Research Article

Prediction of Tourism Demand in Liuzhou Region Based on Machine Learning

Wenqing Li

College of Humanities and Public Administration, Baise University, Baise 533000, Guangxi, China

Correspondence should be addressed to Wenqing Li; lwq7210@bsuc.edu.cn

Received 11 April 2022; Revised 11 May 2022; Accepted 26 May 2022; Published 16 June 2022

Academic Editor: Abid Yahya

Copyright © 2022 Wenqing Li. 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.

Liuzhou is a rich tourism city in China. It is well known for its ethnic and prehistoric culture, folk songs, rare stones, and urban landscape. The demand of tourism in Liuzhou city is increasing day by day. Therefore, a system is needed to accurately predict the increasing demand of tourism in Liuzhou city. For this reason, based on the examination of historical visitor data in natural sites especially in Liuzhou city, this research employs an important machine learning and deep learning approach. The purpose of this study is to identify the consumption patterns and improve prediction, prejudgment, and preparation abilities by incorporating them into an intelligent tourism service platform, preventing visitors from travelling at inconvenient times and providing appropriate suggestions to scenic site managers. In this paper, the Sparse Principal Component Analysis-Long-Short-Term Memory (SPCA-LSTM) model has been updated to the Sparse Principal Component Analysis Convolutional Neural Network Long-Short-Term Memory (SPCA-CNNLSTM) model to accurately predict the tourist traffic during the holidays. Convolutional and pooling layers are added to the network topology to extract the local characteristics of the input data. The data of passenger traffic and influencing factors of Liuzhou scenic area from the months of Sept. 2015 to Nov. 2019 were used as the data set for the experiments. A hybrid-forecasting model is also proposed in the paper, which first removes noise from international crude oil data using compression-aware denoising preprocessing and then combines compression-aware denoising with machine learning using artificial neural network (ANN) and support vector regression (SVR). The experimental result shows that the SPCA-CNNLSTM model predicts better values than the SPCA-LSTM model.

1. Introduction

China’s economy is now expanding rapidly, as well as its consumption and living standards are steadily rising, which is an encouraging sign for the development of China [1]. The tourist sector has also entered into a new era of rapid growth. Tourism and leisure industry in China has entered a new age of large-scale tourism; tourism has emerged as a significant source of pleasure as well as relaxation for modern Chinese and is gradually becoming an intrinsic part of daily life [2]. Liuzhou is a rich tourism city in China, which is known for its own ethnic culture, prehistoric culture, Liu Zongyuan culture, folk song culture, rare stone culture, and urban landscape, among other things. Liuzhou, Guangxi’s second-largest city, has a national-level A-level scenic location. Guangxi’s tourism population and revenues rank third. In the Guangxi area, it is a popular tourist attraction. There are various high-quality tourist resources in the city, including Liuzhou Park and Dalongtan Picturesque Area, which are both national important park scenic places. In Liuzhou, tourism demand is growing with each day; thus, the tourism department requires a powerful and efficient system to forecast tourism demand and respond intelligently [3].

The administration of tourism services in China is still in its early phases of development. There are issues such as insufficient infrastructural development, a lack of resource integration and coordination, and a lack of intellectual and industrial support [4]. It is necessary to expand infrastructure construction in order to overcome these difficulties. It is needed to install intelligent equipment in scenic areas, convert user behavior and information into data resources, increase the scientific management of the whole industry, and provide visitors with a higher-quality, intelligent all-round tourism experience [5].
One of the most significant study topics is the analysis and prediction of visitor flow in tourist sites [6]. In recent years, the number of tourists to China has risen, with the majority of passengers opting for holidays, resulting in an imbalanced allocation of tourism resources in terms of time and space [7]. A large number of people visit to popular sites during peak season, resulting in “underload” in certain appealing regions and “overload” in others [8]. Several picturesque areas appear to be experiencing “overcrowding” and trampling. This sort of occurrence has a serious influence on passengers [9]. As a result, developing an efficient statistical model for the future is essential. It can provide critical support and confidence to picturesque locations and travelers making decisions or plans, particularly during the vacation season. [10]. Visitors may check the current traffic flow of each scenic location as well as the predicted traffic flow for the forthcoming period online while making travel plans [11]. They can prepare ahead of time for guiding, management, and security operations at picturesque locations during peak hours based on predicted passenger flow and make improvements depending on passenger flow changes [12]. An attractive location passenger flow projection might help the administration and relevant agencies make timely, informed judgments. For scenic area service providers, they may be able to plan for supplies and appropriately manage their labor and materials, among other things [13].

