ISSN 2616-5775 Vol. 6, Issue 9: 72-80, DOI: 10.25236/AJCIS.2023.060911

Tourism demand forecasting using PCA-BPNN

Xin Zhao[1]

[1]*College of Economics, China-ASEAN Institute of Financial Cooperation, Guangxi University, Nanning, 530004, China*

Abstract: *Accurate prediction of tourism demand is critically important for the efficient allocation of resources in scenic areas and managing sudden events. This paper presents a new tourism demand prediction model, PCA-BPNN neural network model. It utilizes Principal Component Analysis (PCA) to reduce the dimensionality of the collected Baidu Index data and mitigate overfitting issues. The model then constructs a backpropagation neural network (BPNN). Empirical research demonstrates that PCA-BPNN effectively identifies the nonlinear relationship between search keywords and the number of tourist arrivals and outperforms all benchmark models in terms of predictive performance.*

Keywords: *Tourism demand forecasting; Search engine data; PCA-BPNN; PCA*

1. Introduction

In recent years, China's tourism industry has experienced rapid development. From 2012 to 2019, international tourism foreign exchange revenue increased from CNY 5,028 billion to CNY 131,254 billion, with an overall increase of 162.4% and an average annual growth rate of 7.2%. During the same period, domestic tourism spending increased from CNY 22,706.20 billion to CNY 57,250.92 billion, with an overall increase of 152.1% and an average annual growth rate of 6.2%. Despite a significant decline in tourism revenue in 2020 due to the pandemic, there has been a noticeable recovery trend in 2021. Both international and domestic tourism have seen a fast growth.

The rapid development of the tourism industry is closely linked to the large number of tourists visiting attractions. There was a corresponding increase in the number of domestic and inbound tourists from 2012 to 2019 by 103.1% and 9.7%, respectively. The fluctuation of tourist numbers from 2012 to 2019 and the sharp decline in 2020 caused by the COVID-19 pandemic have created unmet expectations among various stakeholders such as tourism sites, catering service providers, government, and individuals. The aftermath includes the closure of numerous businesses, sharp decline in tourism site revenue and government finances, personal travel restrictions, and serious disruptions to various aspects of the economy and people's lives, affecting their satisfaction.

Therefore, predicting the tourism demand for attractions or destinations accurately is of practical significance. Reasonable tourist number forecasts can help tourism sites and other locations involved in tourism activities have reasonable expectations and adjust their reception capacity to prevent overcrowding. This can also help businesses forecast future customer demand, prepare accordingly, and avoid conflict between optimistic traffic expectations and low tourist arrivals, leading to a surplus of inventory and loss in business profits. Ensuring normal life and traffic order in tourism and urban areas during holiday peak seasons and ensuring the safety of citizens and tourists are crucial for governments. Reasonable tourism demand predictions can help the government to organize personnel and transportation facilities to ensure a stable and orderly social environment during tourist activities. For individuals, excessive tourist flow can cause congestion, a shortage of catering and accommodation resources, and a reduction in tourist experience. Effective tourism demand forecasting can help individuals select peak traveling periods or adjust their travel destinations and routes to fully utilize tourism resources.

Traditional data forecasting methods often use existing data from government statistical departments or relevant databases for analysis. However, such data often lack timeliness, a low level of personalization in statistical indicators, limited data scope, and insufficient quantity of data leading to limited efficacy of models and capabilities in forecasting. Information technology development, particularly the widespread use of search engines among Internet users, has created a significant improvement in people's daily lives and the richness of Internet data information. Due to the broad information sources, rich content coverage, timely and effective information acquisition, and large-scale

ISSN 2616-5775 Vol. 6, Issue 9: 72-80, DOI: 10.25236/AJCIS.2023.060911

data of Internet data, they are highly suitable for information prediction, which can improve predictive accuracy and enrich predictive information.

This article intends to use the search information data of tourism attractions in the Baidu search engine for Internet users, and through the PCA-BPNN model, accurately predict the tourism demand for attractions. The model's validity will be verified by sample data and adjusted accordingly to provide a scientific and effective tourism demand expectation reference for various stakeholders in the tourism industry, allowing for scientifically reasonable arrangements and decisions.

2. Literature review

2.1. Methods for Forecasting

Accurate forecasting of tourism demand is essential for effective planning and resource allocation in the tourism industry. Researchers have employed various methods to predict tourism demand, according to Song and Li (2008) [1] forecasting methods can be divided into classic time series models, econometric models, and artificial intelligence models.

