Annals of Tourism Research 83 (2020) 102923

Contents lists available at ScienceDirect

Annals of Tourism Research

journal homepage: www.elsevier.com/locate/annals

Research article

Daily tourism volume forecasting for tourist attractions

Jian-Wu Bi[a], Yang Liu[b], Hui Li[a,c,*]

[a] College of Tourism and Service Management, Nankai University, Tianjin 300350, China
[b] School of Business Administration, Northeastern University, Shenyang 110167, China
[c] MOE-Tianjin Collaborative Innovation Center of Modern Tourism Development, Nankai University, Tianjin 300350, China

ARTICLE INFO

Associate editor: Haiyan Song

Keywords:
Tourism volume forecasting
Long short-term memory networks
Search engine data
Weather data
Multivariate time series forecasting

ABSTRACT

A novel approach based on long short-term memory (LSTM) networks that can incorporate multivariate time series data, including historical tourism volume data, search engine data and weather data, is proposed for forecasting the daily tourism volume of tourist attractions. The proposed approach is applied to forecast the daily tourism volume of Jiuzhaigou and Huangshan Mountain Area, two famous tourist attractions in China. Through these two applications, the validity of the proposed approach is verified. In addition, the forecasting power of the approach with historical data, search engine data and weather data is stronger than that without search engine data or without both search engine data and weather data, which provides evidence that search engine data and weather data are of great significance to tourism volume forecasting.

Introduction

As one of the important research areas in tourism management, tourism volume forecasting has attracted much attention of both academics and practitioners (Song & Hyndman, 2011; Song & Li, 2008). Accurate tourism volume forecasting is an important part of efficient business operations and destination management (Palmer, Montano, & Sesé, 2006). Hitherto, many tourism volume forecasting models have been proposed (Song, Qiu, & Park, 2019). From the perspective of prediction granularity, most of these models focus on long-term forecasting (e.g., monthly, quarterly and annual) of relatively large areas (e.g., provinces, countries and regions). The prediction results of these studies can provide a reference for macro policy development. However, for a smaller area (e.g., a tourist attraction), short-term (e.g. daily) and high-frequency prediction is more important (Divino & McAleer, 2010; Pan & Yang, 2017). Based on the short-term (e.g. daily) prediction results, on the one hand, tourism managers can formulate tourism packages or pricing strategies to increase tourist arrivals in periods of low demand (Divino & McAleer, 2010); on the other hand, tourism managers can make staffing arrangements and develop emergency plans to prevent the occurrence of tourist detention in periods of high demand (Li, Chen, Wang, & Ming, 2018). However, studies on short-term volume forecasting of a relatively small area are very rare.

In this study, we mainly focus on daily tourism volume forecasting for tourist attractions. To obtain more accurate forecasting results, two factors should be considered, i.e., the predictor variable used in the forecasting, and the method or technology used in the forecasting. In terms of the first factor, historical tourism volume data (thereafter called historical data) are the most commonly used predictor variable in tourism volume forecasting, and acceptable predictions can often be obtained using this kind of predictor variable. However, the accuracy of the forecasted results using historical data is usually limited to a certain extent when unexpected events occur or on the special days (Yang, Pan, & Song, 2014). In recent years, some studies have verified a significant correlation

* Corresponding author at: College of Tourism and Service Management, Nankai University, Tianjin 300350, China.
E-mail addresses: jwbi@nankai.edu.cn (J.-W. Bi), liuy@mail.neu.edu.cn (Y. Liu), lihuihit@gmail.com (H. Li).

https://doi.org/10.1016/j.annals.2020.102923
Received 3 September 2019; Received in revised form 7 April 2020; Accepted 16 April 2020

between search engine data and tourism volume (Dergiades, Mavragani, & Pan, 2018; Yang et al., 2014; Yang, Pan, Evans, & Lv, 2015). On this basis, some scholars have tried to use the search engine data to forecast tourism volume, and relatively good results are achieved in these studies (Bangwayo-Skeete & Skeete, 2015; Gunter & Önder, 2016; Huang, Zhang, & Ding, 2017; Pan, Wu, & Song, 2012; Sun, Wei, Tsui, & Wang, 2019; Yang et al., 2015). Besides, for a tourist attraction, the weather data have a certain influence on the tourist volume (Falk, 2013; Falk, 2014; Meyer & Dewar, 1999; Otero-Giráldez, Álvarez-Díaz, & González-Gómez, 2012). For example, in general, the tourist volume of a tourist attraction during normal weather (e.g., sunny, cloudy) would be significantly greater than that in extreme weather (e.g., heavy rain). Therefore, the weather data should be considered when forecasting the daily tourism volume of a tourist attraction (Álvarez-Díaz & Rosselló-Nadal, 2010). However, the studies on simultaneously using of historical data, search engine data and weather data to forecast the daily tourism volume of a tourist attraction have not been found. In terms of the second factor, hitherto, the commonly used method or technology for tourism volume forecasting mainly includes three categories, i.e., time series analysis approaches, econometric approaches, and artificial intelligence approaches (Song & Li, 2008). When these approaches or technologies are used, the time lags of observations required to make predictions need to be diagnosed and specified from detailed analysis in a fixed form. However, the temporal dependence may change over different time periods in some cases, and the above traditional approaches can only use a fixed overall global temporal dependence of the observation. Obviously, the temporal dependence information is not fully utilized by the above traditional approaches or technologies, especially when multiple independent variables (or observations) are used. By contrast, long short-term memory (LSTM) networks are a kind of deep learning technology that can automatically learn the temporal context of input sequences to make better predictions. In other words, the time lags of observations no longer must be diagnosed and specified as in the above traditional approaches or technologies. Instead, the time lags of observations can be automatically learned, even the changes to the time lags can also be automatically learned. Therefore, LSTM networks are a powerful tool for tourism volume forecasting (Law, Li, Fong, & Han, 2019). Based on the analysis of the above two factors, if the approach based on LSTM networks that can simultaneously use historical data, search engine data and weather data is invented, then more accurate forecasting results could be expected.

In this study, a novel approach based on LSTM networks that can simultaneously use historical data, search engine data and weather data is proposed for forecasting the daily tourism volume of tourist attractions. The approach is composed of four stages, i.e., stage 1, collecting data; stage 2, preprocessing data; stage 3, training LSTM networks; and stage 4, predicting tourist volumes. In stage 1, the relevant data for tourism volume forecasting are collected. In stage 2, the collected data are converted into the data form that can be directly used by LSTM networks. In stage 3, the LSTM networks are trained based on the obtained preprocessed data. In the final stage, the future tourism volume can be predicted based on the trained LSTM networks. The proposed approach is applied to forecast the daily tourism volume of Jiuzhaigou and Huangshan Mountain Area (HMA), two famous tourist attractions in China. Through these two applications, the validity of the proposed approach is verified. In addition, the forecasting power of the approach with historical data, search engine data and weather data is stronger than that without search engine data or without both search engine data and weather data, which provide evidence that search engine data and weather data are of great significance to tourism volume forecasting.

Literature review

Models for tourism forecasting

Hitherto, a variety of tourism volume forecasting approaches have been proposed, such as time series analysis approaches (Athanasopoulos & Hyndman, 2008; Beneki, Eeckels, & Leon, 2012; Chan, Lim, & Mcaleer, 2005; Cho, 2001; Du Preez & Witt, 2003; Fildes, Wei, & Ismail, 2011; Goh & Law, 2002; Lim & McAleer, 2002), econometric approaches (Assaf, Li, Song, & Tsionas, 2018; Cao, Li, & Song, 2017; De Mello & Fortuna, 2005; Shan & Wilson, 2001; Shen, Li, & Song, 2009; Song, Witt, & Jensen, 2003; Turner & Witt, 2001), and artificial intelligence (AI) approaches (Song & Li, 2008), etc. Since the main purpose of this study is to propose an AI approach for forecasting the tourism volume of tourist attractions, this subsection only summarizes the related studies on tourism volume forecasting based on AI approaches. The introduction and review of time series analysis approaches and econometric approaches are not given here, which can be found in Song et al. (2019).

With the rapid development and wide application of AI approaches, they have also emerged in the field of tourism forecasting in recent years. The core idea of AI approaches is to train the AI algorithm using tourism volume related data, then a predictor that can fit the functional relationship between the actual tourism volume and its influencing factors can be obtained. Based on the obtained predictor, the future tourism volume can be forecasted (Silva, Hassani, Heravi, & Huang, 2019). The main advantage of AI approaches is that they not only do not require any assumptions of the data, but also have the characteristics of adaptability and nonlinearity that especially suitable for nonlinear prediction (Song & Li, 2008). According to the principle in forecasting, AI approaches can be further divided into two categories, i.e., shallow learning approaches and deep learning approaches. Shallow learning approaches refer to a kind of AI algorithms with relatively few hidden layers, which usually require manual rules to construct data features (Kon & Turner, 2005). The commonly used shallow learning approaches in tourism volume forecasting are back propagation neural networks (BPNN) (Burger, Dohnal, Kathrada, & Law, 2001; Law, 2000), support vector machines (SVM) (Claveria, Monte, & Torra, 2016; Sencheong & Turner, 2005), rough set approach (Au & Law, 2000; Law & Au, 2000), and extreme learning machine (Sun et al., 2019), etc. In contrast to the shallow learning approaches, the deep learning approaches refer to a kind of AI algorithms with relatively more hidden layers and complex structures, which can automatically learn the features from the data (Hochreiter & Schmidhuber, 1997). Although deep learning approaches have been successfully applied in many fields, studies on tourism forecasting using deep learning approaches are still very scarce. Only limited literatures related to this issue can be found

(Law et al., 2019). For example, Law et al. (2019) proposed a deep learning approach to meet the challenges in tourism forecasting when massive amounts of search intensity indices are adopted as tourism demand indicators. To verify the validity of the proposed approach, an empirical study is conducted based on the monthly Macau tourist arrival volumes. The results show that the proposed approach significantly outperforms BPNN and SVM. Since deep learning approaches have a good performance in tourism volume forecasting, and the research on tourism volume forecasting based on deep learning approaches is still very scarce, thus more tourism volume forecasting approaches based on deep learning need to be developed.

Tourism forecasting with search engine data

Information search is an important part of the traveler's decision-making process. With the development of the Internet, tourists can acquire tourism information more actively, anytime and anywhere. Web search data are gradually becoming one of the main sources of tourism information, which are valuable for analyzing tourists' decision-making processes and future behaviors. Therefore, more and more attentions have been paid to the application of web search data for tourism forecasting (Dergiades et al., 2018; Yang et al., 2014; Yang et al., 2015). The commonly used web search engine data are Google Trends or Baidu Index. On the one hand, studies have shown that the use of Google Trends can significantly improve the precision of tourist forecasting (Bangwayo-Skeete & Skeete, 2015; Gunter & Önder, 2016; Pan et al., 2012; Rivera, 2016). For example, Pan et al. (2012) investigated the usefulness of Google search data in forecasting demand for hotel rooms. On the other hand, Baidu is the most commonly search engine used in China, and studies have shown that the use of Baidu Index can significantly improve the precision of tourist forecasting in China (Huang et al., 2017; Li, Pan, Law, & Huang, 2017; Sun et al., 2019; Yang et al., 2015). For example, Yang et al. (2015) used search engine data to forecast the tourist arrivals to the Hainan Province of China. The results indicate that both Google Trends and Baidu Index can significantly improve forecasting performance. Besides, these types of search engine data are also compared, and they found that Baidu Index performed better than Google Trends since Baidu has a larger market share in China.

