Journal of Humanities, Arts and Social Science, 2023, 7(3), 529-539
https://www.hillpublisher.com/journals/jhass/
ISSN Online: 2576-0548
ISSN Print: 2576-0556

Short and Long Term Tourism Demand Forecasting Based on Baidu Search Engine Data

Mei Li[1,*], **Tiaorong Xu**[2]

[1]Economics School, Guangxi University, Nanning, Guangxi, China.
[2]School of Economics and Management, Changsha University of Science and Technology, Changsha, Hunan, China.

How to cite this paper: Mei Li, Tiaorong Xu. (2023) Short and Long Term Tourism Demand Forecasting Based on Baidu Search Engine Data. *Journal of Humanities, Arts and Social Science*, 7(3), 529-539.
DOI: 10.26855/jhass.2023.03.010

Received: March 2, 2023
Accepted: March 30, 2023
Published: April 27, 2023

*__Corresponding author:__ Mei Li, Economics School, Guangxi University, Nanning, Guangxi, China.

Abstract

With the increasing contribution of tourism in economic development, how to improve the forecasting accuracy has become the focus of researchers and practicer. Statistical data can hardly reflect the useful information in tourism demand, and its role in forecasting tourism demand is very limited. In order to reflect tourists' tourism behavior more comprehensively, many studies have proved the role of online search data in tourism demand forecasting. Further, this study aims to forecasting the volume of tourists from mainland China to Macao, Hong Kong and Thailand based on Baidu Search Index, and applies seasonal autoregressive integrated moving average with explanatory variable (SARIMAX), so the forecasting model called Principal Component Analysis-Seasonal Autoregressive Integrated Moving Average with exogenous variable (PCA-SARIMAX) model is proposed, and the empirical results show that the proposed model has better forecasting effect compared with the benchmark models, and search engine data plays a positive role in forecasting tourism demand.

Keywords

Tourism demand forecasting, Baidu Search Index, Internet search data, PCA-SARIMAX

Introduction

Tourism is an important part of human activities, since the 21st century, with the improvement of living standards, people have more and more tourism activities. Because the rapid growth of tourism around the world, the role of tourism has become more and more important in economics, accurate tourism demand prediction is important reference for decision-makers and planner. However, there are various fluctuations in the tourism market due to it was affected by some internal and external factors, so accurate prediction of tourism demand has become the concerns of researchers and decision-makers (Song, Qiu, & Park, 2019).

With the rapid popularization of mobile internet technology, people's daily life is flooded by the Internet, which also makes Internet search data an important data source for tourism demand prediction. Consequently, traditional research that only introduces some macroeconomic variables, such as income level, GDP and exchange rate (Wong et al., 2007) into the tourism demand forecasting model is not enough to reflect the behavior of tourists, so their contributions to the accuracy of tourism demand forecasting is limited. Baidu Search Index is a record and relevant information statistics of the user's search traces when searching on the Baidu Index website. Tourism related Baidu search reflects the tourists' desire for the tourism destination. Compared with the existing research relying only on the original internal data, the extensive use of Baidu Search provides a more comprehensive and novel data source and new ideas for researching human behavior analysis. In the era of rapid growth of science and technology, tourists often use Baidu search engine to query and collect information when making a series of decisions such as accommodation, tourist resort and activities (Pan, Litvin, & Goldman, 2006), which makes a correlation between Baidu search data and the volume of tourists arrivals. This behavior information and correlation of tourists' search can reflect tourists' travel intentions to a certain extent. This will provide a basis for tourism researchers to analyze the tourist volume of tourist destinations and use it for tourism prediction.

We compared the performance of SARIMAX model and three classic time series models including seasonal autoregressive integrated moving average (SARIMA), SNAIVE and ETS model to verify the usability of Baidu search index in tourism demand forecasting. This research will be helpful to the existing research in two aspects: first, we will verify the impact of Baidu search index on tourism demand and its act on tourism demand prediction; The second is to compare the prediction accuracy of SARIMAX model before and after synthetizing principal component analysis to provide a reference for whether Baidu search index needs to be synthetized.

1. Literature review

1.1 Tourism demand forecasting

Tourism demand forecasting has attracted the attention of many researchers, who mainly use time series, econometric and artificial intelligence model to model and forecast tourism demand (Song & Li, 2008; Song, Qiu, & Park, 2019).

