

Research Article

An Intelligent Passenger Flow Prediction Method for Pricing Strategy and Hotel Operations

Tianyang Wang 

City University of Macau, Macau, China

Correspondence should be addressed to Tianyang Wang; t20091100208@cityu.mo

Received 1 February 2021; Revised 11 February 2021; Accepted 3 March 2021; Published 18 March 2021

Academic Editor: Abd E.I.-Baset Hassanien

Copyright © 2021 Tianyang Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Hospitality industry plays a crucial role in the development of tourism. Predicting the future demand of a hotel is a key step in the process of hotel revenue management. Hotel passenger flow prediction plays an important role in guiding the formulation of hotel pricing and operating strategies. On the one hand, hotel passenger flow prediction can provide decision support for hotel managers and effectively avoid the waste of hotel resources and loss of revenue caused by the loss of customers. On the other hand, it is the guarantee of the priority occupation of business opportunities by hotel enterprises, which can help hotel enterprises adjust their operation strategies reasonably to better adapt to the market situation. In addition, hotel passenger flow prediction is helpful to judge the overall operating condition of the hotel industry and assess the risk level of the hotel project to be built. Hotel passenger flow is affected by many factors, such as weather, environment, season, holidays, economy, and emergencies, and has the characteristics of complex nonlinear fluctuation. The existing demand predicting methods include linear methods and nonlinear methods. The linear prediction methods rely on the stability of environment and time series, so they cannot completely simulate the complex nonlinear fluctuations characteristics of hotel passenger flow. Traditional nonlinear prediction methods need to improve the prediction accuracy, and they are difficult to deal with the increasing data of hotel passenger flow. Based on the above analysis, this paper constructs a deep learning prediction model based on Long Short-Term Memory (LSTM) to predict the number of actual monthly arrival bookings. The number of actual monthly arrival bookings can reflect the actual monthly passenger flow of a hotel. The prediction model can effectively reduce the loss caused by cancellation or nonarrival of bookings due to various reasons and improve the hotel revenue. The experimental part of this paper is based on the booking demand dataset of a resort hotel in Portugal from July 1, 2015, to August 31, 2017. Artificial neural network (ANN) and support vector regression (SVR) are built as benchmark models to predict the number of actual monthly arrival bookings of this hotel. The experimental results show that, compared with the benchmark models, the LSTM model can effectively improve the prediction ability and provide necessary reference for the hotel's future pricing decision and operation mode arrangement.

1. Introduction

The development of tourism plays an increasingly important role in promoting world economic growth and cultural prosperity. China's tourism has grown steadily in recent years. According to the "Basic Situation of Tourism Market in 2019" released by Chinese Tour Research Institute (<http://www.ctaweb.org>), the total tourism revenue in China in 2019 was 6.63 trillion yuan, accounting for 11.05% of the total GDP, and the domestic tourism reception scale was 6.006 billion person-times. With the vigorous development of tourism, the demand of tourists for accommodation and

environmental requirements are also increasing. Hotel operation is facing both opportunities and challenges.

At present, revenue management is a modern scientific operation management mode widely used in hotel management. It refers to the market-oriented analysis and prediction of the consumer behavior of each market segment in the process of hotel management to determine the optimal price and the optimal stock allocation model, so as to optimize the hotel profitability [1]. Revenue management mainly includes hotel demand prediction, pricing strategy, overbooking strategy, and room allocation strategy [1, 2], which is an important tool for making hotel strategic

decisions. Hotel demand prediction is a key step in the process of revenue management. By collecting historical data and establishing prediction model, hotel managers can understand and master the market demand, so as to make differential pricing and adopt different revenue management strategies in different periods to maximize revenue. Hotel demand prediction can provide basic information for the subsequent development planning and decision-making of the hotel. Therefore, accurate prediction of hotel demand is of great significance to the operation and management of hotels.

Hotel passenger flow not only can reflect the current development trend of a hotel but also is an important embodiment of whether the local tourism is developed. Correct analysis and prediction of hotel passenger flow are the core of hotel demand prediction, which plays a crucial role in hotel operation decisions [3] and is the key to hotel revenue management [4]. Hotel passenger flow is affected by many factors, such as weather, environment, season, holidays, economy, and emergencies [5], and has the characteristics of complex nonlinear fluctuation [6]. Therefore, it is often difficult for hotel managers to accurately estimate the passenger flow, leading to unreasonable decision-making, which leads to the waste of hotel resources and loss of revenue. The randomness of passenger flow makes it difficult to analyze and predict. Finding a suitable prediction method is the key and difficult point of hotel passenger flow prediction. Traditional linear predicting methods cannot accurately fit hotel passenger flow, such as Autoregressive Integrated Moving Average (ARIMA), which is the most widely used time predicting method. Zhang [1] used the ARIMA method to model and predict the hotel occupancy rate data in three regions in the research of predicting hotel occupancy rate. On this basis, she used the Ensemble Empirical Mode Decomposition (EEMD) to make up for the shortcomings of time-series models that cannot accurately capture the characteristics of data fluctuations to improve the prediction accuracy of the ARIMA model. Most of the existing nonlinear prediction methods use traditional machine learning methods to build passenger flow prediction models, which cannot automatically extract feature information and adapt to the growing experimental data. Besides, they are easy to fall into the problem of local optimization and overfitting. Ji [7] used Singular Spectrum Analysis (SSA) to model and predict the hotel monthly occupancy times and income, based on the hotel occupancy history data of Victoria Hotel from January 1980 to June 1995. Wu [3] used the SVR model to predict the monthly occupancy, also based on the statistics of the occupancy situation of Victoria Hotel, so as to provide a meaningful reference for the arrangement of the future business model of Victoria Hotel.

Based on the above analysis, this paper constructs LSTM model with good prediction ability for complex time series to predict the monthly passenger flow of a resort hotel. We predict the number of actual monthly arrival bookings to provide reference for the formulation of hotel pricing strategy and the adjustment of operation mode.

Compared with the existing work, the contribution of this paper can be summarized as follows:

- (i) A deep learning algorithm based on automatic extraction of nonlinear complex sequence characteristics is proposed for the first time to predict the number of actual arrival bookings of a hotel.
- (ii) A hotel passenger flow prediction deep learning model based on LSTM is constructed to predict the actual arrival bookings in a certain month in the future. This model has better performance than traditional hotel demand prediction methods.
- (iii) ANN and SVR are constructed as benchmark models and compared with LSTM model in this paper. Mean absolute error, root mean squared error, and mean absolute percentage error are used as evaluation metrics to evaluate the prediction effect of these three models.

