and time = t_{18}
12	Location = l_2 and time = t_{16} → service = s_4
13	Location = l_2 and service = s_4 → time = t_{16}
14	Location = l_5 and service = s_3 → user = u_3 time = t_{18}
15	User = u_1 and device = d_2 and location = l_1 → service = s_2

Appendix

Questionnaire

Your age: —

Gender: male/female

(1) To what extent do you think the user will influence the behavior?

(A) Strongly disagree
(B) Disagree
(C) Neutral
(D) Agree
(E) Strongly agree

(2) To what extent do you think the location will influence the behavior?

(A) Strongly disagree
(B) Disagree
(C) Neutral
(D) Agree
(E) Strongly agree

(3) To what extent do you think the time will influence the behavior?

(A) Strongly disagree
(B) Disagree
(C) Neutral
(D) Agree
(E) Strongly agree

(4) To what extent do you think the device will influence the behavior?

(A) Strongly disagree
(B) Disagree
(C) Neutral
(D) Agree
(E) Strongly agree

(5) To what extent do you think the service will influence the behavior?

(A) Strongly disagree
(B) Disagree
(C) Neutral
(D) Agree
(E) Strongly agree

(6) To what extent do you think the network bandwith will influence the behavior?

(A) Strongly disagree
(B) Disagree
(C) Neutral
(D) Agree
(E) Strongly agree

(7) To what extent do you think the temperature will influence the behavior?

(A) Strongly disagree
(B) Disagree
(C) Neutral
(D) Agree
(E) Strongly agree

(8) To what extent do you think the law will influence the behavior?

(A) Strongly disagree
(B) Disagree
(C) Neutral
(D) Agree
(E) Strongly agree

(9) Others factors will influence the behavior, such as —.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (Grant nos. 71071140 and 71301070005), the National Natural Science Foundation of Zhejiang Province (Grant no. Y1090617), the Key Innovation Team of Zhejiang Province (Grant no. 2010R50041), the Soft science key research project of Zhejiang Province (Grant no. 2013C25053), the Zhejiang Gongshang University Graduate Student Scientific Research Project (1130XJ1512168), and the Modern Business Centre of Zhejiang GongShang University.

References

[1] M. Qing, *Research on Constructing Personalized Tourism Service Supply Chain in the Context of Modern Information Technology*, Shandong University of Finance and Economics, 2012.

[2] A. Farmaki, "An exploration of tourist motivation in rural settings: the case of Troodos, Cyprus," *Tourism Management Perspectives*, vol. 2-3, pp. 72–78, 2012.

[3] C. A. Martin and S. F. Witt, "Tourism demand forecasting models: choice of appropriate variable to represent tourists' cost of living," *Tourism Management*, vol. 8, no. 3, pp. 233–246, 1987.

[4] C. Smallman and K. Moore, "Process studies of tourists' decision-making," *Annals of Tourism Research*, vol. 37, no. 2, pp. 397–422, 2010.

[5] S. S. Kim, D. J. Timothy, and J. Hwang, "Understanding Japanese tourists' shopping preferences using the Decision Tree Analysis method," *Tourism Management*, vol. 32, no. 3, pp. 544–554, 2011.

[6] B. Schilit, N. Adams, and R. Want, "Context-aware computing applications," in *Proceedings of the Workshop on Mobile Computing Systems and Applications*, pp. 85–90, December 1994.

[7] P. J. Brown, J. D. Bovey, and X. Chen, "Context-aware applications: from the laboratory to the marketplace," *IEEE Personal Communications*, vol. 4, no. 5, pp. 58–64, 1997.

[8] Y. A. Rakotonirain, S. W. Iokey, and G. Fitzpatrick, "Context-awareness for the mobile environment," 2008-10-02.

[9] D. Snowdon and A. Grasso, "Providing context awareness via a large screen display," in *Proceedings of the CHI 2000 Workshop on "The What, Who, Where, When, Why and How of Context-Awareness"*, 2000, report no. 2000/011.

[10] J. Z. Gu, "Context aware computing," *Journal of East China Normal University: Natural Science*, vol. 5, pp. 1–20, 2009.

[11] R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," in *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, pp. 207–216, May 1993.

[12] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*, China Machine Press, Beijing, China, 3rd edition, 2012.

[13] K. J. Kang, B. Ka, and S. J. Kim, "A service scenario generation scheme based on association rule mining for elderly surveillance system in a smart home environment," *Engineering Applications of Artificial Intelligence*, vol. 25, pp. 1355–1364, 2012.

[14] Z. Y. Zhang, Y. P. Zhang, J. Y. Zhang, and X. J. Zhang, "Application of improved association rule algorithm in college teaching management," *Computer Engineering*, vol. 38, no. 2, pp. 75–78, 2012.

[15] V. S. Tseng and K. W. Lin, "Efficient mining and prediction of user behavior patterns in mobile web systems," *Information and Software Technology*, vol. 48, no. 6, pp. 357–369, 2006.

[16] Z. B. Ma, T. J. Lu, and H. Q. Li, "Mining of temporal sequence mobile access patterns based on context awareness," *Transaction of Beijing Institute of Technology*, vol. 28, no. 10, pp. 937–940, 2008.

[17] T. S. Chen, Y. S. Chou, and T. C. Chen, "Mining user movement behavior patterns in a mobile service environment," *IEEE Transactions on Systems, Man, and Cybernetics A*, vol. 42, no. 1, pp. 87–101, 2012.

[18] Y. Q. Hu, *Operational Research*, Tsinghua University Press, Beijing, China, 3rd edition, 2007.

Data-Driven Spatial Econometric Analysis Model for Regional Tourism Development

Xuemei Zhao and Guofeng Yu

Anhui Business and Technology College, Hefei 231131, China

Correspondence should be addressed to Xuemei Zhao; 11261201@qq.com

Academic Editor: Adrian Neagu

Taking 16 cities in Anhui province as research units, based on the research perspective of spatial econometrics, using spatial autocorrelation analysis, this paper investigates the spatial correlation of regional tourism industry in Anhui province, indicating that regional per capita tourism income in Anhui province has obvious positive spatial autocorrelation and relatively obvious local spatial cluster characteristics. To further explore the influence factors of the development level of the regional tourism industry in Anhui and construct the spatial econometric model, the model results show that the Anhui tourism industry development has been accompanied by spatial agglomeration process, per capita GDP, the number of star-rated hotels, fixed assets investment, and employment in the tertiary industry which play a significant role for tourism development. Finally, some countermeasures and suggestions are put forward to promote the development of regional tourism in Anhui province.

1. Introduction

As a multifunctional and comprehensive industry, tourism is an economic issue, a livelihood issue, and an issue of social harmony. In recent years, Anhui province has been committed to becoming a major tourism province, and the tourism economy has become a new growth point for Anhui's economic development. In 2017, Anhui province received 5.49 million inbound tourists, a year-on-year increase of 13.07%; it received 626 million domestic tourists, a year-on-year increase of 19.88%, reaping a total tourism revenue of 619.7 billion yuan, a year-on-year increase of 25.64%. According to preliminary estimates, the added value of the tourism industry accounted for 6.6% and 16% of the province's gross domestic product (GDP) and service industry, respectively, with contribution rates of 10.4% and 13.3%, respectively. Although the tourism economy in Anhui province has been developing rapidly, regional tourism development is not coordinated, and the pattern that some regions are developed whereas others are underdeveloped has not changed. Therefore, how to promote the coordinated development of regional tourism in Anhui province and improve the level of regional tourism development in Anhui is currently an important issue in the development of tourism.

By analyzing the evolution characteristics of the spatial pattern of tourism development in Anhui province and the main factors of spatial-temporal pattern evolution of tourism development, indicating that regional per capita tourism income in Anhui province has obvious positive spatial autocorrelation and relatively obvious local spatial cluster characteristics, it is helpful to summarize the evolution law, construct the spatial pattern of regional tourism development, and promote the development of regional tourism in Anhui province.

The spatial autocorrelation of regional tourism development level has rarely been addressed in China. Liu et al. [1] analyzed the spatial correlation between the provincial tourism industry and tourism economic growth in China's 31 provinces through the spatial autocorrelation index and revealed that the development of China's tourism industry has always been accompanied by a process of spatial agglomeration and that the tourism economy growth exhibited a significant positive spatial correlation. Sun and Dai [2] investigated the spatial distribution patterns of per capita tourism income in China's 31 provinces and their evolution and found that China's regional tourism economy exhibits a strong spatial dependence and a long-term polarization pattern of high-value agglomerations and low-value agglomerations at the local level. Wu [3] conducted a spatial correlation analysis on Chinese provincial tourism economic growth and demonstrated that the provincial tourism economic growth has a spatial dependence, in addition to estimating the contributions of capital and labour to tourism economic growth and the spatial spillover effect in the process of tourism economic growth using a spatial panel econometric model. Wang [4] conducted a spatial analysis of the linkage and difference in the tourism industry development in the Greater Pearl River Delta region and demonstrated that there is a negative spatial correlation in the overall tourism economy of the Delta region. Based on these past studies, in this study, we analyzed the spatial characteristics and influencing factors of the tourism industry of Anhui province by combining the current tourism situation in Anhui province to find appropriate countermeasures suitable for the development of tourism in Anhui Province and promote the healthy development of the tourism industry in the province.

About the tourism impact on regional economy and society, foreign scholars mainly focus on the applied aspects of the analysis. Yang and Fik [5] examined two types of spatial effects in regional tourism growth: spatial spillover and spatial heterogeneity. A spatial growth regression framework is used to model the growth in regional tourism and identify the economic and spatial factors that explain the variability in tourism growth across 342 prefectural-level cities in China from 2002 to 2010. Jackson and Murphy [6] investigated the applicability of cluster theory in supporting the movement from comparative advantage to competitive advantage for four regional towns located on the Murray River in Australia. Chaabouni [7] investigated tourism efficiency and its determinants using a two-stage double bootstrap approach for a global panel of 31 Chinese provinces over the period 2008–2013. Bias-corrected data envelopment analysis (DEA) efficiency scores were first calculated by employing the smoothed homogeneous bootstrapped procedure. They were then regressed on a set of explanatory variables using the double-truncated regression approach. . For other background and applications in mathematical model aspects, see [8–22].

2. Research Methods and Data Sources

2.1. Descriptive Analysis. First of all, descriptive analysis is conducted from the time dimension, mainly using some commonly used indicators and coefficients to measure regional development differences. In this paper, the coefficient of variation is used for analysis. The coefficient of variation is a normalized measure of the degree of dispersion of probability distribution, which is defined as the ratio of standard deviation to average. The coefficient of variation eliminates the influence of data level and measurement unit, measures the relative discrete degree of data, and can be used to compare the discrete degree of different groups of data. The larger the coefficient of variation is, the greater the difference degree of variable value is. The smaller the coefficient of variation is, the smaller the difference degree of variable value is. In practical application, the coefficient of variation is generally used to compare the discrete degree of different groups of data:

$$V_\sigma = \frac{\sigma}{\overline{x}} \times 100\%, \tag{1}$$

where V_σ is the coefficient of variation, σ is the standard deviation, and $\overline{x}$ is the mean.

2.2. Spatial Autocorrelation. From the dimension of space, we analyze and test the spatial correlation from the global spatial autocorrelation analysis. Whether the data have a spatial dependence must be examined when using spatial econometric methods. If a spatial dependence is absent, then the standard econometric method is adopted. Otherwise, the spatial econometric method is used.

Autocorrelation of spatial sequences is rather complicated, and spatial autocorrelation refers to the fact that regions with similar locations have similar variable values. If high-value aggregation and low-value aggregation are present, it indicates the presence of "positive space autocorrelation," whereas if high values and low values are distributed completely randomly, it indicates the absence of any spatial autocorrelation.

Given the complexity of spatial autocorrelation, a series of methods for measuring spatial autocorrelation have been proposed in the past, such as $Moran\prime s\, I$, Garyy's C, and Getis, the most common of which is the $Moran\prime s\, I$ [8]. The global $Moran\prime s\, I$is often used to analyze the correlation index of the whole region, and the local $Moran\prime s\, I$ is used to analyze the correlation index of each regional unit in the region. The global $Moran\prime s\, I$ is calculated as follows:

$$I = \frac{W_{11}(Y_1-\overline{Y})(Y_1-\overline{Y}) + \cdots + W_{1n}(Y_1-\overline{Y})(Y_n-\overline{Y}) + \cdots + W_{n1}(Y_n-\overline{Y})(Y_1-\overline{Y}) + \cdots + W_{n1}(Y_n-\overline{Y})(Y_n-\overline{Y})}{s^2(W_{11} + \cdots W_{1n} + W_{21} + \cdots + W_{nn})}, \tag{2}$$

in $s^2 = (Y_1-\overline{Y})^2 + \cdots + (Y_n-\overline{Y})^2/n$, where

$$\begin{aligned} y &= X\beta + \varepsilon, \\ \varepsilon &= \lambda w_y + \mu, \end{aligned} \tag{3}$$

is the sample variance and W_{ij} is the spatial weight matrix (i, j) (for measuring the distance between the measurement regions i and j), which is generally constructed based on the distance and adjacency relationship of geographic features. In this study, the binary matrix ($W_{ij} = \begin{vmatrix} W_{11} & \cdots & W_{1n} \\ \vdots & & \vdots \\ W_{n1} & \cdots & W_{nn} \end{vmatrix}$, in which when regions i and j are adjacent, the matrix has a value of 1. Otherwise, it takes the value 0) was constructed based on the adjacency relationship.

The calculation result of $Moran's\, I$ is between −1 and 1. When it is greater than 0, it indicates the presence of a positive spatial correlation. When it is less than 0, it indicates the presence of a negative spatial correlation. When it is close to 0, it indicates that the adjacent regions are independent of each other, and thus no spatial autocorrelation among them is present.

The global $Moran's\, I$ only examines the spatial agglomeration of the entire region and cannot reveal which regions are high-value agglomeration regions or low-value agglomeration regions in the entire region. To elucidate the spatial agglomeration near a certain region, it is necessary to further examine local spatial autocorrelation, which can be performed through the local $Moran's\, I$, with the following calculation formula:

$$I = \frac{\left(Y_i - \overline{Y}\right)\left(W_{i1}\left(Y_1 - \overline{Y}\right)^2 + \cdots + W_{in}\left(Y_n - \overline{Y}\right)^2\right)}{s^2}, \quad (4)$$

in $s^2 = (Y_1 - \overline{Y})^2 + \cdots + (Y_n - \overline{Y})^2/n$, where Y_j is the observation value of Region j and W is the spatial weight matrix. Like the global $Moran's\, I$, in the case of the local $Moran's\, I$, if $Moran's\, I > 0$, it indicates the adjacent regions have similar feature values, assuming the "high value-high value" or "low value-low value" distributions; if $Moran's\, I < 0$, it indicates the adjacent regions have non-similar feature values, assuming the "high value-low value" or the "high value-low value" distributions, which indicates high spatial heterogeneity among the regions.

2.3. Spatial Econometric Model. The use of traditional econometric models on explanatory variables that have passed the test for spatial autocorrelation will lead to unreasonable estimation results; in this case, it is necessary to add a spatial weight matrix to modify the model and construct a spatial econometric model for analysis. Spatial econometric models are mainly categorized into two types: one is the spatial lag model (SLM), which is mainly used to study the behaviour of neighbouring regions and the spatial impacts on the behaviour of other regions in the whole system, such as diffusion or spillover. The SLM is thus suitable for estimating the intensity of spatial interactions. It has the following calculation formula:

$$y = \rho W_y + X\beta + \varepsilon, \quad (5)$$

where y is the explained variable, X is exogenous explanatory variable matrix of $n \times K$, ρ is the spatial regression coefficient, W is the space weight matrix of order $n \times n$, W_y is the explained variable of spatial lag, and $\varepsilon$is the vector of the random error term.

Another is the spatial error model (SEM), in which the relationship between the regions is represented by the random interference term, with the following calculation formula:

$$\begin{aligned} y &= X\beta + \varepsilon, \\ \varepsilon &= \lambda W_y + \mu, \end{aligned} \quad (6)$$

where ε is the vector of the random error term, λ is the spatial error coefficient of the vector of the explained variable of $n \times 1$ cross section, and μ is the random error vector of normal distribution.

3. Spatial Econometric Analysis of the Regional Tourism Economy

To mine the spatial heterogeneity and spatial dependence information regarding the regional tourism economy of Anhui province, we used the per capita tourism income as the raw data, on which descriptive analysis was performed. The global spatial correlation was analyzed using the global $Moran's\, I$, and the local spatial correlation patterns of the regional tourism economy were further determined using the local $Moran's\, I$ and scatter plots.

3.1. Descriptive Analysis. First, we performed a descriptive analysis of the regional tourism economy of Anhui province using the mean, standard deviation, and coefficient of variation. The results are reported in Table 1 and Figure 1.

Overall, the gap between the per capita tourism revenues of regions in Anhui province has gradually narrowed. The People's Government of Anhui province issued the "Implementation Opinions on Further Accelerating the Development of Tourism Industry," which adopts rational layout plans and promotion of coordinated development of the regional tourism to vigorously develop the tourism economy. It has effectively orchestrated regional tourism development and enabled the regions with relatively undeveloped tourism to develop rapidly, thereby gradually narrowing the regional development gap. However, the coefficient of variation was still greater than 1, indicating that the dispersion in per capita tourism revenues in various regions is still high and that a development imbalance persists among the regions.

3.2. Spatial Autocorrelation Analysis. According to the calculation formula of the spatial autocorrelation, the per capita tourism income of each region was used in the calculation, and the results are shown in Table 2. It indicates that the values of $Moran's\, I$ of per capita tourism income of different regions of Anhui province were all greater than 0. Their normal statistics passed the significance test at the significance level of 0.1, indicating that, in terms of the spatial distribution, the per capita tourism revenues of different regions in Anhui province exhibited a significant positive spatial autocorrelation, that is, the regions with high per capita tourism income were adjacent to each other, as were those with low per capita tourism income (the presence

Table 1: Descriptive indicators of regional per capita tourism revenue of Anhui province from 2011 to 2017.

Year	Mean/yuan/person	Standard deviation/yuan/person	Coefficient of variation
2011	4400.046	5,793.691	1.316734
2012	5975.69	7515.343	1.257653
2013	6734.163	8121.312	1.205987
2014	7568.442	8919.375	1.178495
2015	8553.612	9823.695	1.148485
2016	10028.46	10580.35	1.055032
2017	11929.42	12,514.78	1.049069

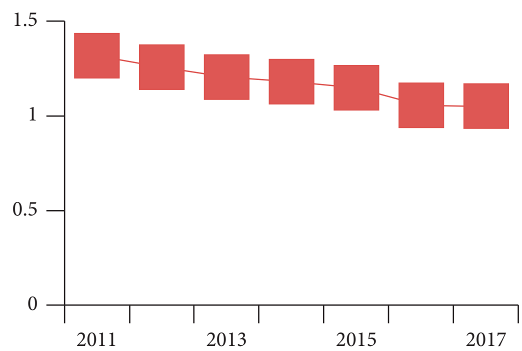


Figure 1: The coefficient of variation of regional tourism.

Table 2: The global $Moran's\, I$ of regional per capita tourism income of Anhui province from 2011 to 2017.

Year	$Moran's\, I$	Z value	*P* value
2011	0.225	2.527	0.011
2012	0.246	2.597	0.009
2013	0.261	2.656	0.008
2014	0.273	2.696	0.007
2015	0.266	2.522	0.012
2016	0.305	2.722	0.006
2017	0.297	2.688	0.007

of spatial agglomeration among the regions). Table 2 indicates that the values of $Moran's\, I$ of different regions in the province from 2011 to 2016 were all greater than 0.2 and have exhibited a rising trend, indicating that, during the period, the spatial dependence of tourism development in various regions has gradually increased, and the trend that regions with high per capita tourism income levels and those with low per capita tourism income levels cluster in their respective regions has also intensified. In 2017, the values of $Moran's\, I$ slightly decreased, indicating that the agglomeration of per capita tourism avenues of different regions in Anhui province in 2017 was slightly weakened. At the same time, the coefficient of variation of per capita tourism revenues of different regions in 2017 was significantly greater than that in 2016, indicating that the difference in per capita tourism income in different regions increased.

We selected two time sections of 2015 and 2017 and generated scatter plots of the local $Moran's\, I$ values of per capita tourism revenues of different regions of Anhui province; the study shows that most regions clustered in the third quadrant, indicating that the low-value agglomeration type dominates and thus the existence of a positive spatial correlation. The "core regions" of high value clustered in the first quadrant included Huangshan and Chizhou Cities, and it is worth noting that Wuhu City caught up and entered the first quadrant, making the south Anhui tourism cluster centred in Huangshan and Chizhou slightly move towards central Anhui (Wuhu), which is largely related to the commercial development, cultural exchange, and the driving and developing of the Fangte Happy World in Wuhu. Xuancheng, Tongling, Anqing, and Maanshan Cities were in the third quadrant (high value-low value), whereas Hefei, the provincial capital, has been always in the fourth quadrant (low value-high value), mainly because, among the regions, Hefei is a comprehensive development area, with inadequate dominant tourism resources, and its tourism revenues are derived mainly from businesses, exhibitions, and conferences.

The above results show that the regional tourism levels in Anhui province were characterized by spatial heterogeneity, profound local spatial clustering characteristics, and local spatial autocorrelation.

3.3. Spatial Econometric Model. Due to the presence of significant spatial dependence of regional per capita tourism revenues in Anhui province, the spatial distribution of the tourism economy in various regions has changed from 2011 to 2017. Thus, traditional econometric regression models are not appropriate for examining the spatial correlations between regions and not able to comprehensively reveal the effects of the factors, making it impossible to obtain complete and rational conclusions. To further analyze factors that have a significant impact on the per capita tourism revenue, we adopted a spatial econometric regression model to analyze the causes of spatial differences to better interpret the spatial correlations and impacts.

Tourism development is affected by many factors; theoretically, it is affected by macroeconomic factors, resource factors, population factors, policy systems, and other factors. To further analyze Anhui's regional tourism development, we chose the indicator of per capita tourism income as the explained variable of the model to directly reflect the operation status and effectiveness of the regional tourism economy and the development healthiness of the tourism market; we used the per capita GDP (X 1), the number of star-rated hotels (X 2), the fixed assets investment (X 3), and the number of employees in the tertiary industry (X 4) as explanatory variables to perform the spatial regression analysis.

Table 3: Results of the spatial error model.

Model parameter	Coefficient	P value
X 1	0.2003439	0.000
X 2	859.4962	0.000
X 3	−5.690162	0.000
X 4	−59.6762	0.000
C	3,370.822	0.013

Based on spatial econometric-related theories, we used the Stata software package to generate the spatial lag model and the spatial error model. The lambda value of the autoregressive coefficient of the error term passed the significance test, whereas rho, the spatial autoregressive coefficient of the spatial lag model, did not, indicating that the spatial error model is able to fully extract and express spatial correlation information and thus superior to the spatial lag model. Therefore, we adopted the spatial error model to examine the influencing factors of the regional per capita tourism income, and the results are reported in Table 3.

Among the influencing factors of per capita tourism income, the per capita GDP, the total number of star-rated hotels, the fixed assets investment, and the number of employees in the tertiary industry all passed the significance test at the significance level of 1%, indicating that they have played a significant role in tourism development.

4. Results

The results are as follows:

(1) The regression coefficient of per capita GDP was positive, indicating that the impact of the local economic development level on tourism development is positive as a promoting effect, and the regression coefficient is small, indicating that local economic development level is one of the main factors affecting tourism development.

(2) The regression coefficient of the number of star-rated hotels was positive and the highest, indicating that the number of star-rated hotels in each region has a positive effect on tourism development and is the most important factor affecting tourism development.

(3) Both the fixed assets investment and the number of employees in the tertiary industry passed the significance test, indicating that both are factors that affect tourism development, but their regression coefficients were negative, which is contradictory to the previous notion that increasing fixed assets investment and the number of employees of the tertiary industry can promote tourism income. This result is mainly due to the limitation of the data used here, in which the amount of fixed assets investment in the development of tourism industry and the number of employees in the tourism industry were lacking, which affected the model estimation results. Once the abovementioned data become available, the model can be further improved.