Time series models are regarded as the benchmark models. Among time series models, autoregressive integrated moving average (ARIMA) model proposed by Box and Jenkins (1970) is the most widely applied. Since then, some variants of ARIMA model are used by researchers, for example, autoregressive integrated moving average with exogenous variable (ARIMAX) model comes into being when various explanatory variables affecting tourism demand are added to the time series model, seasonal autoregressive integrated moving average (SARIMA) model is used for forecasting when seasonality is considered. Tsui et al. (2014) apply SARIMA model and ARIMAX model to forecast airport traffific demand, they found that SARIMA model and ARIMAX model demonstrates good forecasting performance for different forecasting periods.

Econometric models are widely used when the relationship between tourism demand and its influencing factors needs research. Static regression (SR) is first used for tourism demand forecasting (Morley, 1992). Later, some advanced econometric models were proposed to remedy the limitations of SR, such as the distributed lag (DL) model (Guizzardi & Stacchini, 2015), the auto regressive distributed lag model (ADLM) (Blunk, Clark, & McGibany, 2006), the error correction model (ECM) (Nadal, Font, & Rosselló, 2004), Vector Autoregressive (VAR) (Song & Witt, 2006). Mixed-data sampling (MIDAS) model (Bangwayo-Skeete & Skeete, 2015; Wen et al., 2021) is used to explain the case that the variable and the explained variable have different frequencies.

Many artificial intelligence (AI) models have been used to the field of tourism management in an effort to promote in-depth research on tourism demand. Law and Au (1999) used multiple models to forecast Japanese demand for travel to Hong Kong, the results demonstrated that Neural Network (NN) model have smaller error than

NAIVE, moving average, exponential smoothing and multiple regression. Other AI models include the rough sets approach (Goh, Law, &Mok, 2008), SVR (Support Vector Regression) (Hong et al., 2011), the fuzzy time series and grey theory (Wang, 2004) are also used to forecast tourism demand, and have proved to have a good forecasting effect.

1.2 Forecasting with search query data

The search engine data currently used are mainly Baidu Search and Google Trends. Feng, Sun, and Li (2019) forecast China inbound tourism based on Google Trends, they found that Random Forest (RF) model has high prediction accuracy. Park, Lee, and Song (2017) use Google trend data as explanatory variables to construct SARIMAX model to forecast tourists, the results show that the SARIMAX model has better prediction performance than benchmark model. Önder (2017) compares four forecasting models with Google Trends, the research results confirm the effect of Google Trends. In general, the role of Google trend as an explanatory variable in improving the prediction accuracy has been confirmed (Önder & Gunter, 2016; Yang, Pan, & Song, 2014).

Internet big data sources include Baidu search index in addition to Google trend, Baidu Index has the largest market in China and has higher prediction accuracy than Google Trend (Yang et al., 2015). Tang et al. (2020) integrate the bivariate empirical mode decomposition with Baidu search index to forecast the tourist volume of Hainan in China, and they found that Baidu search index is conducive to improving the forecasting accuracy. Wen, Liu, and Song (2019) took Baidu search engine as the Internet big data source to explore a composite search index, the findings indicated that principal component analysis (PCA) helps to improve prediction accuracy, and the effectiveness of PCA has also been proved by other scholars (Li et al., 2015; Li et al., 2018).

Considering the share of Baidu search index in the Chinese market and its role in tourism demand forecasting, SARIMAX model can be extended to take Baidu index as explanatory variable. In this paper, the model called Principal Component Analysis-Seasonal Autoregressive Integrated Moving Average with exogenous variable (PCA-SARIMAX) was proposed, the principal component analysis (PCA) method is used to synthesize Baidu search engine into an index, and the composite index is introduced into the SARIMAX model to forecast tourism demand, and compared with the classic time series model.

2. Methodology

This research forecasts the volumes of tourists arrivals from mainland China to Macao based on Baidu search index. Besides, In order to verify the robustness of the conclusion, this paper selects different test sets and forecasting steps respectively, it also proves the role of Baidu search engine in tourism demand forecasting. The research framework of this paper is shown in Figure 1.