Subsequent parts are organized as follows: Section 2 provides an overview of existing demand prediction methods and previews previous efforts to predict hotel passenger flow. It discusses various types of prediction models. Section 3 analyzes the hotel demand data from a resort hotel in Portugal. Section 4 introduces the LSTM network and the hotel passenger flow prediction model based on LSTM in detail. Section 5 describes the experiments carried out in this paper and analyzes the results; lastly, a summary is discussed in Section 6.

2. Literature Review

As early as the 1960s, experts and scholars had already begun researches on demand prediction methods [8]. The existing demand prediction methods include linear and nonlinear methods [6]. These methods can be divided into four categories: classical time-series methods, econometric methods, machine learning methods, and comprehensive prediction methods [8, 9].

Among the demand prediction methods, classical time-series and econometric methods are representative linear time-series prediction methods. Classical time-series prediction methods are widely used. They mostly use the linear relationship between variables for modeling and predicting [8], which contain a lot of types, mainly including Moving Average, Exponential Smoothing, Differential Autoregression, Regression Prediction Model, and Incremental methods [8, 9]. Andrew et al. [10] built two time-series models based on Box-Jenkins and Exponential Smoothing to predict monthly occupancy rates for hotels and verified that these two models can be very useful in actual hotel operations and other applications such as yield management; Pfeifer et al. [11] described the application of Space-Time ARMA modeling to demand-related data from eight hotels from a single hotel chain in a large US city; Weatherford et al. [4] used data from Choice Hotels and Marriott Hotels to test a variety of prediction methods and to determine the most accurate one; in-depth study using the Marriott Hotel data showed that models based on Exponential Smoothing, pickup, and Moving Average were the most robust; Pan et al. [12] first proposed the value of search query volume data in predicting hotel room demand in the field of tourism and

hospitality research; they used search volume data on five related queries to predict demand for hotel rooms in a specific tourist city and proved that three ARMAX models consistently outperformed their ARMA counterparts; Zhang [1] proposed an ARIMA-EEMD hybrid model and concluded that this model has a better effect than the single ARIMA model in predicting hotel occupancy; Yu [5] constructed a Hotel Management Decision Support System based on Mann-Kendall, Moving Average, Exponential Smoothing, and fuzzy comprehensive evaluation modeling method, which realized the prediction of short-term passenger flow; Yang [13] designed a tourist flow prediction system of resort hotel based on Logistic model and proved that the system has high prediction accuracy, which helps to improve the management and scheduling decision-making ability of resort hotel; Sun [14] realized the demand prediction for cruise revenue management based on non-incremental, classical incremental combined, and advanced incremental combined prediction method. Econometric method can explain the causal relationship between target variables and influencing factors from the perspective of economics, but the influencing factors of econometrics can only be speculated based on theory, which is difficult to clarify [8]. Goh et al. [15] used the time-series Seasonal ARIMA (SARIMA) and Multivariate ARIMA (MARIMA) models to predict a tourism demand in Hong Kong and got highest accuracy compared with eight other time-series models; and then Goh and Law [16] used error correction model to analyze the influence of climate on tourism demand; Choi [17] identified key economic indicators of the hospitality industry in USA and built synthetic indicators to predict the US hotel demands successfully.

Machine learning is the most widely used method in traditional demand prediction. Ji [7] used SSA model to predict the monthly occupancy times and monthly revenue of the hotel based on the hotel occupancy history data of Victoria Hotel from January 1980 to June 1995; Wu [3] used SVR model to predict the monthly occupancy time, also based on the statistical data of Victoria hotel occupancy; Hong et al. [18] presented an SVR model with Chaotic Genetic Algorithm (CGA), namely, SVRCGA, to predict the tourism demands; Rashad et al. [19] developed a fuzzy-rule-based system model for hotel occupancy prediction by analyzing 40 months' time-series data and applying fuzzy c-means clustering algorithm; Sun et al. [20] proposed a new framework integrating machine learning and Internet search index to predict tourist volume and proved that the proposed Kernel Extreme Learning Machine model is more stable and effective according to accuracy and robustness analysis than Least Squares SVR (LSSVR), SVR, Artificial Neural Network (ANN), and ARIMAX.

Comprehensive prediction methods are produced to solve the shortcomings of traditional single prediction methods, including combined prediction, integrated prediction, and hybrid prediction methods [8]. Ke-wei et al. [21] established a combined model based on BP neural network and ARIMA to comprehensively analyze and predict the change trend of China's inbound tourists; Gong and Huang et al. [22] established a demand prediction model based on

Grey Theory and Exponential Smoothing method to predict the demand of a certain model of automobile company; according to the nonlinear characteristics of the hotel occupancy rate, ZHANG et al. [23] took Beijing travel-related consumer search data from January 2011 to April 2017 as the input set and constructed a hybrid model integrating consumer search data and SVR. They used Bat Algorithm (BA) to optimize the parameters of SVR, which effectively improves the prediction accuracy of the model.

Through the analysis of the above researches, we find that preview researches were mainly based on linear models to predict hotel demand [1, 4, 5, 11–13, 17, 19]. However, the linear model relies on the stability of the time series and the economic environment, and it is difficult to effectively simulate the nonlinear characteristics of hotel demand. Among the existing nonlinear hotel demand prediction models, SVR [3, 18, 23] is widely used. It is good at using small sample data for prediction and has good processing ability for nonlinear data, but the selection of parameters has a great influence on the prediction results. In addition, the traditional nonlinear prediction models are often unable to deal with the increasing data in practical applications and lack the ability to automatically extract data features.

At present, some scholars also build deep learning models for demand prediction. For example, Chang and Tsai [24] addressed the problem faced by neural network and SVR and proposed the deep learning neural network to predict the tourist arrivals; the result showed that the deep learning applied neural network with feature selection attained the best testing accuracy. Although the deep learning network proposed in this research can carry out feature selection, its prediction accuracy needs to be improved. Zhang et al. [6] did experiments based on a deep learning framework and search index from August 2008 to May 2019 to predict the overnight passenger flows for hotels accommodation in Hainan Province, China, and then constructed an LSTM model incorporating Internet search index to handle the prediction problem in the hotel accommodation demands, which shows good performance in improving the prediction performance. However, not all customers will know the relevant information of the tourist destination through the search engine before booking the hotel, and this method can only predict the number of bookings but cannot predict the actual occupancy, which has certain incompleteness and uncertainty.

The above research methods on hotel demand prediction mainly have one or more of the following problems:

- (1) They rely too much on the stability of environment and time series
- (2) Characteristics of complex nonlinear fluctuations of hotel passenger flow cannot be extracted automatically
- (3) They cannot deal with a large amount of sample data in practical application
- (4) It is incomplete and uncertain to estimate hotel passenger flow based on customer search data

Neural network models with multiple hidden layers show strong superiority in learning characteristic information and correlation of complex datasets. Compared with other models, the prediction model based on deep learning methods can make demand prediction more accurately. In addition, LSTM model has more significant advantages than other deep learning models in terms of prediction with sequences as inputs [25, 26].