5. Conclusions

Based on the above analyses, the global spatial autocorrelation (dependency) of regional per capita tourism income levels in Anhui province was strong, exhibiting spatial agglomerations. The regions with "high value-high value" agglomeration mainly included Huangshan, Chizhou, and Wuhu, and those with "low value-low value" agglomeration mainly included Chuzhou, Lu'an, Fuyang, Suzhou, Huainan, and Bengbu, whereas Hefei, the capital city, has always been in the region with "low value-high value" agglomeration. Based on the spatial characteristics of tourism development in Anhui province, we provide the following recommendations:

(1) The key to developing the regional tourism industry [7] and increasing tourism revenue is to vigorously develop the local economy. After the local economy is developed, people's disposable income will be increased, which will stimulate tourism demand and promote the development of the tourism industry. Anhui province's tourism investment has been continuously growing, and the implementation of the "335" tourism construction action plan has achieved an accumulated investment of 804.5 billion yuan during the "Twelfth Five-Year Plan" period, which was 5.3 times of the total investment during the "Eleventh Five-Year Plan" period, accounting for 8.9% of the province's total accumulated fixed asset investment. Further adjusting the proportion structure of fixed assets investment in the tourism industry, formulating reasonable investment plans, and forming a more sophisticated investment system can continuously improve the investment efficiency. At the same time, the relevant departments should use funds dedicated to the key points of tourism industry development to maximize capital effectiveness, such that the waste of funds can be minimized.

(2) The differences among different regions of Anhui province were profound. Due to the limitation of geographical location, the agglomeration effect of southern Anhui has had very little driving and spatial spillover effects on central and northern Anhui. Southern Anhui should emphasize its agglomeration effect and spatial spillover effect to drive the development of its neighbouring regions [8], such that tourism can be developed more rapidly in regions with inexplicit resources.

We should encourage each region to engage in regional cooperation and promote tourism for all, actively implement various tourism marketing strategies to increase the publicity of tourist attractions [9], strengthen the focus of tourism promotion on regions neighbouring known scenic spots (e.g., Huangshan and Jiuhuashan), and increase the investment in the tourism infrastructure and supporting facilities, especially the transportation construction, thereby making the access to tourist attractions unimpeded and consequently achieving coordinated development of tourism.

(3) Because the spatial location factor is an important factor affecting the development of the regional tourism industry in Anhui province, the regional tourism industry development in the province is affected not only by the economic development level of the region and the number of star-rated hotels but also by the development level of tourism in the neighbouring regions, and the levels of development of different regions are interdependent to some extent. Improvement of the tourism level or improvement of economic development in a certain region has positive effects on the tourism development in neighbouring regions. Therefore, to develop regional tourism in Anhui province and to increase regional tourism revenue, we should simultaneously grasp the development statuses of the region and neighbouring regions and make full use of all favourable factors. We should also strengthen the exchange of experience in the development of the tourism industry in neighbouring regions, especially information sharing regarding infrastructure and tourism talents. In areas with rich explicit tourism resources, we should utilize the resources and geographical advantages to drive the tourism development in adjacent areas, thereby forming a large-scale tourism industry cluster.

(4) To address the spatial dependence problem of regional tourism development in Anhui province, tourism management departments should coordinate the development of regional tourism in the province, and the government should actively develop effective and adaptive tourism development plans and construct "smart Anhui tourism," using cutting-edge information and big data methods to achieve integrated tourism across the province. Moreover, the tourism industry in different regions should be integrated to conduct unified guidance and management regarding tourism product development, infrastructure construction, and resource and environment protection to continuously deepen the coordinated development of regional tourism while ensuring sustainable tourism development in Anhui province.

Acknowledgments

This work was supported by 2020 Anhui Quality Engineering Project: Accounting and Statistics Teaching Team (2020jxtd031), 2019 Anhui Business and Technology College Teaching Innovation Team Project (2019xjjxtd01), 2018 Anhui Natural Science Research Project (KJ2018A084), 2017 Anhui Natural Science Research Project (KJ2017B004), 2017 Anhui Vocational and Adult Education Association Project (azjxh17109), and 2016 Anhui Provincial Tourism Bureau Key Research Project (2016FACJ2985-8).

References

[1] J. Liu, J. Zhao, and G. Zhang, "Spatial econometric analysis of the relationship between China's tourism industry clustering and tourism economic growth," *Economic Geography*, vol. 33, no. 4, pp. 186–192, 2013.

[2] P. Sun and X. Dai, "Spatial statistics analysis on China's tourism economy difference," *Tourism Science*, vol. 28, no. 2, pp. 35–48, 2014.

[3] Y. Wu, "Spatial panel econometric analysis of tourism economic growth and its spillover effect," *Tourism Tribune*, vol. 29, no. 2, pp. 16–24, 2014.

[4] L. Wang, "Empirical analysis of tourism economy in the Greater Pearl River Delta based on spatial econometri," no. 4, pp. 159-160, 2016.

[5] Y. Yang and T. Fik, "Spatial effects in regional tourism growth," *Annals of Tourism Research*, vol. 46, pp. 144–162, 2014.

[6] J. Jackson and P. Murphy, "Clusters in regional tourism an Australian case," *Annals of Tourism Research*, vol. 33, no. 4, pp. 1018–1035, 2006.

[7] S. Chaabouni, "China's regional tourism efficiency: a two-stage double bootstrap data envelopment analysis," *Journal of Destination Marketing and Management*, vol. 11, pp. 183–191, 2019.

[8] H. Zhou, X. Mao, and N. Li, "Space-time variation trend of wind power in China," *Environmental Science and Management*, vol. 45, no. 12, pp. 148–152, 2020.

[9] J.-B. Liu and X.-F. Pan, "Minimizing Kirchhoff index among graphs with a given vertex bipartiteness," *Applied Mathematics and Computation*, vol. 291, pp. 84–88, 2016.

[10] J. Liang and C.-S. Chan, "Local cultural vicissitudes in regional tourism development: a case of Zhuhai," *Tourism Management Perspectives*, vol. 25, pp. 80–92, 2018.

[11] L. Medeiros de Araujo and B. Bramwell, "Partnership and regional tourism in Brazil," *Annals of Tourism Research*, vol. 29, no. 4, pp. 1138–1164, 2002.

[12] F. Valente, D. Dredge, and G. Lohmann, "Leadership and governance in regional tourism," *Journal of Destination Marketing and Management*, vol. 4, no. 2, pp. 127–136, 2015.

[13] J. Jackson, "Developing regional tourism in China: the potential for activating business clusters in a socialist market economy," *Tourism Management*, vol. 27, no. 4, pp. 695–706, 2006.

[14] A. Zahra and C. Ryan, "From chaos to cohesion-Complexity in tourism structures: an analysis of New Zealand's regional tourism organizations," *Tourism Management*, vol. 28, no. 3, pp. 854–862, 2007.

[15] B. Ahn, B. Lee, and C. S. Shafer, "Operationalizing sustainability in regional tourism planning: an application of the limits of acceptable change framework," *Tourism Management*, vol. 23, no. 1, pp. 1–15, 2002.

[16] A. I. B. Josep, "Regional tourism planning in Spain: evolution and perspectives," *Annals of Tourism Research*, vol. 31, no. 2, pp. 313–333, 2004.

[17] H. Li, J. L. Chen, G. Li, and C. Goh, "Tourism and regional income inequality: evidence from China," *Annals of Tourism Research*, vol. 58, pp. 81–99, 2016.

[18] A. P. F. Lopes, M. M. Muñoz, and P. Alarcón-Urbistondo, "Regional tourism competitiveness using the PROMETHEE approach," *Annals of Tourism Research*, vol. 73, pp. 1–13, 2018.
[19] M. Carrillo and J. M. Jorge, "Multidimensional analysis of regional tourism sustainability in Spain," *Ecological Economics*, vol. 140, pp. 89–98, 2017.
[20] A. Guizzardi and A. Stacchini, "Real-time forecasting regional tourism with business sentiment surveys," *Tourism Management*, vol. 47, pp. 213–223, 2015.
[21] B. Doolin, L. Burgess, and J. Cooper, "Evaluating the use of the Web for tourism marketing: a case study from New Zealand," *Tourism Management*, vol. 23, no. 5, pp. 557–561, 2002.
[22] P. Del Vecchio and G. Passiante, "Is tourism a driver for smart specialization? Evidence from Apulia, an Italian region with a tourism vocation," *Journal of Destination Marketing & Management*, vol. 6, no. 3, pp. 163–165, 2017.

Spatiotemporal Simulation of Tourist Town Growth Based on the Cellular Automata Model: The Case of Sanpo Town in Hebei Province

Jun Yang,[1,2] Peng Xie,[1] Jianchao Xi,[1] Quansheng Ge,[1] Xueming Li,[2] and Fanqiang Kong[2]

[1] *Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China*
[2] *Liaoning Key Laboratory of Physical Geography and Geomatics, Dalian 116029, China*

Correspondence should be addressed to Peng Xie; xiepenggis@163.com

Academic Editor: Jianhong (Cecilia) Xia

Spatiotemporal simulation of tourist town growth is important for research on land use/cover change under the influence of urbanization. Many scholars have shown great interest in the unique pattern of driving urban development with tourism development. Based on the cellular automata (CA) model, we simulated and predicted the spatiotemporal growth of Sanpo town in Hebei Province, using the tourism urbanization growth model. Results showed that (1) average annual growth rate of the entire region was 1.5 Ha^2 per year from 2005 to 2010, 4 Ha^2 per year from 2010 to 2015, and 2.5 Ha^2 per year from 2015 to 2020; (2) urban growth rate increased yearly, with regional differences, and had a high degree of correlation with the Euclidean distance of town center, traffic route, attractions, and other factors; (3) Gougezhuang, an important village center in the west of the town, demonstrated traffic advantages and increased growth rate since 2010; (4) Magezhuang village has the largest population in the region, so economic advantages have driven the development of rural urbanization. It showed that CA had high reliability in simulating the spatiotemporal evolution of tourist town, which assists the study of spatiotemporal growth under urbanization and rational protection of tourism resources.

1. Introduction

Tourist towns have become increasingly important in tourism development and rural urbanization in recent years. Many researchers have focused on the practical use of tourism resources to achieve sustainable development of tourist towns [1, 2]. Such studies have conducted qualitative analysis on the role of various factors in sustainable tourist town development, so it is difficult to objectively predict the future development of these towns. Therefore, it is important to effectively simulate the development trend of tourist towns and suggest reasonable development plans.

The cellular automata (CA) model was first introduced by John von Neumann in the 1940s under the inspiration of mathematician and physicist Stanislaw Ulam during the "Manhattan Project." The cellular automata (CA) model is a mathematical framework discrete in time, space, and state and is a dynamic evolution system consisting of a large number of interacting cells. Wolfram [3] played a significant role in promoting early CA studies and laid the theoretical basis for such work. Because the CA model has strong spatial computing power, a complex and global pattern can be formed through simple local operations. This model has been successfully applied to environmental change, landscape pattern replacement, forest fire diffusion, urban expansion, and other simulation studies [4–11].

In urban geography, CA has been applied in the simulation of large-scale urban changes to explore the application of discrete dynamic models in urban land use change [12, 13] and in the simulation of urban systems [12, 14, 15]. Zhou and Chen [16] studied CA and elaborated on its principles, potential problems, and research significance, while Li and Yeh [17] used binding CA in the simulation of sustainable urban development patterns. The above studies have played a large role in promoting the development and application of CA theory in China.

In recent years, new developments in the urban CA model have been driven by progress in computer technology. Lauf et al. [18] studied an improved CA model, which included family and housing factors as the driving force through integrating system dynamics, to explore residential construction land expansion. García et al. [19] analyzed urban expansion in Northern Spain and Galicia by a variety of methods. Mitsova et al. [20] studied urban expansion and the protection of sensitive areas using CA based on land-use changes. Furthermore, a research has been conducted on the acquisition of CA conversion through combining neural network theory, data mining data and genetic algorithm theory, and on the enhanced simulation accuracy of the model [21–23].

The present paper applied the principle of CA to the simulation of tourist town urbanization, selected both suitable and limiting elements based on tourist town requirements for the development and protection of tourism resources, expanded the parameter system of traditional CA, and established the tourist town CA model to simulate Sanpo development from 2010 to 2020.

2. Model and Methods

2.1. Study Area. Sanpo town is located in the Taihang Mountain area, 28 km northwest of the Laishui County town of Baoding in Hebei Province, China. Sanpo is a national natural scenic area with significant regional advantages and is positioned 90 km east of Beijing, 170 km southeast of Tianjin and Langfang, and 90 and 190 km south of Baoding and Shijiazhuang, respectively. The town covers a total area of 200 km^2, is 180 to 1500 m above sea level, and has a total population of 11,887 people. The Beijing-Yuanping Railway, Baoye Highway, and 108 Highway run through the town. Sanpo is the administrative heart of the Yesanpo National Scenic Area, covering most of the scenic area and having many residents. For management purposes, Sanpo consists of a number of smaller villages, such as Gougezhuang and Magezhuang (Figure 1). With the development of tourism in Yesanpo since 1986, Sanpo has gradually evolved into a specialized tourist town. The past 25 years of development provide clear temporal and spatial evidence on the evolution of tourism development in this representative rural tourism town in China.

2.2. Tourism CA Model Framework. Tourism urbanization is a complex process. To simulate the spatio-temporal evolution of a tourist town, the driving factors and comprehensive mechanisms of tourism urbanization evolution must be studied. We first analyzed the type of tourism in the study area, clarified the development of the main tourism industry chains, and collected land use data, topographic maps, remote sensing data, and economic and social statistics of Sanpo over the years. We further investigated the integration of data compilation and multisource data through the area, studied the spatio-temporal evolution and relationship between driving factors using GIS spatial analysis and statistic functions, and combined analytic hierarchy processes (AHP). On this basis, we conducted spatial growth simulation through Python language programming using CA and urbanization growth models with ArcGIS software, as shown in Figure 2.

2.3. Model Structure and Index System. The model was established based on raster data and using the GRID raster data coding format of ArcGIS. The cellular space of the model covered the entire study area with 10 m × 10 m grids. To achieve multivariate data sharing, the model system used the WGS84 coordinate system. Cellular state space was divided into urban land, land that can be used for urbanization, and land that cannot be used for urbanization. Urban land information was provided by the land department of the Sanpo government, and land that cannot be used for urbanization was determined according to the requirements of protected areas and the constraining conditions of the terrain.

As external environment information of the cellular model, driving factor group data directly influences and controls the evolution of tourism urbanization. Tourism urbanization is a complex process with many influencing factors. By analyzing the comprehensive mechanisms of tourism and urbanization, we determined three major tourism urbanization factor groups and determined the weight coefficients of the driving factors by AHP (Table 1). The processing results of spatial data are shown in Figure 3. To achieve sustainable development and effectively protect tourism resources, natural reserves were taken as a spatial limiting factor, and the cell affected by spatial limiting factors could not be developed into a town.

2.4. Neighborhood Structure and Conversion Rules. The study area was a mountainous region with complex terrain. We considered two kinds of urbanization evolution power, specifically the influence of terrain and the influence of tourist factors. Firstly, the terrain on both sides of the valley watercourse was relatively flat with traffic arteries, so it was considered suitable for urban development. Secondly, the area was close to tourist attractions but beyond the safety distance of attraction protection to develop into a town.

2.4.1. Neighborhood Urbanization Function. The degree of urbanization was represented within the current cellular neighborhood, expressed by the proportion of urbanized cells within the scope of neighborhood, as follows:

$$P_{(x,y)} = \frac{1}{N} \sum_{i=1,j=1}^{\Omega} X_{(i,j)}, \quad (1)$$

where $P_{(x,y)}$ represents the percentage of urbanized cells within the neighborhood space, N represents the number of cells within the neighborhood space, $X_{(i,j)}$ represents urbanization cells, and Ω represents neighborhood space.

2.4.2. Cellular Conversion Probability. We determined the integrated value of the three major influencing factor groups

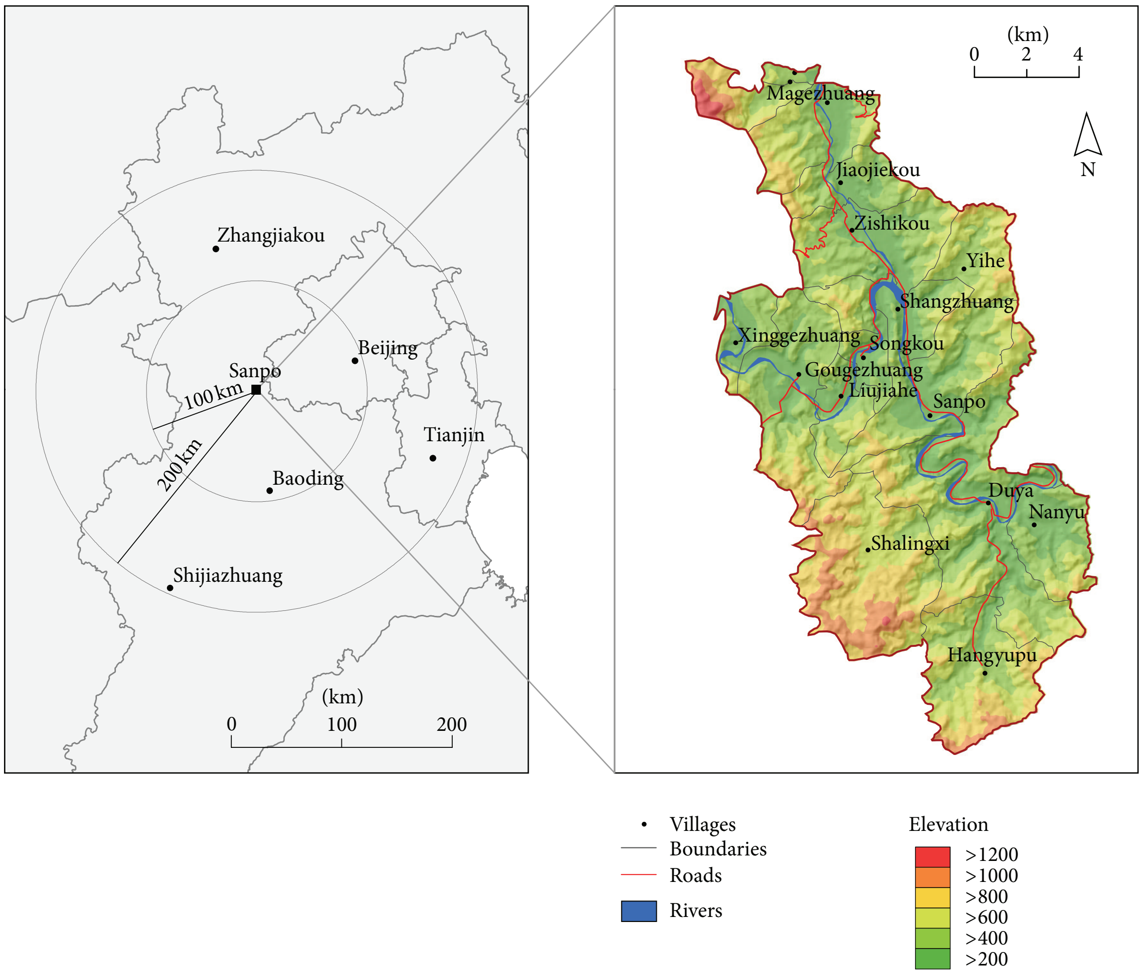


FIGURE 1: Location of study area.

and neighborhood urbanization function, representing the current cost value of cellular conversion, as follows:

$$p_{ij}^{t} = \phi\left(r_{ij}^{t}\right) = \exp\left[\alpha\left(\frac{r_{ij}^{t}}{r_{\max}} - 1\right)\right], \tag{2}$$

where α represents the diffusion coefficient and $r_{\max}$ represents the highest property value. The simple expression of r_{ij}^{t} is

$$r_{ij}^{t} = \left(\sum_{k=1}^{m} F_{ijk}^{t} W_k\right) \prod_{k=m+1}^{n} F_{ijk}^{t}, \tag{3}$$

when $1 \leq k \leq m$, r_{ij}^{t} represents the tourism urbanization driving factor, that is, the factors listed in Table 1; when $m < k$, r_{ij}^{t} represents the spatial limiting factor, referring to natural reserves and rivers, the probability for it to develop into a town is 0.

The simulation of CA was completed by multiple cycles. To express the uncertainty of tourism urbanization, p_{ij}^{t} (probability of developing into urban land) and $p_{\text{threshold}}$ (pregiven threshold) were added in the cycle for comparison to determine whether the current cell can develop into a town, that is,

$$\begin{cases} P_{ij}^{t} \geq P_{\text{threshold}} & \text{converted into urban land} \\ P_{ij}^{t} < P_{\text{threshold}} & \text{not converted into urban land.} \end{cases} \tag{4}$$

3. Results and Analysis

Based on the above models and methods, we conducted dynamic simulation of cellular automata for the spatio-temporal growth of Sanpo. Figure 4(a) represents the actual construction land in 2005 and 2010, and Figure 4(b) represents the simulated construction land in 2010, 2015, and 2020.

Because the study area was located in remote mountains on the north and south sides of the middle reaches of a river, the simulation results showed that the overall urban scale growth trend still did not develop in the flat area around the town. Within the scope of the study period, the towns and villages did not develop into an urban landscape and remained independent and decentralized geographical units.

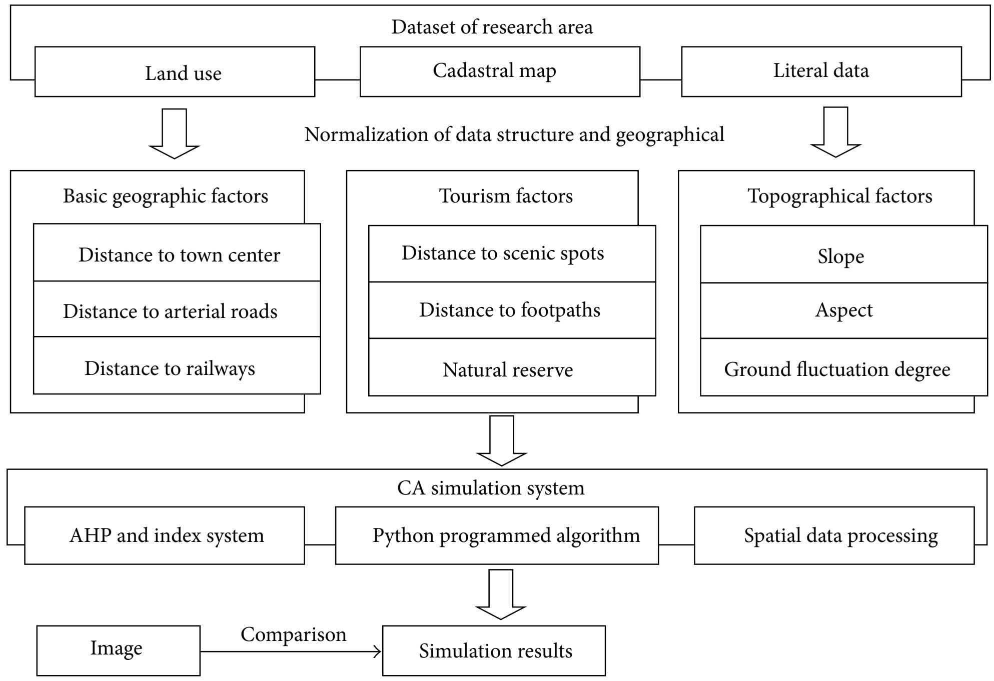


Figure 2: Frame diagram of spatio-temporal growth simulation of tourist town based on CA.

Table 1: Driving factors and weights of spatio-temporal growth of tourist town.