Time series analysis methods have been widely used in tourism demand forecasting. Time series analysis methods, including Autoregressive Integrated Moving Average (ARIMA) models [2], NAÏVE model [3], and Exponential smoothing model (ETS) [4], have been widely employed in the field of tourism demand forecasting. These models leverage historical time series data to identify patterns, trends, and underlying dynamics, enabling accurate predictions of future tourism demand.

One commonly used approach is the ARIMA model, which incorporates autoregressive, moving average, and differencing components. The autoregressive component considers the relationship between past and current values of the time series, capturing the influence of previous demand on future demand. The moving average component takes into account past forecast errors, considering the impact of random shocks. The differencing component helps to remove trends and seasonality, ensuring the stability of the time series data and enhancing the model's predictive capability. Another commonly employed model is the NAÏVE model, which assumes that the future demand will simply replicate the most recent observed value. This method is straightforward and useful when there are no identifiable patterns or trends in the data. It serves as a baseline model that can provide a benchmark for comparison with more sophisticated forecasting methods. The Exponential Smoothing model (ETS) is also widely used in tourism demand forecasting. This model assigns exponentially decreasing weights to past observations, with more recent observations receiving higher weights. It is particularly useful for capturing short-term dynamics and seasonality in the data. In current research, time series models are often used as benchmark models [5].

Econometric models play an important role in predicting consumer decisions. They incorporate factors that influence consumer decisions, such as price, income, and exchange rates, as explanatory variables in the models, thereby improving the accuracy of predictions. Common econometric models include vector autoregressive models (VAR) [4], error correction models (ECM) [6], autoregressive distributed lag models (ADLM) [7], and time-varying parameter models (TVP) [8] . These models, based on historical data and economic theory, provide predictions of consumer decisions by analyzing the relationships between variables and dynamic adjustments. With the development of the Internet, some scholars have started using new data sources, such as search engine data and review data, as explanatory variables to further improve prediction accuracy. These data sources can provide richer information, reflecting changes in consumer behavior and attitudes, and have shown good predictive performance in some studies.

Compared to traditional econometric models, artificial intelligence models are better able to capture the nonlinear relationship between tourist arrivals and influencing variables, and they have been widely applied in the field of tourism demand forecasting. Artificial intelligence models, by training and learning from a large amount of data, can provide more accurate predictions of future tourism demand. Common artificial intelligence models include the BP neural network model (BPNN) [9], support vector machine model (SVM) [10], long short-term memory model (LSTM) [11], gated recurrent unit model (GRU) [12], and convolutional neural network model (CNN) [13]. The BP neural network model is one of the most classic artificial neural network models, capable of learning and predicting through the connections between multiple neurons. In tourism demand forecasting, it can learn the complex nonlinear relationship between historical influencing factor data and tourist arrivals, thus providing more accurate predictions. The long short-term memory model and gated recurrent unit model are a type of recurrent neural network model that can handle time series data and store and utilize long-term memory. In tourism demand

ISSN 2616-5775 Vol. 6, Issue 9: 72-80, DOI: 10.25236/AJCIS.2023.060911

forecasting, these models can capture the time dependencies between influencing factors and consider the importance of historical data in the prediction process, providing accurate forecasts. The convolutional neural network model is a neural network model particularly suitable for processing image and spatial data. In tourism demand forecasting, the convolutional neural network model can extract spatial features of influencing factors through convolution operations and make predictions by learning weights. Law et al. (2019) [14] studied the prediction of monthly tourist arrivals in Macao, and empirical results showed that deep learning methods significantly outperformed support vector regression and artificial neural network models.

2.2. Application of Search Keyword in Tourism Demand Forecasting

Traditional historical data is relatively reliable in terms of information credibility. However, it has limitations such as limited data samples, fixed indicators, and insufficient timeliness, which pose constraints on research in tourism demand forecasting. With the rise of internet technology and the widespread use of internet applications in people's daily lives, online data has become increasingly important in predictive analysis research due to its wide sources, rich information, large quantity, and timely effectiveness, reflecting users' concerns and interests, and thus enabling trend prediction and user behavior prediction. Therefore, many scholars have used Baidu Index for tourism demand forecasting. Unlike Google Trends, which reflects global search popularity and trend data, Baidu Index mainly focuses on Chinese mainland users. Although Google Trends can provide some degree of domestic tourism trend information, its data does not differentiate by region and cannot provide specific search behavior data for Chinese mainland users. Therefore, in predicting domestic tourism, the predictive accuracy of Google Trends may be affected to some extent, and using Baidu Index can achieve better predictive accuracy compared to Google Trends [15]. Önder and Gunter (2016) [4] found that when using Google Trends data to predict tourist numbers in Vienna, Austria, and Belgium, compared to the benchmark model, the inclusion of search engine data improved prediction accuracy.