Tourism forecasting with weather data

Weather is an important factor in the tourism industry (Becken, 2013; Martín, 2005). Weather is a necessary condition and also a key attraction for travel (Pan & Yang, 2017). Studies have shown that there is a certain correlation between weather and tourist volume, and weather data can be used to predict tourist volume (Falk, 2013; Falk, 2014; Meyer & Dewar, 1999; Otero-Giráldez et al., 2012). However, studies on tourism forecasting using weather data are still relatively scarce to date. Only few literatures that directly relate to this issue can be found (Álvarez-Díaz & Rosselló-Nadal, 2010; Pan & Yang, 2017). For example, Álvarez-Díaz and Rosselló-Nadal (2010) fitted a transfer function model and an artificial neural network to forecast monthly British tourism demand for the Balearic Islands based on weather data. The results show that incorporating weather data can increase predictive power.

LSTM networks

Basic concepts

Long short-term memory networks are a kind of recurrent neural networks (RNNs). To better introduce LSTM, we first give a brief description of RNNs. RNNs are an improved version of traditional neural networks specially used to analyze time sequences data. In traditional neural networks, neurons receive information from the input layer and map the received information to the output layer through the hidden layers. In this process, there is no information feedback in the whole network. Due to the structural limitations, each input sample can only be processed separately in traditional neural networks, and there is no correlation between each input sample. Consequently, traditional neural networks cannot capture the dynamic temporal behaviour for time sequences data. Therefore, traditional neural networks are not effective in dealing with time sequences problems. However, many tasks are time sequences related, such as tourism volume forecasting. To obtain better prediction results, the time feature information of sequential time-series data needs to be mined and utilized. On the basis of the traditional neural networks, RNNs introduce a directional loop, which makes their underlying topology of inter-neuronal connections contain at least one cycle, as shown in Fig. 1(a). It can be seen from Fig. 1(a) that there are loops in RNNs, which allows information to be passed in the network from one step to the next. If the loops in RNNs are unrolled, the RNNs can be regarded as multiple copies of the same normal neural network, where each normal neural network passing a message to the next neural network, as shown in Fig. 1(b) and (c). Unlike normal neural networks, RNNs can process time sequences of inputs using their internal state (memory), which allows RNNs to exhibit dynamic temporal behaviour for time sequences data.

Although RNNs are well suited for processing sequence data, RNNs have their inherent problems in learning long-term dependencies, i.e., the vanishing and exploding gradients. To overcome the inherent problems of RNNs, LSTM networks are designed to learn long-term dependencies. LSTM networks are able to remember information for long periods of time and perform tremendously well on a large variety of practical problems. The LSTM networks have the same chain structure as the basic RNNs. The main difference between the LSTM networks and RNNs is that the module of LSTM networks has four layers, and the information is exchanged in a special way, as shown in Fig. 2. The core of the LSTM networks is the cell state, i.e., the horizontal line between C_{t-1} and C_t, as shown in Fig. 3. The cell state is a bit like a conveyor belt, which runs directly on the entire chain with only a few minor linear interactions. Information can easily flow along it.

With the input of new information, the cell state needs to be updated. The main purpose of updating the cell state is to remove the

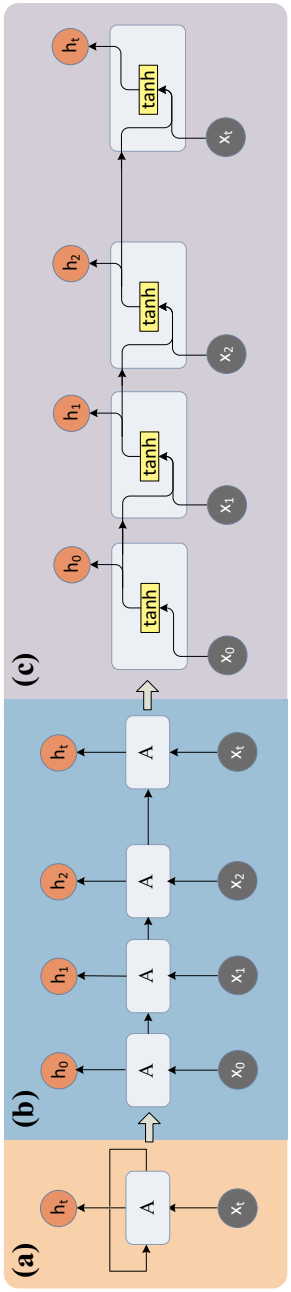

Fig. 1. The structure of RNNs.

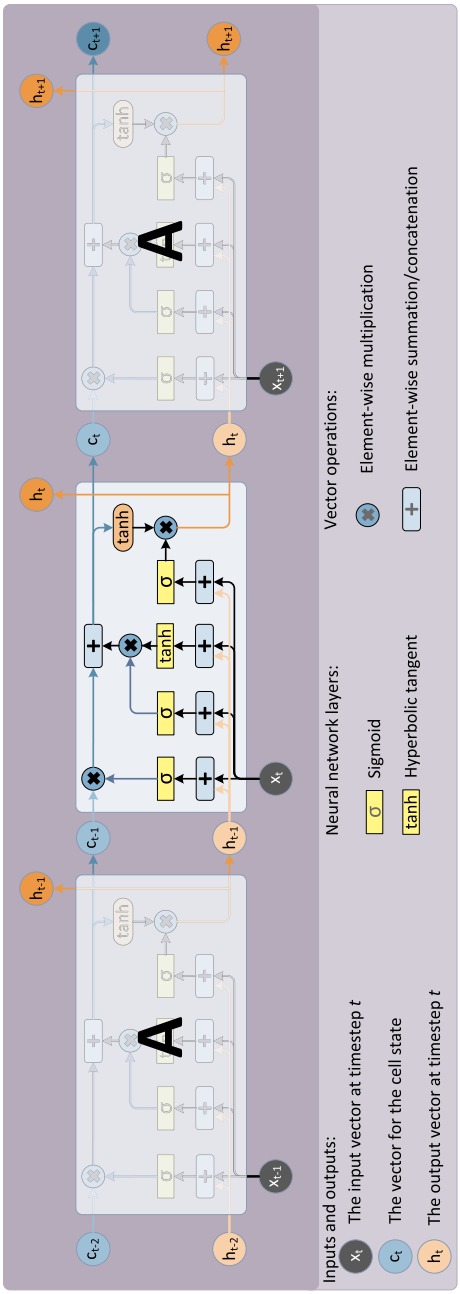

Fig. 2. The structure of LSTM networks.

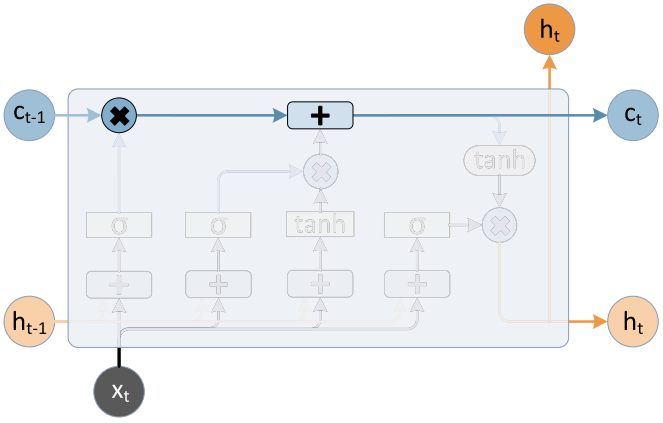

Fig. 3. The cell state of LSTM module.

information that has less impact on the prediction results, and to add the information that has a greater impact on the prediction results, so as to improve the prediction accuracy. To update the cell state, the structures called gates are designed in LSTM module. A gate is composed of a multiplication operation (◉) and a sigmoid neural network layer (▭), allowing information to optionally go through. The outputs of the sigmoid neural network layer are numbers between 0 and 1, describing how much information should be let through, where 0 and 1 respectively mean "let nothing through" and "let everything through".

The process of information processing in LSTM module

Based on the above basic concepts, the process of information processing in LSTM module is given in this subsection. The process is composed of the following four steps, as in Fig. 4.

Step 1. Determining what information should be removed from the cell state C_{t-1}. This process is implemented by a kind of gate in LSTM model called "forget gate", as shown in Fig. 4(a). It can be seen from Fig. 4(a) that the inputs of the sigmoid layer are h_{t-1} and x_t, and output a vector f_t for the activation values of the forget gate. The element in f_t is between 0 and 1, where 0 and 1 respectively represent "completely get rid of C_{t-1}" and "completely keep C_{t-1}". f_t can be calculated by Eq. (1), i.e.,

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{1}$$

where W_t and b_t are the weights matrices and biases vectors for f_t, respectively.

Step 2. Determining what information should be stored in the cell state. This step mainly includes two parts, as shown in Fig. 4(b). The left part is called the "input gate", which outputs a vector i_t for the activation values of the input gate, deciding which values should be updated. The right part determines the vector of new candidate values $\widehat{C}_t$ that could be added to the state by a tanh layer. Based on h_{t-1} and x_t, i_t and $\widehat{C}_t$ can be respectively calculated by Eqs. (2) and (3), i.e.,

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$\widehat{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{3}$$

where W_i and b_i are the weights matrices and biases vectors for i_t; W_C and b_C are the weights matrices and biases vectors for $\widehat{C}_t$, respectively.

Step 3. Updating the old cell state C_{t-1} into the new cell state C_t, as shown in Fig. 4(c). Based on the obtained f_t, i_t and $\widehat{C}_t$, the new cell state C_t can be updated by Eq. (4), i.e.,

$$C_t = f_t * C_{t-1} + i_t * \widehat{C}_t \tag{4}$$

Step 4. Determining the output of the LSTM module. This step includes two parts, as shown in Fig. 4(d). The left part is called the "output gates", which outputs a vector o_t for the activation values of the output gate, determining what parts of the cell state should output. o_t can be obtained by Eq. (5), i.e.,

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$

where W_o and b_o are the weights matrices and biases vectors for o_t, respectively.