Influencing factor	Variable type	Weight coefficient	Variable grading
Basic geographic factors	Distance to town center	4	Grade 1: 0~500 m Grade 2: 501~1000 m Grade 3: 1001~3000 m Grade 4: >3000 m
	Distance to arterial roads	4	Grade 1: 0~100 m Grade 2: 101~500 m Grade 3: 501~1000 m Grade 4: >1000 m
	Distance to railway	2	Grade 1: 0~100 m Grade 2: 101~500 m Grade 3: 501~1000 m Grade 4: >1000 m
Geographic factors for tourism	Distance to footpath	3	Grade 1: 0~100 m Grade 2: 101~300 m Grade 3: 301~600 m Grade 4: >600 m
	Distance to scenic spots	3	Grade 1: 0~500 m Grade 2: 501~1000 m Grade 3: 1001~2000 m Grade 4: >2000 m
Topographical factors	Slope	4	Grade 1: <5 degrees Grade 2: 6 degrees~15 degrees Grade 3: 16 degrees~25 degrees Grade 4: >25 degrees
	Aspect	1	Grade 1: −45 degrees~45 degrees Grade 2: 45 degrees~135 degrees Grade 3: 135 degrees~225 degrees Grade 4: 225 degrees~315 degrees
	Ground fluctuation degree	2	Grade 1: <10 Grade 2: 10~20 Grade 3: 20~40 Grade 4: >40

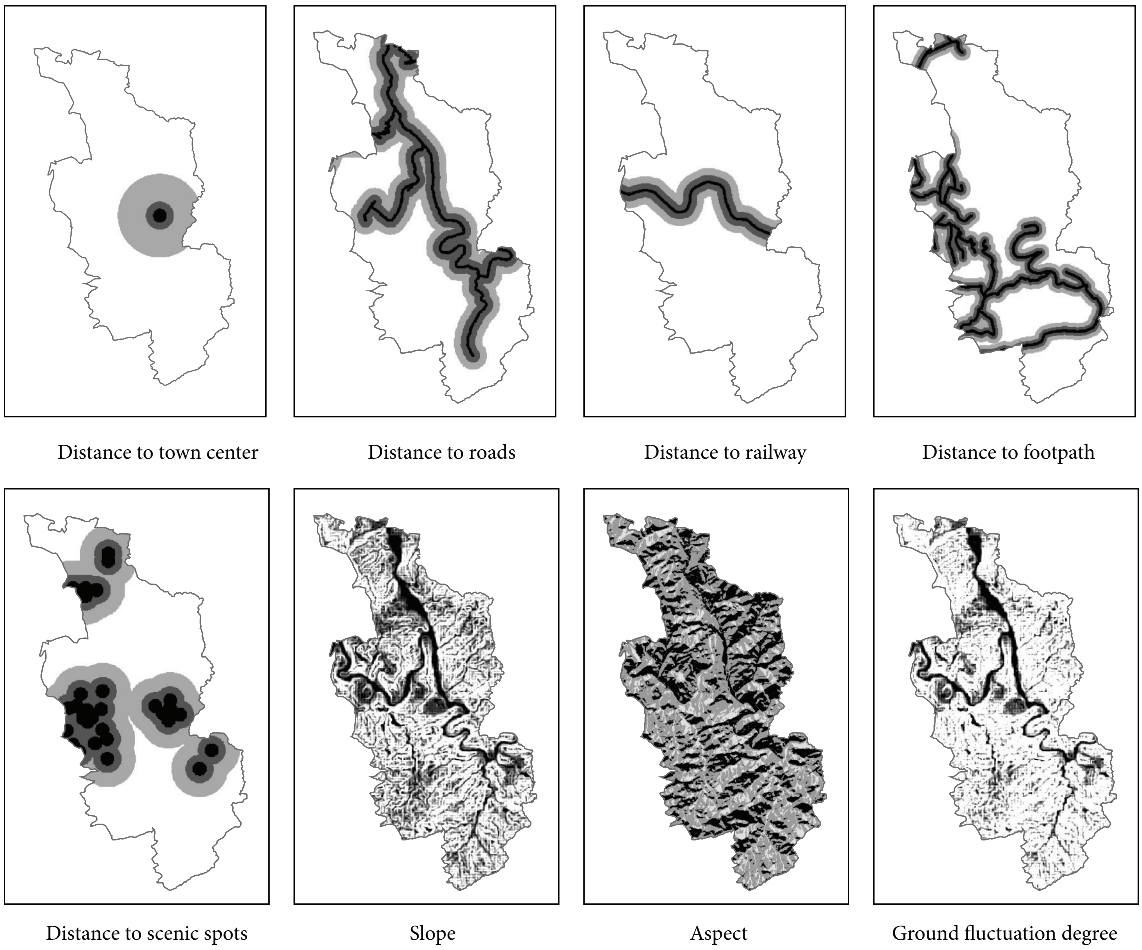


Figure 3: Spatio-temporal growth driving factors of tourist town.

We further analyzed the characteristics of urban growth and its driving mechanism in the Sanpo area regarding changes in time and space patterns.

Differences in urban development speed in different time periods were observed. Urban land increased by 7.5 Ha^2 from 2005 to 2010, with an average increase of 1.5 Ha^2 per year, and was expected to increase by 20 Ha^2 from 2010 to 2015, average increase rate of 4 Ha^2 per year, and by 17.520 Ha^2 from 2015 to 2020, average increase rate of 2.5 Ha^2 per year. These results were consistent with government development planning reports. Urban growth was mainly concentrated in the area near the town center, expanding at a rate of 1 Ha^2 per year, with the urban growth rate of other towns in the region being slow. This may be because the town center of Sanpo, which relies on the Yesanpo National Scenic Area, garnered significant support from development policies and funding in the early period of urban development. Furthermore, because urbanization adapts to the needs of local tourism development, tourism development accelerates urbanization in turn. Town urbanization accelerated after 2010 and was still the highest in the region with a growth rate of 1.3 Ha^2 per year. To the west of the town, Gougezhuang village developed rapidly, with an increase in construction land of 1.1 Ha^2 per year. Due to further demand for economic and tourism development in the Sanpo town center, contradictions existed in relation to sustainable development in the scenic area and land use restrictions, and other villages and towns (such as Gougezhuang and Magezhuang, Figure 4) showed tourism development potential with the support of resources and funding during this period. The simulated increase rate of urban land fell from 2015 to 2020, showing major growth near the town center because the growth area was topographically limited.

From the perspective of spatial patterns, urban growth regions in Sanpo were divided into three parts (right, Figure 4). The first was the northern area composed of South and North Chanfangzhuang and Magezhuang; the second was the western area, including Gougezhuang; and the third was Sanpo town center. Of these, the original urban town center covered the largest area and its growth rate was the fastest within the simulated period, demonstrating that the town center not only serves as a political, economical, and cultural center, but also as a tourist service base for the Yesanpo Scenic Area. The town center has developed at a fast rate due to the influx of people from surrounding towns and

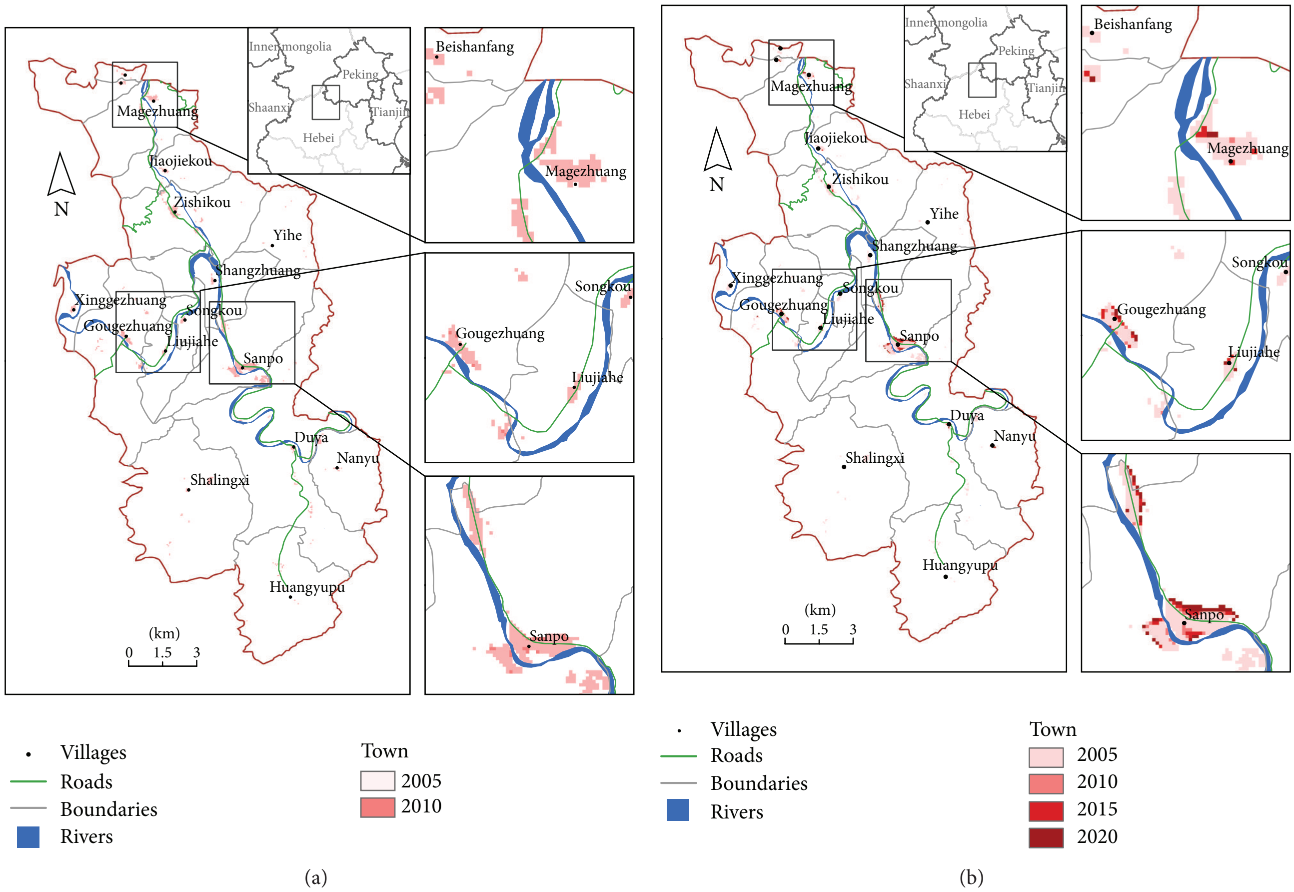


Figure 4: Spatio-temporal growth simulation results of Sanpo.

villages for employment, as well as the accommodation of temporary residents and visitors. Gougezhuang is an important central village to the west of the town, with an established Beijing-Yuanping Railway train station. Furthermore, it plays a role in radiating and driving the surrounding grass-roots villages, so there are unique advantages for tourism trade and service development in this village. In addition, with the largest population in this region, Magezhuang town growth will be inevitably achieved under the driving effect of the economy. For urban growth direction, most areas presented edge expansion; moreover, the urban growth rate of areas at the edge of town with traffic lines was faster than that of others, and the closer the area was to the tourist attraction, the faster its rate of urbanization was. The expansion of urban areas showed an obvious trend towards traffic and the tourism industry.

4. Conclusions and Discussion

This paper combined CA principles and GIS technology, established a tourism-type urban expansion model and simulated urban growth under the driving conditions of tourism factors using land use data of the Sanpo area from 2005 and 2010. The following was concluded.

(1) We analyzed the spatio-temporal growth of Sanpo using GIS technology and CA system principles, which reflected the role of tourism factors in regional urbanization and intuitively predicted the future trend of development of the town, with high reliability in solving spatio-temporal growth of the tourist city.

(2) By comprehensive analysis of the data and model, the Sanpo area was influenced by the natural environment, population growth, economic development, national policies, and other complex factors. The urban growth rate increased yearly, although with regional differences, and had a high degree of correlation with the Euclidean distance of town center, traffic route, attractions, and other factors.

(3) Because the Sanpo terrain was complex and tourism was a leading industry, the growth rates of towns surrounding the tourist attractions were relatively slow in the process of urban growth, achieving rational development, and utilization of tourism resources.

Because urbanization is a complex geographical process influenced by society, culture, and the economy, as well as by national economic policies and other factors, CA is still undergoing research and development and possesses a number of deficiencies [24]. Therefore, it is difficult to accurately simulate and predict urban growth. Some preliminary results have been made in researching the CA based model and simulating the growth process of tourist towns under microscale conditions; however, due to basic data limitations, the simulation accuracy of this study needs to be enhanced and the tourism urbanization CA simulation system requires further improvement.

Acknowledgments

The work described in this paper was substantially supported by the National Natural Science Foundation of China (nos. 40971299, 41171137) and Humanities Social Science Foundation of Ministry of Education (no. 09YJC790135).

References

[1] L. Lu and J. Ge, "Reflection on the research progress of tourism urbanization," *Geographical Research*, vol. 25, no. 4, pp. 741–750, 2006.

[2] Z. Wang and H. Yu, "Study on coupling development between development of tourism industry and small town construction in Zhangjiajie City," *Economic Geography*, no. 7, pp. 165–171, 2012.

[3] S. Wolfram, "Cellular automata as models of complexity," *Nature*, vol. 311, no. 5985, pp. 419–424, 1984.

[4] M. H. Afshar, M. Shahidi, M. Rohani, and M. Sargolzaei, "Application of cellular automata to sewer network optimization problems," *Scientia Iranica*, vol. 18, no. 3, pp. 304–312, 2011.

[5] Y. Feng and Z. Han, "Impact of neighbor configurations on spatially-explicit modeling results," *Geographical Research*, vol. 30, no. 6, pp. 1055–1065, 2011.

[6] S. Kokubo, J. Tanimoto, and A. Hagishima, "A new cellular automata model including a decelerating damping effect to reproduce Kerner's three-phase theory," *Physica A*, vol. 390, no. 4, pp. 561–568, 2011.

[7] E. A. Silva and K. C. Clarke, "Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal," *Computers, Environment and Urban Systems*, vol. 26, no. 6, pp. 525–552, 2002.

[8] V. Spicer, A. A. Reid, J. Ginther, H. Seifi, and V. Dabbaghian, "Bars on blocks: a cellular automata model of crime and liquor licensed establishment density," *Computers, Environment and Urban Systems*, vol. 36, no. 5, pp. 412–422, 2012.

[9] F. Wu, "Calibration of stochastic cellular automata: the application to rural-urban land conversions," *International Journal of Geographical Information Science*, vol. 16, no. 8, pp. 795–818, 2002.

[10] H. Yu, Z. He, and X. Pan, "Wetlands shrink simulation using cellular automata: a case study in Sanjiang Plain, China," *Procedia Environmental Sciences*, vol. 2, pp. 225–233, 2010.

[11] H. Zhang, Y. Zeng, X. Jin, C. Yin, and B. Zou, "Urban land expansion model based on multi-agent system and application," *Acta Geographica Sinica*, vol. 63, no. 8, pp. 869–881, 2008.

[12] M. Batty, H. Couclelis, and M. Eichen, "Urban systems as cellular automata," *Environment and Planning B*, vol. 24, no. 2, pp. 159–164, 1997.

[13] H. Couclelis, "Cellular worlds: a framework for modeling micro-macro dynamics," *Environment and Planning A*, vol. 17, no. 5, pp. 585–596, 1985.

[14] M. Batty and Y. Xie, "From cells to cities," *Environment and Planning B*, vol. 21, supplement 21, pp. 531–548, 1994.

[15] M. Batty and Y. Xie, "Possible urban automata," *Environment and Planning B*, vol. 24, no. 2, pp. 175–192, 1997.

[16] Y. Zhou and Y. Chen, "Cellular automata and simulation of spatial complexity of urban systems: history, present situation and future," *Economic Geography*, vol. 20, no. 3, pp. 35–39, 2000.

[17] X. Li and A. G. O. Yeh, "Constrained cellular automata for modelling sustainable urban forms," *Acta Geographica Sinica*, vol. 54, no. 4, pp. 289–298, 1999.

[18] S. Lauf, D. Haase, P. Hostert, T. Lakes, and B. Kleinschmit, "Uncovering land-use dynamics driven by human decision-making: a combined model approach using cellular automata and system dynamics," *Environmental Modelling & Software*, vol. 27-28, pp. 71–82, 2012.

[19] A. M. García, I. Santé, M. Boullón, and R. Crecente, "A comparative analysis of cellular automata models for simulation of small urban areas in Galicia, NW Spain," *Computers, Environment and Urban Systems*, vol. 36, no. 4, pp. 291–301, 2012.

[20] D. Mitsova, W. Shuster, and X. Wang, "A cellular automata model of land cover change to integrate urban growth with open space conservation," *Landscape and Urban Planning*, vol. 99, no. 2, pp. 141–153, 2011.

[21] X. Li, "Emergence of bottom-up models as a tool for landscape simulation and planning," *Landscape and Urban Planning*, vol. 100, no. 4, pp. 393–395, 2011.

[22] X. Li, Q. Yang, and X. Liu, "Discovering and evaluating urban signatures for simulating compact development using cellular automata," *Landscape and Urban Planning*, vol. 86, no. 2, pp. 177–186, 2008.

[23] X. Liu, X. Li, X. Shi, S. Wu, and T. Liu, "Simulating complex urban development using kernel-based non-linear cellular automata," *Ecological Modelling*, vol. 211, no. 1-2, pp. 169–181, 2008.

[24] P. M. Torrens and D. O'Sullivan, "Cellular automata and urban simulation: where do we go from here?" *Environment and Planning B*, vol. 28, no. 2, pp. 163–168, 2001.

Understanding Visitors' Responses to Intelligent Transportation System in a Tourist City with a Mixed Ranked Logit Model

Yang Liu, Jing Shi, and Meiying Jian

Department of Civil Engineering, Tsinghua University, Beijing 100084, China

Correspondence should be addressed to Jing Shi; jingshi@tsinghua.edu.cn

Academic Editor: Zhi-Chun Li

One important function of Intelligent Transportation System (ITS) applied in tourist cities is to improve visitors' mobility by releasing real-time transportation information and then shifting tourists from individual vehicles to intelligent public transit. The objective of this research is to quantify visitors' psychological and behavioral responses to tourism-related ITS. Designed with a Mixed Ranked Logit Model (MRLM) with random coefficients that was capable of evaluating potential effects from information uncertainty and other relevant factors on tourists' transport choices, an on-site and a subsequent web-based stated preference survey were conducted in a representative tourist city (Chengde, China). Simulated maximum-likelihood procedure was used to estimate random coefficients. Results indicate that tourists generally perceive longer travel time and longer wait time if real-time information is not available. ITS information is able to reduce tourists' perceived uncertainty and stimulating transport modal shifts. This novel MRLM contributes a new derivation model to logit model family and for the first time proposes an applicable methodology to assess useful features of ITS for tourists.

1. Introduction

Tourist cities in China have witnessed an impressive growth in tourism industry over the last decade. In particular, this growth should considerably owe to the surge of the "Free Independent Tourists" (FIT), a widely used terminology referring to an individual or a small group who travel and vacation with a self-booked and deeply customized itinerary. Unlike traditional guided tourists, FIT can choose the transport to destination and local transport by themselves with the help of various tourism information and/or tips from open information sources (e.g., websites). Many of them prefer to drive individual vehicles (IV) when traveling. In this research, they are called FIT who travel by individual vehicles (FIT by IV). FIT by IV include both the drivers and accompanying tourists in private cars or rental cars that are not changed during their whole journey. Alongside the FIT by IV, other FITs may use noncar modes (like plane, train, and coach) to arrive at their tourism destinations and do not rent a car there. Instead, they would prefer to depend on local public transport services. Thus, these FIT without personal or rental cars are herein distinguished as Free Independent Tourists who travel by public transport (FIT by PT). The developments of FIT by IV and FIT by PT have both flourished China's tourism industry.

Coinciding with substantial economic benefits from the influx of tourists, traffic congestion and environmental pressure have posed serious threat to popular tourist destinations at the same time, which would adversely affect the sustainability of tourism. Traditional solutions, such as building more infrastructures to meet the growing number of FIT by IV, cannot fundamentally mitigate those problems and may even deteriorate the situation as the personal vehicle use may be stimulated. Another solution is to improve public transport services and shift more tourists from personal cars to green public transport modes to alleviate roadway congestion and reduce vehicle emissions [1–3]. The latter provides more smart and feasible strategies that could enhance the transport system in an efficient and ecofriendly way without weakening the mobility of tourists.

However, according to the report issued by Tourism Research Center of Chinese Academy of Social Sciences, in the year of 2015, approximately 58.5% of the total Chinese domestic tourists arrived in tourism destinations by car

[4]. Meanwhile, China National Tourism Administration revealed that about 75% of the total domestic tourists were FIT in the same year [5]. These two data indicate a high proportion of FIT by IV and a relatively low ridership of public transit among FIT in China. The possible reasons are the following: (1) FIT by IV in China may regard a driving tour as a new fashion with comfort and convenience; (2) tourist city may lack adequate public transport services for visitors; (3) unlike daily commutes, travel routes in a tourist city are often irregular; thus tourists may find difficulties using public transit in an unfamiliar city if local transportation information is limited. For example, when traveling for tourism activities, 36% of the visitors in the city of Changchun would take buses, while this ratio among local residents was 61% [6]. With empirical evidence from four tourism attractions in Germany, Gronau and Kagermeier [7] identified several key factors for successful leisure and tourism public transport, including transparency and quality of the public transport service, restricting use of private car, and market promotion strategies (e.g., combined tickets for both tourism attraction and public transport). In particular, the first factor regarding "transparency" was in line with the dimension of simplicity proposed by Frima et al. [8] and "ease of use" which was identified by Schiefelbusch et al. [9]. From this enlightening perspective, more information on public transport services, especially in the context of tourism, should be offered to improve public awareness and acceptance. Kenyon and Lyons [10] also suggested that presentation of integrated multimodal traveler information for a journey in response to a single enquiry could influence drivers' perceived utility of noncar modes, which can contribute to a modal shift with overcoming the underlying habitual and psychological barriers.

With widespread adoption of varied sensors, or the concept of "Internet of Things", large-scale transport data can be collected by Intelligent Transport System (ITS) to analyze the operation state of a transport system. Meanwhile, information and communication technologies are accelerating the development of ITS applications. From early real-time arrival screens at transit stops to recent apps in mobile devices that release transit information [11], ITS tools are capable of delivering integrated multimodal traveler information to tourists in more cost-effective and user-friendly ways. Therefore, as a potential approach, ITS enables the integration of travel services (e.g., intelligent buses, public bicycles) and information services (e.g., traffic condition, parking fee, and waiting time for public transit), which can better balance travel demands generated from tourists and transport supply from tourist cities.

Previous studies have rarely investigated the impacts of ITS on tourism transport users, while urban commuters' responses to real-time transport information have been extensively observed and analyzed. Transit information users are observed to have a number of important responses, including shortenings in both perceived and actual waiting time with the help of detailed arrival information [12–15], increased satisfaction with transit services [16–18], and increased ridership and transfers [15, 17–19]. Long-period observation and corresponding regression analysis can help to investigate the change in commuters' ridership with influence from various factors, and then most of these findings were obtained from longitudinal data [16, 19, 20]. However, adopting this method in the study of tourists seems to make little sense. Different from daily commutes, visitors' routes in a tourist city are not frequently repeated. As a result, setting up a control group without instant access to transit information to explore the benefits of ITS applications in a natural experiment could be hard to implement. Moreover, several studies found no significant changes in commuters' ridership with the introduction of transit information [20] or merely "modest" [19], because daily commute trips are relatively fixed and inelastic. Therefore, different levels of familiarity with the local transport system and different travel patterns between tourists and commuters should be noted. A novel methodology is required when probing into the impacts of ITS applications on tourists.

To go beyond the limitations of longitudinal data, stated preference (SP) survey is employed in this research to collect tourists' potential choices on transportation mode by imitating possible scenarios that they may face in the future. As a growing number of tourists currently inquire transportation information via mobile devices [11] and this trend seems to expand, SP survey is designed for tourists to complete it on mobile devices to cater to their habits in tourism information inquiry. SP survey that is conducted on mobile devices can help to record the responses of these potential ITS information users. At the same time, data is collected in a more efficient way especially compared to traditional paper-based questionnaire. Correspondingly, with the data from mobile devices, a novel Mixed Ranked Logit Model (MRLM) will be gradually developed to estimate how ITS-supported travel services influence transport modes choices of the tourists who are accessible to ITS information, which is the objective of this research. The rest of this paper is organized as follows. In Section 2, the formulation of MRLM is introduced and the reason for developing such a model in this study is explained. After that, in Section 3 of a case study, data collecting strategy and the sample characteristics are described. Then estimation results of MRLM are presented in Section 4 with adoption of simulation-based estimation procedure. In Section 5, potential changes brought by ITS on tourism-related transportation mode choices are predicted. Finally, the main conclusions are highlighted in the last section with some suggested directions for future research.