In addition, scholars have found that combining search data with other types of data can also improve the accuracy of substitute predictions compared to using search data alone as explanatory variables. Li et al. (2020) [16] combined search data with weather data, holiday data, and seasonal data in predicting tourist arrivals in Jiuzhaigou and Gulangyu. The empirical research found that incorporating multiple data sources, compared to using only search data, improved the accuracy of the predictions. In addition, some scholars have combined search engine data with review data, economic variables, and confirmed COVID-19 cases to predict tourist arrivals.

When it comes to setting initial search keywords, most scholars consider the six elements of tourism, namely, accommodation, food, transportation, sightseeing, shopping, and entertainment. They search these aspects as keywords in search engines and identify keywords with less missing data and those that are not yet indexed. These selected keywords form the final set of keywords to be used.

2.3. Principal Component Analysis

In order to avoid the problems of multicollinearity and overfitting, it is necessary to identify which keywords can be used for prediction. Principal Component Analysis (PCA) is a statistical method used to analyze and reduce the dimensionality of a dataset. It achieves this by identifying the underlying structure of the data and transforming it into a set of uncorrelated variables called principal components. The first principal component captures the largest amount of variance in the data, followed by the second component and so on. PCA is particularly useful when dealing with high-dimensional datasets, as it allows for a simplified representation of the data while retaining important information. It helps to identify the most meaningful variables that contribute to the overall variation in the dataset.

When predicting monthly automobile sales, Zhang at al. (2022) [17] used sentiment index, Baidu search data, and historical sales as input data. They built a predictive model called PCA-DSFOA-BPNN by combining Principal Component Analysis (PCA), Backpropagation Neural Network (BPNN), and an improved Fruit Fly Optimization Algorithm (DSFOA). Empirical research indicated that the proposed method outperformed the benchmark model in terms of prediction accuracy. Li at al. (2018) [18] employed Principal Component Analysis (PCA) to preprocess the inputs for Backpropagation Neural Network (BPNN) in order to predict tourist arrivals in Beijing and Hainan, China. This approach addressed the issue of correlating different search keywords. Additionally, an Adaptive Differential Evolution algorithm (ADE) was employed to optimize the performance of BPNN, leading to the construction of the PCA-ADE-BPNN model. Empirical findings indicated the effectiveness of the PCA-

ISSN 2616-5775 Vol. 6, Issue 9: 72-80, DOI: 10.25236/AJCIS.2023.060911

ADE-BPNN model in predicting tourism demand. Additionally, in addition to utilizing PCA for dimensionality reduction of search data, researchers also employ various techniques such as Random Forest (RF) [19], Recursive Feature Elimination (RFE) [20], Kernel Principal Component Analysis (KPCA) [21], and Generalized Dynamic Factor Model (GDFM) [22]. These methods contribute to enhancing the analysis of the data and improving the modeling process.

2.4. Rationale of this study

From the perspective of research progress, in the field of tourism demand forecasting, scholars are gradually turning to the use of artificial intelligence methods, which have shown better predictive results compared to traditional research methods in many aspects. However, within the realm of artificial intelligence methods, there are also many further subdivided models that often exhibit different explanatory power in different research problems, requiring selective discrimination based on actual circumstances. On the other hand, with the popularization of the internet and the abundance of online data, internet data has also become an important source of information for scholars in predicting tourism demand. There exists a close correlation between the quantity of keyword searches in domestic and foreign search engines and the actual visitor flow of tourist attractions. Scholars have incorporated web search data into their predictive models and achieved good forecasting results.

Therefore, this study selects Gulangyu Island in Xiamen as a case study and uses Principal Component Analysis (PCA) to process the Baidu keyword search data related to the tourist attraction. Based on this, a PCA-BPNN model is constructed to predict the daily visitor arrivals to Gulangyu Island. The forecasting results are then compared with those of the Seasonal Autoregressive Integrated Moving Average (SARIMA) model, Seasonal Autoregressive Integrated Moving Average with Explanatory Variables (SARIMAX) model, and the Seasonal Naïve model (SNAÏVE), in order to assess the superiority of different models in predicting weekly visitor flow in Gulangyu Island.