Based on the above analysis, we use the historical booking demand dataset of a resort hotel in Portugal [27] and establish a deep learning model of hotel passenger flow prediction based on LSTM to predict the number of actual monthly arrival bookings of this hotel. This model can automatically and effectively extract the complex nonlinear characteristics of the hotel passenger flow data and make more accurate analysis and prediction.

3. Booking Demand Data Analysis

The dataset used in this paper is from paper [27]. It contains booking demand data for a resort hotel from July 1, 2015, to August 31, 2017. A total of 31 variables describe the 40060 observations of the resort, including “is_canceled,” “lead_time,” “arrival_date_year,” “arrival_date_month,” “arrival_date_week_number,” “reservation_status,” and “reservation_status_date.” “reservation_status” includes three categories: “Canceled,” which represents that the booking was canceled by the customer, “Check-Out,” which represents that the customer has checked in but already departed, and “No-Show,” which represents that the customer did not check in and did inform the hotel of the reason why. The paper predicts the number of actual monthly arrival bookings of hotel, that is, the number of “Check-Out” occurrences of the “reservation_status” in a certain month.

Not all variables are highly associated with the change of “reservation_status.” We use CorrelationAttributeEval as Attribute Evaluator and Ranker as Search Method and select most correlated variables with “reservation_status.” We can see them in Table 1.

In Table 1, “deposit_type” indicates if the customer made a deposit to guarantee the booking. This variable has the greatest impact on the final booking status; it can be divided into three categories: “No Deposit,” which means that no deposit was made, “Non Refund,” which means that a deposit was made in the value of the total stay cost, and “Refundable,” which means that a deposit was made with a value under the total cost of stay. “lead_time” represents the number of days that elapsed between the entering date of the booking and the arrival date; “total_of_special_requests” represents the number of special requests made by the customer; “required_car_parking_spaces” represents the number of car parking spaces required; “country” indicates which country the customer comes from; “distribution_channel” represents booking distribution channel, where “TA” represents Travel Agents and “TO” represents Tour Operators; “assigned_room_type” is the room type assigned to the customer; the value of this variable may be different from the room type that the customer has booked, and this is

TABLE 1: Correlation with “reservation_status.”

Variable	Correlation
deposit_type	0.47636
lead_time	0.29455
total_of_special_requests	0.23246
required_car_parking_spaces	0.19226
Country	0.17389
distribution_channel	0.1676
assigned_room_type	0.15666
booking_changes	0.14338
customer_type	0.1254
previous_cancellations	0.11008

because sometimes the room type assigned will be changed due to overbooking or customer requirements; “booking_changes” represents the number of changes made to booking from the moment the booking was entered until the moment of check-in or cancellation; “customer_type” represents the type of the customer; it has four categories: “Contract,” “Group,” “Transient,” and “Transient-party”; “previous_cancellations” represents the number of cancellations before this booking; it has the least influence on the final booking status of customers.

Table 2 shows some booking demand data of the resort hotel in July 2017. “d_t,” “l_t,” “t_o_s_r,” “r_c_p_s,” “c,” “d_c,” “a_r_t,” “b_c,” “c_t,” and “p_c” represent 10 variables in turn in Table 1; “a_d_y” represents “arrival_date_year,” “a_d_m” represents “arrival_date_month,” and “r_s” represents “reservation_status.”

From Table 2, we can see that, in July 2017, the hotel had a total of 9 bookings, among which 5 bookings were canceled and 1 booking did not arrive. The number of actual arrival bookings was 3. Usually, the number of hotel bookings can reflect its development trend. However, the booking may be canceled by the customer, or the customer fails to check in. As shown in Table 2, due to the change of customer status, the hotel may have 6 spare rooms, which has a negative impact on the hotel room allocation and pricing and ultimately leads to the loss of hotel revenue. If the number of bookings that actually arrive at the hotel can be estimated in advance, it can help the hotel to prejudge the allocation of rooms, so as to make a more reasonable pricing decision and operation strategy. Therefore, predicting the number of actual monthly arrival bookings has practical application value. The focus of this paper is to establish a prediction model to predict the number of actual arrival bookings in a certain month in the future.

4. Prediction Model

4.1. LSTM Network. What we want to predict in this paper is the number of actual arrival bookings in a certain month in the future, that is, the actual monthly passenger flow; this is a typical nonlinear time-series predicting problem. Time-series prediction analysis refers to using the time characteristics of an event in the past to predict the characteristics of the event in a certain period of time in the future, that is, predicting the future changes of an object according to the existing time-series data.

TABLE 2: Booking demand data.

a_d_y	a_d_m	d_t	l_t	t_o_s_r	r_c_p_s	c	d_c	a_r_t	b_c	c_t	p_c	r_s
2017	July	No deposit	59	0	0	USA	TA/TO	G	0	Transient	0	Cancelled
2017	July	No deposit	52	0	0	PRT	Corporate	D	0	Transient	0	Cancelled
2017	July	No deposit	17	0	0	PRT	Corporate	A	0	Transient	0	No-show
2017	July	No deposit	52	2	0	GBR	TA/TO	E	0	Transient	0	Check-out
2017	July	No deposit	3	1	0	ESP	TA/TO	A	0	Transient-party	0	Check-out
2017	July	Non-refund	17	0	0	PRT	Corporate	A	0	Transient-party	0	Cancelled
2017	July	Non-refund	0	0	0	GBR	TA/TO	E	0	Transient-party	0	Check-out
2017	July	Refundable	36	1	0	PRT	TA/TO	A	0	Transient	0	Cancelled
2017	July	No deposit	0	1	0	PRT	Direct	D	0	Transient	0	Cancelled

Recurrent Neural Network (RNN) is a commonly used algorithm in time-series prediction. It is a type of neural network with short-term memory capabilities. The connection of RNN can have loop structure, which can improve the accuracy of time behavior modeling in time series, text, audio, and other fields. A connection method introduced by RNN can take the input of hidden layer neurons as output and connect with neurons in the same hidden layer, so that the input can be obtained from the previous time step as part of the incoming neuron information. Therefore, the output of the network is not only related to the current input but also related to the output of the previous moment, which enables RNN to have the short-term memory ability when processing the time-series data of any length. Figure 1 shows the loop structure of RNN.