In the “5G” era, technology will further disrupt traditional tourism management and marketing techniques. To advance the tourism sector in the direction of intelligence, mobility, and customization, the growth of all elements of China’s tourist industry knowledge, as well as the radiation of people’s trip, is promoted [14]. The majority of current research on scenic location traffic forecasts concentrates on long-term traffic, such as annual or monthly visitor arrivals. Short-term traffic forecasting, such as weekly or daily traffic predictions, has received less attention [15]. In terms of practical application, there is a dearth of a scientific, effective, and intuitive daily traffic forecasting platform for scenic areas. Data from relevant agencies are occasionally delayed, making it difficult to effectively advise and guide travelers when they make travel arrangements [16]. More exact traffic forecasts for tourist locations in the future are still a problem in terms of research. Firstly, the forecasting scenic spot traffic is a multivariate time-series forecasting problem, and tourist traffic trends are impacted by a range of external factors [17]. Many external factors affect the number of people who visit the beautiful region, such as weather conditions and holidays. These impacting factors are still in the exploratory and research stages of analysis and mining. Secondly, traditional approaches cannot achieve long-term accurate forecasts due to the complicated non-linear features of scenic site traffic and their affecting factors. Moreover, current models are frequently limited to a single or finite set of circumstances [18]. It is difficult to promote those models widely, and more study is needed to predict passenger flow in tourist destinations. Artificial neural networks and deep learning research have advanced significantly in many areas of practice and theory in recent years [19]. Many areas have benefited from the use of these methodologies. The ability of neural networks to fit time-series predictions is considerably superior to prior methods [20].

As a result, this study, which is based on the research and mining of historical visitor data of beautiful sites, employs important machine learning and deep learning theory methodologies. We develop a mathematical model based on an examination of the spatial and temporal distribution aspects of daily visitor flow in attractive areas, as well as the influencing factors that affect it. The regional and temporal distribution aspects of the daily visitor flow were examined. We build a mathematical model to anticipate the short-term visitor flow of picturesque regions, as well as associated influencing variables. The idea is to use the data to uncover a consumption pattern and improve prediction abilities. The goal is to use data to detect consumption patterns, improve prediction, prejudgment, and preparation abilities, and apply them to the intelligent tourist service platform to prevent visitors from travelling at inconvenient times and give scenic site managers suggestions or recommendations.

The rest of the research paper is organized as follows: Section 2 of the paper will explain the related works pertaining to tourist demand forecasting using machine learning. Section 3 discusses the accuracy of tourist day traffic prediction, new challenges, prediction algorithms, and their performance. Similarly, Section 4 explains the SPCALSTM algorithm with CNN model construction process used in this research. Finally, Sections 5 and 6 describe the experimental procedure and data analysis, as well as concluding remarks.

2. Related Work

From qualitative prediction to more precise quantitative prediction, from econometric models to machine learning approaches, tourism demand research has garnered attention one after the other. Each of these factors has resulted in the formation and development of approaches that have expanded tourism forecasting research findings and contributed to rational tourist planning and, as a result, economic growth. Machine learning has been applied to enhance performance in a variety of time-series forecasting problems, including tourism demand forecasting. Tourism volumes, according to research, are directly related to the level of consumption at the destination as well as the currency exchange rate between the origin and destination of tourists. The Error Correction Model (ECM) outperforms the Seasonal Autoregressive Integrated Moving Average (SARIMA) and Naive techniques significantly, according to Kulendran and Wilson’s investigation. The Time-Varying Parameter (TVP) and ECM give considerable forecasting benefits in various econometric alternatives and time-series models [21].

In addition to econometric models and time-series models, several scientists have begun to investigate fresh quantitative methods in the field of tourist demand forecasting, such as artificial intelligence techniques. An ANN
assesses the error between real and predicted values by modelling the inputs and outputs, develops the model continuously by backward transmission of errors, and learns the relationship between inputs and outputs through machines. As a result, this technique essentially replaces statistical regression prediction models. According to the findings, ANN beats other time-series models in terms of prediction accuracy. SVR is an artificial intelligence approach for handling nonlinear regression and prediction issues. Time-series and econometric models, in general, rely on data stability and economic structure. Artificial intelligence models, on the other hand, rely on a huge number of data samples and are susceptible to overfitting and other issues [22].