3. Methodology

3.1. PCA-BPNN

3.1.1. PCA

According to Jolliffe and Cadima (2020) [23], Principal Component Analysis (PCA) is a commonly used dimensionality reduction technique, which can transform high-dimensional data into a lower-dimensional space while preserving the most important information from the original data. Each principal component is a linear combination of the original data, and they are sorted in order of importance. The first principal component explains the maximum variance in the data, the second principal component explains the next most significant variance, and the principal components are uncorrelated with each other. There are five steps using PCA to process the search query data.

3.1.2. BPNN

BPNN stands for Backpropagation Neural Network. It is a type of artificial neural network (ANN) that utilizes the backpropagation algorithm to train and adjust the weights of the network's connections. The Backpropagation Neural Network (BPNN) exhibits rapid convergence and can attain a reasonably accurate prediction accuracy [24]. As a result, it has gained wide application in the realm of predictive modeling. The BPNN consists of an input layer, hidden layer(s), and output layer. As for the number of hidden layers, scholars believe that one to two hidden layers are sufficient to handle most problems. Therefore, in our research, we adopted a three-layer neural network, specifically one hidden layer. There are four steps in construct a BPNN model.

Step 1, initializing weights and biases. Randomly initialize the weights and biases in the neural network, typically using small random numbers for initialization.

Step 2, forward propagation. Feed the input data into the network and calculate the output of each neuron by multiplying the input with the weight matrix, adding the bias term, and passing it through an activation function (such as the Sigmoid function, ReLU function, etc.) for non-linear mapping to obtain the final outputs. In our study, we utilize the sigmoid function as the activation function.

Step 3, calculate loss. Compute a loss value based on the difference between the network's output and the true labels. Common loss functions include mean square error (MSE) and cross-entropy.

ISSN 2616-5775 Vol. 6, Issue 9: 72-80, DOI: 10.25236/AJCIS.2023.060911

Step 4, backward propagation. Using the chain rule, calculate the impact of each weight and bias on the loss based on the computed loss value. Then, update the values of each weight and bias according to the gradient descent algorithm to gradually reduce the loss value.

Step 5, repeat steps 2-4. Iterate through forward propagation and backward propagation until the network's loss value reaches a small threshold or the training epochs reach a predetermined number.

3.2. Benchmark models

3.2.1. Seasonal auto-regressive integrated moving average model (SARIMA)

ARIMA, proposed by Box and Jenkins [24], is now widely used for time series forecasting. It is an extension of ARIMA (Auto Regressive Integrated Moving Average) model that can support time series data with seasonal components. In the SARIMA(p, d, q)(P, D, Q, S) model, p represents the autoregressive order, which is the number of lagged terms used for prediction. d represents the differencing order, which is the number of times the original data needs to be differenced to achieve stationarity. When d=0, it means no differencing is required. q represents the moving average order, which is the number of lagged forecast errors used for prediction. P represents the seasonal autoregressive order, similar to p but for seasonal fluctuations with lagged terms. D represents the seasonal differencing order, similar to d but for seasonal differencing. Q represents the seasonal moving average order, similar to q but for seasonal lagged forecast errors. S represents the seasonal period length, which is the length of the data's seasonal cycle. In our study, S is 7. The SARIMA model can be expressed as follows:

$$\phi_p(B)(1-B)^d\Phi_P(B)y_t = \theta_q(B)(1-B)^D\Theta_Q(B)\epsilon_t$$

Where y_t is the time series, ϵ_t is the white noise, p and q are the orders of the autoregressive and moving average terms, P and Q are the orders of the seasonal autoregressive and moving average terms, d and D are the orders of the non-seasonal and seasonal differencing. $\phi_p(B)$ and $\theta_q(B)$ are the polynomial coefficients of the autoregressive and moving average terms, $\Phi_P(B)$ and $\theta_q(B)$ are the polynomial coefficients of the seasonal autoregressive and moving average terms. B is the backshift operator, $B^d(y_t) = y_{t-k}$.