As can be seen from Figure 1, connections exist not only between neurons in adjacent layers (such as Hidden Layer 1 and Hidden Layer 2) but also between neurons in the same hidden layer in temporal dimension (such as neurons in $t=0$ time step and neurons in $t=1$ time step of Hidden layer 1). Suppose that the time step is t , the input of RNN is X_t , the neuron activity value of the hidden layer is Y_t , and the net input vector of the hidden layer is Z_t . Y_t is not only related to X_t but also related to the activity value Y_{t-1} of hidden neurons in the previous time step:

$$\begin{aligned} Z_t &= \tan h(W_Z X_t + R_Z Y_{t-1} + b_Z), \\ Y_t &= f(Z_t), \end{aligned} \quad (1)$$

where W is the rectangular input weight matrix, R is the square cyclic weight matrix, b is the bias vector, and f is the nonlinear activation function, usually set to $\tan h$ or sigmoid function. In equation (1), when $t=0$, $Y_t=0$.

Although RNN has some advantages in the field of time series, its long-term memory ability is weak. The gradient vanishing or gradient explosion that occurs when optimizing RNN in certain time steps makes it difficult to model the long-term structural dependence of the input dataset [6].

LSTM network is the most commonly used variant of RNN, which was proposed by Hochreiter and Schmidhuber in 1997 [25]. LSTM network is better at capturing long-term dependencies than regular RNN models. The LSTM network provides a solution for fusing memory cells, allowing learning of previously forgotten hidden cells, and updating the hidden cells based on new information [6]. LSTM network is composed of many LSTM cells, the main body of

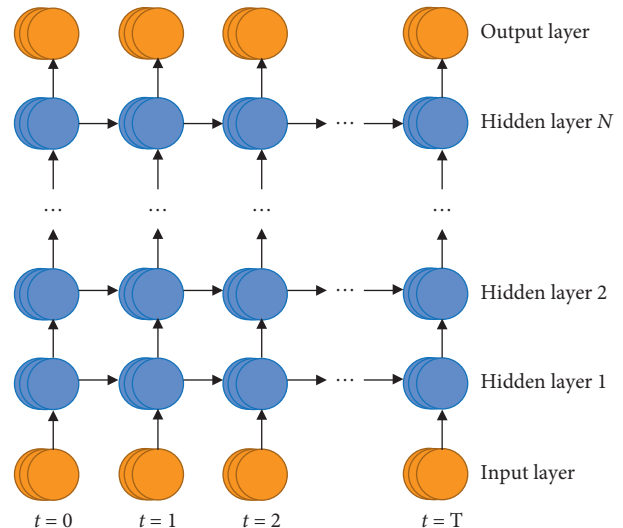


FIGURE 1: RNN structure unfolded along time axis.

which is LSTM block, including input gate, forgetting gate, output gate, fast input, memory cell, output activation function, and peephole connection. Memory cell, forgetting gate, and input gate are the key components of LSTM network. The content of memory cell is adjusted by forgetting gate and input gate. When both gates are closed, the content of memory cell will remain unchanged from one time step to the next. The gate structure allows information to be retained across multiple time steps and also allows gradients to be transmitted across multiple time steps, so that the LSTM network overcomes the gradient vanishing problem of RNN [28]. Refer to [28]; Figure 2 shows the structure of LSTM block.

The output and the input of the LSTM block are cyclically connected with the input gate, output gate, and forgetting gate. In Figure 2, the orange rectangle represents the input activation function, and the blue rectangle represents the output activation function; they are usually $\tan h$. ● (small black circle) represents the branch point, ⊗ represents dot product, ⊕ is on behalf of all the input, the thin line represents the connection with no weight, the thick line represents the connection with weight, and the dotted line represents the connection with time delay. The three green rectangles represent the input gate, the output gate, and the forgetting gate. They usually use sigmoid activation function to restrict $[0,1]$, where the activation output of 0 means

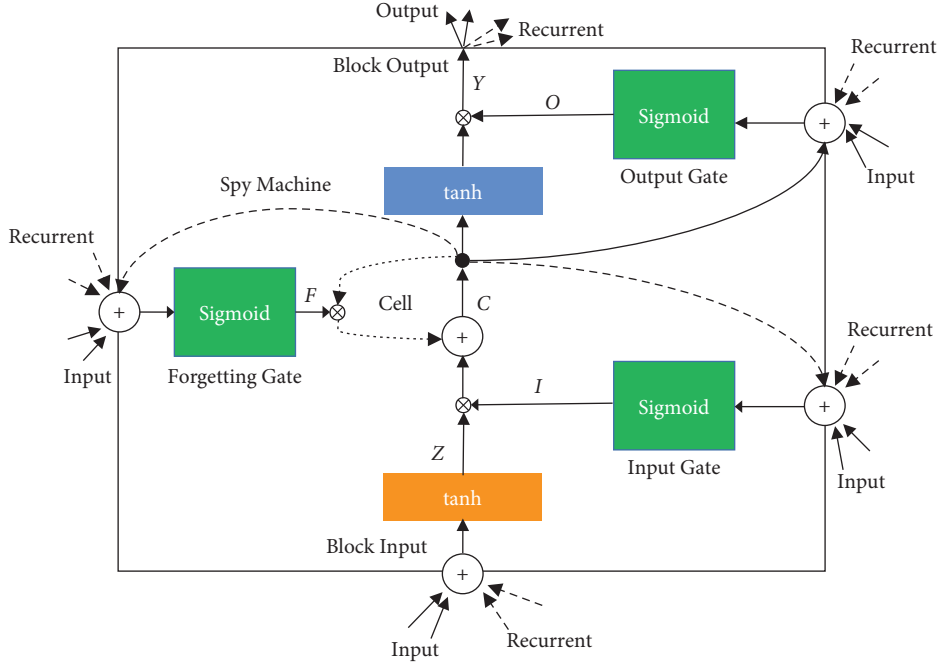


FIGURE 2: Structure of LSTM block.

“forget everything” and the activation output of 1 means “remember everything.”

References [6, 26, 28] give the vector equations of forward transfer of LSTM layer as follows:

$$\begin{aligned}
 I_t &= \sigma(W_I X_t + R_I Y_{t-1} + P_t \otimes C_{t-1} + b_I), \\
 F_t &= \sigma(W_F X_t + R_F Y_{t-1} + P_F \otimes C_{t-1} + b_F), \\
 C_t &= I_t \otimes Z_t + F_t \otimes C_{t-1}, \\
 O_t &= \sigma(W_O X_t + R_O Y_{t-1} + P_O \otimes C_t + b_O), \\
 Y_t &= O_t \otimes \tanh(C_t),
 \end{aligned} \tag{2}$$

where P represents the weight vector of the peephole, \tanh represents activation function, and $\tanh(x) = ((e^x - e^{-x}) / (e^x + e^{-x}))$. Equation (2) represents the input gate, which remembers some current information and determines which value will be updated, protecting the cell from unrelated input events. Equation (2) represents the forgetting gate, which controls how much data is discarded from the current memory state and helps the cell forget the previous memory content. Cell state C_t determines how much information to add or remove from the memory state of the previous time step through sigmoid activation function $\sigma(x) = (1 / (1 + e^{-x}))$ and dot product definition layer. Equation (2) represents the output gate, which controls whether the contents of the memory cell are exposed in the LSTM cell. $O_t \otimes \tanh(C_t)$ controls how much memory data will be used in the next phase of the update.