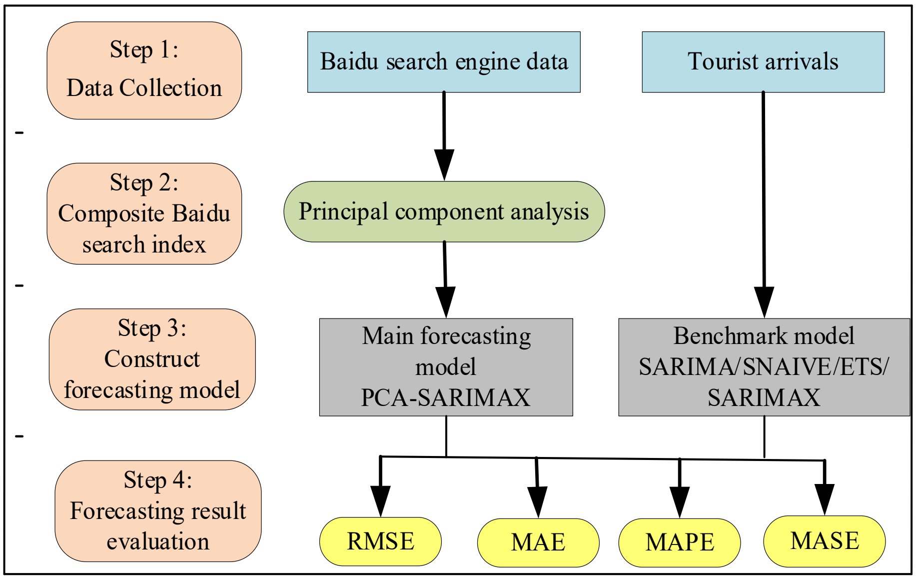

Figure 1. Research framework.

For verifying the validity of Baidu search engine data, this study constructs the PCA-SARIMAX model. The principal component analysis and SARIMAX model used in this paper will be introduced below.

2.1 Principal component analysis

With *n* months of sample data and *m* Baidu search engines keywords related to tourism demand, the original sample matrix is:

$$\mathrm{X} = \begin{pmatrix} \mathrm{A}_{\mathrm{m}1} & \cdots & \mathrm{A}_{1\mathrm{n}} \\ \vdots & \ddots & \vdots \\ \mathrm{A}_{\mathrm{m}1} & \cdots & \mathrm{A}_{\mathrm{mn}} \end{pmatrix} = (\mathrm{X}_{\mathrm{pq}})_{\mathrm{m}\times\mathrm{n}}$$

In this matrix, $\mathrm{p} = (1,2,\cdots\mathrm{m})$ represents row *p* of the original sample matrix and $\mathrm{q} = (1,2,\cdots\mathrm{n})$ denotes column *q* of the original sample matrix. Principal component analysis is mainly divided into the following four steps:

Step 1: Calculate the correlation coefficient matrix.

$$\mathrm{R} = (\mathrm{r}_{\mathrm{pq}})_{\mathrm{m}\times\mathrm{n}}$$

$$r_{pq} = \frac{1}{n}\sum_{p=1}^{m}\frac{(x_{pq} - x_p)((x_{pq} - x_q)}{\delta}$$

where, δ represents sample variance.

Step 2: Calculate the eigenvalues and eigenvectors of R. According to the expression of characteristic equation $|R-\gamma I|=0$, it can be calculated γ value. The expression R represents the correlation coefficient matrix and I represents the identity matrix. According to γ in descending order, can get $\gamma_1, \gamma_2, \cdots, \gamma_n$, at the same time, an eigenvector can be calculated.

Step 3: Calculate contribution rate (e_i) and cumulative contribution rate (E_m).

$$e_i = \frac{\gamma_i}{\sum_{p=1}^{m} \gamma_i}$$

$$E_m = \frac{\sum_{i=1}^{m} \gamma_i}{\sum_{p=1}^{m} \gamma_i}$$

Step 4: Calculate principal components.

$$Z_i = A_{iq} x_q$$

Step 5: Comprehensive analysis.

Determine the number of principal components by calculating the cumulative contribution rate, and select all factors>70% as the principal components. Finally, according to the obtained indicators, build an indicator system of Baidu search keywords affecting tourism demand.

2.2 SARIMAX model

SARIMA consists of AR, MA, seasonal AR and seasonal MA. SARIMA *(p,d,q) (P,D,Q)*$_m$, The SARIMA model of Box and Jenkins (1970) is formulated as:

$$\Phi(B^m)\varphi(B)(1-B^m)^D(1-B)^d y_t = c + \Theta(B^m)\theta(B)\varepsilon_t$$

where B is the backshift operator; m denotes the seasonal cycle;$\Phi(x)$ and $\Theta(x)$ are polynomials of orders P and Q, respectively; $\varphi(x)$ and $\theta(x)$ are polynomials of orders p and q, respectively; ε_t is a white noise process with a mean of zero and a variance of σ^2; and d and D refer to the rank of difference, which is determined by a unit root test.