3.2.2. Seasonal Autoregressive Integrated Moving Average models with exogenous factor model (SARIMAX)

SARIMAX is an extension of SARIMA that improves the model's explanatory power by incorporating exogenous variables. The formula for SARIMAX is as follows:

$$\phi_p(B)(1-B)^d\Phi_P(B)y_t = \sum \beta_k x_k + \theta_q(B)(1-B)^D\Theta_Q(B)\epsilon_t$$

Where x_k represents the k th explanatory variable, and β_k represents the coefficient of that explanatory variable.

3.2.3. Seasonal Naïve model (SNAÏVE)

Seasonal Naïve (SNAÏVE) model is a simple time series forecasting model that is specifically designed to capture seasonal patterns in the data, and assumes future observations will be equal to the corresponding historical observations from the same season.

$$\hat{y}_t = y_{t-s}$$

Where $\hat{y}_t$ represents the forecasted value for period t, while y_{t-s}represents the observed value from the previous period.

3.3. Evaluation

In our research, we utilize the metrics mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE) to assess the predictive performance of the proposed model.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

ISSN 2616-5775 Vol. 6, Issue 9: 72-80, DOI: 10.25236/AJCIS.2023.060911

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Where $\hat{y}_i$ is the predicted value, and y_i is the actual value.

4. Empirical study

4.1. Data collection

4.1.1. Tourist arrivals data

Located in Siming District, Xiamen City, Fujian Province, China, Gulangyu is a well-known tourist attraction and a national 5A-level scenic spot. Renowned as a famous tourist destination and a national 5A-level scenic spot, it attracts a significant number of visitors to Xiamen. Dubbed the "Garden on the Sea," Gulangyu captivates visitors with its distinctive architectural style, vibrant cultural ambiance, and breathtaking natural scenery. We collected daily tourist arrival data for Gulangyu from July 4, 2016, to December 20, 2019, and then summed it up to obtain weekly data. Considering the significant fluctuations in the weekly data, we applied averaging to represent the weekly tourist arrival by using the average value. In the end, we collected a total of 181 weeks of data. The first 159 weeks were used for training, the next 8 weeks were used to search for the hyperparameters of PCA-BPNN, and the final 24 weeks were used to evaluate the performance of the PCA-BPNN model. The number of weekly visitor arrivals on Gulangyu Island is depicted in Figure 1.

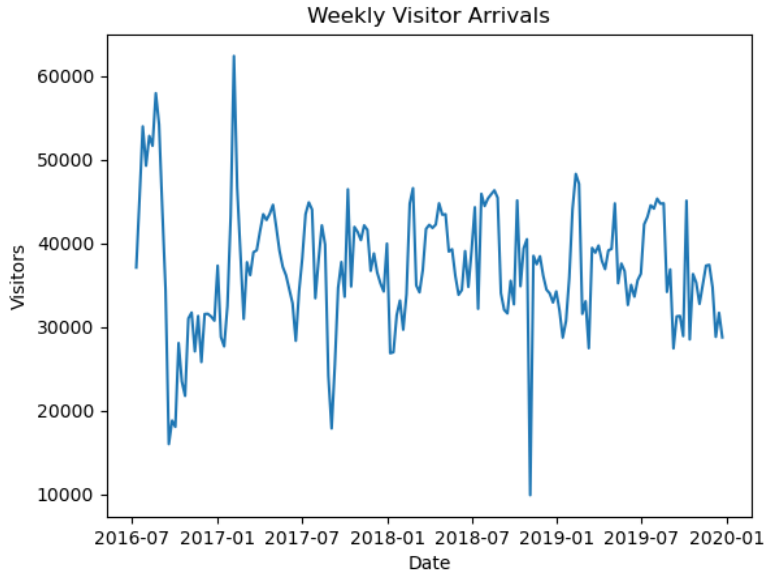

Figure 1: Weekly visitor arrivals in Gulangyu Island

4.1.2. Search engine data

According to Yang (2015) [15], Baidu Index performs better than Google Trends in predicting the arrival of domestic tourists. Therefore, when predicting the arrival of tourists to Gulangyu Island, we use Baidu Index to search for suitable keywords. We set the initial keywords based on the six elements of travel (eating, accommodation, transportation, sightseeing, shopping, and entertainment). According to the demand graph provided by Baidu Index, we iterate the original keywords, delete those with less data and those that are not provided, and finally collect 12 keywords. Considering that different search keywords may have different impacts on tourists' decisions during different search periods, we lag the search keywords by 1-12 periods. Then, we use PCA to reduce the dimensionality of the search keywords and finally set the threshold to 0.8. Figure 2 shows the cumulative contribution rate of the principal components.