4.2. Prediction Model Construction. We build a deep learning prediction model with two hidden layers of LSTM. The activation function of all layers is \tanh , which has a more stable gradient and is often used for regression problem [6].

Sigmoid is selected as the gate activation function of the two LSTM hidden layers. The numbers of outputs of the LSTM layers are both 150. The stochastic gradient descent algorithm is used for optimization, and the ADAM algorithm is used for parameter iteration. To solve the overfitting problem of the model, we use the Dropout algorithm developed by Hinton et al. [29] and set the dropout of all LSTM hidden layers as 0.5. Dropout algorithm is a powerful tool to solve the overfitting problem of deep learning models.

Figure 3 shows the basic structure of the prediction model of hotel’s actual monthly passenger flow constructed in this paper.

The input layer is specified by the “input_shape” of the first hidden layer of the LSTM model, and its data is a three-dimensional data array. These three dimensions are as follows: sample (a time series), time step (an observation point in the sample represents a time step), and feature (an observation within a time step). In the prediction model in this paper, we set the time steps of the input layer and the feature to 1.

We use Output Layer as the output layer of the prediction model, which has a built-in fully connected Dense Layer. The number of outputs of Output Layer is set to 1, and the function Loss $l1$ is selected as loss function. Loss $l1$ is also called Minimum Absolute Value Deviation and Minimum Absolute Value Error; its purpose is to minimize the sum S of absolute differences between actual value x_i and estimated value \hat{x}_i .

$$S = \sum_{i=1}^n |x_i - \hat{x}_i|. \tag{3}$$

5. Experiments

5.1. Data Process. Our original dataset is booking demand data from a resort hotel in Portugal from July 1, 2015, to

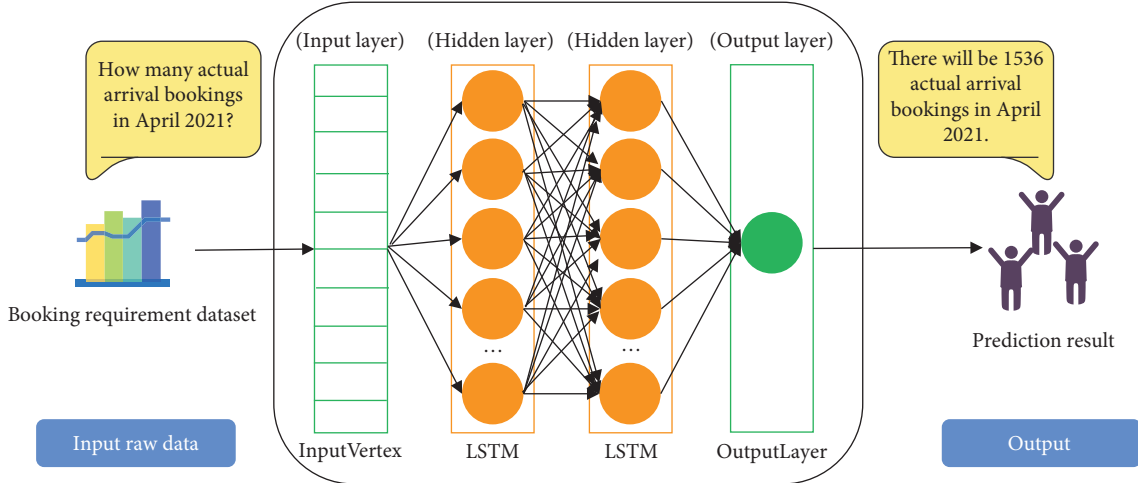


FIGURE 3: Prediction model structure.

August 31, 2017 [27]. To predict the number of actual arrival bookings in a certain month, firstly, we slice the raw data by time. The size of the time slice is set as one month, starting from July 1, 2015, to July 31, 2015, and ending from August 1, 2017, to August 31, 2017. We calculate the number of all bookings, actual arrival bookings, canceled bookings, and nonarrival bookings in each time slice. Then, we use April 1, 2017, as the time point to split the training set and the testing set; that is, 80% of the dataset are the training set and 20% are the testing set.

5.2. Benchmark Models and Experimental Setup. In order to evaluate the prediction effect of LSTM network under the dataset for hotel booking demand, ANN and SVR models are constructed as the benchmark models. The ANN model adopts a relatively simple single hidden layer network structure. SVR model is one of the most representative nonlinear prediction models. The introduction of SVR model proves that deep learning prediction model in this paper has more advantages than traditional machine learning prediction models in dealing with complex nonlinear data.

We build LSTM, ANN, and SVR models on WEKA 3.8.5 platform of Windows 10 system. In the experiment, we adjust the parameters of all models to a better state to ensure that they can achieve better prediction accuracy. The initial learning rate for all models was set to 0.001. Because the processed dataset is small, we set the batch size of the models to 4. The number of hidden layers in ANN model is 1. The kernel function of SVR model is set as RBF kernel, which contains two important parameters: c and γ . c is the penalty coefficient, which represents the tolerance to error. When c is larger, overfitting is more likely to occur, and when c is smaller, underfitting is more likely to occur. γ implicitly determines the distribution of data mapped to a new feature space. When it is larger, the number of support vectors is smaller, which affects the speed of training and testing. Here, we set c to 1.0 and γ to 0.05.

5.3. Evaluation Metrics. In this paper, we use the three following common model evaluation metrics to evaluate the prediction performance of LSTM, ANN, and SVR. They are mean absolute deviation (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE). MAE can well reflect the actual situation of the predicted value error. The larger the MAE is, the worse the effect of the model is; RMSE measures the deviation between the predicted value and the actual value, which is more affected by outliers than the other two evaluation metrics; the value range of MAPE is $[0, +\infty)$. When MAPE is 0%, the model is perfect; when MAPE is more than 100%, the model is poor.

Their calculation equations are as follows:

$$\begin{aligned} \text{MAE} &= \frac{1}{n} \sum_{i=1}^n |x_i - \hat{x}_i|, \\ \text{RMSE} &= \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2}, \\ \text{MAPE} &= \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - \hat{x}_i}{x_i} \right| \times 100\%, \end{aligned} \quad (4)$$

where x_i is the actual value, \hat{x}_i is the predicted value, and n is the number of test samples.