The right part is to convert values in the vector of the cell state to -1 and 1 by a tanh function. By multiplying o_t and the converted values of cell state, the output h_t can be calculated by Eq. (6).

$$h_t = o_t * \tanh(C_t) \tag{6}$$

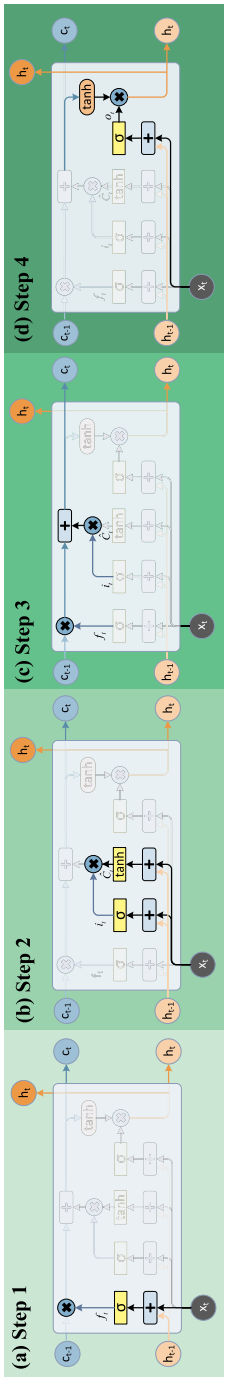

Fig. 4. The process of information processing in LSTM module.

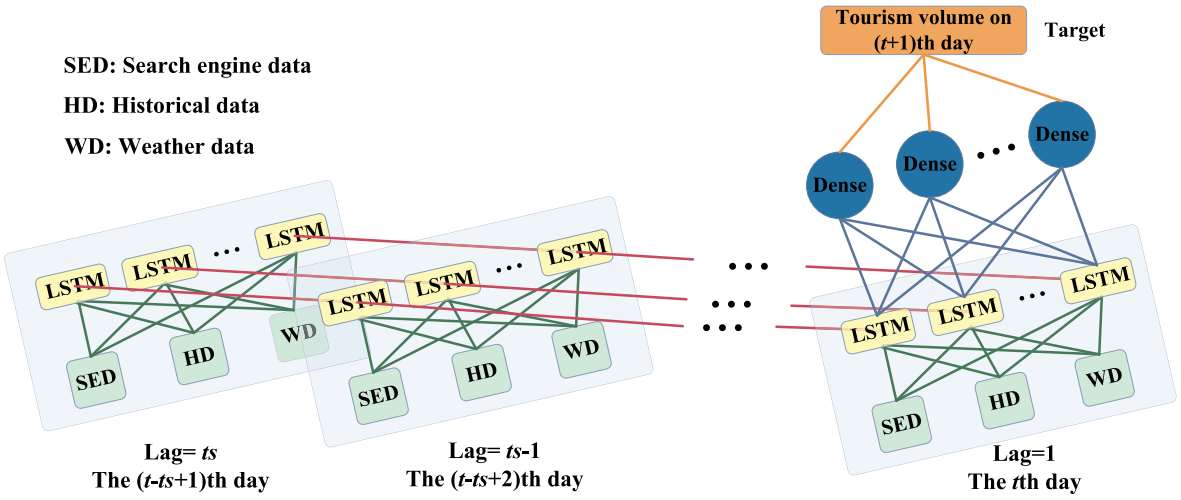

Fig. 5. The LSTM architecture for daily tourism volume forecasting.

Methodology

LSTM architecture for daily tourism volume forecasting

The historical data, search engine data and weather data are simultaneously considered in daily tourism volume forecasting for tourist attractions in this study. Based on the three kinds of data, a LSTM architecture for daily tourism volume forecasting is given, as shown in Fig. 5. In the figure, the output is the tourism volume on the $(t + 1)$th day, the inputs are the historical data, search engine data, and weather data from the $(t\text{-}ts + 1)$th day to the tth day, where the ts denotes the time step of the LSTM architecture, indicating that the number of the lagged days used in training the model. In the tsth time step (the $(t\text{-}ts + 1)$th day), the historical data, search engine data, and weather data of the day are connected to multiple LSTM modules at the same time, and each LSTM module will output a cell state according to the input data. This cell state can capture the contribution of current tourism volume, search data and weather data to the tourism volume on the day to be forecasted. Each cell state will carry some useful information to each subsequent LSTM module. On this basis, the cell state of each LSTM module on the $(t\text{-}ts + 2)$th day will be update based the inputted cell state and the tourism volume, search data and weather data on the $(t\text{-}ts + 2)$th day. Repeat this process until the tth day. Further, the cell state of each LSTM module on the tth day is connected to a dense layer. Finally, the dense layer is connected to the output layer, which is used to fit the tourism volume on the $(t + 1)$th day.

The promise of the LSTM architecture for daily tourism volume forecasting

As mentioned in the previous subsection, the historical data, search engine data and weather data should be simultaneously used in daily tourism volume forecasting for tourist attractions. When the above three kinds of data are considered at the same time, the data for daily tourism volume forecasting usually have the following characteristics: (1) They have some nonlinearity; (2) They are simultaneously affected by multiple explanatory variables, and the lag order or temporal dependence between each explanatory variable and the actual tourism volume may not be the same; (3) The lag order or temporal dependence between each explanatory variable and the actual tourism volume may vary with circumstance. Considering the above characteristics, some existing models, such as ARIMA, artificial neural networks (ANN), and support vector regression (SVR), cannot fully capture and mine the complex relationship between the explanatory variables and the actual tourism volume. As a result, these models may not be ideal for daily tourism volume forecasting. The reasons are discussed as follows. (1) The above three kinds of data need to be considered simultaneously in daily tourism volume forecasting, which will involve many explanatory variables. As the number of explanatory variables increases, this will make it difficult for some forecasting models to learn from training samples (Guyon & Elisseeff, 2003; Zhang, Zhang, & Yang, 2003). (2) When the above models are used to forecast daily tourism volume, it is necessary to determine the lead or lag order between each explanatory variable and the actual tourism volume. As mentioned above, there are many explanatory variables involved in daily tourism volume forecasting. Therefore, if the lead or lag order between each explanatory variable and the actual tourism volume is detected one by one, then extensive human effort will be required (Sun et al., 2019; Yang et al., 2015). (3) A fixed lead or lag order (temporal dependence) between each explanatory variable and the time series data is adopted in the above models (Law et al., 2019). However, the lead or lag order may vary with circumstance. For example, when an emergency occurs, the lag relationship between tourism volume and explanatory variables may be different from that of normal days. However, the above models can only use a fixed average lead or lag order to forecast daily tourism volume (Gers, Eck, & Schmidhuber, 2002; Sutskever, Vinyals, & Le, 2014). Obviously, the existing models cannot fully capture and mine the complex relationship between the explanatory variables and the actual tourism volume. It is necessary to point out that the problem of high dimension of independent variables may

be coped by the reduction dimension methods, such as factor decomposition and dynamic decomposition factors (Bräuning & Koopman, 2014). However, these methods can neither automatically identify the lead or lag order between each explanatory variable and the actual tourism volume nor fully utilize the changes of the lead or lag order. Therefore, considering the characteristics of daily tourism volume data, it is necessary to develop new forecasting models.

It should be noted that the LSTM architecture can not only capture the nonlinear relationship between explanatory variables and daily tourism volume, but also learn a mapping function for the explanatory variables over time to daily tourism volume. In addition to these general benefits, the LSTM architecture can also learn the temporal dependence from the data through the cell state, and a fixed set of lagged observations does not need to be specified (Hochreiter & Schmidhuber, 1997). More importantly, the lead or lag order that varies with circumstance can also be learned (Sutskever et al., 2014). Therefore, the LSTM architecture is a promising approach for daily tourism volume forecasting.

The approach for daily tourism volume forecasting

In this section, a novel approach based on the LSTM networks that can simultaneously use historical data, search engine data and weather data is proposed for forecasting the daily tourism volume. The forecasting approach is given in Fig. 6, which composed of four stages, i.e., stage 1, collecting data; stage 2, preprocessing data; stage 3, training LSTM networks; and stage 4, forecasting tourist volumes. The detailed descriptions of the four stages are given below.

Stage 1. Collecting data

This stage is mainly to collect the relevant data for tourism volume forecasting. The data mainly include the following three kinds, i.e., historical data, search engine data, and weather data. The collection process of these three kinds of data is given as follows.

(1) The collection of historical data.

Usually the destination or tourist attraction will count the daily tourist volume, thus the historical data can be obtained directly. Without loss of generality, we assume that the tourism volume on the $(T + 1)$th day is unknown and to be predicted. Then the historical tourist volumes from 1th day to Tth day are known, denoted as $H = (h_1, h_2, \ldots, h_T)$, where h_t represents the historical tourism volume of tth day, $t = 1, 2, \ldots, T$. Take the historical tourism volumes in Stage 1 of Fig. 6 as an example, $H = (h_1, h_2, \ldots, h_T) = (3130, 1523, \ldots, 10019)$.

(2) The collection of search engine data.

Tourists usually collect the relevant information of destinations before making travel decisions through the Internet, such as transportation, hotel and weather. Web search data are gradually becoming one of the main sources of travel information, which are valuable for analyzing tourists' decision-making processes and future behaviors.

The collection process of search engine data mainly includes three steps, i.e., (1) determining the keyword sets $KW = \{kw_1, kw_2, \ldots, kw_K\}$, where kw_k denotes the kth keyword, $k = 1, 2, \ldots, K$; (2) entering keyword kw_k into search tools such as Google Trends or Baidu Index to get the historical search volume of keyword kw_k, denoted as $S_k = (s_1^k, s_2^k, \ldots, s_T^k)$, where s_t^k represents the search volume of keyword kw_k on the tth day, $k = 1, 2, \ldots, K$, $t = 1, 2, \ldots, T$; (3) calculating the Pearson correlation coefficient between S_k and H with different lag periods, and eliminating the keywords with lower correlation coefficient. Finally, the set of keywords that are highly relevant to the historical tourist volumes can be obtained, denoted as $KW' = \{kw_1, kw_2, \ldots, kw_{K'}\}$, where $kw_{k'}$ denotes the k'th keyword in KW', $k' = 1, 2, \ldots, K'$, $K' \le K$. The historical search volume of keyword $kw_{k'}$ is $S_{k'} = (s_1^{k'}, s_2^{k'}, \ldots, s_T^{k'})$, where $s_t^{k'}$ represents the search volume of keyword $kw_{k'}$ on the tth day, $k' = 1, 2, \ldots, K'$, $t = 1, 2, \ldots, T$, $K' \le K$.

(3) The collection of weather data.

The history and future weather data of the tourist attractions can be obtained directly from the relevant weather website, such as World weather (https://www.worldweatheronline.com/) and Tianqi (http://lishi.tianqi.com/). In this study, two kinds of weather data are used, i.e., the weather condition and the temperature. The weather condition is denoted as $W_w = (w_1^w, w_2^w, \ldots, w_T^w, w_{T+1}^w)$, where w_t^w indicate the weather state on the tth day, $t = 1, 2, \ldots, T$. The temperature is described by two variables, i.e., the minimum temperature $(W_{mint} = (w_1^{mint}, w_2^{mint}, \ldots, w_T^{mint}, w_{T+1}^{mint}))$ and the maximum temperature $(W_{maxt} = (w_1^{maxt}, w_2^{maxt}, \ldots, w_T^{maxt}, w_{T+1}^{maxt}))$, where w_t^{mint} and w_t^{maxt} respectively indicate the minimum temperature and the maximum temperature on the tth day, $t = 1, 2, \ldots, T$. Since the weather conditions are nominal data, they cannot directly perform mathematical calculations. Before preprocessing the collected weather condition data, we first convert them into dummy variables according to their suitable degree for travel. The conversion rules are defined as shown in Table 1.