When exogenous variables are included in SARIMA, the SARIMAX model can be expressed as:

$$\Phi(B^m)\varphi(B)(1-B^m)^D(1-B)^d y_t = c + \sum_i \sum_k \beta_{i(t-k)} X_{i(t-k)} + \Theta(B^m)\theta(B)\varepsilon_t$$

Some parameters of the above formula have the same meanings as those contained in the SARIMA model. In addition, X_i represents the *I* th exogenous variable, which is composite index of Baidu index in this study and the corresponding coefficient.

3. Data

3.1 Data sources

We collect data on the monthly tourists arrivals from mainland China to Macao from Wind Database from January 2011 to June 2019. Daily Baidu search engine data is collected from Baidu Index website, the time range is January 1, 2011 to December 31, 2019. The correlation between monthly tourists arrivals and Baidu search keywords is shown in Table 1, the bold keywords are the 10 most relevant keywords, which are the 10 baidu search engine keywords finally selected in this article.

This study use principal component analysis to synthesize Baidu search engine data, and the variance contribution rate of the first principal component is as high as 71.22%, it is regarded as the final explanatory variable.

3.2 Basic analysis

After synthesizing Baidu search index, it is necessary to analyze the relationship between explanatory variables and explained variables, Figure 2 simply shows the trend between the number of tourists arrivals and the first principal component, from the figure, there is a consistent trend between the number of tourists arriving and the synthetised principal component.

Table 1. Correlation between search query keywords and tourists arrivals

Search query keywords	Lag order	Correlation	Search query keywords	Lag order	Correlation
Dining			New Yaohan	9	0.584
Macau Restaurant	10	-0.238	AUD to RMB	7	0.328
Macau Food Festival	7	-0.211	Macao Handletter	0	0.658
Macao specialty	0	0.662	**Macao DFS**	**0**	**0.778**
Macao Doulao	1	-0.600	Tour		
Macao cuisine	1	0.387	**Macao weather forecast**	0	0.721
Traffic			**Macao weather**	12	0.790
Satellite Map of Macao	5	-0.630	Macao Free Travel	12	0.139
Macao 3D Map	0	-0.660	One day tour of Macao	0	0.629
Macao Tourist Map	2	-0.093	Macao Travel	1	0.544
Full Map of Macao	0	0.505	Macao tourism guide	1	-0.234
Map of Macao	7	-0.303	Macao guide	0	0.574
Macau Airlines official website	12	0.701	Attraction		
Macao visa	0	0.350	**Macao tourist attractions**	**0**	**0.745**
Macao Airlines	**1**	**0.803**	**Macao Attractions**	0	0.809
Macao airport	**0**	**0.768**	Macao Venice	7	0.699
Macau car rental	3	0.684	Macau Lisboa Casino	4	0.303
Macao endorsement	**12**	**0.779**	Macao Sauna	12	0.484
Macao Grand Triple Bus	**12**	**0.826**	Macao Casino	1	0.226
Hong Kong to Macao	**6**	**0.833**	Is Macao fun	7	0.493
Lodging			Macao Tourism	0	0.149
Macao accommodation guide	5	0.327	The Venetian Macau	11	-0.313
Macao hotel	12	0.627	Macao new Lisboa	12	0.604
Macao hotel book	2	-0.314	Macao Lisboa	0	0.681
Macaohotelrecommendation	12	0.505	Macau Gambling	2	0.303
Macao accommodation	12	0.666	Macau tourism Tower	0	0.313
MGM Macaohotel	12	0.143	Macao tower	12	0.630
Macao wynn hotel	10	0.658	Macao Sightseeing Tower	12	0.193
Macao galaxy hotel	12	0.557	Macao Entertainment	5	0.613
Macao Parisian Hotel	12	0.691	Macau Casino Pictures	0	-0.393
The Venetian Macao-Resort-Hotel	2	-0.488	What are the casinos in Macao	0	0.574
Shopping			Gambling City	0	0.384
Macao shopping guide	12	0.551			