5.4. Result Analysis. We train three models on the training dataset. Table 3 shows the 10-month prediction results of LSTM, ANN, and SVR in the training set, and the best prediction values of each month are shown in red. As can be seen from Table 3, in the 10-month prediction results, LSTM model performs best, SVR performs second best, and ANN performs worst. The optimal prediction results of LSTM, ANN, and SVR are 5 months, 1 month, and 4 months, respectively.

In order to more intuitively compare the prediction effects of the three models on the training set, we draw the prediction curve of each model in Figure 4. The fitting effect

TABLE 3: Training predictive results of each model.

Time	The number of actual arrival bookings	LSTM	ANN	SVR
July 2016	985	980.4311	990.0974	986.3615
August 2016	1107	1106.7570	1115.8225	1096.7277
September 2016	1068	1069.1288	1106.3889	1069.8033
October 2016	1417	1270.1794	1361.7197	1253.9491
November 2016	1056	1057.2378	1069.8413	1056.4465
December 2016	1056	1056.2752	1076.8899	1055.1749
January 2017	1103	1103.1509	1051.7895	1102.1410
February 2017	1195	1192.0812	1163.5447	1193.9570
March 2017	1164	1161.6975	1229.6771	1164.1725
April 2017	1205	1203.0363	1187.0714	1219.8078

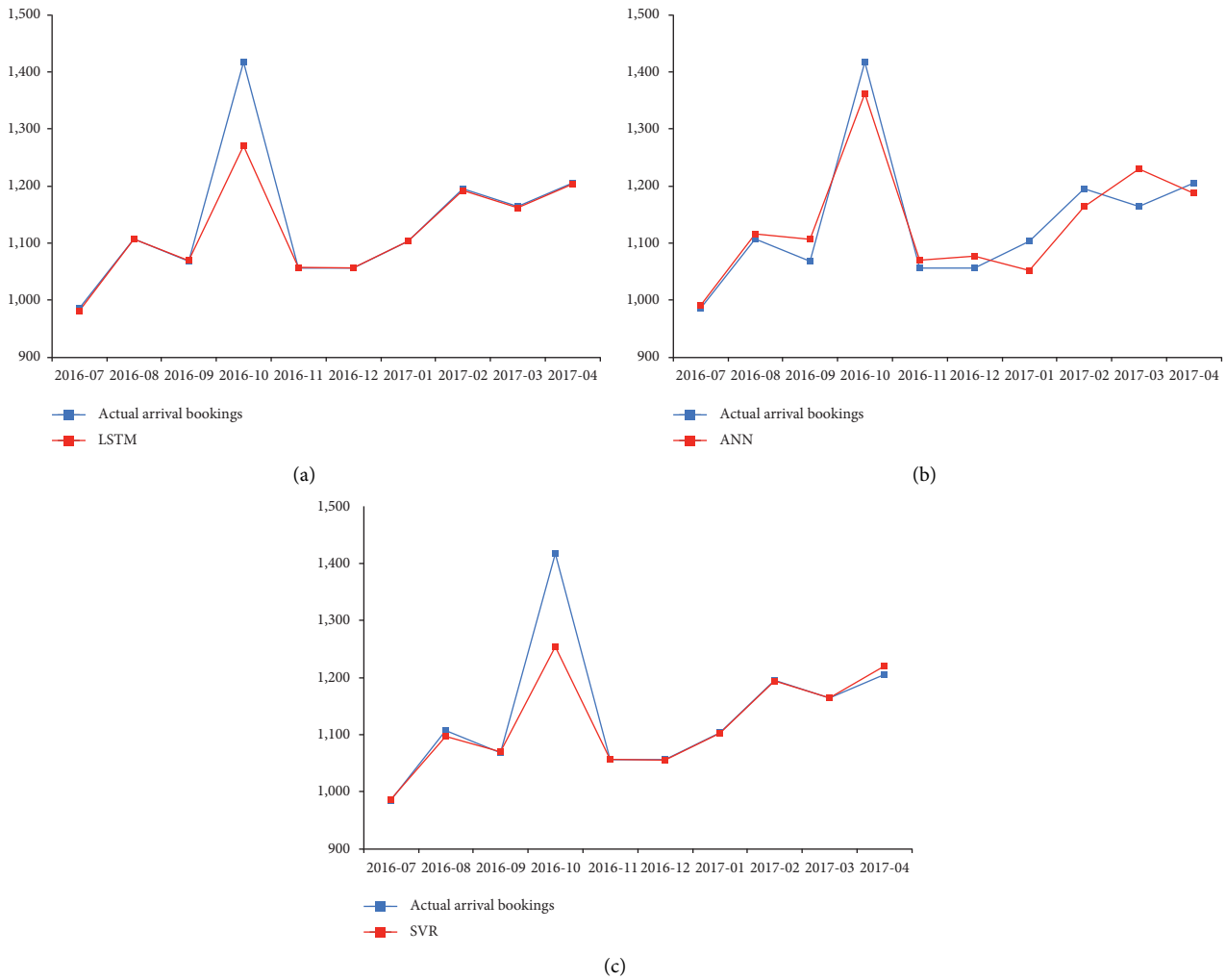


FIGURE 4: Comparison of the fitting curves of each model in training stage.

of LSTM and SVR model in October 2016 is worse than that of ANN model, but the fitting effect is good in the other nine months. The performance of SVR in August 2016 and April 2017 is slightly worse than that of LSTM model. The fitting effect of ANN model in July, August, and November 2016 is good, but, in January and March 2017, it shows a completely opposite trend to the actual passenger flow.

In general, LSTM model is more consistent with the dynamic characteristics of the actual arrival bookings per

month, followed by SVR model, and ANN model is the worst.

In order to evaluate the generalization ability of the model, the optimal structure of the model obtained after training was used as the prediction model for the prediction test. Table 4 shows the 5-month prediction results of LSTM, ANN, and SVR models on the testing set, and the best prediction value of each month is shown in red. As can be seen from Table 4, in the prediction results of 5 months, the

TABLE 4: Testing predictive results of each model.