It should be noted that the way of using weather data is different from that of the using of historical data and search engine data in forecasting. The future weather data can be directly obtained through the weather forecast, which will have a certain degree of influence on tourism volume. Therefore, when making predictions, not only the historical weather data but also the future weather data are used.

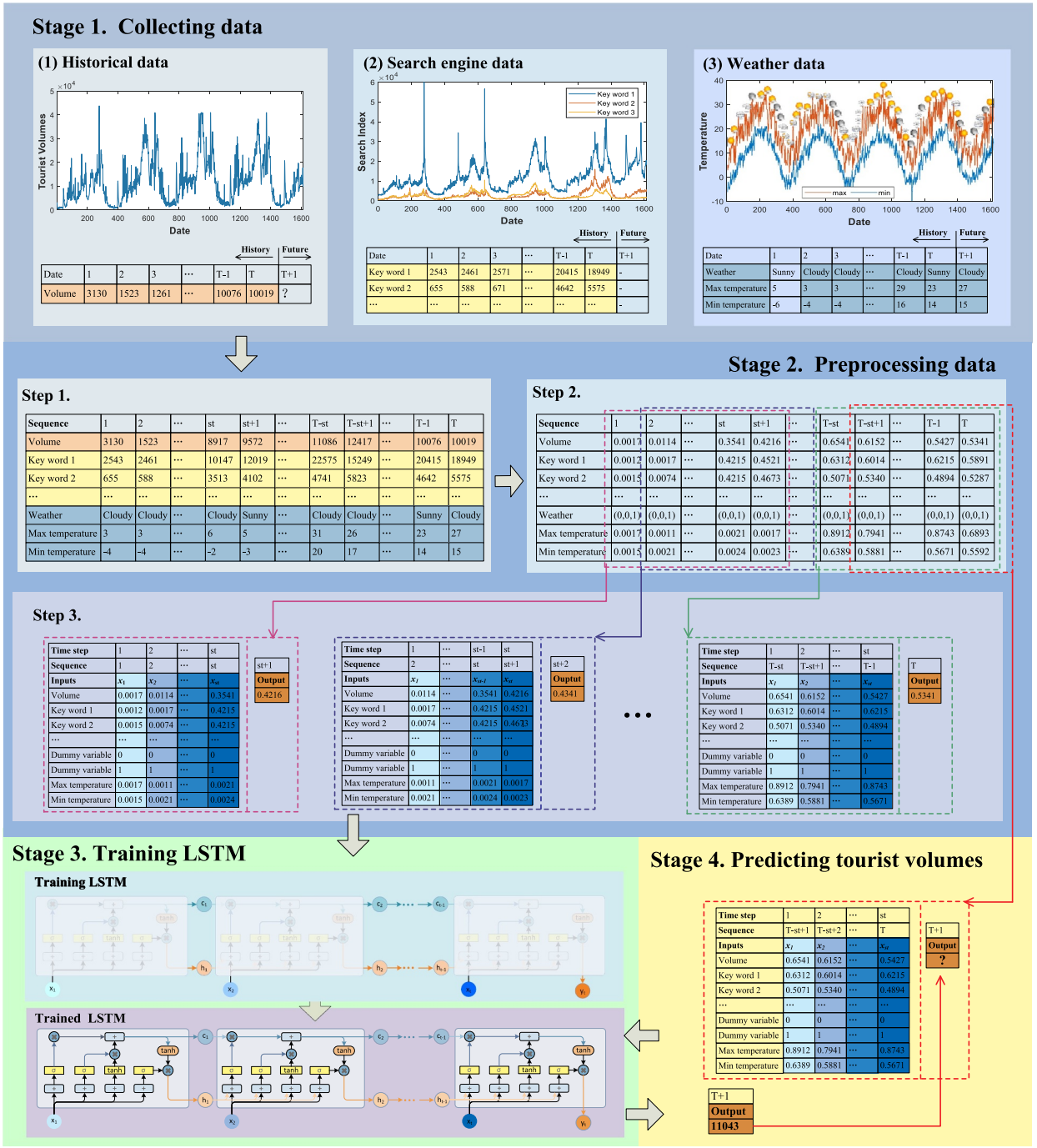

Fig. 6. The forecasting approach.

Table 1
The conversion rules of converting the weather condition data into dummy variables.

Weather condition	Dummy variables
Heavy/moderate rain/snow	(1,0,0)
Light rain/snow, shower and snow shower	(0,1,0)
Sunny, cloudy and overcast	(0,0,1)

Stage 2. Preprocessing data

This stage is mainly to convert the collected data into the data form that can be directly used by the LSTM networks, which mainly includes the following three steps.

(1) Combining the three kinds of data into one matrix $D_{J\times T}$, i.e.,

$$D_{J\times T}=\begin{bmatrix} d_1\\ d_2\\ d_3\\ \vdots\\ d_{J-2}\\ d_{J-1}\\ d_J \end{bmatrix}=\begin{bmatrix} H\\ S_1\\ S_2\\ \vdots\\ W_{\mathrm{w}}\\ W_{\min \mathrm{t}}\\ W_{\max \mathrm{t}} \end{bmatrix}=\begin{bmatrix} h_1, & h_2, & \ldots, & h_T\\ s_1^1, & s_2^1, & \ldots, & s_T^1\\ s_1^2, & s_2^2, & \ldots, & s_T^2\\ \vdots & \vdots & \vdots & \vdots\\ w_2^{\mathrm{w}}, & w_3^{\mathrm{w}}, & \ldots, & w_{T+1}^{\mathrm{w}}\\ w_2^{\min \mathrm{t}}, & w_3^{\min \mathrm{t}}, & \ldots, & w_{T+1}^{\min \mathrm{t}}\\ w_2^{\max \mathrm{t}}, & w_3^{\max \mathrm{t}}, & \ldots, & w_{T+1}^{\max \mathrm{t}} \end{bmatrix}$$

where d_j represents the jth predictor valuable in $D_{J\times T}$, J denotes the total number of predictor variables, $j = 1, 2, \ldots, J$. It should be noted that when constructing the matrix $D_{J\times T}$, the historical data and search engine data from the 1th day to the Tth day are used, while the weather data from the 2th day to the $(T + 1)$th day are used.

(2) Data normalization

Since the obtained three kinds of data belong to different types of information, they are incommensurate. To improve the forecasting accuracy, the data need to be normalized. Specifically, each line d_j in $D_{J\times T}$ can be normalized by the following Eq. (7), i.e.,

$$\widehat{d}_{jt}=\frac{d_{jt}-\min(d_j)}{\max(d_j)-\min(d_j)},\ j=1,\ 2,\ \ldots,J,\ t=1,\ 2,\ \ldots,T \tag{7}$$

where d_{jt} represents the tth element in d_j, $\widehat{d}_{jt}$ represents the normalized result of d_{jt}.

Based on the obtained $\widehat{d}_{jt}$, $D_{J\times T}$ can be further expressed as $\widehat{D}_{J\times T}$, i.e.,

$$\widehat{D}_{J\times T}=\begin{bmatrix} \widehat{d}_1\\ \widehat{d}_2\\ \widehat{d}_3\\ \vdots\\ \widehat{d}_{J-2}\\ \widehat{d}_{J-1}\\ \widehat{d}_J \end{bmatrix}=\begin{bmatrix} \widehat{H}\\ \widehat{S}_1\\ \widehat{S}_2\\ \vdots\\ \widehat{W}_{\mathrm{w}}\\ \widehat{W}_{\min \mathrm{t}}\\ \widehat{W}_{\max \mathrm{t}} \end{bmatrix}=\begin{bmatrix} \widehat{h}_1, & \widehat{h}_2, & \ldots, & \widehat{h}_T\\ \widehat{s}_1^1, & \widehat{s}_2^1, & \ldots, & \widehat{s}_T^1\\ \widehat{s}_1^2, & \widehat{s}_2^2, & \ldots, & \widehat{s}_T^2\\ \vdots & \vdots & \vdots & \vdots\\ \widehat{w}_2^{\mathrm{w}}, & \widehat{w}_3^{\mathrm{w}}, & \ldots, & \widehat{w}_{T+1}^{\mathrm{w}}\\ \widehat{w}_2^{\min \mathrm{t}}, & \widehat{w}_3^{\min \mathrm{t}}, & \ldots, & \widehat{w}_{T+1}^{\min \mathrm{t}}\\ \widehat{w}_2^{\max \mathrm{t}}, & \widehat{w}_3^{\max \mathrm{t}}, & \ldots, & \widehat{w}_{T+1}^{\max \mathrm{t}} \end{bmatrix}$$

(3) Constructing training samples.

Let ts denote the time step of the LSTM networks. According to the obtained $\widehat{D}_{J\times T}$, $T - ts$ training samples can be constructed, denoted as TD_1, TD_2, …, TD_{T-ts}, where TD_g represents the gth training sample, $g = 1, 2, \ldots, T - ts$. TD_g can be further expressed in the form of a matrix, i.e.,

$$TD_g=\begin{array}{c} 1,\ 2,\ \ldots,\ ts\ \ |ts+1\\ \left[\begin{array}{cccc|c} \hat{h}_g, & \hat{h}_{g+1}, & \ldots, & \hat{h}_{g+ts-1} & \\ \hat{s}_g^1, & \hat{s}_g^1, & \ldots, & \hat{s}_{g+ts-1}^1 & \\ \hat{s}_g^2, & \hat{s}_g^2, & \ldots, & \hat{s}_{g+ts-1}^2 & \hat{h}_{g+ts}\\ \vdots & \vdots & \vdots & \vdots & \\ \hat{w}_{g+1}^{\mathrm{w}}, & \hat{w}_{g+2}^{\mathrm{w}}, & \ldots, & \hat{w}_{g+ts}^{\mathrm{w}} & \\ \hat{w}_{g+1}^{\min \mathrm{t}}, & \hat{w}_{g+2}^{\min \mathrm{t}}, & \ldots, & \hat{w}_{g+ts}^{\min \mathrm{t}} & \\ \hat{w}_{g+1}^{\max \mathrm{t}}, & \hat{w}_{g+2}^{\max \mathrm{t}}, & \ldots, & \hat{w}_{g+ts}^{\max \mathrm{t}} & \end{array}\right] \end{array},\ g=1,\ 2,\ldots,T-ts$$

where the data corresponding to the 1th to tsth column are the inputs of the LSTM networks, and the data corresponding to the $(ts + 1)$th column is the output of the LSTM networks.