2. Methodology

SP survey is a widely used methodology in analyses of travel choices. Commonly, three responses are designed in SP surveys: (1) the single most-preferred alternative, (2) ranking of alternatives, and (3) rating of alternatives [21]. Among all these dimensions of response, single choice causes the least burden on respondents, but has the least preference information and consequently requires a relatively larger sample size. Rating of alternatives, on the contrary, is the most demanding to respondents, but provides the richest preference data. However, the ranking of alternatives asks respondents to decide the whole or partial preferred order from the option set, then it can generate more data than just a single choice. Meanwhile, despite getting less information than quantitative evaluations of preference in the rating approach, the ranking of alternatives can help to avoid heavy burden imposed on respondents, and controls the possible influences on response reliability that may be caused by respondents' impatience. Therefore, the ranking method will be employed in this study as it can balance the sample size and data quality in SP surveys compared to other two methods.

Beggs et al. [22] and Chapman and Staelin [23] first applied the ranked data in several logit model formulations. They exploited the information of ranked data with an "explosion procedure" of the entire ranking. The entire ranking of a set of J alternatives can be "exploded" into a sequence of $J-1$ independent choosing steps. The first choosing step is to select the perceived best alternative from all the J available alternatives. The second step is to select the second-best one from the remaining $J-1$ alternatives excluding the chosen one in the preceding step, and so on until the $(J-1)$th step in which an alternative is to be assigned rank $J-1$ from the last two remaining alternatives that have not been ranked before. Although ranking of all the available alternatives can provide much preference information, Hausman and Ruud [24] argued that respondents might give less consideration to decide lower ranks compared to higher ranks as the variance in the ranking of less preferred alternatives increased. In order to reduce the negative effects of lower ranks, this paper focuses on the ranking of best two alternatives that most respondents pay attention to [25], and at the same time respondents will not find much difficulty in ranking. It should be noted that early logit models for analyzing ranked data were mostly derived from standard multinomial logit model (MNL) [22, 23]. However, standard MNL exhibits independence from irrelevant alternatives (IIA), which implies proportional substitution across alternatives. Moreover, the assumption that each choosing step in different ranking level was an independent observation might not reflect the reality [26, 27].

As a model that contains random coefficients, mixed logit can relax the restrictive independence assumption, then it obviates limitations of standard MNL by capturing the correlations across different ranking levels and allowing for unrestricted substitution patterns [28]. Layton [25] and Calfee et al. [29] reported in their surveys that using random-coefficient models to analyze ranked preference data could improve the precision of estimation. Another ability of random coefficients is to describe some particular distributions in logit models. Since tourists who arrive at an unfamiliar city are more likely to have limited transport information than traveling in the city they live in, it is of particular concern in this research to explain the role of information in tourists' travel decision-making. The uncertainty of information is more likely to be described as a random distribution rather than a fixed value, and then mixed logit model with random coefficients can help to achieve this goal. Detailed description of random coefficients in this study will be presented in the next sections.

Combining the advantages of ranked logit model and mixed logit model, MRLM can be developed. MRLM not only avoids IIA, but also provides opportunities to build random coefficients as distributions rather than constants. Therefore, this model is especially suitable for this research to describe tourists' perception of uncertainty in information.

2.1. Model Specification. The impacts of ITS are quantified from tourists' ranked choices on transport modes. In every choosing step of the entire ranking under a particular scenario, tourists' perceived utility of each alternative transport mode is assumed to be

$$U_{ni} = \alpha_{ni} X_{ni} + \omega_{ni} + \varepsilon_{ni}. \tag{1}$$

In this equation, U_{ni} is the perceived utility of alternative mode i for tourist n, X_{ni} is a vector of parameters (observed attributes of alternative i and observed characteristic of tourist n), and α_{ni} is a vector of coefficients associated with parameters X_{ni}. In this study, some coefficients in α_{ni} are assumed to be random ones. If coefficient α_{ni} is a constant, its probability density function (PDF) can be deemed as $f(\alpha_i = \text{constant}) = 1$ across individuals. If coefficient α_{ni} is a random coefficient, α_{ni} is assumed to follow a particular PDF $f(\alpha_i)$ across individuals. The function of $f(\alpha_i)$ commonly needs other parameters θ_i to describe. For example, the PDF of a lognormal distribution $f(\alpha_i) = \exp(-(\ln \alpha_i - \mu)^2/2\sigma^2)/(\alpha_i \sigma \sqrt{2\pi})$ is decided by two parameters μ and σ, then these two parameters are collectively included in θ_i. Therefore, the total utility does not depend on the values of α_i, but the functions of θ_i actually.

ω_{ni} is an error term to depict the individual-specific unobserved factors of alternative i. For each alternative i, ω_{ni} is assumed to be independently and normally distributed across individuals.

With the introduction of random error term ω_{ni}, the restrictive independence assumption in early ranked logit models can be relaxed. Meanwhile, the normal distribution $f(\omega_{ni})$ of alternative i and the PDF $f(\alpha_{ni})$ are both assumed to be independently and identically distributed (IID) across individuals, that is $f(\omega_i)$ and $f(\alpha_i)$.

ε_{ni} is an IID extreme-value type one error term across the individuals and the alternatives. This term is assumed to be Gumbel-distributed.

Respondents participating in this research are asked to choose and rank two alternatives in each scenario. If a respondent thinks alternative $i^{(1)}$ is the best one and alternative $i^{(2)}$ is the second-best one from the alternatives set J, the conditional probability (conditional on α_{ni} and ω_{ni}, $i \in J$) of the respondent' ranking will be the product of the probability of each choosing step:

$$\begin{aligned} P_n &\left[i^{(1)}, i^{(2)} \mid \alpha_{ni}, \omega_{ni}, i \in J \right] \\ &= \frac{\exp\left(\alpha_{ni^{(1)}} X_{ni^{(1)}} + \omega_{ni^{(1)}}\right)}{\sum_{m \in J} \exp\left(\alpha_{nm} X_{nm} + \omega_{nm}\right)} \\ &\quad \cdot \frac{\exp\left(\alpha_{ni^{(2)}} X_{ni^{(2)}} + \omega_{ni^{(2)}}\right)}{\sum_{k \in J, k \neq i^{(1)}} \exp\left(\alpha_{nk} X_{nk} + \omega_{nk}\right)}. \end{aligned} \tag{2}$$

Finally, the unconditional probability of the ranking $[i^{(1)}, i^{(2)}]$ is calculated as

Table 1: ITS-related variables in Mixed Ranked Logit Model for different types of tourists.

ITS-supported travel services	Variable description	Variable type	Logit model of FIT by PT			Logit model of FIT by IV			
			Bicycling	Bus	Taxi	Bicycling	Bus	Taxi	Car
Road traffic condition	In-vehicle travel time	Continuous		$X_1^{(P)}$	$X_1^{(P)}$		—[a]	$X_1^{(I)}$	$X_1^{(I)}$
	Uncertainty in in-vehicle travel time	Binary		$X_2^{(P)}$	$X_2^{(P)}$		—[a]	$X_2^{(I)}$	$X_2^{(I)}$
Parking management	Vacant parking space	Binary							$X_3^{(I)}$
	Uncertainty in vacant parking space	Binary							$X_4^{(I)}$
	Parking fee	Continuous							$X_5^{(I)}$
Provision of public transport and level of services	Vacant public bicycle	Binary	$X_3^{(P)}$			—[b]			
	Uncertainty in vacant public bicycle	Binary	$X_4^{(P)}$			—[b]			
	Vacant seats on bus	Binary		$X_5^{(P)}$			$X_6^{(I)}$		
	Uncertainty in vacant seats on bus	Binary		$X_6^{(P)}$			$X_7^{(I)}$		
	Waiting time for bus	Continuous		$X_7^{(P)}$			$X_8^{(I)}$		
	Uncertainty in waiting time for bus	Binary		$X_8^{(P)}$			$X_9^{(I)}$		
	Waiting time for taxi	Continuous			$X_9^{(P)}$				
	Uncertainty in waiting time for taxi	Binary			$X_{10}^{(P)}$				

Superscript $^{(P)}$ or $^{(I)}$ denotes that corresponding variable is in the logit model of FIT by PT or the model of FIT by IV, respectively; in the SP surveys for FIT by IV: [a]Dedicated bus lanes are in operation; thus the travel time will not change. [b]Public bicycles are always available.

$$P_n\left[i^{(1)}, i^{(2)}\right] = \int_{\omega_{ni}, i \in J} \int_{\alpha_{ni}, i \in J} \left\{P_n\left[i^{(1)}, i^{(2)} \mid \alpha_{ni}, \omega_{ni}, i \in J\right]\right\} \cdot f\left(\alpha_{ni} \mid \theta_i\right) d\alpha_{ni} f\left(\omega_{ni}\right) d\omega_{ni}. \quad (3)$$

2.2. ITS-Related Variables in MRLM. Referring to previous studies [7–9], three noteworthy categories of factors are considered as ITS-related variables in MRLM: (1) provision of real-time information, including information about traffic condition, vacant parking space near scenic spots, and waiting time for a bus/taxi, (2) individual vehicle restriction policies, including adjusting parking fee and operating park-and-ride hubs, and (3) quality of public transport services, including in-vehicle environment and travel time. Tourists' perceived uncertainty of certain transport information is considered in corresponding transport modes. Those variables are categorized, designed and described in Table 1 for the logit model of FIT by PT and the logit model of FIT by IV.

The variables regarding "uncertainty" indicate tourists' inaccessibility to particular transport information. The values of these variables are equal to 1 if corresponding information is not available for tourists, otherwise 0. Therefore, these variables are binary. Consequently, the respondents of the SP surveys can belong to both experimental group and control group under the same hypothetical scenario according to different accessibility to different ITS-supported information.

The design idea of estimating the change due to ITS is explained with example of variables in logit model of FIT by PT. Variable $X_1^{(P)}$ denotes the travel time in motorized vehicles, and the values of those variables depend on real-time road condition information released by ITS technology. If information of traffic condition has been released, tourists will take the total decreased utility of $\alpha_1^{(P)} X_1^{(P)}$ into consideration as the impacts caused by in-vehicle travel time on their modal choices. Under this situation, $X_2^{(P)}$ will be correspondingly set as 0 to indicate that uncertainty does not exist in tourists' awareness of traffic condition. Conversely, if tourists do not receive real-time information of traffic condition, the value of $X_1^{(P)}$ will be set as 0, and $X_2^{(P)}$ will be set as 1. The total decreased utilities as the impacts of in-vehicle travel time then all generate from the lack of real-time information and will be $\alpha_2^{(P)} X_2^{(P)}$. Since the purpose is to mine tourists' psychological responses when they lack relevant information, it is necessary to set up a reference standard to quantify the impacts. To this end, a covariate T_{free} is introduced into the calculation of $\alpha_2^{(P)}$ by defining $\alpha_2^{(P)} = \beta_2^{(P)} T_{\text{free}}$, where T_{free} denotes the travel time in smooth traffic and takes the role of reference standard and $\beta_2^{(P)}$ is the associated coefficient of that covariate. By then, the total decreased utility caused by the lack of real-time information is $\beta_2^{(P)} T_{\text{free}} X_2^{(P)}$. At the same time, tourists are assumed to perceive a length of travel time even though they have no idea about the real traffic condition. This assumption is just like the situation that tourists also receive a released real-time information although it is perceived by themselves, which inspires us to calculate the total decreased utility as the product of tourists' perceived travel time (denoted by T_{per}) with coefficient $\alpha_1^{(P)}$ (since $\alpha_1^{(P)}$ means decreased utility caused by every extra minute). In this way, $\alpha_1^{(P)} \times T_{\text{per}} = \beta_2^{(P)} T_{\text{free}} \times X_2^{(P)}$. Because $X_2^{(P)}$ equals 1 when real-time information is not provided, $\beta_2^{(P)}/\alpha_1^{(P)} (= T_{\text{per}}/T_{\text{free}})$ will denote the multiple of travel time in smooth traffic perceived by tourists, which is used to reflect tourists' psychological perception.

$X_3^{(P)}$ is equal to 1 if a vacant public bicycle remains in the nearest bicycle docking station and the respondent also recognizes that, otherwise, the value is 0. $X_4^{(P)}$ concerns the uncertainty of whether a vacant bicycle is available. If the respondent is not accessible to the information of public bicycle system, the value of $X_4^{(P)}$ is 1, otherwise 0. The values of $X_5^{(P)}$ (vacant seats on bus) and $X_6^{(P)}$ (uncertainty in vacant seats on bus), $X_3^{(I)}$ (vacant parking space) and $X_4^{(I)}$. (uncertainty in vacant parking space) are similar to the definitions of $X_3^{(P)}$ and $X_4^{(P)}$. $X_7^{(P)}$ and $X_9^{(P)}$ denote waiting time for bus and taxi, respectively, while $X_8^{(P)}$ and $X_{10}^{(P)}$ indicate corresponding uncertainty in waiting time. The method to estimate the perceived waiting time by tourists is similar to the idea dealing with coefficient of $X_2^{(P)}$. The only difference is that $X_8^{(P)}$ and $X_{10}^{(P)}$ do not contain any covariates. Since $\alpha_7^{(P)} \times T_{\text{per\_wait}} = \alpha_8^{(P)} \times 1$, the value of $\alpha_8^{(P)}/\alpha_7^{(P)}$ (for bus) or $\alpha_{10}^{(P)}/\alpha_9^{(P)}$ (for taxi) indicates the perceived waiting time by tourists.

As for the PDFs of coefficients, coefficients of variables regarding "uncertainty" in waiting time for bus/taxi are assumed to be lognormal distributions so that the signs keep unchanged in the entire domain. This is because waiting time, as well as monetary cost, only causes negative effect in utility. Besides, other coefficients are assumed to be fixed values because hypotheses of random coefficients do not pass the tests.

This study focuses on tourists' transport choices (especially modal shifts of FIT by IV) after they have already arrived in the tourism destination. Without a car, FIT by PT can only rely on local public transport services when traveling. As a result, FIT by PT and FIT by IV will face different set of available transport choices. Moreover, different type of tourists have different sensitivity on different factors. For instance, FIT by IV are mostly sensitive about parking fee and available parking place, while FIT by PT do not take these factors into consideration. Furthermore, if bus service does not have advantages in the travel time when the traffic is congested, FIT by IV can hardly leave from their cars. So in the SP survey for FIT by IV, it is assumed that dedicated bus lanes are in operation and public bicycles are always available so that other factors that may lead to modal shifts can be considered particularly. Therefore, in order to analyze the factors that may be most concerned by different type of tourists, the variables of some transportation mode in each logit model are partially different. It should be pointed out that the set of variables will not influence the generalized application of MRLM. The set of variables in each MRLM depends on its corresponding research objective.

2.3. Mixed Ranked Logit Model in Tourism. From the practical perspective, Mixed Ranked Logit Model that developed in this research also has other advantages in the response analysis among tourists. (1) Various hypothesized scenarios designed in SP surveys help to acquire more data about tourists' response than natural experiments. (2) Cognitive bias in tourists' self-reported perceived waiting time and other feelings can be avoided, since MRLM mines these psychological responses of ITS from tourists' final choices on transport modes. (3) Future possible scenarios can be simulated with MRLM to provide better guidance for the applications of ITS in other tourist cities.

3. Data Collection

In order to test the effectiveness of MRLM, SP surveys are conducted in Chengde, a representative tourist city in China. Chengde is approximately 180 km away from China's capital Beijing. Its UNESCO world cultural heritage site, Mountain Resort and Outlying Temples, attracts a great number of domestic and overseas tourists. With a rapid increasing number of FIT by IV, Chengde is suffering terrible traffic problem these years, especially during peak tourist seasons [30]. To date, Chengde only operates a traditional bus system with about 30 bus lines and real-time transport information system is not available. Aiming to become a smart tourist city, Chengde has launched the plan of Slow Traffic System, which aims to apply ITS technologies to restrict the use of individual vehicles by providing public bicycles and other public transit services for city dwellers and tourists.

To obtain adequate suitable data from real tourists, an on-site questionnaire survey and a subsequent web-based SP survey were carried out within the same sample of tourists who have traveled in Chengde. Respondents stated the attributes of their tourist group during on-site survey and later were classified by their transport modes into two types, namely, FIT by IV and FIT by PT. Different type of tourists would face different scenarios to choose their preferred transport modes in the web-based SP surveys.

3.1. On-Site Questionnaire Survey in Chengde. On-site survey was conducted in Chengde's tourism attractions during China's National Day Holidays (Oct. 1~7) in 2014, when high-density tourist flow occurs within a short period. The respondents are randomly selected. A total of 706 valid on-site questionnaires were collected, including 521 FIT by IV and 185 FIT by PT, which indicates a high proportion of FIT by IV in total FIT, reaching about three-quarters (73.8%). The surveyed ratio of FIT by IV to FIT by PT is close to the ratio released by Chengde tourism administration.

Table 2 shows a high percentage of FIT by IV who take children (<10 years old) or older people (>60 years old) in their journeys, reaching to approximately 60%. In contrast, 87.03% of FIT by PT groups contain neither children nor older people. The distributions of group size also differ between FIT by IV and FIT by PT. As demonstrated in Figure 1, the group size of FIT by IV is more likely to be 3 to 5, maybe for the reason that a driving trip is always a family trip and a private car usually have five seats maximum. Lower than FIT

Table 2: Distribution of children and older people in tourist group.

Participants in a tourist group	FIT by IV	FIT by PT
Presence of children, absence of older people	24.26%	2.16%
Presence of older people, absence of children	13.73%	7.03%
Presence of both children and older people	20.59%	3.78%
Absence of both children and older people	41.42%	87.03%

Child: less than 10 years old; older people: over 60 years old.

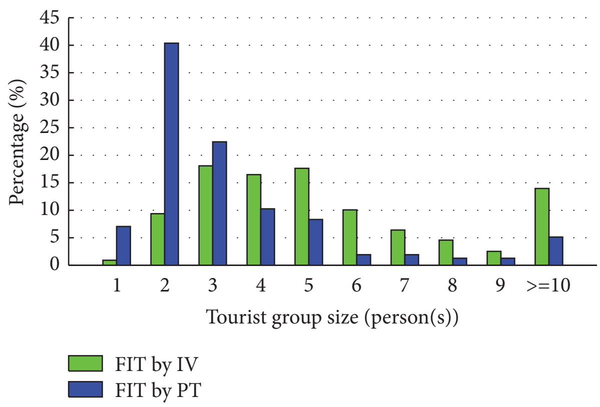


Figure 1: Distribution of tourist group size.

Figure 2: Interface of SP survey on tablet PC and mobile phone (translated from Chinese).

by IV, the group size of FIT by PT is more likely to be 2 to 3. This phenomenon may be explained with the fact that young couples and college students used to travel without personal cars.

3.2. Web-Based Stated Preference Survey. It can be inferred from on-site survey that FIT by IV and FIT by PT may have distinct attitudes and preference. Therefore, in the next step, different versions of SP surveys were designed for different types of tourists. SP surveys were carried out on the web. We asked tourists for their e-mail addresses during the on-site survey and sent the link of the web-based questionnaire via e-mail when they finished their journey. Tourists were encouraged to complete the survey on their mobile phones and tablet PCs by scanning the QR code in the e-mail, since 67.77% of FITs by IV and 70.65% of FITs by PT stated in the on-site survey that they would search tourism travel information via smartphones. The process of answering questionnaire on mobile devices is in accordance with their habits in tourism information inquiry. For the same reason, Chengde's tourism ITS plans to distribute information on mobile devices.

Each version of SP survey consists of 48 scenarios with 5 factors, and each factor has 4 levels. As was discussed before, factors were partially different in each version as listed in Table 3. All the scenarios were orthogonally designed with these factors. In an orthogonal experimental design, orthogonality can guarantee that the effect of one factor or interaction can be estimated separately from the effect of any other factor or interaction in the model. By this means, all pairs of factorial levels appear together the same number of times for each pair of factors, ensuring the coverage of more potential scenario. Consequently, the impacts of each factor can be estimated more efficiently and effectively. An orthogonal experimental design with five 4-level factors could generate 16 scenarios in total. Furthermore, the impacts of travel distance will be considered and three different lengths of travel distance are added to the surveys: 1.5 km, 3 km, and 4.5 km. Since each distance contained 16 scenarios, the total number of scenarios in a survey would be 48.

Figure 2 shows the interface of the web-based SP survey with examples of hypothesized scenarios. All the transport information was integrated in a card-shaped picture, which implements the suggestion proposed by Kenyon and Lyons [10] so that respondents can soon generate a full and clear image of all the transport modes. In each card-shaped picture, information of a particular transport mode will be described in its corresponding column. Specially, the shading of traffic condition information will be drawn as grey, green, yellow or red when the traffic condition is uncertain (unknown), smooth, slow-moving or congested, respectively. The predicted travel time in each traffic condition is in accordance with Table 4. The values in this table are set according to real travel experiences along with the calculation of average waiting time in signed road intersections. As urban area in Chengde has always been developing around the Mountain Resort and along the ancient imperial road, most tourism

TABLE 3: Description of variables and levels in hypothetical scenarios.

Version of SP survey	Factors in surveys	Level 1	Level 2	Level 3	Level 4
For both surveys	Road traffic condition	Uncertain	Smooth	Slow-moving	Congested
	Vacant seats on bus	Uncertain	Enough	Only a few	None left
	Waiting time for bus (minute)	Uncertain	5	10	15
For FIT by IV	Vacant parking space	Uncertain	Enough	Only a few	None left
	Parking fee (RMB yuan)	10	20	30	40
For FIT by PT	Vacant public bicycle	Uncertain	Enough	Only a few	None left
	Waiting time for taxi (minute)	Uncertain	5	10	15

TABLE 4: Assumed in-vehicle travel time in different traffic condition (minute).

Road traffic	In-vehicle travel time (1.5 km)		In-vehicle travel time (3.0 km)		In-vehicle travel time (4.5 km)	
Condition	Bus	Taxi/car	Bus	Taxi/car	Bus	Taxi/car
Smooth	10	6	14	10	20	13
Slow-moving	15	11	24	18	32	25
Congested	25	20	38	30	48	40

attractions and other buildings are mixed in the downtown area. Therefore, the roads linking tourism attractions are mostly narrow urban streets and the average velocity is relatively low in Chengde downtown area.

Respondents need imagine that they are still traveling in Chengde with the same group members and choose the best two transport modes under each scenario. In order to simulate the real-world scenes experienced by tourists, all the hypothesized travel routes are designed according to real popular routes in Chengde. Furthermore, to avoid respondents' impatience and ensure the effectiveness of the answers, of the whole 48 scenarios, only 12 will be randomly selected and faced by the respondents.

4. Estimation Results

Simulated maximum-likelihood procedure is employed here to estimate the parameters θ_i and then the coefficients α_i and ω_i. The probabilities in (3) can be approximated through simulation with different realization of α_i and ω_i for any given values of θ_i [28]. (1) Draw a value for each α_i and ω_i from their PDF $f(\alpha_i \mid \theta_i)$ and $f(\omega_i)$ $(i \in J)$, respectively, and mark them as α_i^r and ω_i^r with the superscript $r = 1$ to denote the first draw. (2) Use this draw to calculate the ranking probability $P_n[i^{(1)}, i^{(2)} \mid \alpha_i^r(\theta_i), \omega_i^r, i \in J]$. (3) Repeat the first and the second steps many times and obtain the average. Then the average is the simulated probability:

$$\breve{P}_n\left[i^{(1)}, i^{(2)}\right] = \frac{1}{R}\sum_{r=1}^{R} P_n\left[i^{(1)}, i^{(2)} \mid \alpha_i^r\left(\theta_i\right), \omega_i^r, i \in J\right]. \quad (4)$$

In this equation, R is the number of draws of ω_i and α_i $(i \in J)$. For each ω_i and α_i, we use $R = 2000$ draws of Halton sequence [31]. Then $\breve{P}_n[i^{(1)}, i^{(2)}]$ is an unbiased estimator of $P_n[i^{(1)}, i^{(2)}]$ by construction. The simulated probabilities are inserted into the log-likelihood function to give simulated log likelihood, and the estimator is the value of θ_i and ω_i that maximizes this function:

$$\mathrm{SLL}^* = \sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{k=1}^{K}\delta_{kt}\ln\breve{P}_n\left[i^{(1)}, i^{(2)}\right], \quad (5)$$

where $\delta_{kt} = 1$ if the best two alternatives are chosen and ranked as $[i^{(1)}, i^{(2)}]$ under the scenario t, and otherwise $\delta_{kt} = 0$; $K = A_J^2$ is the number of all possible permutations of ranking 2 items from J alternatives; J is the number of all the scenarios that respondent n will face in a survey, $T = 12$ in both models for FIT by PT and FIT by IV, and N is the number of total respondents in each model.