ISSN 2616-5775 Vol. 6, Issue 9: 72-80, DOI: 10.25236/AJCIS.2023.060911

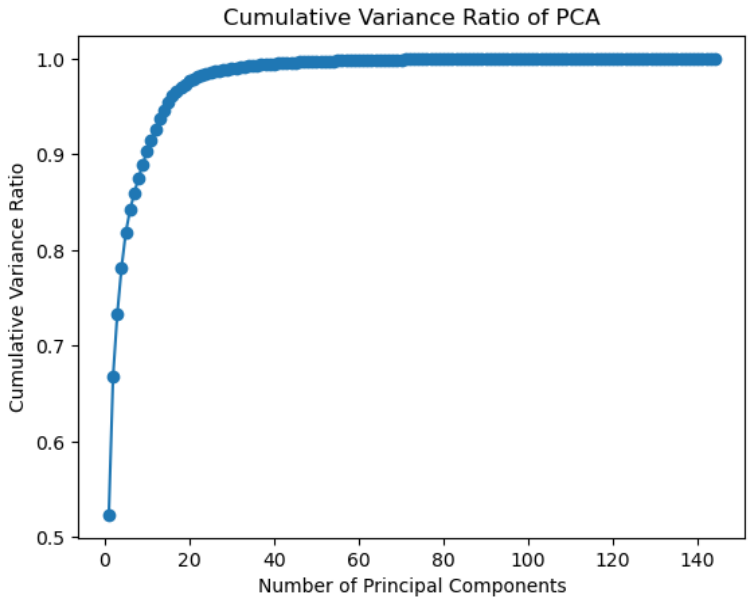

Figure 2: Cumulative Variance Ratio of PCA

4.2. Model selection

Standardization can ensure that all input features are computed on the same scale, avoiding the impact of differences in feature values on gradient descent. Standardization can accelerate the convergence speed of the model and improve training efficiency. Additionally, it can also ensure fair weight updates and prevent unreasonable situations from occurring. Standardization accelerates model convergence and improves training efficiency. Additionally, it ensures fair weight updates and prevents undesired situations.

$$x_t' = \frac{x_t - x_{min}}{x_{max} - x_{min}}$$

Where x_t' represents the normalized data, x_t represents the observe data, and x_{min} and x_{max} represent the minimum and maximum values of the explanatory variable sequence. By using grid search and cross-validation, we searched for appropriate parameters on the validation set. As a result, the BPNN model was built with one hidden layer, six neurons, a threshold set at 0.006, and a learning rate of 0.01.

4.3. Forecasting results

Table 1: Forecasting results

	MAE	MAPE	RMSE
PCA-BPNN	**3918.8230**	**10.5638**	**5272.7993**
SRRIMA	4184.2690	12.1601	5366.4689
SNAÏVE	4029.3012	12.0780	6153.3287
SARIMAX	4525.5501	12.9490	5831.6043

As shown in Table 1, the forecasting accuracy of PCA-BPNN outperformed all benchmark models in terms of prediction accuracy, with an average improvement of 7.2975% in MAE, 14.6944% in MAPE, and 8.5459% in RMSE. Specifically, compared to the SARIMAX model, PCA-BPNN is more effective in identifying the nonlinear relationship between tourist arrivals and search keywords, resulting in better predictive performance. In contrast, SARIMA and SNAIVE rely on historical data for prediction without considering the influence of other variables.

5. Conclusion

In this study, we used weekly visitor flow data from Gulangyu Scenic Area and online data from Baidu search engine, specifically Baidu Index for keywords, to make predictions. We applied PCA to

ISSN 2616-5775 Vol. 6, Issue 9: 72-80, DOI: 10.25236/AJCIS.2023.060911

reduce the dimensionality of the search data, set a threshold, and extracted principal components to construct the PCA-BPNN model. The predictive performance of this model was compared with the SARIMA, SARIMAX, and Naive models. Empirical research revealed that the PCA-BPNN model outperformed the SARIMA and Naive models, which did not incorporate online data. This finding demonstrates the advantages of online data in terms of its broad sources, diverse indicators, large quantity, and strong timeliness.