Taking the data in Stage 2 of Fig. 6 as an example, if $g = 1$, then TD_g can be expressed as

$$TD_1 = \begin{array}{c} 1, \quad 2, \quad \ldots, \quad ts \quad | ts+1 \\ \left[\begin{array}{cccc|c} 0.0017, & 0.0114, & \ldots, & 0.3541 & \\ 0.0012, & 0.0017, & \ldots, & 0.4215 & \\ 0.0015, & 0.0074, & \ldots, & 0.4215 & 0.4216 \\ \vdots & \vdots & \vdots & \vdots & \\ 0, & 0, & \ldots, & 0 & \\ 1, & 1, & \ldots, & 1 & \\ 0.0017, & 0.0011, & \ldots, & 0.0021 & \\ 0.0015, & 0.0021, & \ldots, & 0.0024 & \end{array}\right] \end{array}$$

where the 0.4216 corresponding to the $(ts + 1)$th column is the output of the LSTM networks, and the data corresponding to the 1th to tsth column are the inputs of the LSTM networks.

Stage 3. Training LSTM networks

In accordance with the obtained TD_1, TD_2, …, TD_{T-ts}, the LSTM networks can be trained. The inputs of the LSTM networks are the data corresponding to the 1th to tsth column in each training sample, and output of the LSTM networks is the data corresponding to the $(ts + 1)$th column in each training sample. By training the LSTM networks using the T-ts constructed training samples, the LSTM networks with actual prediction capabilities can be obtained. The detailed training principle and process of the LSTM networks can be found in Hochreiter and Schmidhuber (1997).

Stage 4. Forecasting tourism volumes

In the third step of Stage 2 in Fig. 6, not only the $T - ts$ training samples, but also the input data (ID) for forecasting tourist volumes of the $(T + 1)$th day can be obtained. The form of ID is the same as that of TD_g, the only difference is that the value corresponding to the $(ts + 1)$th column in ID is unknown and needs to be forecasted. Take the data in Fig. 6 as an example, input data for forecasting the tourist volumes of the $(T + 1)$th day are

$$ID = \begin{array}{c} 1, \quad 2, \quad \ldots, \quad ts \quad | ts+1 \\ \left[\begin{array}{cccc|c} 0.6541, & 0.6152, & \ldots, & 0.5427 & \\ 0.6312, & 0.6014, & \ldots, & 0.6218 & \\ 0.5071, & 0.5340, & \ldots, & 0.4894 & ? \\ \vdots & \vdots & \vdots & \vdots & \\ 0, & 0, & \ldots, & 0 & \\ 1, & 1, & \ldots, & 1 & \\ 0.8912, & 0.7941, & \ldots, & 0.8743 & \\ 0.6389, & 0.5881, & \ldots, & 0.5671 & \end{array}\right] \end{array}$$

By inputting the ID into the trained LSTM networks in Stage 3, the tourism volumes of the $(T + 1)$th day can be obtained.

Experimental study

This paper takes two famous tourist attractions in China (Jiuzhaigou and Huangshan Mountain Area) as examples to verify the effectiveness of the proposed approach. The processes of the data collection and preprocessing, the performance measures, the parameter selection, the experimental results analysis and the model discussion will be given in the subsequent subsection, respectively.

Data collection and preprocessing

The collection and preprocessing processes of historical data, search engine data and weather data concerning the two tourist attractions are given as follows.

(1) Historical data.

We used the daily tourism volume released by the official website as the sample source. For Jiuzhaigou, a total of 1642 samples were collected from January 1, 2013 to June 30, 2017, as shown in Fig. 7. For Huangshan Mountain Area (HMA), a total of 1826 samples were collected from January 1, 2014 to December 31, 2018, as shown in Fig. 8.

(2) Search engine data.

According to the travel needs of tourists, several initial keywords are first selected, such as “Jiuzhaigou”, “Jiuzhaigou Weather”, “Jiuzhaigou Hotel” and “Jiuzhaigou Airport”. Then, the initial keywords are extended using the keywords' automatic recommendation technology of Baidu. Further, the Pearson correlation coefficients between the Baidu Indexes of obtained keywords and tourism volume with different lag periods are calculated. By comparing the obtained Pearson correlation coefficients with the

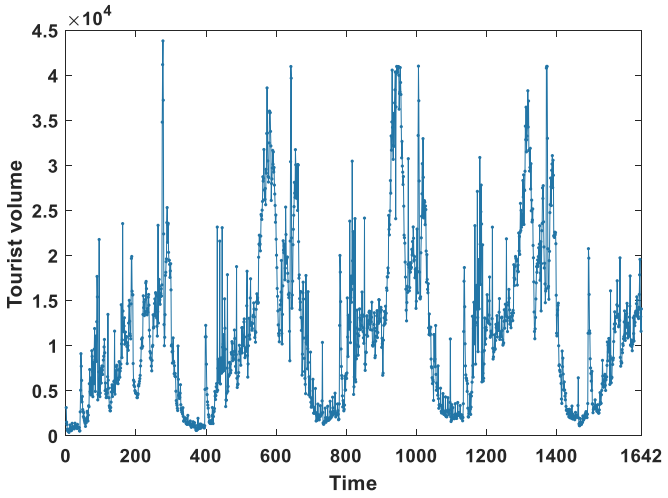

Fig. 7. Daily tourism volume of Jiuzhaigou from January 1, 2013 to June 30, 2017.

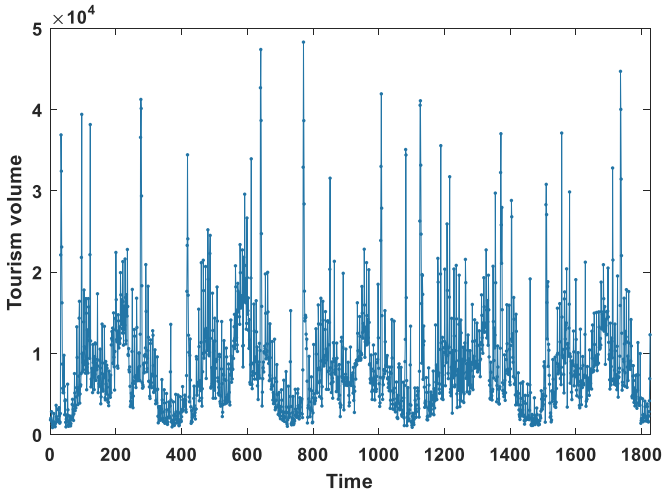

Fig. 8. Daily tourism volume of HMA from January 1, 2014 to December 31, 2018.

preset threshold value 0.6, finally seven keywords are obtained, i.e., “Jiuzhaigou”, “Jiuzhaigou weather”, “Jiuzhaigou map”, “Jiuzhaigou tourism strategy”, “Jiuzhaigou lodging”, “Jiuzhaigou hotel”, and “Jiuzhaigou tourism”. The Baidu indexes for these seven keywords from January 1, 2013 to June 30, 2017 are shown in Fig. 9. Similarly, the search engine data of HMA from January 1, 2014 to December 31, 2018 can be obtained, which are shown in Fig. 10.

(3) Weather data.

The weather data of the two tourist attractions were collected from the Tianqi website. The weather data of Jiuzhaigou from January 1, 2013 to June 30, 2017 and the weather data of HMA from January 1, 2014 to December 31, 2018 were obtained, which are shown in Figs. 11 and 12, respectively. In the figures, Figs. 11 (a) and 12 (a) are the weather condition data, Figs. 11 (b) and 12 (b) are the temperature data. Since the weather conditions cannot directly perform mathematical calculations, we first convert them into dummy variables using the conversion rules shown in Table 1.

Since the obtained three types of data are incommensurate, to improve the prediction accuracy, the data are normalized by Eq. (7). Then, according to the process of constructing training samples in Stage 2 of the proposed approach, the samples for training the LSTM networks can be constructed.

Performance measures

To evaluate the performance of different models, the mean absolute deviation (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE) are adopted in this study. The formulas for calculating MAE, RMSE and MAPE can be respectively

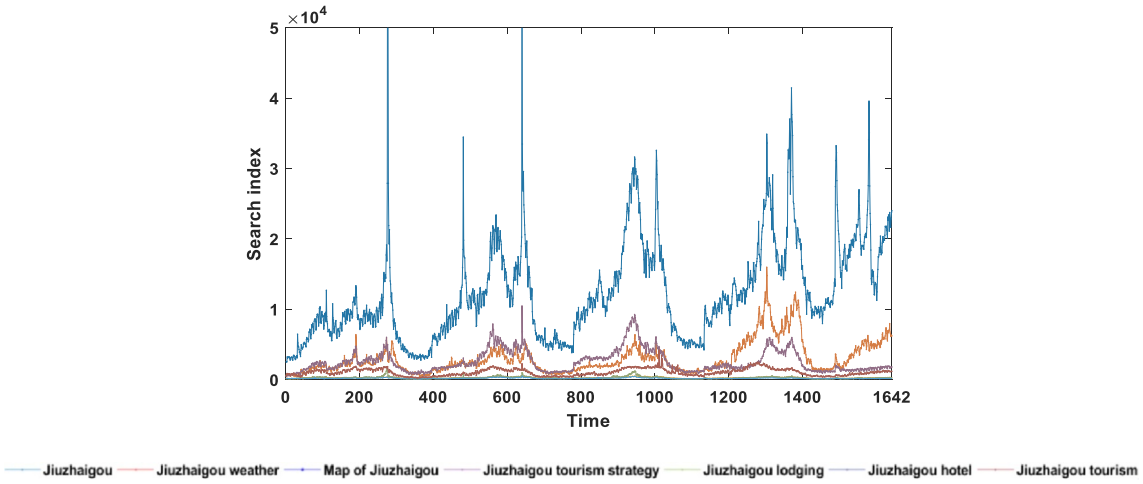

Fig. 9. The Baidu indexes for the keywords concerning Jiuzhaigou from January 1, 2013 to June 30, 2017.

represented by Eqs. (8)–(10), i.e.,

$$MAE = \frac{1}{N}\sum_{n=1}^{N}|x_n - x'| \tag{8}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(x_n - x_n')^2} \tag{9}$$

$$MAPE = \frac{1}{N}\sum_{n=1}^{N}\left|\frac{x_n - x_n'}{x_n}\right| \tag{10}$$

where x_n and x_n' are respectively denote the actual tourism volume and the forecasted tourism volume, N denotes the number of observations. Obviously, the smaller the values of MAE, RMSE and MAPE, the better the performance of the approach.

Parameter selection

For the two datasets, we respectively use the last one month as the test set to verify the performance of the proposed approach. In other words, for Jiuzhaigou, we use the data from January 1, 2013 to May 31, 2017 as training and validating samples to forecast the tourism volume for the next 30 days (June 1, 2017 to June 30, 2017); for HMA, we use the data from January 1, 2014 to November 30, 2018 as training and validating samples to forecast the tourism volume for the next 31 days (December 1, 2018 to December 31, 2018). When training the LSTM networks, several key parameters need to be set in advance, such as time step, number of units, batch size and epochs. In these key parameters, time step and number of units have a certain impact on forecasting performance of the LSTM networks. To find best values of the two parameters, an exhaustive grid search technique is adopted (Bergstra & Bengio, 2012). Here, we take Jiuzhaigou as an example to illustrate the process of determining the optimal parameters. The data from January 1, 2013 to April 30, 2017 are used as training samples and the data from May 1, 2017 to May 31, 2017 are used as validating samples. The training samples are used to training the LSTM networks with different combinations with respect to time steps and number of units, and the validating samples is used to evaluate the performance of LSTM networks with different combinations with respect to time steps and number of units.