Tourists may cancel their planned routes if the traffic is terrible. Therefore, "canceling" are added as an alternative as other transport modes. The alternative "canceling" does not include any attributes, and its utility is assumed to be zero (for theoretical and empirical modeling considerations). Consequently, J equals 5 for FIT by PT, that is walking, bicycling, bus, taxi and canceling, and J equals 6 for FIT by IV, that is walking, bicycling, bus, taxi, car and canceling. Of the total 185 FIT by PT and 521 FIT by IV who answered the on-site questionnaires, 121 FIT by PT and 235 FIT by IV completed the web-based SP surveys. Every respondent needs to make decisions in 12 randomly selected scenarios, and every scenario can be seen as a single observation. Therefore the total number of observations in MRLM of FIT by PT and MRLM of FIT by IV will be 1452 (= 121 × 12) and 2820 (= 235 × 12), respectively.

All the variables in each MRLM, including ITS-related variables, attribute variables of tourist groups and other public variables (e.g., travel distance) are estimated together. The estimation are calculated with programming language of MATLAB. The results are listed in Table 5.

As discussed in Section 2, the ranking response can provide more preference data than answers of single choice from the same scale of respondents. Therefore, ranked data can improve the precision of estimation with Mixed Ranked Logit Model compared to mixed logit model. Furthermore, based on respondent ranking of alternatives, ranked logit model with fixed coefficients (i.e., standard logit model that restricts ω_{ni} and α_{ni} in the MRLM to be constants across individuals) is also estimated. The log-likelihood values at convergence in the Mixed Ranked Logit Models are −1538.64 (model for FIT by PT) and −3548.04 (model for FIT by IV),

Table 5: Estimated coefficients in Mixed Ranked Logit Model.

Variables	Distribution hypothesis	FIT by PT: Estimated parameter θ: μ (Std. error)	FIT by PT: Estimated parameter θ: σ (Std. error)	FIT by PT: Estimated coefficient (Std. error)	FIT by IV: Estimated parameter θ: μ (Std. error)	FIT by IV: Estimated parameter θ: σ (Std. error)	FIT by IV: Estimated coefficient (Std. error)
In-vehicle travel time	Constant			−0.0514 (0.0058)***			−0.0690 (0.0064)***
Uncertainty in in-vehicle travel time (covariate β)	Constant			−0.0765[a] (0.0126)***			−0.1671[a] (0.0169)***
Vacant parking space	Constant						0.2534 (0.1566)
Uncertainty in vacant parking space	Constant						0.2037 (0.1755)
Parking fee	Constant						−0.0225 (0.0058)***
Vacant seats on bus	Constant			0.1171 (0.1228)			0.5094 (0.1271)***
Uncertainty in vacant seats on bus	Constant			0.0401 (0.1246)			0.1565 (0.1372)
Vacant bicycles	Constant			0.6075 (0.1212)***			
Uncertainty in vacant bicycles	Constant			−0.2957 (0.1425)**			
Waiting time for bus	Constant			−0.0346 (0.0140)**			−0.0305 (0.0157)*
Waiting time for taxi	Constant			−0.0364 (0.0179)**			
Uncertainty in waiting time for bus	Lognormal	−0.5963 (0.3567)*	0.6236 (0.4214)	−0.6700[b] [0.4644]	−0.9289 (0.6657)	0.7121 (0.5176)	−0.5084[b] [0.4157]
Uncertainty in waiting time for taxi	Lognormal	−0.5662 (0.5193)	0.7308 (0.3983)*	−0.7688[b] [0.6484]			
Walking	Normal	5.6425 (0.3369)***	1.7960 (0.1551)***	5.6425[c] [1.7960]	4.8023 (0.5471)***	2.2538 (0.2157)***	4.8023[c] [2.2538]
Bicycling	Normal	3.8593 (0.3684)***	1.3643 (0.1138)***	3.8593[c] [1.3643]	3.8529 (0.4123)***	1.8593 (0.1696)***	3.8529[c] [1.8593]
Bus	Normal	5.3887 (0.3925)***	1.0621 (0.1236)***	5.3887[c] [1.0621]	3.2231 (0.4064)***	0.9798 (0.1440)***	3.2231[c] [0.9798]
Taxi	Normal	3.0769 (0.3805)***	1.5384 (0.1371)***	3.0769[c] [1.5384]	2.9592 (0.4003)***	1.5093 (0.1227)***	2.9592[c] [1.5093]
Car	Normal				4.8422 (0.4809)***	2.1400 (0.1751)***	4.8422[c] [2.1400]
Estimated coefficients of public variables for each transport mode are shown in Table 7							
Summary statistics			MRLM of FIT by PT			MRLM of FIT by IV	
Number of observations ($N \times T$)			1452			2820	
Likelihood ratio index ρ			0.3565			0.3118	
SLL at convergence SLL($\breve{\theta}$)			−1538.64			−3548.04	

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$; [a]estimated coefficient of covariate β; [b]the estimated coefficient α is the mean of the lognormal distribution, which is calculated by $e^{\mu+\sigma^2/2}$, and the value in the bracket is the standard deviation, which is calculated by $e^{\mu+\sigma^2/2}\sqrt{e^{\sigma^2}-1}$; [c]the estimated coefficient α is the same as μ, and the estimated standard deviation in the bracket is the same as σ.

respectively. The corresponding values in the ranked standard logit models are −1487.05 (model for FIT by PT) and −3459.13 (model for FIT by IV), respectively. Then the log-likelihood values clearly indicate that MRLM can fit the response data better than the ranked standard logit models.

In the estimation result of MRLM, the standard deviations of the individual-specific error terms are discovered to be statistically significant, which reveals that individual-specific unobserved factors that influence personal utility for each transportation mode strongly exist.

Table 6: Characteristics of equivalent perceived waiting time.

Equivalent perceived waiting time	Waiting time for bus (FIT by IV)	Waiting time for bus (FIT by PT)	Waiting time for taxi (FIT by PT)
Mode (most probable value, $e^{\mu-\sigma^2}$)	7.80 min	10.79 min	9.46 min
Mean (average value, $e^{\mu+\sigma^2/2}$)	16.67 min	19.36 min	21.08 min

Note: equivalent perceived waiting time is lognormally distributed as $\ln N(\mu, \sigma^2)$.

In the following sections, estimation results of ITS-related variables and other variables will be described in detail.

4.1. ITS-Related Variables. The lack of real-time information heightens tourists' perceived risk of longer travel time. Comparing the values between −0.0514 and −0.0765, 0.0765/0.0514 = 1.49, which indicates that the impact of lacking traffic condition information is equal to the situation that FIT by PT are told to spend about 1.49 times of travel time that is needed in smooth traffic. As for FIT by IV, their perceived travel time will be 2.42 (0.1671/0.069) times of that. This result maybe a consequence of frequent traffic congestion experienced by FIT by IV. They concern road traffic condition much more than FIT by PT and then overestimate their risk at a higher level.

In terms of the variables relevant with parking management, receiving the information that vacant parking space is available will stimulate the use of private car (0.2534). However, raising the parking fee can significantly reduce the utility of driving a private car, with the coefficient of −0.0225 (per RMB yuan), which provides an evidence to employ economic method to adjust the usage of cars.

Vacant seats on bus, or more comfortable on-bus environment, will attract more FIT by IV to take a bus (0.5094). But in the model of FIT by PT, vacant seats do not affect the utility of bus significantly.

Longer waiting time means less utility for tourists, and the impacts of waiting time on bus and taxi are almost the same (−0.0346 and −0.0364 for FIT by PT, resp.). As mentioned above, the coefficients of variables "uncertainty in waiting time for bus/taxi" are assumed to be lognormal distributions. Therefore, two parameters, mean and standard deviation, should be estimated to describe tourists' perceived waiting time which is also lognormal distributed. For FIT by PT, the values of mean of the final estimated coefficients α are −0.67 and −0.7688 for bus and taxi, respectively. If dividing the values of mean by the coefficients of released waiting time, that is −0.0346 and −0.0364, respectively, the equivalent expected waiting time in average can be calculated. −0.67/−0.0346 = 19.36, which indicates the average equivalent waiting time for bus is expected to be 19.36 minutes, is higher than the maximum released waiting time in SP survey-15 minutes. This phenomenon also occurs in the transport mode of taxi. If tourists do not know the waiting time for taxi, their equivalent perceived waiting time will be average 21.08 (−0.7688/−0.0364) minutes. Similarly, we can also calculate equivalent waiting time for bus perceived by FIT by IV in average: 16.67 (−0.5084/−0.0305) minutes. The main characteristics of equivalent waiting time for bus and taxi of FIT by PT and FIT by IV are listed in Table 6. "Mode" in this table indicates the most probable equivalent perceived waiting time, while "mean" is the average equivalent perceived waiting time. They are both crucial indices that reflect the impact of uncertainty on the utility of bus and taxi.

The notion "equivalent" used here tries to explain that the impacts of lacking information about waiting time may consist of two parts. One part is the real waiting time while the other part is tourists' perception of the risk. People tend to overestimate the risk under uncertainty; thus the value of information is present in this case: more information can reduce the uncertainty and then ensure the tourists' decisions to be more reasonable.

4.2. Travel Distance and Attributes of Tourist Group. A range of tourists' attributes are tested in model and finds that "gender" and "individual annual income" did not show significant effects on tourists' choice in transport modes. The reason may be that the final decision is more related with attributes of tourist group rather than individual attributes. Table 7 shows the coefficients of public variables for each mode.

The utility of walking will decrease dramatically with the increase in distance, followed by that of bicycling and bus. Generally, the utility of each mode will decrease if a child or an older person exists in a group of FIT by PT, except taxi. This result reveals that taxi will attracts more children and older people for its convenience and safety. As for FIT by IV, they are unwilling to let children or older persons in their tourist group to walk or bicycle, though this effect is not so significant among FIT by PT. The size of tourist group affects the mode choices in a special way. The more persons a tourist group of FIT by PT has, the higher utility of choosing a taxi would be, since it is more worthwhile to share the taxi fare. FITs by IV have shown a similar habit, but larger scale tourist group of them prefer to drive their personal cars and it is uneasy to make a change.

5. Transport Modal Choice Prediction

Once each coefficient of Mixed Ranked Logit Model has been estimated, choice probability of transport mode i can be calculated with (6) by applying simulation method:

$$\check{P}(i) = \sum_n w_n \times \frac{1}{R} \sum_{r=1}^{R} \frac{\exp\left(\alpha_{ni}^r X_{ni} + \omega_{ni}^r\right)}{\sum_{m \in J} \exp\left(\alpha_{nm}^r X_{nm} + \omega_{nm}^r\right)}, \quad (6)$$

where w_n is the percentage of a particular type of tourists n.

To fully assess the effect of application of ITS on tourists' transport choices, we predict the changes brought from

Table 7: Estimated coefficients of public variables for each transport mode.

Variables	Estimated coefficients for FIT by PT (standard error)				Estimated coefficients for FIT by IV (standard error)				
	Walking	Bicycling	Bus	Taxi	Walking	Bicycling	Bus	Taxi	Car
Distance (km)	−1.5094 (0.1907)***	−0.6389 (0.1804)***	−0.0804 (0.0431)*	−0.1575 (0.1853)	−1.4618 (0.0933)***	−0.8038 (0.0613)***	−0.1441 (0.0497)***	−0.1161 (0.1911)	−0.0821 (0.1783)
Presence of children (<10)	−0.3705 (0.2985)	−0.3115 (0.2643)	−0.4118 (0.3145)	0.2604 (0.1653)	−0.5578 (0.2303)**	−0.3181 (0.1547)**	−0.2078 (0.1787)	−0.1160 (0.1007)	−0.1276 (0.2689)
Presence of older people (>60)	−0.3489 (0.3171)	−0.3791 (0.3217)	−0.2874 (0.2195)	0.2249 (0.1915)	−0.8619 (0.3560)**	−0.4772 (0.2322)**	−0.1502 (0.1289)	−0.2191 (0.1903)	−0.1213 (0.1747)
Group size[a]	Over 2 persons				3~5 persons				
	0.3412 (0.5074)	0.2891 (0.1980)	0.2929 (0.1906)	0.5306 (0.3317)	0.5378 (0.4039)	0.2974 (0.2576)	0.8249 (0.4248)*	1.0137 (0.4738)**	1.3124 (0.5262)**
					Over 5 persons				
					0.3267 (0.2466)	0.2653 (0.2371)	0.4445 (0.3588)	1.1042 (0.6367)*	1.5204 (0.6999)**

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$; [a]1~2 person(s) is the base category of group size.

ITS-supported information services and dedicated bus lanes separately among FITs by PT, and further investigate these influences on FITs by IV. In the simulated scenarios, all the transport information are assumed to be accessible if ITS-supported tourism information system has been applied, such as waiting time for a bus/taxi (ranging from 0 to 15 minutes in a uniform distribution), whether a rental bicycle is available nearby (the probability a tourist can rent a bicycle is assumed to be 80%), road traffic condition, and so forth. Two-thirds of tourists will use these real-time information, just as what the surveys discovered. In addition, the probability of having a seat on bus is set as 80%.

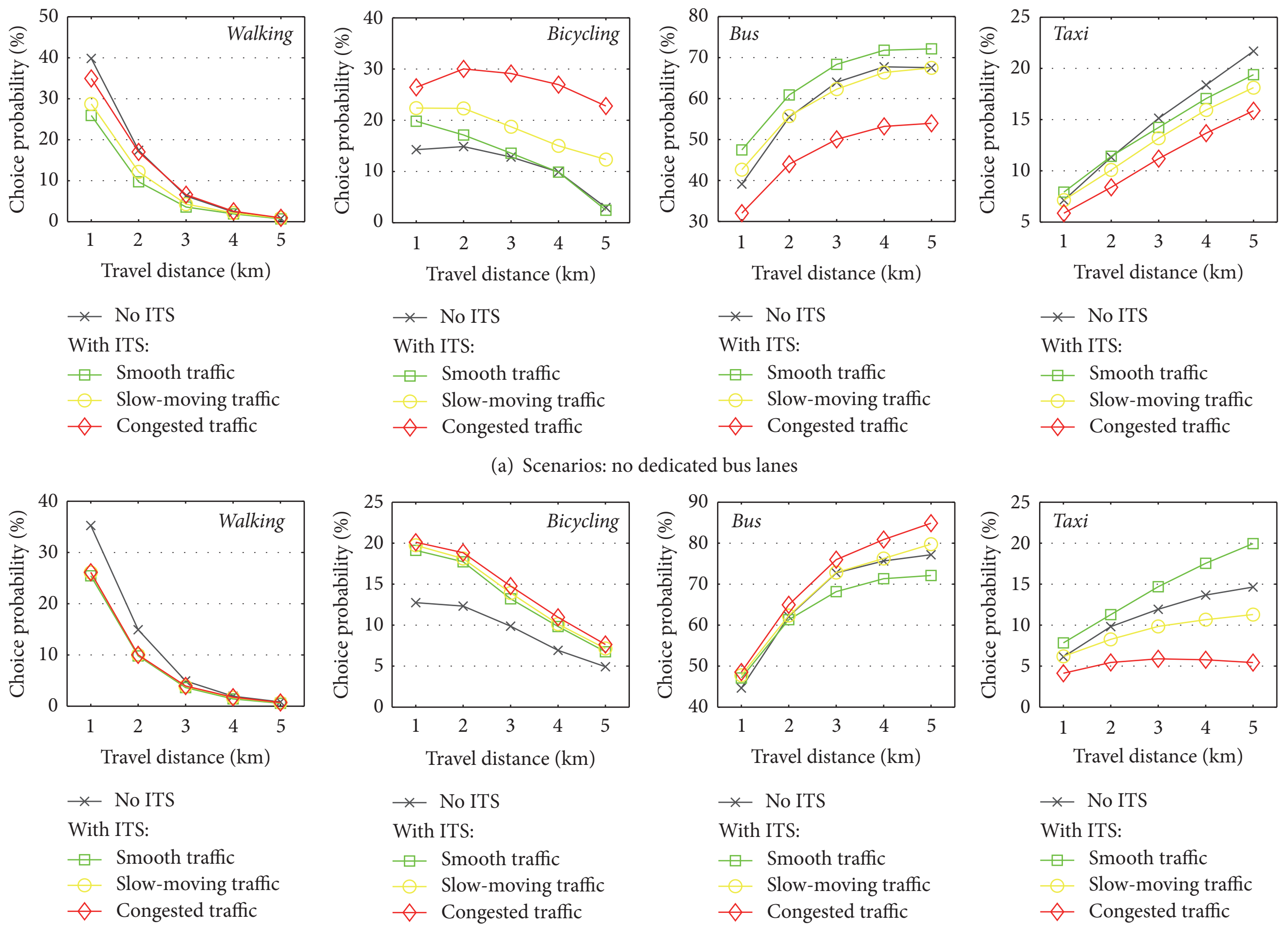
5.1. Transport Modal Choice Prediction for FIT by PT. The prediction results of FIT by PT are illustrated in Figure 3, some changes in transport modal choices occur with the application of ITS-supported information services. Without dedicated bus lanes, the choice probability of bicycling meets a significantly growth after providing multimodal real-time information but the share of bus declines when traffic is congested. In contrast, operation of dedicated bus lanes can keep the ridership of bus at a high level, especially in longer travel distances and under congested traffic condition. Meanwhile, the participant rate of bicycling stays at a more reasonable level within the distance of 3 km. These results indicate that the integration of improved public transport and corresponding public transport information services can lead to a more efficient tourism transport market.

5.2. Transport Modal Choice Prediction for FIT by IV. FITs by IV generally strongly rely on their own private car and can hardly transfer to other alternative transport modes if high-quality public transport service is lacking. Therefore, in the following scenarios, both public bicycle system and bus lanes network are assumed to have entered services. The predicted choice probabilities of FIT by IV are demonstrated in Figure 4. The most significant modal shift appears in the share of bus: when roads are congested, buses that move on the dedicated lanes will attract a considerable portion of FITs by IV.

6. Conclusions and Discussions

Previous studies have addressed the support of ITS-supported information for the travel of urban dwellers, however, which primarily focused on daily commutes. To go beyond from these attentions to the ITS technological evolution and fill a research gap for the flourishing leisure and tourism industry, this paper investigates both psychological and behavior responses from tourists to ITS-supported travel services.

It is noted that one technical challenge lies in that the randomicity of the tourists' travel routes has substantial difficulties in recording the behavioral change of a certain tourist before and after using ITS-supported services. This particular constraint requires a distinct research approach than longitudinal observation that is commonly used for analyzing commuters' behaviors. Therefore, this study has developed a MRLM to overcome the hurdle in collecting panel data. With the special design in random coefficients, our MRLM not only obviates independence from irrelevant alternatives, but also successfully estimate the influence of uncertainty in information. In the corresponding SP surveys, tourists' perception of waiting time and other evaluations are mined from their final choices on transport modes rather than self-reported psychological responses, which could improve the reliability of the survey results by avoiding cognitive bias towards the questions among tourists. Moreover, the impact due to a particular type of information can be isolated as respondents can be regarded as experimental group and control group of different tested variables in the same scenario. One first major finding obtained from MRLM reveals that tourists would perceive the risk of traffic congestion and overestimate their waiting time without real-time traffic information. For instance, when accurate information of traffic condition is inaccessible, FITs by PT expect the



(a) Scenarios: no dedicated bus lanes

(b) Scenarios: dedicated bus lanes in operation

FIGURE 3: Predicted choice probabilities of FIT by PT.

in-vehicle travel time to be 1.49 times of that in smooth traffic, and for FITs by IV the estimate is 2.42 times.

Second, this study also indicates that ITS-supported travel services can encourage modal shifts from individual vehicles to public transit among tourists. Since tourists generally overestimate the out-vehicle waiting time and in-vehicle travel time, provision of both high-quality public transit services and ITS information are able to reduce tourists' perceived risk caused by information uncertainty. Meanwhile, for FIT by PT, awareness of good in-vehicle environment can increase their willingness to use public transit. These observations also suggest some other improvements (e.g., real-time adjusting departure intervals of buses) can be made for the public transit to attract more FIT by IV in the future. Some characteristics of particular tourist groups also influence the choices in transport modes, which suggests that differentiated market policies should be adopted to promote public transit.

This study focuses tourists' travel mode choices after they have already arrived in tourism destinations. In the SP survey, FIT by PT and FIT by IV would face different transportation alternatives in SP surveys. However, limited by the mathematical characteristics of logit model that is applied in SP surveys, the difference in transportation alternatives and resulting different associated variables of alternatives will cause difficulty in comparing the estimated coefficients of different models. Therefore, the different perceptions towards some travel services between FIT by PT and FIT by IV are hard to be observed. Moreover, responses of travelers without using mobile devices were not collected in this web-based SP survey. This is a limitation caused by survey methodology, because travelers without smartphones may receive some single-modal transport information from other sources (e.g., real-time information board at bus stop). Additional survey is needed if we want to further study the responses of travelers without mobile devices.

Another limitation is the results obtained from SP surveys, as they might not be in line with the reality after actual implementation of ITS, and requires further validation by other practical experiments. One solution is to track actual responses of tourists by launching revealed preference (RP) surveys when planned functions of ITS applications are realized step by step in the future. Meanwhile, a joint SP/RP mixed logit model can be developed to reduce the

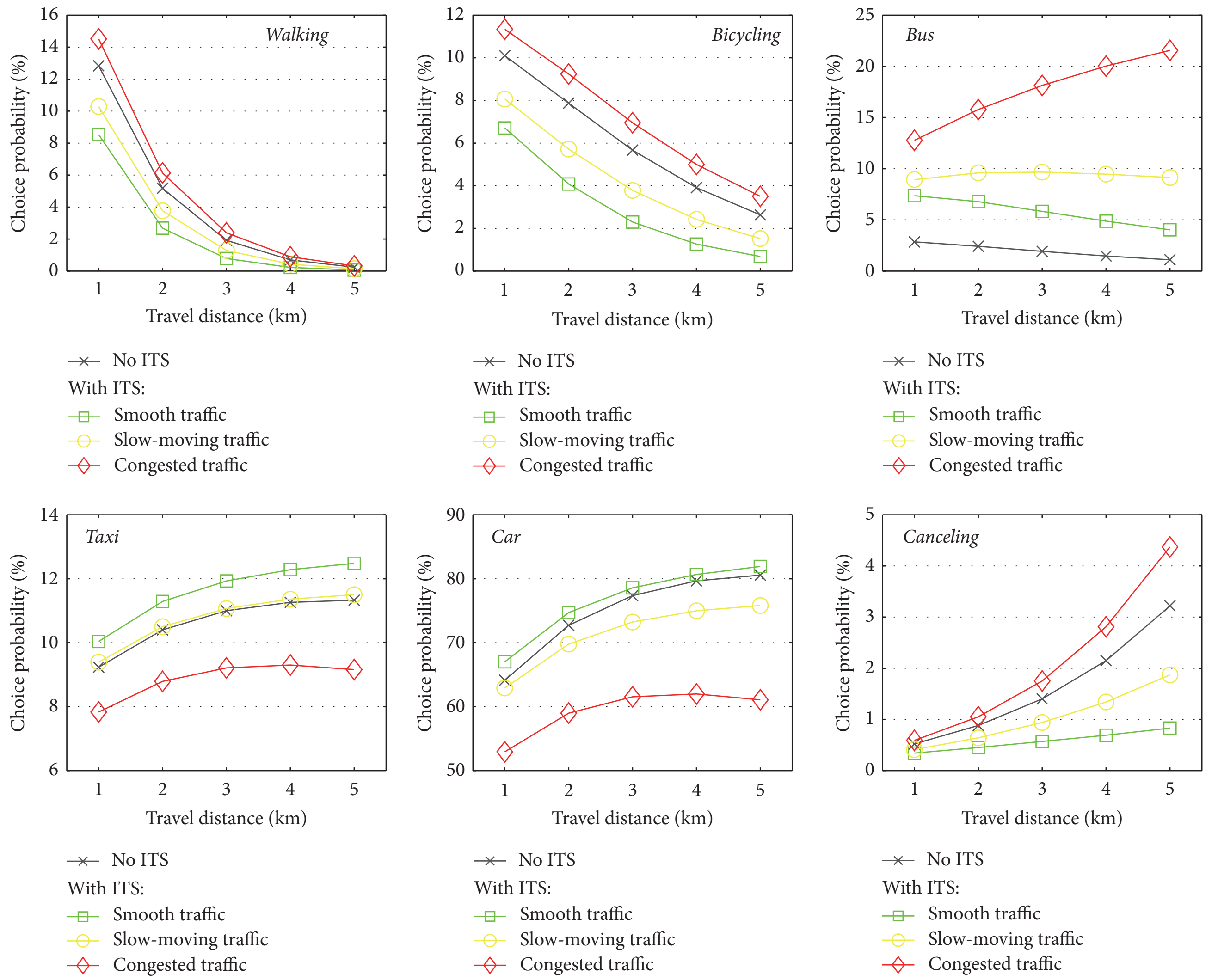


Figure 4: Predicted choice probabilities of FIT by IV.

potential bias in coefficients estimation and modal choice prediction caused by pure SP data (see, e.g., Hensher et al. [32]). Besides, when simulating the process of releasing transportation information in this research, each hypothetical scenario only contains a single route. While in fact, tourists' travel behavior may be associated with their entire activity chains [33]. This limitation remains to be addressed in future work. Notwithstanding, our study can lead better insights into visitors' potential respondents to ITS-supported travel services in a tourist city, and proposals a novel methodology to assess the impacts of ITS on tourists and their tourism destinations.