A hybrid forecasting model is proposed, in which compression-aware denoising preprocessing is used to eliminate noise from international crude oil data and then compression-aware denoising is combined with machine learning using ANN and SVR. Deep learning approaches are presented to investigate the complex correlations between oil price and other parameters, utilizing a stacked denoising autoencoder (SDAE) as the foundation model. A SDAE is used as the foundation model to investigate the complex relationships between oil prices and other variables [1].

Multiobjective particle swarm optimization-deep belief net (MOPSO-DBN) is a new hybrid prediction model that improves the parameters of deep belief net (DBN) for short-term traffic flow prediction using a multiobjective particle swarm optimization approach [2]. To capture nonlinear traffic dynamics, the LSTM neural network is used and the model is trained. The LSTM neural network’s efficacy for short-term traffic prediction tasks was confirmed using data from remote sensing microwave sensors in Beijing [23].

### 3. Problem Analysis

We choose and downscale the features in the SPCA-LSTM model, and the learning process of the SPCA-LSTM model primarily relies on the forward transmission and reverse updating of network parameters, as well as the input of features and labels. The SPCA-LSTM model generates better prediction results than other comparison algorithms, and it can remember a large amount of long-term information. However, the holiday forecast results are not accurate enough, and the model’s performance has to be improved. The problem of gradient disappearance in Recurrent Neural Network (RNN) is not entirely solved by LSTM, and when feature points occur at long intervals, a pure LSTM model may perform poorly (e.g., 1000 magnitudes). In order to increase the model’s accuracy even more, this research studies a tourist day traffic prediction model based on SPCA-CNN-LSTM and confirms the model with high performance by adding a convolutional layer to the SPCA-LSTM model to extract the feature points in the sequence and enhance the network structure and improved accuracy by combining with comparison experiments. In this article, a deep network structure is applied in this model. Deep neural networks have more complicated structures, more parameters, and a longer coordination involvement and training process as compared to single-layer or shallow-layer neural networks. Its advantages include improved characterization and data fitting capabilities, as well as a focus on feature extraction, which is critical for time-series prediction issues and will become more essential in the future. This is significant for time-series prediction issues and will be a future research topic. Convolutional Neural Networks (CNNs) are well known for their benefits in feature extraction and are mostly employed in image processing and speech recognition; nevertheless, they are now being gradually applied to time-series prediction issues, with some success gained. In this paper, we add convolutional layers to the network structure to extract features (patterns) in time series, identify the underlying logic behind the data from the difficult data, and then train LSTM models based on these feature vectors to get improved prediction outcomes.

### 4. SPCA-CNNLSTM Model

#### Construction Process

This section is composed of several subsections such as CNN-LSTM and Network Parameter Selection.

#### 4.1. CNN-LSTM

In this research, we extend the SPCA-LSTM model with CNN to extract local features from time series, which necessitates the addition of convolutional and pooling layers to the original. The basic network structure is added with the convolutional layer and pooling layer. Multiple convolutional kernels are used in the convolutional layer to extract local features from the data by convolving the input data. To lower the dimensionality of the data, the pooling layer uses the pooling function to do downsampling. As shown in Figure 1, the CNN-LSTM network constructed in this paper has the following layers: the input layer, the one-dimensional convolutional layer, the pooling layer, the LSTM layer, the fully connected layer, and the output layer.

However, for time-series prediction tasks, a one-dimensional convolutional layer is commonly utilized, i.e., convolutional operations are performed in a single spatial dimension by the convolutional kernel and the input of this layer, as seen in Figure 2.

The network learning impact is affected by the convolutional kernel’s window size in the convolutional layer. A small window size may easily extract local features, but it is possible that it will reduce the correlation of local features. A wide size of the window would increase computation and miss the local feature.

\[
OH = \frac{H + 2P - FH}{S} + 1, \quad (1)
\]

\[
OW = \frac{H + 2P - FW}{S} + 1.
\]

In the convolution layer, the parameters to be set are the number of convolution kernels \( n \), the activation function \( f \), the convolution kernel size \((FH, FW)\), and the step size \( S \). The commonly useful activation functions are Tanh, Sigmoid, Scaled Exponential Linear Unit (SELU), Rectified
Linear Unit (ReLU), etc. In this paper, we used ReLU activation function, because of its linear and nonsaturated form; the ReLU function can avoid the gradient disappearance problem and improve up training speed, and it is expressed as follows:

\[ f(x) = \begin{cases} 
  x, & (x > 0), \\
  0, & (x \leq 0). 
\end{cases} \tag{2} \]

The output feature vector after the convolutional layer operation looks as follows: output = (batch, new_steps, filters). new_steps indicates that the input data’s steps value may change as a result of padding or step size, and the output has a new step size; the number of convolutional kernels in the convolutional layer is known as filters, filters is equal to the number of output feature vectors, and output is the new feature vector.