In the experiment, the parameters of the LSTM networks are set as follows: time step $\in\{1,2,\ldots,12\}$, number of units $\in$ {1,5,10, …, 95,100}, batch size = 16 and epochs = 200. Since the training of the LSTM networks has a certain randomness, different forecasting results may be obtained by the trained LSTM networks with the same parameter combination. Therefore, to find best values of the two parameters, we run five times for each parameter combination. The average values of RMSE concerning different parameter combinations are calculated, as shown in Fig. 13.

In Fig. 13, the closer the color of the image is to dark blue, the smaller the RMSE of the LSTM networks trained using the parameter combination, i.e., the better the prediction results of the parameter combination. It can be seen from Fig. 13 that the RMSE shows a trend of decreasing first and then increasing with the time step and number of units increase. Finally, the best parameter combination of the proposed approach concerning Jiuzhaigou is obtained, i.e., time step = 5 and number of units = 55. Similarly, the best parameter combination of the proposed approach concerning HMA can also be obtained, i.e., time step = 7 and number of units = 50.

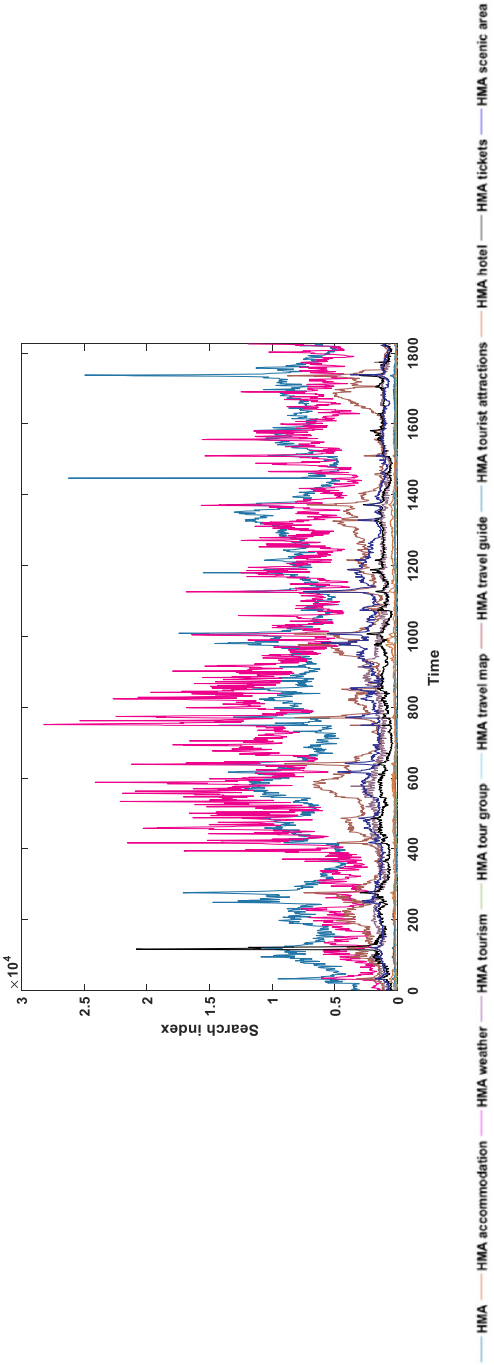

Fig. 10. The Baidu indexes for the keywords concerning HMA from January 1, 2014 to November 30, 2018.

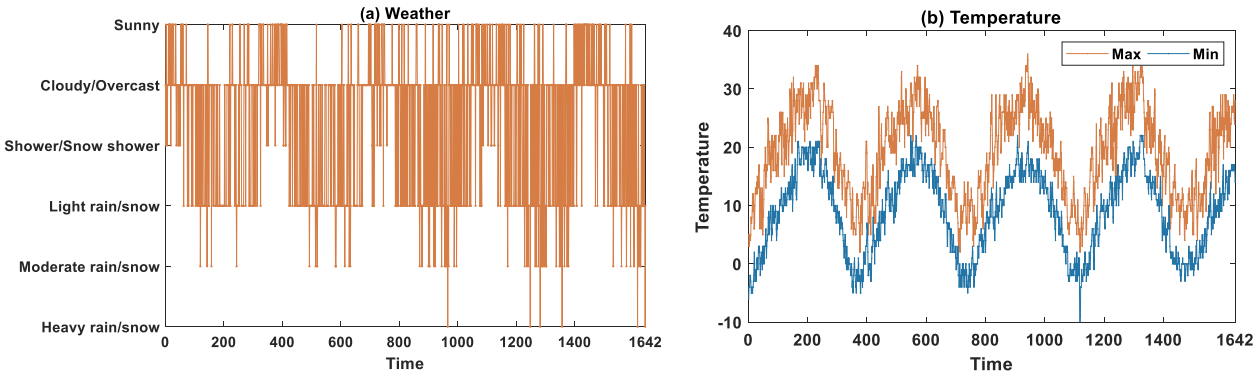

Fig. 11. The weather data of Jiuzhaigou from January 1, 2013 to June 30, 2017.

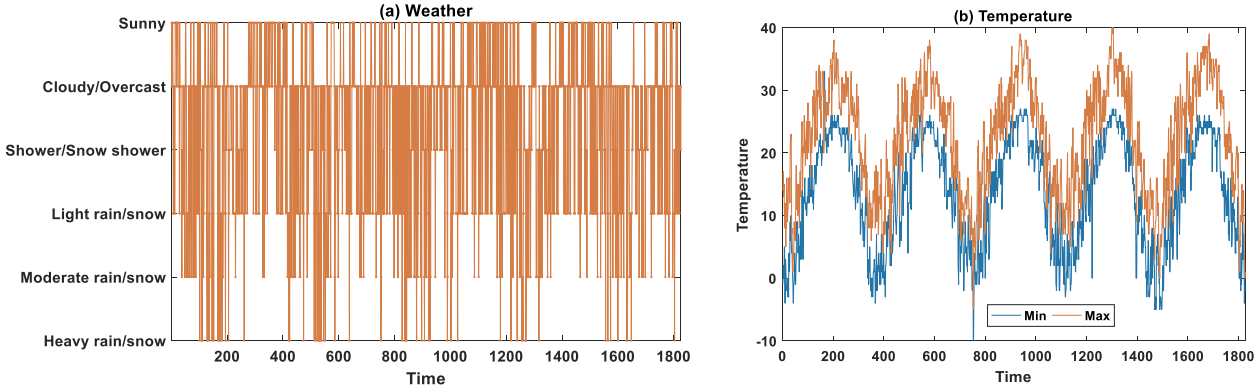

Fig. 12. The weather data of HMA from January 1, 2014 to November 30, 2018.

Experimental results analysis

Using the obtained parameters concerning Jiuzhaigou, the LSTM networks are trained based on the training samples of Jiuzhaigou. According to the trained LSTM networks, the tourism volume of Jiuzhaigou from June 1, 2017 to June 30, 2017 can be forecasted. Similarly, the tourism volume of HMA from December 1, 2018 to December 31, 2018 can also be forecasted. In addition, to verify the effectiveness of the proposed approach, ARIMAX, SVR and ANN are selected as the baselines. Similar to the process of the parameter selection in the LSTM networks, the optimal parameters of the SVR and ANN concerning the two tourist attractions are determined by the exhaustive grid search technique (Cho, 2003). Additionally, the optimal parameters of ARIMAX are determined through the statistical test method. The obtained parameters are shown in Table 2.

In accordance with the optimal parameters, the three baselines concerning Jiuzhaigou can be trained. On this basis, the tourism volume of Jiuzhaigou from June 1, 2017 to June 30, 2017 can be respectively forecasted using the above three baselines. In addition, the naïve method is also adopted to forecast the tourism volume of Jiuzhaigou. The forecasted tourism volume using different models are shown in Fig. 14. Similarly, the tourism volume of HMA from December 1, 2018 to December 31, 2018 can be forecasted using different models, as show in Fig. 15.

On the whole, compared with the baselines, the forecasted results of the LSTM networks concerning the two tourist attractions are closer to the actual tourism volumes, especially when the tourist volumes fluctuate significantly. To evaluate the performance of each model more accurately, the MAE, RMSE and MAPE of each model concerning the two tourist attractions are respectively calculated, as shown in Table 3. It can be seen from Table 3 that the proposed approach outperforms the baselines in terms of the MAE, RMSE and MAPE, which verifies the validity of the proposed approach. It should be noted that regarding the forecasting results of HMA, the MAPEs are relatively high for all the models, this is mainly due to the volatility of the data.

To verify whether the search engine data and weather data can improve the prediction accuracy of the LSTM networks, we further respectively use “historical data” (HD) and “historical + search engine data” (HSD) as independent variables to forecast the tourism volume of the two tourist attractions. The MAE, RMSE and MAPE of the LSTM networks concerning the two tourist attractions are calculated with respect to the two kinds of independent variables. Finally, the MAE, RMSE and MAPE of the LSTM networks concerning the two tourist attractions with respect to HD, HSD and “historical + search engine + weather data” (HSWD) are shown in Table 4.

It can be seen from Table 4, on the one hand, the performance of “HSD” is better than that of “HD” in terms of the MAE, RMSE and MAPE; on the other hand, the performance of “HSWD” is better than that of “HSD” in terms of the MAE, RMSE and MAPE. In other

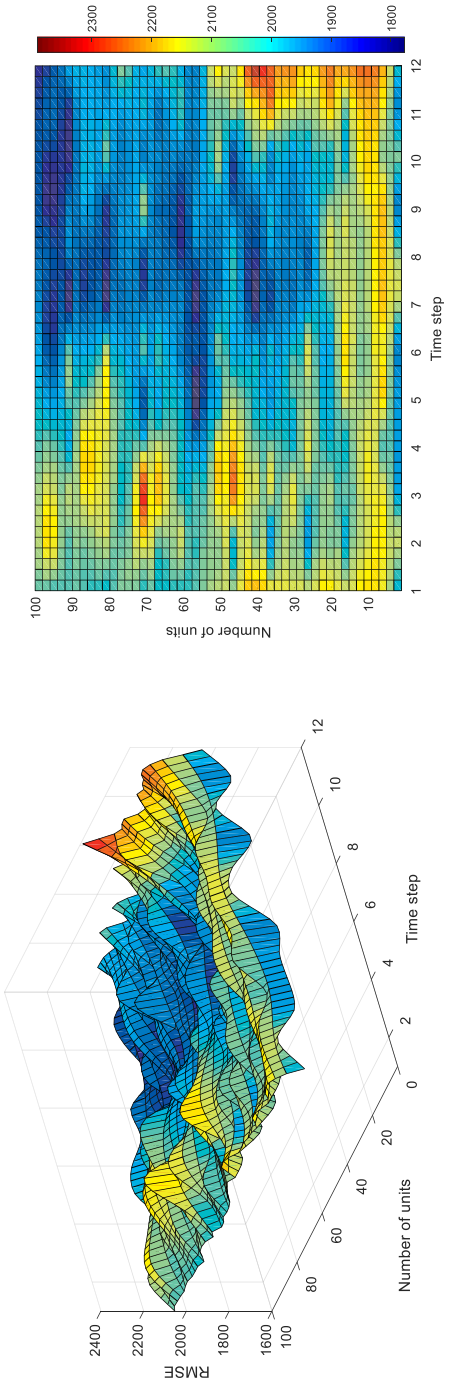

Fig. 13. The average values of RMSE concerning different parameter combinations.