Time	Actual arrival bookings	LSTM	ANN	SVR
April 2017	1205	1203.0363	1187.0714	1219.8078
May 2017	1212	1260.8806	1161.1423	1201.0106
June 2017	1045	1088.5516	1074.7438	1143.3589
July 2017	1094	1091.2047	1076.9058	1107.4611
August 2017	1107	1122.6816	1118.6439	1153.5216

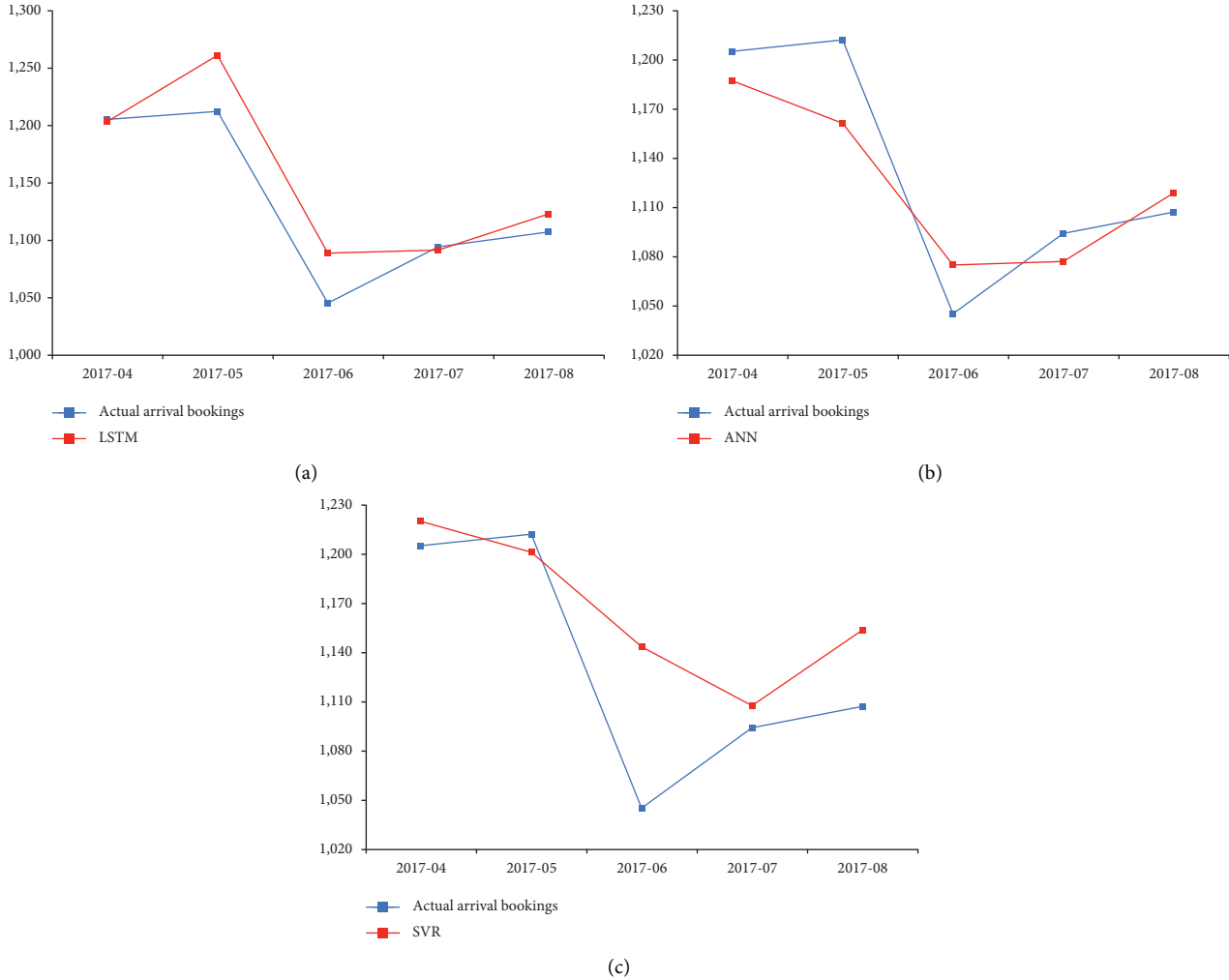


FIGURE 5: Comparison of the fitting curves of each model in testing stage.

LSTM and ANN models perform the best, while the SVR model performs the worst. The optimal prediction results of LSTM, ANN, and SVR are 2 months, 2 months, and 1 month, respectively.

In order to compare the prediction effects of LSTM, ANN, and SVR on the testing set more intuitively, we draw the prediction curve of each model in Figure 5. In general, the prediction effect of the three models on the testing set is worse than that on the training set. LSTM model has the best fitting effect in April 2017 and July 2017, ANN model has the best fitting effect in August 2017, and SVR model has the best fitting effect in May 2017 and the worst fitting effect in June 2017. As can be seen from Figure 5, from April 2017 to May

2017, the number of actual arrival bookings showed a slight upward trend, while the predicted results of ANN model and SVR model showed a downward trend, which is completely contrary to the reality. From June 2017 to July 2017, the actual arrival bookings showed an obvious upward trend, but the prediction results of the three models in this period showed poor fitting effect, and the prediction results of SVR model are completely opposite to the actual ones.

In general, LSTM and ANN models show better prediction effect than SVR model in testing stage.

In order to further explore the predictive ability of the LSTM model, MAE, RMSE, and MAPE are used to compare and evaluate the three models in the training stage and

TABLE 5: Evaluation metric scores for each model on the experimental dataset.

Model	MAE		MAPE (%)		RMSE	
	Training	Testing	Training	Testing	Training	Testing
LSTM	17.7385	22.5746	1.2835	2.0071	48.9829	30.1451
ANN	32.2959	25.4518	2.7796	2.2288	38.2780	29.0524
SVR	19.9816	36.8277	1.4491	3.3962	54.4661	49.7191

testing stage. Table 5 records the evaluation metric score of the three models on the experimental dataset, and the best score of each evaluation metric is shown in red.

It can be seen from Table 5 that LSTM model has the best scores on MAE and MAPE but not on RMSE. The overall fitting effect of ANN model on experimental datasets is poor. We can notice that the prediction accuracy of LSTM model in the testing set is close to that in the training set, which indicates that it has the best generalization ability. On the contrary, the performance of SVR model in the testing set is far from that in the training stage.

In the training stage, the score of MAE and MAPE of LSTM model is the lowest, followed by SVR model, and ANN model is the highest. In the testing stage, the scores of MAE and MAPE of LSTM model are still the lowest, followed by ANN model, and SVR model is the highest. As for RMSE, we know that RMSE is very sensitive to outliers, and if there is a predicted value that is very different from the actual value, the score will be very high. In the training stage, ANN model has the lowest score in this metric, the score of LSTM is slightly higher than that of ANN, and SVR is the highest. In the testing stage, the SVR has the highest score on this metric. As can be seen from Figures 4 and 5, in the training stage, the predicted values of LSTM and SVR in October 2016 deviate greatly from the actual values; in the testing stage, the predicted value of SVR in June 2017 deviates greatly from the actual value. Therefore, the RMSE scores of LSTM and SVR are higher than that of ANN model.