Acknowledgments

The authors would like to appreciate the financial support from TOSHIBA Corporation and the National Natural Science Foundation of China (no. 51578319). They also extend their appreciation to Chengde Travel & Tourism Administration and Chengde Municipal Institute of City Planning & Design for the assistance in their surveys. They thank two anonymous referees for their valuable comments on the improvement of this paper.

References

[1] H. D. Regnerus, R. Beunen, and C. F. Jaarsma, "Recreational traffic management: the relations between research and implementation," *Transport Policy*, vol. 14, no. 3, pp. 258–267, 2007.

[2] J. Guiver, L. Lumsdon, and R. Weston, "Traffic reduction at visitor attractions: the case of Hadrian's Wall," *Journal of Transport Geography*, vol. 16, no. 2, pp. 142–150, 2008.

[3] V. Filimonau, "Carbon management in tourism: mitigating the impacts on climate change," *Tourism Management*, vol. 46, pp. 62–63, 2015.

[4] Tourism Research Center of Chinese Academy of Social Sciences, *China's Self-Driving Tourism Development Analysis and Forecast (2015-2016)*, China Travel & Tourism Press, Beijing, China, 2016.

[5] China National Tourism Administration, "Report on China's Tourism Development," 2016 http://www.cnta.gov.cn/xxfb/hydt/201605/t20160516_770884.shtml.

[6] F. Liu, *Research on the development of public transportation of tourism in Changchun based on tourists' experience [M.S. thesis]*, Northeast Normal University, Changchun, China, 2014.

[7] W. Gronau and A. Kagermeier, "Key factors for successful leisure and tourism public transport provision," *Journal of Transport Geography*, vol. 15, no. 2, pp. 127–135, 2007.

[8] M. Frima, B. Edvardsson, and T. Gaerling, "Perceived service quality attributes in public transport: inferences from complaints and negative critical incidents," *Journal of Public Transportation*, vol. 2, no. 1, pp. 67–89, 1998.

[9] M. Schiefelbusch, A. Jain, T. Schäfer, and D. Müller, "Transport and tourism: roadmap to integrated planning developing and assessing integrated travel chains," *Journal of Transport Geography*, vol. 15, no. 2, pp. 94–103, 2007.

[10] S. Kenyon and G. Lyons, "The value of integrated multimodal traveller information and its potential contribution to modal change," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 6, no. 1, pp. 1–21, 2003.

[11] N. Gammer, T. Cherrett, and C. Gutteridge, "Disseminating real-time bus arrival information via QR code tagged bus stops: a case study of user take-up and reaction in Southampton, UK," *Journal of Transport Geography*, vol. 34, pp. 254–261, 2014.

[12] R. Mishalani, M. McCord, and J. Wirtz, "Passenger wait time perceptions at bus stops: empirical results and impact on evaluating real—time bus arrival information," *Journal of Public Transportation*, vol. 9, no. 2, pp. 89–106, 2006.

[13] K. Dziekan and K. Kottenhoff, "Dynamic at-stop real-time information displays for public transport: effects on customers," *Transportation Research Part A: Policy and Practice*, vol. 41, no. 6, pp. 489–501, 2007.

[14] K. E. Watkins, B. Ferris, A. Borning, G. S. Rutherford, and D. Layton, "Where is my bus? Impact of mobile real-time information on the perceived and actual wait time of transit riders," *Transportation Research Part A: Policy and Practice*, vol. 45, no. 8, pp. 839–848, 2011.

[15] C. Brakewood, G. S. Macfarlane, and K. Watkins, "The impact of real-time information on bus ridership in New York City," *Transportation Research Part C: Emerging Technologies*, vol. 53, pp. 59–75, 2015.

[16] F. Zhang, Q. Shen, and K. J. Clifton, "Examination of traveler responses to real-time information about bus arrivals using panel data," *Transportation Research Record*, no. 2082, pp. 107–115, 2008.

[17] B. Ferris, K. Watkins, and A. Borning, "OneBusAway: results from providing real-time arrival information for public transit," in *Proceedings of the 28th Annual CHI Conference on Human Factors in Computing Systems (CHI '10)*, pp. 1807–1816, Atlanta, Ga, USA, April 2010.

[18] A. Gooze, K. Watkins, and A. Borning, "Benefits of real-time transit information and impacts of data accuracy on rider experience," *Transportation Research Record*, no. 2351, pp. 95–103, 2013.

[19] L. Tang and P. V. Thakuriah, "Ridership effects of real-time bus information system: a case study in the city of Chicago," *Transportation Research Part C: Emerging Technologies*, vol. 22, pp. 146–161, 2012.

[20] C. Brakewood, S. Barbeau, and K. Watkins, "An experiment evaluating the impacts of real-time transit information on bus riders in Tampa, Florida," *Transportation Research Part A: Policy and Practice*, vol. 69, pp. 409–422, 2014.

[21] D. A. Hensher, "Stated preference analysis of travel choices: the state of practice," *Transportation*, vol. 21, no. 2, pp. 107–133, 1994.

[22] S. Beggs, S. Cardell, and J. Hausman, "Assessing the potential demand for electric cars," *Journal of Econometrics*, vol. 17, no. 1, pp. 1–19, 1981.

[23] R. G. Chapman and R. Staelin, "Exploiting rank ordered choice set data within the stochastic utility model," *Journal of Marketing Research*, vol. 19, no. 3, pp. 288–301, 1982.

[24] J. A. Hausman and P. A. Ruud, "Specifying and testing econometric models for rank-ordered data," *Journal of Econometrics*, vol. 34, no. 1-2, pp. 83–104, 1987.

[25] D. F. Layton, "Random coefficient models for stated preference surveys," *Journal of Environmental Economics and Management*, vol. 40, no. 1, pp. 21–36, 2000.

[26] M. Ben-Akiva, T. Morikawa, and F. Shiroishi, "Analysis of the reliability of preference ranking data," *Journal of Business Research*, vol. 24, no. 2, pp. 149–164, 1992.

[27] K. Sælensminde, "Inconsistent choices in stated choice data; use of the logit scaling approach to handle resulting variance increases," *Transportation*, vol. 28, no. 3, pp. 269–296, 2001.

[28] K. E. Train, *Discrete Choice Methods with Simulation*, Cambridge University Press, London, UK, 2nd edition, 2009.

[29] J. Calfee, C. Winston, and R. Stempski, "Econometric issues in estimating consumer preferences from stated preference data: a case study of the value of automobile travel time," *Review of Economics and Statistics*, vol. 83, no. 4, pp. 699–707, 2001.

[30] Chengde Travel & Tourism Administration, *Chengde's Tourism is Entering the Era of Free Independent Tourists*, 2013, http://www.cdtour.gov.cn/zhijian/cdlyzjw/xydt/363.html.

[31] C. R. Bhat, "Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model," *Transportation Research Part B: Methodological*, vol. 35, no. 7, pp. 677–693, 2001.

[32] D. A. Hensher, J. M. Rose, and W. H. Greene, "Combining RP and SP data: biases in using the nested logit "trick"—contrasts with flexible mixed logit incorporating panel and scale effects," *Journal of Transport Geography*, vol. 16, no. 2, pp. 126–133, 2008.

[33] A. Lew and B. McKercher, "Modeling tourist movements: a local destination analysis," *Annals of Tourism Research*, vol. 33, no. 2, pp. 403–423, 2006.

Design of a Cultural Tourism Passenger Flow Prediction Model in the Yangtze River Delta Based on Regression Analysis

Jian Xu

School of Economics and Management, Hefei University, Hefei 230601, Anhui, China

Correspondence should be addressed to Jian Xu; xujian@hfuu.edu.cn

Academic Editor: Shah Nazir

Cultural tourism has gained much attention in the last decade and has promoted the preservation of a variety of tangible and intangible assets of culture. In order to accurately predict the cultural tourism passenger flow in the Yangtze River Delta and improve its economic benefits, this paper designs the prediction model of cultural tourism passenger flow in the Yangtze River Delta based on regression analysis. Taking the competitiveness of passenger flow as the core, this paper selects 28 indexes from four aspects of cultural tourism brand resources, cultural tourism support and protection, and urban tourism market income to build the evaluation index system of influencing factors of passenger flow. The principal component analysis method is used to simplify many related factors into a few uncorrelated factors to eliminate the multicollinearity caused by too many dependent variables; on this basis, the principal component regression model is constructed, and the determination coefficient is used to test the model fitting. Taking 15 cultural tourism cities in the Yangtze River Delta as the research object, the results show that the designed model has a good fitting degree, and the average error is only 0.41%, which can meet the needs of the prediction of cultural tourism passenger flow in the Yangtze River Delta. After the application of the prediction model, the foreign exchange earning amount of each cultural tourism city can be increased by more than 12%. The study has revealed good results.

1. Introduction

The improvement of living standards makes the tourism industry gradually rise. Generally speaking, tourism cities have developed commercial economy, and there are many tourist attractions and classic snacks, which have attracted a large number of tourists. The development and prosperity of cultural tourism in the near future will be due to the dual role of industry consciousness and system design [1]. Promoting the consciousness, self-confidence, and self-improvement of regional cultural construction is not only the main tone of current cultural construction but also the fundamental driving force for the rapid development of cultural tourism [2]. Different cities may have the same mode and path of tourism development, but the connotation of urban cultural tourism is unique and competitive. The tourism development mode dominated by cultural tourism often has more vitality and durability.

The Yangtze River Delta region is an important growth pole of China's economic development and also one of the regions with the highest level of urbanization in China [3, 4]. As an important strategic base of China's tourism industry, the Yangtze River Delta region is not only the main force and vanguard of China's tourism development but also the most successful region of China's cross regional tourism cooperation and the region with great potential to become a world-class tourism destination. In the process of implementing the national strategy of regional integration in the Yangtze River Delta, cultural tourism needs to play a greater role. The Yangtze River delta needs a complete image in the process of representing China to participate in global competition and cooperation. Cultural tourism is the advance force and core backbone of regional international image publicity. At present, regional integration has become a consensus of China's future development. Through resource sharing, complementary advantages, and market interaction, we can break the regional, spatial, and institutional barriers, create barrier-free tourism areas, and realize the all-round development of social economy. At the same time, the Yangtze River Delta region is rich in tourism resources, with a complete range of natural and cultural landscapes. It is an important concentrated distribution area of tourism resources in China and has the inherent advantages of developing regional tourism integration. Therefore, it has attracted a large number of tourists to travel all the time.

Tourist flow refers to the number of people and the flow pattern of tourists from the source to the destination [5]. According to the different travel time, it can be divided into four types: daily passenger flow, monthly passenger flow, quarterly passenger flow, and annual passenger flow. Tourist flow is an important indicator of the development level of tourism industry, an important part of tourism planning by national or regional tourism authorities, an effective guarantee to improve the quality of tourism products, and an important basis for the development of tourism resources and the construction of reception facilities such as hotels. Accurate prediction of tourist flow is related to the successful operation of an international and regional tourism project, which will directly affect the scientific decision making of the tourism project and is an important part of urban tourism development planning [6, 7].

There are many factors that affect the cultural tourism passenger flow in the Yangtze River Delta, among which economy, politics, education level, resources, and transportation will have an impact on it [8]. Taking transportation as an example, the places with convenient transportation with the Yangtze River Delta will have a large tourist flow to the Yangtze River Delta. In addition, psychological factors and personal preferences will also have an impact on the tourist flow of the Yangtze River Delta. Therefore, many factors must be fully considered in the process of analyzing the tourist flow in the Yangtze River Delta.

Scientific and reasonable prediction of tourist flow is of great guiding significance for the efficient use of tourism resources and local economic development. It can also help the government to formulate tourism development planning and tourism emergency plan, improve the quality and level of tourism service, and then, improve the tourist satisfaction and sense of experience [9]. The interdisciplinary research on "tourism passenger flow" and "passenger flow prediction" has developed rapidly. It has gone deep into systems science, computer science and technology, and other disciplines and has derived a number of interdisciplinary themes.

At present, the methods of predicting tourist flow are mainly divided into quantitative prediction and qualitative prediction. Qualitative prediction is generally based on qualitative analysis combined with empirical judgment [10], with low prediction accuracy. Quantitative prediction is mainly to establish a quantitative prediction model through mathematical methods. It is a widely used prediction method with high prediction accuracy. There are three kinds of models for tourism passenger flow prediction by using quantitative prediction methods, namely, the permeability model, gravity model, and GM (1, 1) model. The penetration model is an intuitive model with strong subjectivity [11], which is based on the interviewees' willingness to visit as the main data, combined with the population base and carrying coefficient to make an intuitive inference of the passenger flow of tourist attractions. However, the model has its own regional bias due to the willingness to visit, so the permeability model is only used for interval estimation of the willingness to visit and the passenger flow. The gravity model is a commonly used international method to predict passenger flow under normal conditions [12]. However, in the actual prediction process, this method only considers a single or a small number of factors affecting passenger volume, which is lack of comprehensiveness, resulting in biased passenger volume prediction. In the practical application of the GM (1, 1) model, too many factors may be considered, resulting in multicollinearity among factors [13], and the regression coefficient cannot pass the significance test, and even the sign of some regression coefficients is inconsistent with the actual economic significance.

In view of the problems existing in the abovementioned models, a prediction model of cultural tourism passenger flow in the Yangtze River Delta based on regression analysis is designed on the basis of the gravity model and permeability model. In the parameter selection of the multiple regression model, the model comprehensively considers many factors that affect the cultural tourism passenger flow in the Yangtze River Delta, so compared with the gravity model and GM (1, 1) model, it considers the factors affecting passenger flow more comprehensively, overcomes the influence of multicollinearity, and improves the accuracy of passenger volume prediction.

The organization of the paper is as follows: the materials and methods of the paper are presented in Section 2 with details. Section 3 of this paper shows the results of the paper. The conclusion of this paper is given in Section 4.

2. Materials and Methods

The following sections briefly present the materials and methods used in this study.

2.1. Index Selection. Compared with other traditional types of industries or industries, cultural tourism has its own unique characteristics in the development connotation, development context, and social effects [14, 15]. First of all, in the world, the government vigorously promotes the cultural tourism industry, making it an increasingly important way to cultivate national cultural identity and national identity. Cultural tourism shoulders the sacred function of nurturing citizens. Secondly, the system structure and operation mechanism of cultural tourism embody the characteristics of the binary compound system of culture and tourism integration. The correct understanding of the concept of culture and the reasonable mining of tourism elements in the cultural field are the primary problems of cultural tourism transformation from resources to products, which provides an important reference for identifying the elements of cultural tourism competitiveness. Thirdly, cultural tourism reflects the close interaction between the host and the guest. Cultural tourism space is the combination of geographical space, cultural space, and social space of cultural tourism destination. It is a field of interaction between the host and the guest with clear geographical space, which also defines a clear spatial scope for clarifying the system of cultural tourism competitiveness. Therefore, the development of theory and practice shows that the competitiveness of urban cultural tourism passenger flow in the Yangtze River Delta can carry out more scientific index screening and system construction.

Based on the theory of cultural tourism passenger flow competitiveness in the Yangtze River Delta region, the following three principles are considered and followed in the design process of the specific factor index:

(i) The combination of comprehensiveness and operability: this paper selects and analyzes the connotation of cultural tourism development, the structure of tourist flow competitiveness, and the influencing factors in the Yangtze River Delta, so as to present the overall situation of tourist flow competitiveness of cultural tourism in the Yangtze River Delta as far as possible. At the same time, the feasibility and reliability of the index data sources are fully considered. According to the different grades of the Yangtze River Delta, in the design of the index system, besides quoting some necessary total indexes reflecting the scale effect, the strength indexes are considered as much as possible.

(ii) Systematic and hierarchical: by fully combining the competitiveness of cultural tourism passenger flow in the Yangtze River Delta with the regional reality, the differences and emphases of cultural and tourism industry in different regions of the Yangtze River Delta are very different. In the evaluation, we should consider the problems from the actual situation of different regions as far as possible, so that the differences can be reflected in the evaluation index system. The hierarchy is reflected in that the competitiveness of cultural tourism passenger flow is mainly composed of three-level indexes. Besides the general goal and target decomposition level, it is explained and evaluated by element indexes [16].

(iii) The evaluation objectives and methods are consistent. This paper focuses on the theme of the evaluation of cultural tourism passenger flow competitiveness in the Yangtze River Delta region, designs the indexes around the center as far as possible, and always defines the direction and goal of the evaluation system. In terms of evaluation methods, after solving the basic problems through qualitative methods, we use scientific mathematical methods to calculate and screen the indexes needed by the research according to the quantitative relationship of each indicator and then carry out comprehensive measurement and evaluation.

Based on the abovementioned competitiveness theory and index system design principles, this paper decomposes the evaluation objectives of cultural tourism passenger flow competitiveness in the Yangtze River Delta region into four subobjectives: cultural tourism brand resources, cultural performance and creativity, cultural tourism support and protection, and urban tourism market income. At the same time, 28 quantitative indexes are designed to form the Yangtze River Delta region's cultural tourism. The evaluation index system of influencing factors of passenger flow is shown in Table 1.

2.2. Construction of the Prediction Model. Before the establishment of the model, the passenger volume is y, the sample size is n, and the observation value of the i-th index is $x_i\,(i = 1, 2, \ldots, 27)$.

2.2.1. Multicollinearity Diagnosis. The dependent variable y and the independent variable $x_1, x_2, \ldots, x_{27}$ are used to establish the regression model. SPSS statistical software is used to select the variables by the backward regression method, and p variables are set to enter the regression analysis model, which are $x_{s1}, x_{s2}, \ldots, x_{sp}\,(p \leq n)$, called initialization variables.

From the regression results to find the variance expansion factor (VIF), if $VIF_i \geq 10$, it means that there is a serious multicollinearity between the independent variables.

Multicollinearity has a great influence on the regression coefficient [17], which can be processed by principal component analysis.

2.2.2. Principal Component Analysis.

(1) Data standardization

In order to eliminate the influence of different orders of magnitude and dimensions, it is necessary to standardize the original data. The standardized formula is as follows:

$$x_{ij}' = \frac{x_{ij} - \overline{x}_j}{s_j}, \tag{1}$$

where x_{ij}' is the standardized data and $\overline{x}_j$ and $(j = 1, 2, \ldots, n)$ represent the mean value and standard deviation of the j-th index sample, respectively.

(2) Calculation of the correlation coefficient matrix: after processing the original data, the standardized data matrix $(x_{ij}')_{p\times n}$ is obtained, and the corresponding correlation coefficient matrix is calculated.

$$R = \left(r_{ij}\right)_{n\times n}. \tag{2}$$

In formula (2), R is a symmetric matrix of order n.

(3) The eigenvalues and eigenvectors of correlation coefficient matrix R are calculated.

The eigenvalue $\lambda_i\,(i = 1, 2, \ldots, n)$ of R and its corresponding eigenvector $u_i\,(i = 1, 2, \ldots, n)$ are solved,

Table 1: Evaluation index system of influencing factors of cultural tourism passenger flow in the Yangtze River Delta.

General objective	Target decomposition	Element description
Influencing factors of cultural tourism passenger flow in the Yangtze River Delta	Brand resources of cultural tourism	Number of world cultural heritages
		Number/item of abnormal intangible culture at the national level or above
		Number of national key cultural relics protection units
		Number of famous historical and cultural cities in China
		Number of famous historical and cultural towns and villages in China
		Number of 4A tourist attractions
		Number of red tourist attractions
		Number of local folk festivals and special activities
	Cultural performance and creative tourism	Creative tourism performance project
		Number of provincial and national cultural industry demonstration bases
		Number of art performing groups
		Number of cultural centers, art performance centers, and mass art centers
		Number of libraries
		Number of open museums
	Cultural tourism support	Number of accommodation facilities
		Number of beds
		Number of rooms
		Number of travel agencies
		Number of food blocks
		Number of passengers sent by civil aviation
		Scale of highway passenger transport
		The scale of railway passengers
	Regional tourism market revenue	Number of inbound tourists
		Domestic tourists
		Foreign exchange earned by tourism
		Total tourism revenue
		The proportion of total tourism revenue in total economic output

and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, where λ_i is the variance of the main component F_i, and the greater the variance is, the greater the contribution to the total variance is [18].

(4) The contribution rate is calculated, and the principal component is determined.

Formula (3) is defined as the contribution rate of main component F_i.

$$e_i = \frac{\lambda_i}{\sum_{i=1}^{n} \lambda_i} \times 100\%. \tag{3}$$

In formula (3), $\sum_{i=1}^{n} e_i$ is the contribution rate of cumulative variance. Generally, m principal components with $\sum_{i=1}^{n} e_i$ greater than or equal to 85% are selected for comprehensive analysis. Therefore, n factors are reduced to m principal components, and the main factors are selected [19].

2.2.3. Principal Component Regression Model. Multiple regression analysis is conducted between F_r and dependent variable Y (standardized value of y) to obtain the standardized regression equation.

$$Y = \sum_{r=1}^{m} B_r F_r, \tag{4}$$

In formula (4), B_r is the standardized partial regression coefficient of the r-th principal component F_r.

By synthesizing formula (2), (3), and (4), the standardized regression equation is obtained.

$$Y = \sum_{j=1}^{p} b_j' X_{s_j}. \tag{5}$$

In formula (5), b_j' is the j-th standardized partial regression coefficient of the standardized regression equation.

Through the abovementioned analysis, we can get the principal component regression model as follows:

$$y = \sum_{j=1}^{p} b_j x_{s_j} + b_0, \tag{6}$$

where

$$b_j = b_j' \sqrt{\frac{L_{yy}}{L_{x_{ij}x_{ij}}}}, \tag{7}$$

$$b_0 = \overline{y} - \sum_{j=1}^{p} b_j \overline{x}_{s_j}. \tag{8}$$

In the abovementioned formula, b_j is the j-th partial regression coefficient of the general linear regression equation; L_{yy} is the sum of squares of deviation of y; $L_{x_{ij}}$ is the sum of squares of deviation of x_{sj}; $\overline{y}$ is the mean value of y; $\overline{x}_{s_j}$ is the mean value of x_{sj}; and b_0 is the constant of the general linear regression equation.

By substituting the observed values of each index in the forecast year into formula (6), the cultural tourism passenger flow of the Yangtze River Delta in the forecast year can be obtained.

2.3. Model Test. The determination coefficient R^2 is used to test the model fitting. The formula for determining coefficient R^2 is as follows:

$$R^2 = 1 - \frac{\sum e_t^2}{\sum (y_t - \overline{y})^2}. \tag{9}$$

In formula (9), $\sum e_t^2$ is the sum of squares of residuals.

In the multiple linear regression model, the number of variables in each regression model may not be the same [20]. It is not appropriate to use the size of R^2 as a measure of fitting quality. Therefore, the coefficient of determination R^2 of modified degrees of freedom is often used. The calculation formula is as follows:

$$\overline{R}^2 = 1 - \frac{\left(\left(\sum e_t^2\right)/n - p\right)}{\left(\left(\sum (y_t - \overline{y})^2\right)/n - 1\right)} = 1 - \frac{n-1}{n-p}\left(1 - R^2\right). \tag{10}$$

In formula (10), n is the sample size, and p is the number of regression coefficients.