Table 2
The optimal parameters of the three baselines.

Model	Parameters	
	Jiuzhaigou	HMA
ARIMAX	p = 0, d = 1, q = 9	p = 0, d = 0, q = 8
SVR	Kernel = RBF, c = 1, g = 0.01, Degree = 3	Kernel = RBF, c = 100, g = 0.05, Degree = 3
ANN	Learning rate = 0.01, Number of hidden layers = 6, epochs = 1000	Learning rate = 0.1, number of hidden layers = 20, epochs = 1000

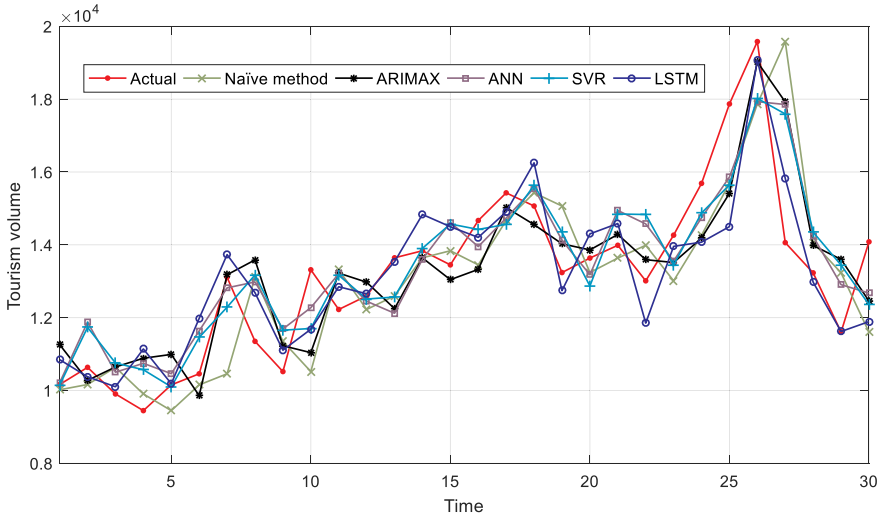

Fig. 14. The actual and predicted tourism volume of Jiuzhaigou from June 1, 2017 to June 30, 2017.

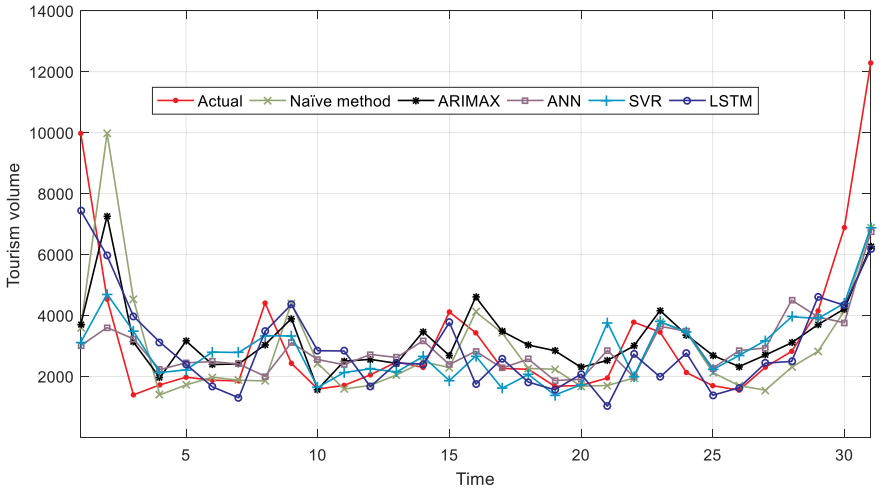

Fig. 15. The actual and predicted tourism volume of HMA from December 1, 2018 to December 31, 2018.

words, using search engine data and weather data can improve the performance of the LSTM networks.

Model discussion

Historical tourism volume data are the commonly used predictor variable in tourism volume forecasting (Yang et al., 2015). Using historical data to forecast the future tourism volume recognizes the ductility of the trend of changes in tourism volume (Du Preez & Witt, 2003). However, this may limit the forecasting accuracy when unexpected events occur or the economic structure changes (Yang et al., 2014). To further improve the forecasting accuracy, some studies suggest introducing new types of predictor variables in developing the forecasting model (Hubbard, 2011; Yang et al., 2014). On the one hand, with the rapid development of the

Table 3
The MAE, RMSE and MAPE of each model.

Model	Jiuzhaigou			HMA		
	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)
Naïve method	1228.701	1641.417	9.158	1336.968	2140.166	37.931
ARIMAX	1046.324	1336.726	8.115	1254.564	1895.373	37.114
ANN	1047.717	1261.991	8.014	1246.956	1950.803	37.379
SVR	1089.716	1306.267	8.320	1166.849	1877.900	35.455
LSTM	**883.963**	**1154.882**	**6.667**	**1040.497**	**1581.516**	**34.321**

The significance of bold indicates the best prediction results obtained by these models.

Table 4
The MAE, RMSE and MAPE of the LSTM networks with respect to different kinds of variables.

Variable	Jiuzhaigou			HMA		
	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)
HD	1229.02	1615.477	9.172	1224.666	1878.547	40.183
HSD	1059.923	1250.224	8.313	1125.794	1769.978	39.945
HSWD	**883.963**	**1154.882**	**6.667**	**1040.497**	**1581.516**	**34.321**

The significance of bold indicates the best prediction results obtained by different kinds of variables.

information technology, more and more people tend to search relevant tourism information through the Internet before making the travel decision (Dergiades et al., 2018). Search engine data are gradually becoming one of the main sources of tourism information, which can reflect the tourists' intention to visit (Yang et al., 2014; Yang et al., 2015). Additionally, search engine data have been validated by many studies as a kind of effective predictor variable in tourism volume forecasting (Li et al., 2017; Yang et al., 2014; Yang et al., 2015). On the other hand, tourists' final travel decision is also affected by some constraints, which may also affect the actual tourism volume (Woodside & Lysonski, 1989). Tourists usually check the weather conditions of the destination through the weather forecast before traveling, and decide whether to travel according to the actual weather conditions (Meyer & Dewar, 1999). Thus, weather data can reflect the constraint for tourists in making the final travel decision. Additionally, studies have also shown that there is a certain correlation between weather and tourist volume, and weather data can be used to forecast tourist volume (Falk, 2013; Falk, 2014; Meyer & Dewar, 1999; Otero-Giráldez et al., 2012). Therefore, search engine data and weather data are two kinds of predictor variables that are suitable for tourist volume forecasting. If these two kinds of predictor variables can be introduced into the forecasting of tourist volume, then better forecasting results could be expected. It is necessary to point out that when the above three kinds of data (historical tourism volume data, search engine data and weather data) are considered at the same time, the data for daily tourism volume forecasting usually have the following characteristics: (1) They have some nonlinearity; (2) They are simultaneously affected by multiple explanatory variables, and the lag order or temporal dependence between each explanatory variable and the actual tourism volume may not be the same; (3) The lag order or temporal dependence between each explanatory variable and the actual tourism volume may vary with circumstance. In view of the above characteristics of the tourist volume data, the LSTM networks are introduced into daily tourist volume forecasting. The LSTM networks can not only capture the nonlinear relationship between explanatory variables and daily tourism volume, but also learn the temporal dependence from the data through the cell state, and a fixed set of lagged observations does not need to be specified. In addition, the lead or lag order of the data that varies with circumstance can also be learned. Therefore, if data with similar characteristics are encountered in future tourist volume forecasting, LSTM may be a good alternative.

Conclusions

The objective of the study is to forecast the daily tourism volume of tourist attractions. For this, a novel approach based on LSTM networks that can simultaneously use historical data, search engine data and weather data is proposed. The proposed approach is applied to the daily tourism volume forecasting of Jiuzhaigou and HMA, two famous tourist attractions in China. Through these two applications, the validity of the proposed approach is verified. The major contributions and management applications are discussed as follows.

Theoretically, first, a novel approach based on LSTM networks that can simultaneously use historical data, search engine data and weather data is proposed for forecasting the daily tourism volume of tourist attractions. To our knowledge, it is the first attempt to simultaneously incorporate historical data, search engine data and weather data into the development of the approach for daily tourism volume forecasting. Meanwhile, this is the first study that attempt to apply the LSTM networks to tourism volume forecasting based on multiple data sources. In addition, the proposed approach has better universality and expansibility, which can be served as a basic approach for integrating more kinds of predictors into the development of tourism forecasting model. Second, although the existence of the effect of weather conditions on tourism volume has been verified, the study on daily tourism volume forecasting of

tourist attractions considering weather data is very limited. This paper presents a way to process weather data and to use them in daily tourism volume forecasting. The experimental results show that using the processed weather data can improve the performance of the LSTM networks in daily tourism volume forecasting. Third, LSTM networks are a kind of deep learning technology that can automatically learn the temporal context of input sequences. In addition, LSTM networks are able to almost seamlessly model the problem of multivariate time series forecasting, which is very suitable for tourism volume forecasting considering multiple predictors simultaneously. This study provides an in-depth guide on data collection and processing, as well as the development, training, and deployment of LSTM networks for daily tourism volume forecasting. This laid a good foundation for the application of LSMT networks in tourism volume forecasting and tourism management. Fourth, the effectiveness of the proposed approach has been verified by applying it in the daily tourism volume forecasting of Jiuzhaigou and HMA. In addition, the forecasting power of the approach with historical data, search engine data and weather data is stronger than that without search engine data or without both search engine data and weather data, which may provide evidence that search engine data and weather data are of great significance to tourism volume forecasting.

Accurate forecasting of daily tourism volume is also crucially important for management decisions. Based on the forecasting results, on the one hand, tourism operators can formulate tourism packages or pricing strategies to increase tourist arrivals in periods of low demand (Divino & McAleer, 2010); on the other hand, tourism operators can make staffing arrangements and develop emergency plans to prevent the occurrence of tourist detention in periods of high demand (Li et al., 2018). In addition, accurately forecasting the daily tourism volume may contribute to a variety of industries that directly or indirectly depend on tourism (Bi, Liu, Fan, & Zhang, 2020; Sun et al., 2019).

The study also has some limitations, which may serve as avenues for future research. To further verify the superiority of the proposed approach, specialized studies focus on error analysis against other methods are needed (Makridakis, Spiliotis, & Assimakopoulos, 2018). Additionally, three kinds of data are used in this study. In real life, other factors or predictor valuables may also have influence on the tourism volume (Bi, Liu, Fan, & Zhang, 2019). Therefore, to further improve the forecasting accuracy of tourism volume, extended studies of the proposed approach based on more factors or predictor valuables could be conducted.