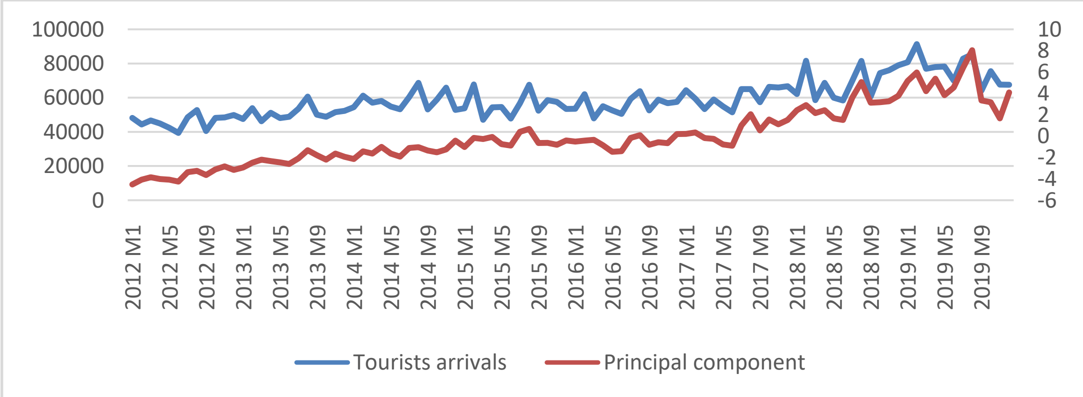

Figure 2. Relationship between the first principal component and tourists arrivals.

The regression relationship between the number of tourists arrivals from mainland China to Macao and the first principal component is shown in Table 2, from the results, there is a significant causal relationship between the number of tourists arrivals and the first principal component with the fitting degree reaching 80%, the increase of Baidu search index has a significant positive effect on the number of tourists arrivals.

Table 2. Regression analysis results of tourist arrival and the first principal component

Dependent Variable	Tourists arrivals
Independent Variable	Generalized least square method
Principal component	3655.7***（19.6）
Standard error	186.6
Adjusted R^2	0.801
Observations	96

4. Empirical results

In this study, RMSE, MAE, MAPE and MASE are used as the evaluation index of forecasting performance. The calculation formula of each evaluation index is as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}|y_i - y_i'|^2} \qquad MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - y_i'|$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - y_i'}{y_i}\right| \qquad MASE = mean(|q_t|) \qquad q_t = \frac{e_t}{\frac{1}{n-M}\sum_{i=M+1}^{n}|y_i - y_{i-M}|},$$

where y_i denotes the actual value of the observation; y_i' is the prediction value; and *N* represents the length of the forecasting period, *M* indicates the number of seasons.

Tables 3 summarises the measurement results of the PCA- SARIMAX, SARIMAX, SARIMA, ETS, and SNAIVE models when forecasting tourist demand from mainland China to Macao. The results show that the PCA-SARIMAX model shows better forecasting performance than benchmark models. The results imply that introducing Baidu search engine data into the forecasting model can improve the prediction accuracy, which are consistent with previous studies (Li et al., 2016; Li et al., 2017).

The tourists arrivals from mainland China to Hong Kong and Thailand are used for robustness test, the results are shown in Table 4. For Hong Kong, the PCA-SARIMAX model based on Baidu Search engine data consistently outperforms the time series models when forecasting multiple-step-ahead. Besides, for benckmark time series forecasting models, PCA-SARIMA model is superior to other time series models in short-term prediction, while long-term prediction is inferior to other time series models. For Thailand, PCA-SARIMAX model show higher prediction accuracy than the time series models in most cases, the prediction performance of the PCA-SARIMAX model is proved again.

Table 3. Forecasting valuation from mainland China to Macao

	Measure			
	RMSE	MAE	MAPE	MASE
h = 1				
SARIMA	5329	4567	6.21	0.75
ETS	6245	5468	7.25	0.90
SNAIVE	10036	7700	10.37	1.27
SARIMAX	5712	4286	5.70	0.71
PCA-SARIMAX	4762	3888	5.24	0.64
h = 2				
SARIMA	5907	5031	6.94	1.23
ETS	6217	5366	7.28	1.40
SNAIVE	10423	9005	11.97	1.27
SARIMAX	7900	6410	8.34	1.06
PCA-SARIMAX	5510	4641	6.19	0.77
h = 3				
SARIMA	7434	6288	8.87	1.30
ETS	8509	7409	10.23	1.47
SNAIVE	8485	7126	9.72	1.27
SARIMAX	7385	5540	7.54	0.91
PCA-SARIMAX	5254	4615	6.23	0.76
h = 6				
SARIMA	11896	10075	14.46	1.33
ETS	15086	12720	18.00	1.65
SNAIVE	7635	6476	9.26	1.27
SARIMAX	7368	5775	7.96	0.95
PCA-SARIMAX	4564	4134	5.60	0.68
h = 9				
SARIMA	16689	15417	22.59	1.70
ETS	18486	17064	25.04	1.83
SNAIVE	14977	13284	19.73	1.27
SARIMAX	7400	5951	8.22	0.98
PCA-SARIMAX	5553	4828	6.51	0.80