3. Results

3.1. Case Profile and Data Sources. The Yangtze River Delta is an urban sprawling area with a high level of tourism development in China. It has rich cultural resources and profound cultural heritage. In view of the different regional coverage of the Yangtze River Delta, this paper takes the core area of the Yangtze River Delta as the research object in the reply of the State Council on the regional planning of the Yangtze River Delta (Guo Han (2010), No. 38), that is, taking Shanghai as the leader, and Nanjing, Suzhou, Wuxi, Changzhou, Zhenjiang, Nantong, Yangzhou, and Taizhou in Jiangsu Province, Hangzhou, Ningbo, and Huzhou in Zhejiang Province, and Jiaxing, Shaoxing, Zhoushan, and Taizhou, a total of 16 cities as both wings. This paper selects the 10-year statistical data from 2008 to 2018 as an example, and the data come from the statistical yearbooks and statistical bulletins of various cities, including the China Urban Statistical Yearbook, China Regional Economic Statistical Yearbook, China Tourism Statistical Yearbook, China Tourism Yearbook, List of Top 100 Travel Agencies, and List of National Star Hotels. Data collection adheres to the principle of combining scientificity, authority, standardization, and data availability. When the data from different channels are not unified, the higher-level government department shall prevail.

3.2. Data Processing. Since the data collected are from the statistical yearbooks and statistical bulletins of various cities, the time span is from 2008 to 2018. Due to the snow disaster, earthquake, financial crisis, and other major events in 2008, the cultural tourism reception in the Yangtze River Delta region is seriously affected, and the number of tourists decreased significantly. In order to improve the accuracy of the prediction model, the data of cultural tourism passenger flow in the Yangtze River Delta in 2008 and 2009 are revised.

Linear interpolation is used to correct the data.

Firstly, the starting year a_1 and the ending year a_2 which are suitable for linear interpolation are selected, and the passenger flow in the starting year and the ending year is expressed by y_1 and y_2, respectively. The tolerance d is determined by formula (11).

$$d = \frac{y_2 - y_1}{a_2 - a_1}. \tag{11}$$

The correction value is calculated by the interpolation equation, and the calculation formula is as follows:

$$y_n = y_1 + (a - a_1)d. \tag{12}$$

In formula (12), n is the year to be corrected and y_n is the correction value of the n-th year.

According to the abovementioned method, the starting year of linear interpolation is 2010, and the ending year is 2013. By substituting the data of these two years into formula (11) and formula (12), we can get the revised value of cultural tourism passenger flow in the Yangtze River Delta in 2008 and 2009. Table 2 shows the revised tourist flow of cultural tourism in the Yangtze River Delta from 2008 to 2018.

3.3. Multicollinearity Diagnosis. Taking the statistical data of the Yangtze River Delta from 2008 to 2018 as an example, this paper makes regression analysis on the cultural tourism passenger flow and 28 influencing factors in the Yangtze River Delta and uses SPSS statistical software to select variables by the backward regression method. Finally, there are 9 variables X_7, X_8, X_{15}, X_{16}, X_{17}, X_{18}, X_{20}, X_{22}, and X_{23} to enter into the regression model, and the VIF values are greater than 10, which indicates that there is still a serious collinear relationship between variables. Therefore, it is necessary to use principal component regression to simplify the analysis. On the premise of retaining all or most of the original information, the abovementioned interrelated variables are transformed into a few independent or

Table 2: Revised tourist flow of cultural tourism in the Yangtze River Delta from 2008 to 2018.

Year/year	Passenger flow/10000 person times
2008	1090.09
2009	1110.62
2010	1128.60
2011	1175.27
2012	1241.94
2013	1268.52
2014	1364.39
2015	1392.41
2016	1379.66
2017	1485.59
2018	1590.33

unrelated variables, and then, these variables are integrated to establish a regression model.

Principal component analysis is performed on X_7, X_8, X_{15}, X_{16}, X_{17}, X_{18}, X_{20}, X_{22}, and X_{23} variables. Using SPSS statistical software, the eigenvalues and eigenvectors are obtained. The cumulative variance contribution rate of the first two eigenvalues has reached 91.041%. It is generally believed that the effective information can be retained when the cumulative contribution rate of principal components reaches 85%. Therefore, this paper only needs to take the first two principal components to reflect most of the information of all indexes:

$$F_1 = 0.149X_7 + 0.180X_8 + 0.227X_{15} - 0.217X_{16} - 0.204X_{17} + 0.199X_{18} + 0.179X_{20} - 0.179X_{22} - 0.210X_{23}, \tag{13}$$

$$F_2 = 0.455X_7 - 0.359X_8 + 0.060X_{15} + 0.137X_{16} + 0.245X_{17} + 0.219X_{18} + 0.342X_{20} + 0.289X_{22} - 0.059X_{23}. \tag{14}$$

The standardized regression equation is obtained by multiple regression analysis between the evaluation values of F_1 and F_2 and the dependent variable Y.

$$Y = 0.973F_1 + 0.113F_2\left(R^2 = 0.972\right). \tag{15}$$

By substituting formula (10) and (13) into (14), the standardized regression equations for the standardized variables X_7, X_8, X_{15}, X_{16}, $X_{17}X_{18}$, X_{20}, X_{22}, and X_{23} are obtained.

$$Y = 0.204X_7 + 0.138X_8 + 0.230X_{15} - 0.200X_{16} - 0.159X_{17} + 0.219X_{18} + 0.220X_{20} - 0.142X_{22} - 0.215X_{23}. \tag{16}$$

Using formula (7), the principal component regression model is obtained as follows:

$$y = 1596.733x_7 + 964.461x_8 + 5.013x_{15} - 576.898x_{16} \; 529.266x_{17} + 0.164x_{18} + 7.999x_{20} - 392158.897x_{22} - 1478326.246x_{23} - 158927.732. \tag{17}$$

3.4. Model Validation. Through the test formula of formula (8) and (9), it can get $R^2 = 0.984$; $\overline{R}^2 = 0.859$, showing that formula (17) has a high degree of fit and can make a reasonable forecast of passenger volume. By substituting the values of X_7, X_8, $X_{15}X_{16}$, X_{17}, X_{18}, X_{20}, X_{22}, and X_{23} from 2008 to 2018 into formula (17), the predicted passenger volume of each year can be obtained. The comparison between the predicted value and the actual value (Table 2) is shown in Figure 1.

It can also be seen from Figure 1 that the predicted value of the principal component regression model has a good fit with the actual value, with the highest error of 1.23%, the lowest error of 0.01%, and the average error of 10 years is only 0.41%, which can basically meet the needs of the prediction of cultural tourism passenger flow in the Yangtze River Delta.

3.5. Model Performance Analysis. In order to analyze the application performance of the model, this paper uses the method to predict the cultural tourism passenger flow in the Yangtze River Delta (Shanghai as the leader, Nanjing, Suzhou, Wuxi, Changzhou, Zhenjiang, Yangzhou, and Taizhou in Jiangsu Province and Hangzhou, Ningbo, Huzhou, Jiaxing, Shaoxing, Zhoushan, and Taizhou in Zhejiang Province, a total of 15 cities as both wings) in 2020. MAPE (average absolute percentage error, which can directly reflect the pros and cons of the prediction effect), MAE (average absolute error, the smaller the value is, the smaller the error is), RMSE (root mean square error, the smaller the value is, the smaller the error is), and EC (equalization coefficient, the higher the value is, the higher the fitting degree is) are taken as evaluation indexes to verify the performance of the model. The results are shown in Table 3. The calculation formula of each evaluation index is as follows:

$$\text{MAPE} = \frac{1}{V}\sum_t \left|\frac{C_p(t) - C_r(t)}{C_r(t)}\right| \times 100\%, \tag{18}$$

$$\text{MAE} = \frac{1}{V}\sum_t \left|C_p(t) - C_r(t)\right|, \tag{19}$$

$$\text{RMSE} = \sqrt{\frac{\sum_t \left(C_p(t) - C_r(t)\right)^2}{V}}, \tag{20}$$

$$\text{EC} = 1 - \frac{\sqrt{\sum_t \left(C_p(t) - C_r(t)\right)^2}}{\sqrt{\sum_t \left(C_p(t)\right)^2} + \sqrt{\sum_t \left(C_r(t)\right)^2}}. \tag{21}$$

In the abovementioned formula, $C_p(t)$, $C_r(t)$, and V are the predicted output value, the measured value of cultural tourism passenger flow of each city in 2020, and the number of predicted samples, respectively. The prediction accuracy of the proposed model and the comparison model (the prediction model based on the gravity model and the prediction model based on GM (1, 1)) are compared fairly to the maximum extent by using the abovementioned indexes. Prediction results of this model are shown in Table 3.

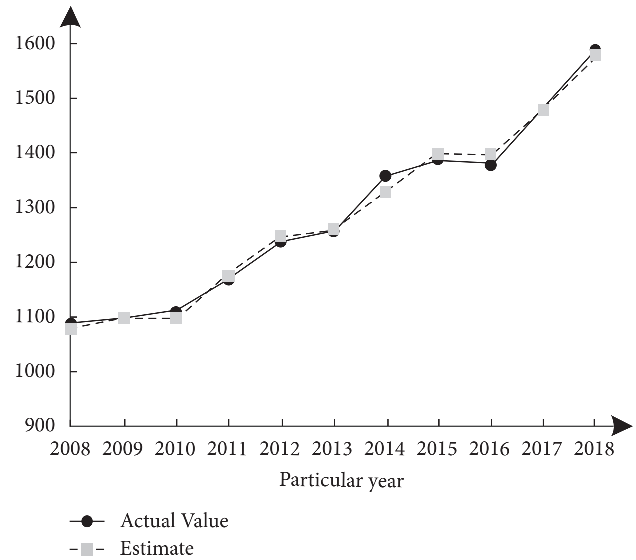


Figure 1: Comparison between the predicted value and actual value.

Table 3: Prediction results of this model.

Cultural tourism cities in the Yangtze River Delta	MAPE	MAE	RMSE	EC
Shanghai	0.57	1.26	2.05	0.97
Nanjing	0.60	1.14	2.05	0.98
Suzhou	0.62	1.30	2.01	0.98
Wuxi	0.35	1.06	1.65	0.98
Changzhou	0.51	1.10	1.79	0.98
Zhenjiang	0.50	1.05	1.81	0.99
Yangzhou	0.40	1.02	1.69	0.98
Taizhou	0.37	1.05	1.77	0.98
Hanzghou	0.69	1.14	1.96	0.97
Ningbo	0.43	1.12	1.72	0.98
Huzhou	0.49	1.11	1.82	0.99
Jiaxing	0.39	1.09	1.93	0.97
Shaoxing	0.44	1.13	2.00	0.98
Zhoushan	0.51	1.06	1.91	0.98
Taizhou	0.56	1.11	2.02	0.98
Average value	0.495	1.116	1.879	0.979

The average values of each index obtained by the algorithm in Table 3 are compared with those obtained by the two comparative prediction models, and the results are shown in Table 4.

After analyzing the prediction performance of the three prediction models in Table 4, it is found that the evaluation indexes of the proposed model are significantly better than those of the two comparison models, and the MAPE is reduced by more than 50 compared with the comparison model, which indicates that this model has higher prediction accuracy compared with the comparison model.

The running time (the larger the value is, the higher the complexity of the method is) is used as the evaluation index of prediction performance. The running time of this model and the two comparison models in the evaluation process is compared, and the results are shown in Table 5.

Analysis of Table 5 shows that the running time of the proposed model and the GM (1, 1) model is significantly better than that of the gravity model. The average running time of the proposed model is 0.71 s, and the average running time of the GM (1, 1) model is 0.74 s. There is no significant difference between the two models. Combined with the data in Table 4, it can be concluded that the proposed model has a significant performance advantage compared with the comparison model.

3.6. Application Test. This paper compares the amount of foreign exchange earned by cultural tourism cities in the Yangtze River Delta after using this model to predict the passenger flow with the amount of foreign exchange earned before using this model (the average of the previous three years is the standard value). The results are shown in Figure 2, in which the numbers 1–15 represent the cultural tourism cities in the Yangtze River Delta.

According to the analysis of Figure 2, after using the proposed model to predict the passenger flow, the amount of

Table 4: Comparison results of prediction performance of different prediction models.

Prediction model	MAPE	MAE	RMSE	EC
Model of this paper	0.495	1.116	1.879	0.979
Prediction model based on the gravity model	1.326	2.735	2.789	0.962
Prediction model based on the GM (1, 1) model	0.866	2.003	2.104	0.941

Table 5: Running time of different prediction models.

Cultural tourism cities in the Yangtze River Delta	Model of this paper	Prediction model based on the gravity model	Prediction model based on the GM (1,1) model
Shanghai	0.71	1.05	0.58
Nanjing	0.74	1.16	0.92
Suzhou	0.77	1.13	0.84
Wuxi	0.70	1.08	0.73
Changzhou	0.66	1.15	0.81
Zhenjiang	0.69	1.02	0.69
Yangzhou	0.72	0.99	0.75
Taizhou	0.73	1.15	0.68
Hanzghou	0.69	1.03	0.75
Ningbo	0.71	1.01	0.72
Huzhou	0.70	0.97	0.67
Jiaxing	0.68	1.08	0.73
Shaoxing	0.68	1.04	0.64
Zhoushan	0.71	1.03	0.79
Taizhou	0.72	1.10	0.80

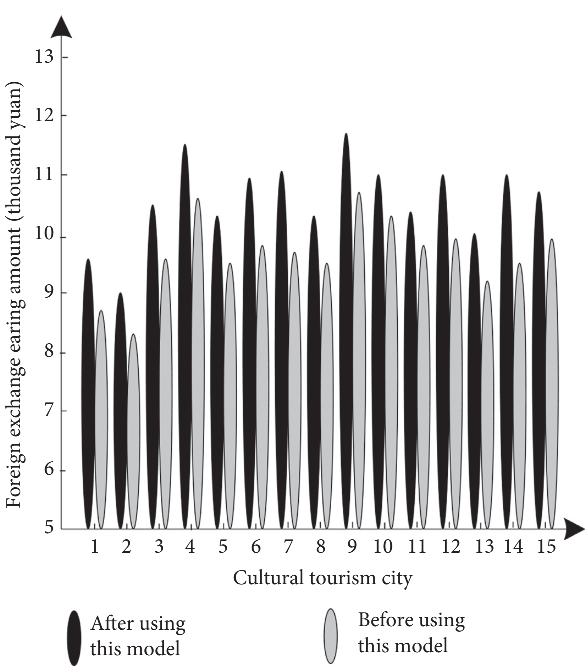


Figure 2: System sales comparison.

foreign exchange earnings of each city shows different increases, and the increase rate is maintained at more than 12%, which shows that the economic benefits of cultural tourism in the Yangtze River Delta can be significantly improved by using this model.

4. Conclusions

Cultural tourism has gained much attention in the last decade and has promoted the preservation of a variety of tangible and intangible assets of culture. The applications of cultural assets for development of tourism has generated various debates such as the matter of whether the intangible values of cultural assets including those of their education, aesthetics, and history can properly be carried out in order to attract tourist. In order to accurately predict the cultural tourism passenger flow in the Yangtze River Delta and improve its economic benefits, this paper designs the prediction model of cultural tourism passenger flow in the Yangtze River Delta based on regression analysis. This paper considers many factors that affect the cultural tourism passenger flow and designs a prediction model of cultural tourism passenger flow in the Yangtze River Delta based on regression analysis. Through the experimental analysis, this model can accurately predict the cultural tourism passenger flow in the Yangtze River Delta in the process of practical application, and significantly improve the cultural tourism amount of foreign exchange in the Yangtze River Delta.

Acknowledgments

This work was supported by University Humanities and Social Sciences Research Project of Anhui Province "Research on the high quality development mode of culture and tourism in the process of Yangtze River Delta integration," Project No. SK2020A0431..

References

[1] X. Wang, E. Yao, and S. Liu, "Urban rail transit passenger flow forecastingfor large special event based on afc data," *Beijing Jiaotong Daxue Xuebao*, vol. 42, no. 1, pp. 87–93, 2018.
[2] C. Lin, K. Wang, D. Wu, and B. Gong, "Passenger flow prediction based on land use around metro stations: a case study," *Sustainability*, vol. 12, no. 17, p. 6844, 2020.
[3] X. Liu, L. Li, X. Liu, and T. Zhang, "Analysis of passenger flow and its influences on hvac systems: an agent based simulation in a Chinese hub airport terminal," *Building and Environment*, vol. 154, pp. 55–67, 2019.
[4] K. Mbbchir, R. Statistics, C. G. Hons, and M. Alain Vuylstekec, "Prediction of patient length of stay on the intensive care unit following cardiac surgery: a logistic regression analysis based on the cardiac operative mortality risk calculator, euroscore," *Journal of Cardiothoracic and Vascular Anesthesia*, vol. 32, no. 6, pp. 2676–2682, 2018.
[5] Y. Liu, C. Lyu, X. Liu, and Z. Liu, "Automatic feature engineering for bus passenger flow prediction based on modular convolutional neural network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 99, pp. 1–10, 2020.
[6] W. Zhou, W. Wang, and D. Zhao, "Passenger flow forecasting in metro transfer station based on the combination of singular spectrum analysis and adaboost-weighted extreme learning machine," *Sensors*, vol. 20, no. 12, p. 3555, 2020.
[7] S. Y. Song, "Analysis and prediction of passenger flow differences in shanghai 3a scenic spots and above based on arima model," *Statistics and Applications*, vol. 08, no. 3, pp. 537–552, 2019.
[8] H. Chen, B. Wang, W. He, and J. Zheng, "Research on passenger flow early warning of urban rail transit station based on system dynamics," *MATEC Web of Conferences*, vol. 308, no. 2, p. 01003, 2020.
[9] C. Li, "Combined forecasting of civil aviation passenger volume based on arima-regression," *International Journal of System Assurance Engineering and Management*, vol. 10, no. 5, pp. 945–952, 2019.
[10] F. Wang, H. Y. Liu, C. F. Shao, and J. J. Zhang, "Prediction model of public transport vehicle allocation based on multiple regression analysis," *Advances in Transportation Studies*, vol. 3, pp. 69–78, 2018.
[11] Y. Liu, C. Liu, and Z. Zheng, "Traffic congestion and duration prediction model based on regression analysis and survival analysis," *Open Journal of Business and Management*, vol. 08, no. 02, pp. 943–959, 2020.
[12] L. Z. Jing, Q. S. Li, J. L. Xu, X. L. Jia, and Y. J. Han, "Average travel time prediction model in basic expressway sections based on v/c ratio and truck percentage," *Chang'an Daxue Xuebao (Ziran Kexue Ban)/Journal of Chang'an University (Natural Science Edition)*, vol. 38, no. 5, pp. 106–113, 2018.
[13] T. Wu, P. Zhang, J. Qin, D. Wu, and Y. Wan, "A flood-discharge-based spatio-temporal diffusion method for multi-target traffic hotness construction from trajectory data," *IEEE Access*, vol. 99, pp. 1-2, 2020.
[14] L. Liu, R.-C. Chen, and S. Zhu, "Impacts of weather on short-term metro passenger flow forecasting using a deep lstm neural network," *Applied Sciences*, vol. 10, no. 8, p. 2962, 2020.
[15] T.-Z. Chen, Y.-Y. Chen, and J.-H. Lai, "Estimating bus cross-sectional flow based on machine learning algorithm combined with wi-fi probe technology," *Sensors*, vol. 21, no. 3, p. 844, 2021.
[16] X. Chang, J. Wu, H. Sun, G. Wang, and X. Bao, "Understanding and predicting short-term passenger flow of station-free shared bike: a spatiotemporal deep learning approach," *IEEE Intelligent Transportation Systems Magazine*, vol. 1, no. 99, pp. 1–3, 2021.
[17] D. Li, L. Deng, and Z. Cai, "Statistical analysis of tourist flow in tourist spots based on big data platform and da-hkrvm algorithms," *Personal and Ubiquitous Computing*, vol. 24, no. 1, pp. 87–101, 2020.
[18] H. J. Heng and P. Ren, "Short-Term passenger flow forecasting for the check-in process in the airport based on time series," *Computer Simulation*, vol. 37, no. 02, pp. 31–37, 2020.
[19] W. Wang, Y. Wang, G. Correia, and Y. Chen, "A network-based model of passenger transfer flow between bus and metro: an application to the public transport system of Beijing," *Journal of Advanced Transportation*, vol. 2020, no. 15, 4 pages, Article ID 6659931, 2020.
[20] L. Qiu, H. Zhou, Z. Wang, S. Zhang, L. Zhang, and W. Lou, "High-speed elevator car air pressure compensation method based on coupling analysis of internal and external flow fields," *Applied Sciences*, vol. 11, no. 4, p. 1700, 2021.

Analysis of the Influence of Quantile Regression Model on Mainland Tourists' Service Satisfaction Performance

Wen-Cheng Wang,[1,2] Wen-Chien Cho,[3] and Yin-Jen Chen[4]

[1] *Department of Business Administration, National Taiwan University of Science and Technology, Taiwan*
[2] *Department of Business Administration, Hwa Hsia Institute of Technology, No. 111 Gongzhuan Road, Zhonghe District, New Taipei City 235, Taiwan*
[3] *Department of Hospitality Management, Hsing Wu University, No. 101, Sec. 1, Fenliao Road, LinKou District, New Taipei City 24452, Taiwan*
[4] *Department of Hotel and Restaurant Management, Ching Kuo Institute of Management and Health, 336 Fu Hsin Road, Keelung, Taiwan*

Correspondence should be addressed to Wen-Cheng Wang; wcwang@cc.hwh.edu.tw

Academic Editors: M. Saberi and R. Zhang

It is estimated that mainland Chinese tourists travelling to Taiwan can bring annual revenues of 400 billion NTD to the Taiwan economy. Thus, how the Taiwanese Government formulates relevant measures to satisfy both sides is the focus of most concern. Taiwan must improve the facilities and service quality of its tourism industry so as to attract more mainland tourists. This paper conducted a questionnaire survey of mainland tourists and used grey relational analysis in grey mathematics to analyze the satisfaction performance of all satisfaction question items. The first eight satisfaction items were used as independent variables, and the overall satisfaction performance was used as a dependent variable for quantile regression model analysis to discuss the relationship between the dependent variable under different quantiles and independent variables. Finally, this study further discussed the predictive accuracy of the least mean regression model and each quantile regression model, as a reference for research personnel. The analysis results showed that other variables could also affect the overall satisfaction performance of mainland tourists, in addition to occupation and age. The overall predictive accuracy of quantile regression model Q0.25 was higher than that of the other three models.

1. Introduction

The opening up of Taiwan to visitors from mainland China can boost Taiwan's tourism industry and periphery industries. Based on the estimate by the Taiwan Visitors Association, if Taiwan opens up to 3,000 mainland visitors per day, there will be 1 million visitors per year. If each mainland visitor stays in Taiwan for seven to ten days and spends 50,000 NTD, this could contribute about 50 billion NTD to Taiwan's tourism industry. Due to the multiplier effect of consumption, the output value of the Taiwan service industry could reach over 100 billion NTD, in industries such as the airline industry, travel industry, tourist hotels, transportation, food, recreation areas, department stores, and native products. This represents good news, and it can increase the number of employment opportunities in Taiwan. However, the business opportunity resulted from opening up to mainland visitors is too large for Taiwan. The aim of this paper was to discuss whether Taiwan has engaged in appropriate planning and relevant industry software and hardware facilities and services.

A great deal of the literature on consumer behavior already exists [1, 2]; however, the literature on mainland visitors traveling to Taiwan is seldom found. Thus, the question items for the satisfaction of mainland visitors with Taiwan's tourism services were used as the research basis in this study. This paper conducted a questionnaire survey of mainland tourists. The questionnaire included basic data and nine satisfaction items. First, grey relational analysis was conducted on the nine question items of satisfaction, and the grey relational grade of the analysis result was regarded as the overall satisfaction performance value through the overall satisfaction performance could be learned. Next, the basic data and the first eight satisfaction items were used as the independent variables (X), and the overall satisfaction performance was used as the dependent variable (Y). The least mean square regression model and quantile regression model analysis were used for analysis. The difference between the two models was compared, and the relationships among the independent variables and the dependent variable were observed under different quantiles. At last, this study discussed the predictive accuracy of the least mean regression model and each quantile regression model, and the results could be provided as a reference for research personnel.

This paper is organized as follows. Section 1 presents the motivation and purpose of the research. Section 2 introduces grey correlation performance analysis, the quantile regression model, and relevant literature. Section 3 involves the sample data and empirical analysis. Section 4 proposes the research conclusion.