Acknowledgements

This work was partly supported by the Humanities and Social Science Fund of Ministry of Education of China under Grant number 20YJC630002, the China Postdoctoral Science Foundation under Grant number 2019M661000, the National Natural Science Foundation of China under Grant number 71971124, 71771043, 71932005.

References

Álvarez-Díaz, M., & Rosselló-Nadal, J. (2010). Forecasting British tourist arrivals in the Balearic Islands using meteorological variables. *Tourism Economics, 16*(1), 153–168.
Assaf, A. G., Li, G., Song, H., & Tsionas, M. G. (2018). Modeling and forecasting regional tourism demand using the Bayesian Global Vector Autoregressive (BGVAR) model. *Journal of Travel Research, 58*(3), 383–397.
Athanasopoulos, G., & Hyndman, R. J. (2008). Modelling and forecasting Australian domestic tourism. *Tourism Management, 29*, 19–31.
Au, N., & Law, R. (2000). The application of rough sets to sightseeing expenditures. *Journal of Travel Research, 39*, 70–77.
Bangwayo-Skeete, P. F., & Skeete, R. W. (2015). Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tourism Management, 46*, 454–464.
Becken, S. (2013). Measuring the effect of weather on tourism: A destination-and activity-based analysis. *Journal of Travel Research, 52*(2), 156–167.
Beneki, C., Eeckels, B., & Leon, C. (2012). Signal extraction and forecasting of the UK tourism income time series: A singular spectrum analysis approach. *Journal of Forecasting, 31*, 391–400.
Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research, 13*, 281–305.
Bi, J. W., Liu, Y., Fan, Z. P., & Zhang, J. (2019). Wisdom of crowds: Conducting importance-performance analysis (IPA) through online reviews. *Tourism Management, 70*, 460–478.
Bi, J. W., Liu, Y., Fan, Z. P., & Zhang, J. (2020). Exploring asymmetric effects of attribute performance on customer satisfaction in the hotel industry. *Tourism Management, 77*https://doi.org/10.1016/j.tourman.2019.104006.
Bräuning, F., & Koopman, S. J. (2014). Forecasting macroeconomic variables using collapsed dynamic factor analysis. *International Journal of Forecasting, 30*(3), 572–584.
Burger, C. J. S. C., Dohnal, M., Kathrada, M., & Law, R. (2001). A practitioners guide to time-series methods for tourism demand forecasting—A case study of Durban, South Africa. *Tourism Management, 22*, 403–409.
Cao, Z., Li, G., & Song, H. (2017). Modelling the interdependence of tourism demand: The global vector autoregressive approach. *Annals of Tourism Research, 67*, 1–13.
Chan, F., Lim, C., & Mcaleer, M. (2005). Modelling multivariate international tourism demand and volatility. *Tourism Management, 26*, 459–471.
Cho, V. (2001). Tourism forecasting and its relationship with leading economic indicators. *Journal of Hospitality & Tourism Research, 25*(4), 399–420.
Cho, V. (2003). A comparison of three different approaches to tourist arrival forecasting. *Tourism Management, 24*(3), 323–330.
Claveria, O., Monte, E., & Torra, S. (2016). Combination forecasts of tourism demand with machine learning models. *Applied Economics Letters, 23*, 428–431.
De Mello, M. M., & Fortuna, N. (2005). Testing alternative dynamic systems for modelling tourism demand. *Tourism Economics, 11*, 517–537.
Dergiades, T., Mavragani, E., & Pan, B. (2018). Google Trends and tourists' arrivals: Emerging biases and proposed corrections. *Tourism Management, 66*, 108–120.
Divino, J. A., & McAleer, M. (2010). Modelling and forecasting daily international mass tourism to Peru. *Tourism Management, 31*(6), 846–854.
Du Preez, J., & Witt, S. F. (2003). Univariate versus multivariate time series forecasting: An application to international tourism demand. *International Journal of Forecasting, 19*(3), 435–451.
Falk, M. (2013). Impact of long-term weather on domestic and foreign winter tourism demand. *International Journal of Tourism Research, 15*(1), 1–17.
Falk, M. (2014). Impact of weather conditions on tourism demand in the peak summer season over the last 50 years. *Tourism Management Perspectives, 9*, 24–35.
Fildes, R., Wei, Y., & Ismail, S. (2011). Evaluating the forecasting performance of econometric models of air passenger traffic flows using multiple error measures. *International Journal of Forecasting, 27*(3), 902–922.
Gers, F. A., Eck, D., & Schmidhuber, J. (2002). Applying LSTM to time series predictable through time-window approaches. *Neural Nets WIRN Vietri-01* (pp. 193–200). .
Goh, C., & Law, R. (2002). Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention. *Tourism Management, 23*(5), 499–510.

Gunter, U., & Önder, I. (2016). Forecasting city arrivals with Google Analytics. *Annals of Tourism Research, 61*, 199–212.
Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*, 1157–1182.
Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780.
Huang, X., Zhang, L., & Ding, Y. (2017). The Baidu Index: Uses in predicting tourism flows-A case study of the Forbidden City. *Tourism Management, 58*, 301–306.
Hubbard, D. W. (2011). *Pulse: The new science of harnessing internet buzz to track threats and opportunities.* Hoboken, NJ: Wiley.
Kon, S. C., & Turner, W. L. (2005). Neural network forecasting of tourism demand. *Tourism Economics, 11*, 301–328.
Law, R. (2000). Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting. *Tourism Management, 21*, 331–340.
Law, R., & Au, N. (2000). Relationship modeling in tourism shopping: A decision rules induction approach. *Tourism Management, 21*, 241–249.
Law, R., Li, G., Fong, D. K. C., & Han, X. (2019). Tourism demand forecasting: A deep learning approach. *Annals of Tourism Research, 75*, 410–423.
Li, S., Chen, T., Wang, L., & Ming, C. (2018). Effective tourist volume forecasting supported by PCA and improved BPNN using Baidu index. *Tourism Management, 68*, 116–126.
Li, X., Pan, B., Law, R., & Huang, X. (2017). Forecasting tourism demand with composite search index. *Tourism Management, 59*, 57–66.
Lim, C., & McAleer, M. (2002). Time series forecasts of international travel demand for Australia. *Tourism Management, 23*, 389–396.
Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The M4 competition: Results, findings, conclusion and way forward. *International Journal of Forecasting, 34*(4), 802–808.
Martín, M. B. G. (2005). Weather, climate and tourism a geographical perspective. *Annals of Tourism Research, 32*(3), 571–591.
Meyer, D., & Dewar, K. (1999). A new tool for investigating the effect of weather on visitor numbers. *Tourism Analysis, 4*(3–4), 145–155.
Otero-Giráldez, M. S., Álvarez-Díaz, M., & González-Gómez, M. (2012). Estimating the long-run effects of socioeconomic and meteorological factors on the domestic tourism demand for Galicia (Spain). *Tourism Management, 33*(6), 1301–1308.
Palmer, A., Montano, J. J., & Sesé, A. (2006). Designing an artificial neural network for forecasting tourism time series. *Tourism Management, 27*(5), 781–790.
Pan, B., Wu, C. D., & Song, H. (2012). Forecasting hotel room demand using search engine data. *Journal of Hospitality and Tourism Technology, 3*(3), 196–210.
Pan, B., & Yang, Y. (2017). Forecasting destination weekly hotel occupancy with big data. *Journal of Travel Research, 56*(7), 957–970.
Rivera, R. (2016). A dynamic linear model to forecast hotel registrations in Puerto Rico using Google Trends data. *Tourism Management, 57*, 12–20.
Sencheong, K., & Turner, L. W. (2005). Neural network forecasting of tourism demand. *Tourism Economics, 11*, 301–328.
Shan, J., & Wilson, K. (2001). Causality between trade and tourism: Empirical evidence from China. *Applied Economics Letters, 8*, 279–283.
Shen, S., Li, G., & Song, H. (2009). Effect of seasonality treatment on the forecasting performance of tourism demand models. *Tourism Economics, 15*(4), 693–708.
Silva, E. S., Hassani, H., Heravi, S., & Huang, X. (2019). Forecasting tourism demand with denoised neural networks. *Annals of Tourism Research, 74*, 134–154.
Song, H., & Hyndman, R. J. (2011). Tourism forecasting: An introduction. *International Journal of Forecasting, 27*(3), 817–821.
Song, H., & Li, G. (2008). Tourism demand modelling and forecasting—A review of recent research. *Tourism Management, 29*(2), 203–220.
Song, H., Qiu, R. T., & Park, J. (2019). A review of research on tourism demand forecasting. *Annals of Tourism Research, 75*, 338–362.
Song, H., Witt, S. F., & Jensen, T. C. (2003). Tourism forecasting: Accuracy of alternative econometric models. *International Journal of Forecasting, 19*, 123–141.
Sun, S., Wei, Y., Tsui, K. L., & Wang, S. (2019). Forecasting tourist arrivals with machine learning and internet search index. *Tourism Management, 70*, 1–10.
Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in neural information processing systems* (pp. 3104–3112). .
Turner, L. W., & Witt, S. F. (2001). Factors influencing demand for international tourism: Tourism demand analysis using structural equation modelling, revisited. *Tourism Economics, 7*, 21–38.
Woodside, A. G., & Lysonski, S. (1989). A general model of traveler destination choice. *Journal of Travel Research, 27*(4), 8–14.
Yang, X., Pan, B., Evans, J. A., & Lv, B. (2015). Forecasting Chinese tourist volume with search engine data. *Tourism Management, 46*, 386–397.
Yang, Y., Pan, B., & Song, H. (2014). Predicting hotel demand using destination marketing organization's web traffic data. *Journal of Travel Research, 53*(4), 433–447.
Zhang, S., Zhang, C., & Yang, Q. (2003). Data preparation for data mining. *Applied Artificial Intelligence, 17*(5–6), 375–381.

Jian-Wu Bi, PhD, is a Postdoctoral Fellow in College of Tourism and Service Management, Nankai University, Tianjin, China. He was previously a Visiting Scholar in School of Computer Science and Engineering at Nanyang Technological University, Singapore. His research interests are tourism information technology and tourist satisfaction analysis. He has published more than ten papers in leading and reputable journals on tourism and information management, such as: *TM*, *INS*, *INFFUS*, *IJPR*, *ESWA*, *IJITDM*, etc.

Yang Liu, PhD, is a professor in School of Business Administration at Northeastern University (NEU), Shenyang, China. His current research interests include decision analysis and operations research. He has published over 20 refereed articles in international leading and reputable journals on tourism and information management, such as: *TM*, *EJOR*, *COR*, *IEEE T SMCA*, *INS*, *INFFUS*, *ESWA*, among others.

Hui Li, PhD, is a Professor at College of Tourism and Service Management, Nankai University, China. He is on the First Level of Tianjin 131 Creative Talent Program. His research focuses on tourism management analysis and prediction. He has published over 80 papers in journals on tourism, management and information science, such as: *ATR*, *TM*, *IJHM*, *IJCHM*, *EJOR*, *COR*, *JORS*, *FOR*, *IEEE TSMCA*, *IAM*, *INS*, *INFFUS*, etc.