4.2. Network Parameter Selection. There are two features in the data set, the input step size is set to 7 × 2, the number of convolutional kernels is \( n \), their size is (\( FH, FW \)), and the number of input data columns is represented by the length of convolutional kernels \( FW \), i.e., \( FW = 2 \), because the 1-
Figure 3: Convergence of the loss function at $n - 8$ and the fit of the prediction results.
Figure 4: Convergence of the loss function at FH = 3 and the fit of the prediction results.
5. Experimental Procedure and Analysis of Results

The purpose of this research is to build an SPCA-CNNLSTM deep neural network to predict the daily passenger flow of tourist attractions. The task of SPCA is to downscale the features, the CNN network layer extracts the local features at each time step, and the LSTM layer predicts future values. The daily passenger flow and its data of affecting variables from September 25, 2015, to November 30, 2019, in Liuzhou scenic area, with a total of 1528 rows, were used in this paper. Daily passenger flow values and 15 influencing factor indicators are included in the original dataset, and following feature selection using correlation analysis, 8 indicators with low correlation are discarded, leaving 6 influencing factor indicators as sample set features as follows: holiday (HO), low temperature (LT), PM2.5 concentration (PM2.5), historical passenger flow (HF), average temperature (MT), and high temperature (HT).

Two main components were derived after SPCA to reduce the dimensionality of the features $u_1$ and $u_2$.

$$ u_1 = -11.035x_1 $$

$$ u_2 = -43.411x_2 - 13.077x_3 + 283.070x_4 + 305.840x_5 + 294.47x_6 $$

The data are then fed into a CNN-LSTM network with a convolutional layer with the number of convolutional kernels $n = 8$ and convolutional kernel width $FH = 2$. The
loss function of the SPCA−CNNLSTM model after training and the function convergence plot are displayed in Figure 5. Figure 5 shows that the final loss value of the training set is approximately 0.031 and the loss value of the test set is around 0.033. The fit of the final prediction findings is represented in Figure 6.

According to Figure 6, the model’s prediction is good and the predicted and actual values may basically overlap, and this is also true for the holiday parameter. Table 1 displays the prediction findings’ error levels.

According to the given data, the SPCA-CNNLSTM model has the greatest overall accuracy. Although the MAPE value is not as low as that of the SPCA-LSTM model, the SPCA-CNNLSTM model predicts better in terms of RMSE, MAE, and R value than the SPCA-LSTM model. This also indicates that the convolutional layer can extract local features in the time step and discover the data’s deep logic. We can observe the advantage of the model studied in this paper in dealing with nonlinear systems, which can handle data with strong nonlinear properties and is extremely suited for the issue of predicting short-term number of tourists, when compared to other machine learning techniques [24].

6. Conclusion

The demand for tourism in Liuzhou is steadily expanding. As a result, a system is required to effectively forecast the growing demand for tourism in Liuzhou. Therefore, in this paper, the SPCA-LSTM model has been improved to SPCA-CNNLSTM model for solving the issue that the SPCA-LSTM model cannot accurately predict the flow of tourists during holidays. After preprocessing the source data of passengers’ traffic and associated influencing factors, the features are reduced to two dimensions via SPCA and then sent to the CNN-LSTM deep network for training and prediction. The grid search technique is used to identify the optimal combination of network parameters. Because it can efficiently extract spatial characteristics in the time step, CNN is ideal for time-series prediction. It is observed that the model outperformed in dealing with nonlinear systems, when compared to other machine learning algorithms, which can handle data with strong nonlinear properties and is particularly suitable for the challenge of predicting short-term number of tourists.

Data Availability

The raw data supporting the results of this paper can be made available without undue reservation.
Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

Acknowledgments

This work was supported by the 2019 University-Level Scientific Research Project of Baise University (No. 2019KS10) and the 2020 Open Project of Humanistic Spirit and Social Development Research Base in the Old Revolutionary Base Area of Yunnan, Guizhou, and Guangxi (No. JDC11).

References