Based on the above analysis, LSTM model has a better performance than ANN and SVR models in predicting the number of actual monthly arrival bookings, which can better capture the complex nonlinear characteristics of hotel passenger flow and achieve a better fitting effect.

6. Conclusion

Hotel passenger flow is affected by weather, season, holidays, environment, and other factors, showing characteristics of a complex nonlinear fluctuation. Considering that the traditional demand prediction methods cannot automatically extract the characteristic information from the passenger flow data and cannot deal with the increasing sample data in the practical application, this paper builds a hotel passenger flow prediction model based on deep learning method. Taking a resort hotel in Portugal as an example, we construct an LSTM model with good predictive ability for complex time series to predict the number of actual monthly arrival bookings for the hotel. In order to explore the prediction ability of this model, we construct ANN and SVR as the

benchmark models in the experimental stage and compare the prediction effects of the three models on the datasets, with MAE, RMSE, and MAPE as the evaluation metrics. The experimental results show that, compared with the benchmark models, the LSTM model can better simulate the dynamic characteristics of hotel passenger flow and effectively improve the prediction performance and can help hotel managers make more accurate and reasonable pricing decisions and adjust operation mode.

Data Availability

All the data used in this study can be available upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

- [1] M. Zhang, *Hotel Occupancy Prediction Based on EEMD-ARIMA*, Shaanxi Normal University, Xi'an, China, 2016.
- [2] C. Xu, "Hotel revenue management: research overview and prospects," *Journal of Management Sciences in China*, vol. 29, no. 6, pp. 72–78, 2003.
- [3] W. Wu, "Prediction of hotel occupancy based on support vector regression," *New Technology & New Products of China*, vol. 3, no. 361, pp. 125–126, 2018.
- [4] L. R. Weatherford and S. E. Kimes, "A comparison of forecasting methods for hotel revenue management," *International Journal of Predicting*, vol. 19, no. 3, pp. 401–415, 2003.
- [5] J. Yu, *Model Established and Analysis of Passenger Flow Forecast in Hotel Management Decision Support System*, Dalian University of Technology, Dalian, China, 2008.
- [6] B. Zhang, Y. Pu, Y. Wang, and J. Li, "Forecasting hotel accommodation demand based on LSTM model incorporating internet search index," *Sustainability*, vol. 11, no. 17, p. 4708, 2019.
- [7] S. Y. Ji, "Hotel occupancy prediction based on singular spectrum analysis," *Information Recording Materials*, vol. 19, no. 3, pp. 49–51, 2018.
- [8] Y.-M. Zhang, "Research summary of demand forecasting based on hotel revenue management," *Sci-Tech Innovation & Productivity*, vol. 26, no. 7, pp. 7–12, 2019.
- [9] B.-R. Zhang, "Research on tourism demand forecast based on consumer search within internet environment," University of International Business and Economics, Beijing, China, 2017.
- [10] W. P. Andrew, D. A. Cranage, and C. K. Lee, "Forecasting hotel occupancy rates with time series models: an empirical analysis," *Hospitality Research Journal*, vol. 14, no. 2, 1990.
- [11] P. E. Pfeifer and S. E. Bodily, "A test of space-time ARMA modelling and forecasting of hotel data," *Journal of Forecasting*, vol. 9, no. 3, pp. 255–272, 1990.
- [12] B. Pan, D. C. Wu, and H. Song, "Forecasting hotel room demand using search engine data," *Journal of Hospitality and Tourism Technology*, vol. 3, no. 3, pp. 196–210, 2012.
- [13] J.-J. Yang, "Design of resort hotel passenger flow prediction system based on logistic model," *Journal of Changchun Normal University*, vol. 39, no. 8, pp. 54–59, 2020.
- [14] X. Sun, *Cruise Line Revenue Management: Demand Forecasting and Revenue Optimization*, Shanghai Jiao Tong University, Shanghai, China, 2011.

- [15] C. Goh and R. Law, "Modeling and forecasting tourism demand for arrivals with stochastic non-stationary seasonality and intervention," *Tourism Management*, vol. 23, no. 5, pp. 499–510, 2002.
- [16] C. Goh, "Exploring impact of climate on tourism demand," *Annals of Tourism Research*, vol. 39, no. 4, pp. 1859–1883, 2012.
- [17] J.-G. Choi, "Developing an economic indicator system (a forecasting technique) for the hotel industry," *International Journal of Hospitality Management*, vol. 22, no. 2, pp. 147–159, 2003.
- [18] W.-C. Hong, Y. Dong, L.-Y. Chen, and S.-Y. Wei, "SVR with hybrid chaotic genetic algorithms for tourism demand forecasting," *Applied Soft Computing*, vol. 11, no. 2, pp. 1881–1890, 2011.
- [19] A. Rashad, S. Sara, and A. Rafig, "Development of fuzzy time series model for hotel occupancy forecasting," *Sustainability*, vol. 11, no. 3, p. 793, 2019.
- [20] S. Sun, Y. Wei, K.-L. Tsui, and S. Wang, "Forecasting tourist arrivals with machine learning and internet search index," *Tourism Management*, vol. 70, pp. 1–10, 2019.
- [21] L. E. I Ke-wei and Y. Sheng, "Forecast of inbound tourists to China based on BP neural network and ARIMA combined model," *Tourism Tribune*, vol. 22, no. 4, pp. 20–25, 2007.
- [22] W. Gong and J. Huang, "A demand forecast model based on the gray theory and exponential smoothing method," *Statistics & Decision*, vol. 33, no. 1, pp. 72–76, 2017.
- [23] B.-R. Zhang, S.-L. Liu, C.-F. Zhang, and Y.-L. Pu, *Forecasting hotel occupancy rate based on consumer search within network environment*, Statistics & Information Forum, vol. 33, no. 3, 2018.
- [24] Y. Chang and C. Tsai, "Apply deep learning neural network to predict number of tourists," in *Proceedings of the 2017 31st International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, pp. 259–264, Taipei, Taiwan, 2017.
- [25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [26] R. Law, G. Li, D. K. C. Fong, and X. Han, "Tourism demand forecasting: a deep learning approach," *Annals of Tourism Research*, vol. 75, pp. 410–423, 2019.
- [27] N. Antonio, A. De Almeida, and L. Nunes, "Hotel booking demand datasets," *Data in brief*, vol. 22, 2019.
- [28] J. Patterson and A. Gibson, *Deep learning: A Practitioner's Approach by Josh Patterson and Adam Permissions*, O'Reilly Media, Inc., Sebastopol, CA, USA, 2017.
- [29] G. Hinton, L. Deng, D. Yu et al., "Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.