2. Methodology

2.1. Grey Mathematics and Grey Relational Analysis. The grey mathematic theory, proposed by Long [3], is used under a situation of system uncertainty and information incompleteness to perform the model setup, relational analysis, and forecasts, so as to understand the system. Before performing grey relational analysis, in order to let the sequence satisfy comparability requirements, a normalization treatment needs to be performed on the data of the sequence. This is called the formation of the grey relation, in which the sequence can be made to satisfy comparability. Before performing grey relational analysis, the reference sequence has to be confirmed first, and then the closeness between other sequences (comparison sequences and son sequences) and the reference sequence can be compared so as to find out the level of grey relation and create the grey relation ordinals. Through the rankings, the advantages and disadvantages can be judged so as to assist in carrying out the decision. The steps of grey relational analysis are as follows.

(A) From the original decision matrix D, find out standard sequence A_0 and inspected sequence A_i. The standard sequence is set $A_0 = (x_{01}, x_{02}, \ldots, x_{0j}, \ldots, x_{0n})$ formed by the ideal target value of each influencing factor and has a total of j terms, wherein $j = 1, 2, \ldots, n$. In addition, the performance value of inspected sequence $A_i = (x_{i1}, x_{i2}, \ldots, x_{ij}, \ldots, x_{in})$, wherein $i = 1, 2, \ldots, m$.

(B) Normalize the data of original decision matrix D.

(C) Calculate grey relational distance Δ_{0ij} and use it to evaluate the difference between each normalized value and normalized reference data:

$$\Delta_{0ij} = \left| x^*_{0j} - x^*_{ij} \right|. \tag{1}$$

(D) Calculate grey relational coefficient γ_{0ij}:

$$\gamma_{0ij} = \frac{\Delta_{\min} + \zeta \Delta_{\max}}{\Delta_{0ij} + \zeta \Delta_{\max}}. \tag{2}$$

(E) Calculate grey relational grade Γ_{0i} by the following equation:

$$\Gamma_{0i} = \sum_{j=1}^{n} \left[\omega_j \times \gamma_{0ij} \right]. \tag{3}$$

Rank the grey relational ordinal and follow grey relational grade value to perform performance ranking.

Grey relational analysis has been applied in many industries, such as business [4], education [5], hospitals [6], and tourism [7]. In this paper, grey relational analysis was performed to discuss the satisfaction of visitors from mainland China traveling to Taiwan.

2.2. Quantile Regression Method. The quantile regression method [8] is an extension of the traditional least square regression method. The difference is that the former can accurately estimate the marginal effect of independent variables on the dependent variable in a specific "condition" quantile, which is superior to the marginal effect of the latter's average trend. The quantile regression method has one more observation dimension than that of the traditional least square regression, and it can therefore analyze and observe the margin effect of each specific quantile condition. Quantile regression has been applied to financial research [9–11]. In this study, it is applied to the tourism industry.

STATA was used to perform quantile regression analysis; its graphics is similar to Figure 1. It was based on the average concept of the traditional least square method but was extended to different quantile positions of the entire distribution interval. From this, the impact of the basic data of the questionnaires and the first eight satisfaction items (X) on the overall satisfaction and the impact of the ninth question item (Y) were observed.

3. Empirical Research

3.1. Sample Data and Variables. There were a total of 144 mainland visitors traveling to Taiwan who participated in the questionnaire survey. The question items in the questionnaires were written using simplified Chinese characters and included basic information items such as gender (A_1), age (A_2), occupation (A_3), and favorite Taiwan food type (A_4). The satisfaction question items included Taiwan tourist attractions (X_1), Taiwanese food (X_2), Taiwanese hotel facilities (X_3), Taiwanese hotel service attitude (X_4), Taiwanese night fair culture (X_5), Taiwanese architectural styles (X_6), Taiwan weather conditions (X_7), Taiwanese road cleanliness (X_8), and whether to visit Taiwan in the next time (X_9). The descriptive statistical values of these satisfaction question items are shown in Table 1.

Table 1: Descriptive statistic value of the nine satisfaction variables.

Satisfaction	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
Max	100	100	100	100	100	100	95	90	100
Min	40	50	40	50	40	40	30	10	30
Avg	80.76	83.26	81.17	81.88	79.84	76.24	70.38	67.19	78.10
Std	14.15	12.74	12.10	11.80	13.90	11.57	13.30	16.80	13.86
N	144	144	144	144	144	144	144	144	144

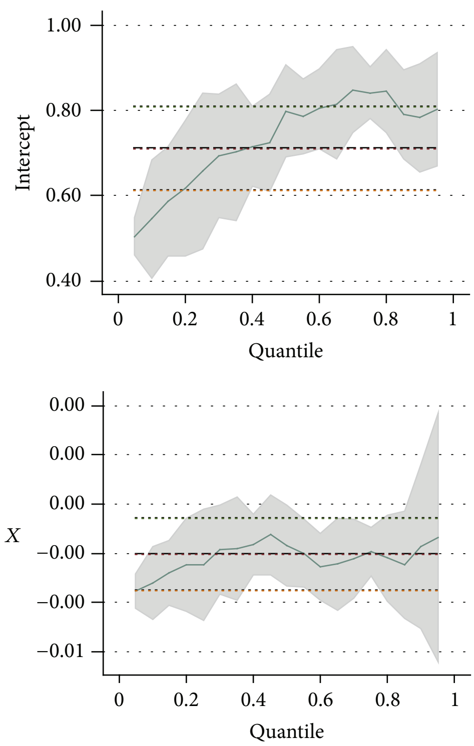


Figure 1: Quantile regression diagram.

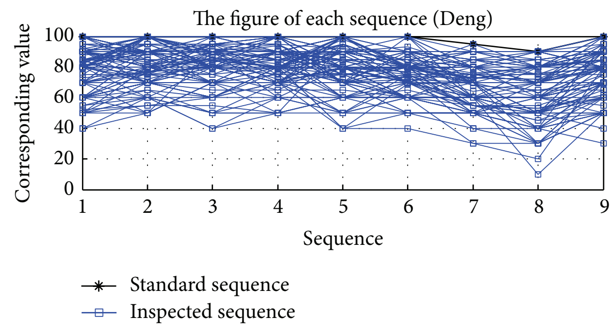


Figure 2: Output chart of grey relational grade of eight satisfaction items.

3.2. Quality Satisfaction Analysis. This paper used the grey relational grade analysis proposed by Deng and the grey relational analysis MATLAB procedure developed by Wen et al. [12] to calculate the grey relational grade and then analyze the satisfaction performance of the nine question items. Higher service satisfaction values of the nine question items are better. In this paper, the maximum value of the nine question items was used as the standard sequence. The grey relational analysis is shown in Figure 2, in which the upper bold black line denotes the standard sequence and the thin solid line denotes the inspected sequence. Each piece of data included nine nodes which denoted the service satisfaction values of the nine question items. If the inspected sequence was closer to the standard sequence, the service quality satisfaction in the questionnaire information would be better; namely, the customers would be more satisfied with the contents in the nine question items. The grey relational analysis result indicated the three respondents to evaluate the overall satisfaction; the highest were the 97th respondent (GRG = 0.921), the 65th respondent (GRG = 0.8988), and the 125th respondent (GRG = 0.8864).The gray relational grade in the analysis result was regarded as the overall satisfaction performance value, namely, the dependent variable (Y). In conjunction with the first eight question items (X), the least square regression model and quantile regression were used for analysis.

3.3. The Past Eight Satisfaction Items Are Used as the Independent Variables; the Overall Satisfaction Performance Is Used as One Dependent Variable for Regression Model Analysis. In this paper, the first eight satisfaction items were used as independent variables (X) and the gray relational analysis result was used as the dependent variable (Y). The least square regression model and quantile regression model were used for analysis. The quantile regression model was established using the 0.25 quantile (1/4 quantile), 0.5 quantile (1/2 quantile), and 0.75 quantile (3/4 quantile). The analysis results are shown in Table 2 and Figure 3.

As shown in Tables 2 and 3, gender had an insignificant impact on the overall satisfaction performance in the high quantiles, but it had a significant impact on the overall satisfaction performance in the low quantiles (0.25 and 0.5). In Figure 3, the long thick dotted line denotes the least regression model, and the upper and lower short fine dotted lines denote the confidence interval. From Figure 3, it could be seen that the quantile regression overestimated the situation under the 0.25 quantile and that the quantile regression underestimated the situation under the 0.25 quantile–0.5 quantile, as compared to the least square regression model. The questionnaires showed that the overall satisfaction of men was lower than that of women, because Taiwanese hotels pay more attention to women. The hotels failed to provide smoking areas for male guests, and they were not allowed to speak loudly or drop litter and cigarette ends. The male guests thus felt inconvenienced, and their overall satisfaction was

TABLE 2: Analysis results of least square regression model and quantile regression model for overall satisfaction performance.

Regression	OLS			Q0.25			Q0.5			Q0.75		
variable	Coefficient	*T* value	Significant	Coefficient	*T* value	Significant	Coefficient	*T* value	Significant	Coefficient	*T* value	Significant
Sex	−0.004	−1.074		−0.0001	−1.77	*	−0.0001	−2.93	***	−0.0001	−1.13	
Occupation	0.001	1.004		0.0013	0.43		−0.0025	−0.63		−0.0057	−0.90	
Age	−0.002	−0.949		−0.0001	−0.24		−0.0001	−0.08		0.0002	0.19	
Tourist attraction	0.001	3.837	***	0.0006	2.64	***	0.0007	2.45	**	0.0008	2.02	**
Food	0.001	6.103	***	0.0015	5.56	***	0.0015	4.18	***	0.0017	4.05	***
Hotel facilities	0.001	2.989	***	0.0012	4.60	***	0.0013	3.46	***	0.0003	0.52	
Hotel service attitude	0.002	6.016	***	0.0013	4.86	***	0.0010	2.84	***	0.0019	3.89	***
Night fair culture	0.001	4.552	***	0.0012	5.64	***	0.0009	3.40	***	0.0007	1.91	*
Architectural style	0.001	5.568	***	0.0015	5.44	***	0.0015	4.43	***	0.0016	3.16	***
Weather conditions	0.002	7.091	***	0.0015	5.33	***	0.0019	4.43	***	0.0021	5.39	***
Street cleanliness	0.000	2.167	**	0.0007	2.61	***	0.0007	2.20	**	0.0004	1.11	
Pseudo R^2		0.9670			0.8744			0.8349			0.7985	

Note: * is significant at the 10% significance level; ** is significant at the 5% significance level; *** is significant at the 1% significance level.

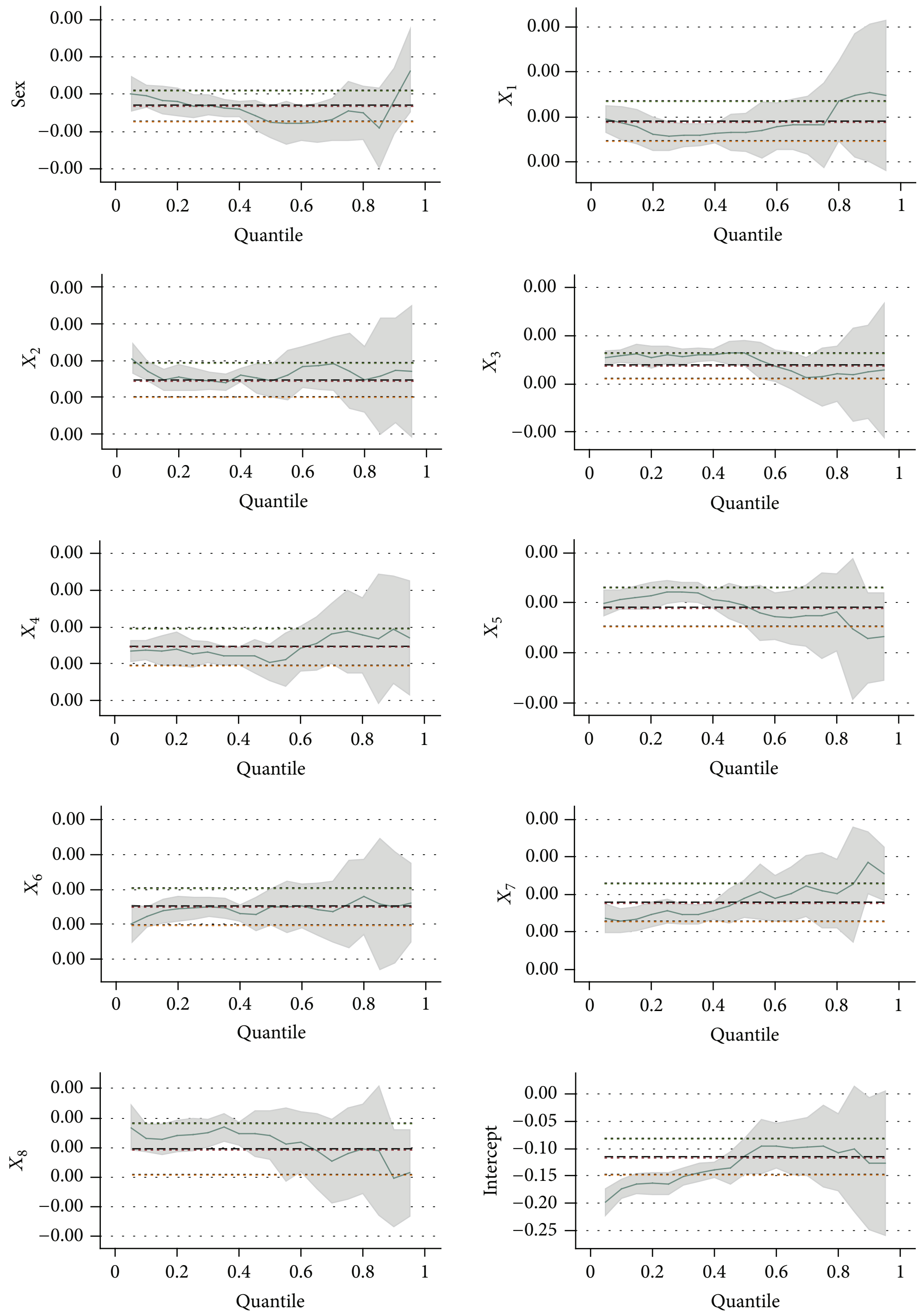


Figure 3: Output of analysis results of least square regression model and quantile regression model for overall satisfaction performance.

lower. Taiwan tourism enterprises and relevant governmental department should work to solve these problems.

From Table 2, it could be found that variable X_1, satisfaction with tourist attractions, reached the significance level. Thus, the satisfaction of Taiwanese tourist attractions could significantly affect the overall satisfaction performance. As shown in Figure 3, if the satisfaction with tourist attractions was estimated using the least square regression model, the high quantile could be underestimated. This revealed that mainland visitors who had high overall satisfaction performance also had higher requirements for tourist attractions. Namely, Taiwanese tourist attractions could affect the evaluation of Taiwan's overall tourism quality by mainland visitors. In Taiwan, some famous scenic spots have complete management planning, such as Ali Mountain and Sun Moon Lake; however, other less-famous scenic spots often trigger

Table 3: Verification result of coefficient difference under high and low quantiles (Q0.25–Q0.75).

Variable	Sex	Occupation	Age	Tourist attraction	Food	Hotel facilities	Hotel service attitude	Night fair culture	Architectural style	Weather conditions	Street cleanliness
F value	3.02	0.11	0.20	3.16	0.12	5.15	0.36	3.41	0.03	1.59	4.59
Different significance	*			*		***		**			***

Note: * is significant at the 10% significance level; ** is significant at the 5% significance level; *** is significant at the 1% significance level.

complaints due to the lack of management planning. Thus, the management of Taiwanese tourist attractions should be enhanced, with the help of relevant government agencies.

From Table 2, it could be found that satisfaction with food reached the significance level under different quantiles. Thus, satisfaction with food could also affect the overall satisfaction performance. From Figure 3, if the least square regression model was used for estimation, the overall satisfaction would be underestimated under different quantiles. This indicated that Taiwanese food was an important indicator to measure whether the mainland tourists were willing to visit Taiwan, regardless of low or high overall satisfaction. Thus, relevant departments should give assistance to enterprises in the food industry. Taiwanese food has unique features and can attract more visitors, thus increasing the inbound tourist flow.

It was found that hotel facilities (X_3) were not significant under the 0.75 quantile, as shown in Tables 2 and 3. If the least square regression model was used, this variable would be underestimated under the low quantiles. This revealed that the worse overall satisfaction performance of the visitors was caused by poor hotel facilities. Hotel service attitude (X_4) reached the significance level under different quantiles; however, if the least square regression model was used, the variable could be overestimated under the low quantiles and be underestimated in the high quantiles. This revealed that the requirements for hotel service attitude would become higher and higher with the improvement of the overall satisfaction performance of mainland tourists. Thus, Taiwan's relevant departments should give assistance to hotels in the improvement of hotel facilities and staff service attitudes.

It was found that night fair culture, architectural styles, and weather conditions (X_5, X_6, and X_7) reached the significance level, as shown in Table 2. Taiwan has many night fairs throughout the island, and many tourists travel to Taiwan in order to taste local food at the night fairs. Taiwan also has many famous scenic spots and historical sites, such as Fort Santo Domingo in New Taipei City, Taipei's National Palace Museum, and Fort Provintia and Fort Zeelandia in Tainan. The architectural style of many famous temples is also an important factor that attracts foreign tourists. In addition, Taiwan is located in the subtropical zone, and the weather is spring-like all year round, which can attract more tourists to come to Taiwan.

Tables 2 and 3 show that street cleanliness (X_8) had a significant difference under the high and low quantiles. It did not reach the significance level under the 0.75 quantile, but it reached the significance level under the 0.25 quantile. Thus, street cleanliness was an important indicator for the mainland tourists with lower satisfaction performance. In Taiwan, street sweepers are responsible for cleaning rubbish on the street, and such rubbish is often caused by road construction. This problem should be solved by the urban development bureaus. In addition, some parts of the sidewalks are illegally occupied. These problems could result in the worse satisfaction performance of mainland tourists. The Taiwanese tourist department should coordinate with the relevant government departments for road improvement.

3.4. Division of Sample Data. In this section, the quantile regression model was used to analyse the reliability of the satisfaction of mainland tourists and determine the accuracy of the quantile regression model. The original sample data were divided into four groups; three groups were used as training data to establish the regression model, and one group was used as the testing data for cross-validation of the prediction methods, so as to test the reliability and predictive ability of the models. In this paper, five assessment indicators and the overall predictive accuracy were used to compare the predictive ability of the four models, as follows:

(1) the root mean squared error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{t=1}^{n} x_t - \hat{x}_t}{N}}, \tag{4}$$

(2) the revised Theil inequality coefficient (RTIC):

$$\mathrm{RTIC} = \left[\frac{\sum_{t=1}^{N}\left(X_t - \widehat{X}_t\right)^2}{\sum_{t=1}^{N} x_t^2}\right]^{1/2}, \tag{5}$$

(3) the mean absolute error (MAE):

$$\mathrm{MAE} = \frac{1}{M}\sum_{l=1}^{M}\left|Z_{t+1} - \widehat{Z}_t(l)\right|, \tag{6}$$

where M is the number of predictive values, Z_{t+1} is observed value of one hour, and $\widehat{Z}_t(l)$ is estimated value of one hour,

(4) the mean absolute percentage error (MAPE):

$$\mathrm{MAPE} = \left(\frac{1}{M}\sum_{l=1}^{M}\left|\frac{\left(Z_{t+1} - \widehat{Z}_t(l)\right)}{Z_{t+1}}\right|\right) \times 100\%, \tag{7}$$

Table 4: Evaluation indicators and result of total predictive accuracy.

Regression	RMSE	RTIC	MAE	MAPE	CE	Accuracy %
OLS	1.788	0.031	1.526	0.017	0.903	91.25
Q0.25	1.024	0.019	0.996	0.011	0.967	97.60
Q0.5	1.332	0.026	1.288	0.015	0.935	94.87
Q0.75	2.156	0.045	1.961	0.030	0.866	89.16

(5) the coefficient of efficiency (CE):

$$\mathrm{CE} = 1 - \frac{\sum \left(x_t - \widehat{x}_t\right)^2}{\sum \left(x_t - \overline{x}_t\right)^2}. \qquad (8)$$

For indicators 1 to 4, a result closer to 0 indicated that the model had higher accuracy. For the fifth indicator, a result closer to 1 indicated that the model had higher accuracy. The analysis and test results are shown in Table 2.

Table 4 shows that, in the $Q0.25$ predictive model, the RMSE was 1.024, the RTIC was 0.019, the MAE was 0.996, and the MAPE was 0.011, which were lower than the results for OLS in $Q0.5$ and $Q0.75$. The CE was 0.967 and the overall accuracy was 97.6%, which was higher than that of the other three models. Thus, the $Q0.25$ predictive model had better predictive ability as compared to the other three models. The analysis results showed that the overall predictive accuracy of the three quantile regression models was almost 90%. In terms of the pseudo R^2 in Table 2, the regression models had high interpretability; thus, the quantile regression had a high confidence level when it was used to analyze the impact of the variables on the overall satisfaction of mainland tourists.

4. Conclusion

The main contribution of this paper is different from the past literature. In this paper, quantile regression and the OLS method were used to discuss the impact of the satisfaction of mainland tourists with service quality on the overall satisfaction performance. The causes for the impact of the variables on the overall satisfaction under different quantiles and improvement solutions were further analyzed. The empirical results showed that gender, tourist attractions, hotel facilities, night fair culture, and street cleanliness affected the overall satisfaction performance of the mainland tourists and had significant differences in the high and low quantiles. Besides work and age, other variables could also affect the overall satisfaction performance of mainland tourists. In addition, quantile regression was used to analyze the reliability of the satisfaction of mainland tourists in this paper. Further, the original data were divided into several groups to establish the prediction models. Five evaluation indicators and the overall predictive accuracy were used to conduct cross verification and analysis of the four models. The findings showed that the overall accuracy of the $Q0.25$ quantile regression predictive model reached a level of 97.6% and that $Q0.5$ and $Q0.75$ reached the 90% level. Thus, the quantile regression model is feasible for analyzing the impact of variables on the satisfaction of mainland tourists traveling to Taiwan.

References

[1] C. Heijes, "Culture, convenience or efficiency: customer behaviour in choosing local or foreign banks in China," *Chinese Management Studies*, vol. 2, no. 3, pp. 183–202, 2008.

[2] M. Mavri and G. Ioannou, "Customer switching behaviour in Greek banking services using survival analysis," *Managerial Finance*, vol. 34, no. 3, pp. 186–197, 2008.

[3] J. Long, "Control problems of grey systems," *Systems and Control Letters*, vol. 1, no. 5, pp. 288–294, 1982.

[4] W.-T. Pan, "The use of genetic programming for the construction of a financial management model in an enterprise," *Applied Intelligence*, vol. 36, no. 2, pp. 271–279, 2010.

[5] Y. Z. Wang and F. M. Lee, "Grey relational analyzing the relationship between the students' records of entrance examination and their future academic performance," *The Journal of Grey System*, vol. 16, no. 2, pp. 155–164, 2004.

[6] C. H. Fang, S. T. Chang, and K. Chen, "A study on important duties of middle management in medical industry—Application of grey relational analysis," *Chinese Management Review*, vol. 14, no. 2, pp. 1–15, 2011.

[7] J. C. Shen, T. H. Shih, and October, "A study travelling risk evaluation model by using grey relation analysis," *Journal of Kaohsiung Hospitality College*, vol. 1, pp. 93–106, 1998.

[8] R. Koenker and G. Bassett, "Regression quantiles," *Econometrica*, vol. 46, pp. 33–50, 1978.

[9] G. Q. Li, S. W. Xu, Z. M. Li, Y. G. Sun, and X. X. Dong, "Using quantile regression approach to analyze price movements of agricultural products in china," *Journal of Integrative Agriculture*, vol. 11, no. 4, pp. 674–683, 2012.

[10] S. Saha and J. J. Su, "Investigating the interaction effect of democracy and economic freedom on corruption: a cross-country quantile regression analysis," *Economic Analysis & Policy*, vol. 42, no. 3, pp. 389–396, 2012.

[11] K. Okada and S. Samreth, "The effect of foreign aid on corruption: a quantile regression approach," *Economics Letters*, vol. 115, no. 2, pp. 240–243, 2012.

[12] K. L. Wen, S. K. Chang-Chien, C. K. Yeh, C. W. Wang, and H. S. Lin, *Apply MATLAB in Grey System Theory*, Chuan Hwa Book CO., LTD, 2006.