This study contributes in two aspects. Firstly, from a theoretical perspective, it was found that PCA-BPNN can effectively identify the nonlinear relationship between explanatory and dependent variables. Additionally, incorporating search data into the model was found to improve predictive accuracy. Therefore, in future studies on data forecasting, the reasonable utilization of online data resources can enhance the interpretability of models and improve their practical effectiveness. Secondly, from an applied standpoint, accurate demand forecasting enables scenic areas to allocate resources according to the demand levels during different time periods. This helps to avoid resource waste and unnecessary cost expenditure, thereby improving the efficiency of resource utilization.

This article still has some limitations. Firstly, it only utilized search engine data and did not consider the comment data from OTA platforms. Secondly, when using search data, there was no differentiation between data from mobile devices and PC devices. Lastly, deep learning models such as Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN) were not employed.

References

[1] Song, H., & Li, G. (2008). Tourism demand modelling and forecasting—A review of recent research. Tourism management, 29(2), 203-220.
[2] Kulendran, N., & Wong, K. K. (2005). Modeling seasonality in tourism forecasting. Journal of Travel Research, 44(2), 163-170.
[3] Guizzardi, A., & Mazzocchi, M. (2010). Tourism demand for Italy and the business cycle. Tourism Management, 31(3), 367-377.
[4] Gunter, U., & Önder, I. (2016). Forecasting city arrivals with Google Analytics. Annals of Tourism Research, 61, 199-212.
[5] Lendasse, A., Oja, E., Simula, O., & Verleysen, M. (2007). Time series prediction competition: The CATS benchmark. Neurocomputing, 70(13-15), 2325-2329.
[6] Moore, W. R. (2010). The impact of climate change on Caribbean tourism demand. Current Issues in Tourism, 13(5), 495-505.
[7] Wan, S. K., & Song, H. (2018). Forecasting turning points in tourism growth. Annals of Tourism Research, 72, 156-167.
[8] Smeral, E., & Song, H. (2015). Varying elasticities and forecasting performance. International Journal of Tourism Research, 17(2), 140-150.
[9] Hu, M., & Song, H. (2020). Data source combination for tourism demand forecasting. Tourism Economics, 26(7), 1248-1265.
[10] Alshanbari, H. M., Mehmood, T., Sami, W., Alturaiki, W., Hamza, M. A., & Alosaimi, B. (2022). Prediction and Classification of COVID-19 Admissions to Intensive Care Units (ICU) Using Weighted Radial Kernel SVM Coupled with Recursive Feature Elimination (RFE). Life, 12(7), 1100.
[11] Adil, M., Wu, J. Z., Chakrabortty, R. K., Alahmadi, A., Ansari, M. F., & Ryan, M. J. (2021). Attention-based STL-BiLSTM network to forecast tourist arrival. Processes, 9(10), 1759.
[12] Zhang, C., & Tian, Y. X. (2022). Forecast daily tourist volumes during the epidemic period using COVID-19 data, search engine data and weather data. Expert Systems with Applications, 210, 118505.
[13] Chen, Y. C., & Huang, W. C. (2021). Constructing a stock-price forecast CNN model with gold and crude oil indicators. Applied Soft Computing, 112, 107760.
[14] 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.
[15] Yang, X., Pan, B., Evans, J. A., & Lv, B. (2015). Forecasting Chinese tourist volume with search engine data. Tourism management, 46, 386-397.
[16] Li, C., Ge, P., Liu, Z., & Zheng, W. (2020). Forecasting tourist arrivals using denoising and potential factors. Annals of Tourism Research, 83, 102943.
[17] Zhang, C., Tian, Y. X., & Fan, Z. P. (2022). Forecasting sales using online review and search engine data: A method based on PCA–DSFOA–BPNN. International Journal of Forecasting, 38(3), 1005-1024.
[18] 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.
[19] Li, X., Li, H., Pan, B., & Law, R. (2021). Machine learning in internet search query selection for

ISSN 2616-5775 Vol. 6, Issue 9: 72-80, DOI: 10.25236/AJCIS.2023.060911

tourism forecasting. Journal of Travel Research, 60(6), 1213-1231.
[20] Darst, B. F., Malecki, K. C., & Engelman, C. D. (2018). Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. BMC genetics, 19(1), 1-6.
[21] Xie, G., Li, X., Qian, Y., & Wang, S. (2021). Forecasting tourism demand with KPCA-based web search indexes. Tourism Economics, 27(4), 721-743.
[22] Li, X., Pan, B., Law, R., & Huang, X. (2017). Forecasting tourism demand with composite search index. Tourism management, 59, 57-66.
[23] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences, 374(2065), 20150202.
[24] Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.