Table 4. Forecasting valuation from mainland China to Hong Kong and Thailand

	Measure			
	RMSE	MAE	MAPE	MASE
Hong Kong				
h = 1				
SARIMA	10686	9092	6.01	0.61
ETS	10561	9468	6.39	0.63
SNAIVE	17558	15531	10.68	1.04
SARIMAX	10733	9534	6.25	0.64
PCA-SARIMAX	10692	9313	6.07	0.62
h = 2				
SARIMA	12828	10586	6.88	0.99
ETS	11541	9797	6.51	1.14
SNAIVE	21449	18390	12.24	1.04
SARIMAX	13756	11696	7.63	0.79
PCA-SARIMAX	11755	10070	6.55	0.67
h = 3				
SARIMA	16006	12958	8.44	0.84
ETS	13377	12053	8.06	0.99
SNAIVE	21626	16958	10.99	1.04
SARIMAX	10605	8604	5.54	0.58
PCA-SARIMAX	11449	9869	6.51	0.66
h = 6				
SARIMA	18610	15962	10.08	1.26
ETS	14539	12600	8.07	1.54
SNAIVE	26332	21504	13.82	1.04
SARIMAX	12392	10647	7.03	0.71
PCA-SARIMAX	11461	9715	6.36	0.65
h = 9				
SARIMA	15223	13569	9.33	1.37
ETS	13380	11056	7.59	1.58
SNAIVE	12731	9198	6.48	1.04
SARIMAX	8901	7527	4.99	0.51
PCA-SARIMAX	9607	7744	5.08	0.52
Thailand				
h = 1				
SARIMA	1525	1153	3.89	0.24
ETS	1769	1365	4.45	0.29
SNAIVE	3861	3214	10.45	0.67
SARIMAX	2667	2247	7.76	0.47
PCA-SARIMAX	2201	1946	6.55	0.41
h = 2				
SARIMA	2525	2075	7.10	0.78
ETS	1998	1824	6.30	0.61
SNAIVE	5720	4916	16.14	0.67
SARIMAX	3601	2910	9.98	0.61
PCA-SARIMAX	1526	1187	3.97	0.25
h = 3				
SARIMA	3255	2686	9.38	1.12
ETS	3031	2418	8.37	0.78
SNAIVE	5865	4804	16.97	0.67
SARIMAX	1666	1183	4.20	0.25
PCA-SARIMAX	1608	1210	4.18	0.25
h = 6				
SARIMA	3496	2878	9.67	0.98
ETS	2661	2046	7.01	0.78
SNAIVE	2788	2340	7.90	0.67
SARIMAX	2507	1750	5.91	0.37
PCA-SARIMAX	2168	1738	5.62	0.37
h = 9				
SARIMA	2587	2548	9.30	1.51
ETS	1313	1177	4.32	1.21
SNAIVE	7268	6208	23.09	0.67
SARIMAX	4721	3188	10.89	0.67
PCA-SARIMAX	2363	1779	6.13	0.37

This paper uses the Improvement Rate (IR) to further evaluate the PCA-SARIMAX model, its calculation formula is as follows:

$$IR = \frac{MAPE(\text{Model}_{SARIMAX}) - MAPE(Model_{PCA-SARIMAX})}{MAPE(\text{Model}_{SARIMAX})} \times 100\%$$

where, $MAPE(Model_{SARIMAX})$ denotes the MAPE value of SARIMAX model, $MAPE(Model_{PCA-SARIMAX})$ denotes the MAPE value of PCA-SARIMAX model, IR greater than 0 indicates that PCA-SARIMAX model has improved effect compared with SARIMAX model. Figure 3 shows the forecasting effect improvement of PCA-SARIMAX model when forecasting the tourism demand from Macao, Hong Kong and Thailand to mainland China.

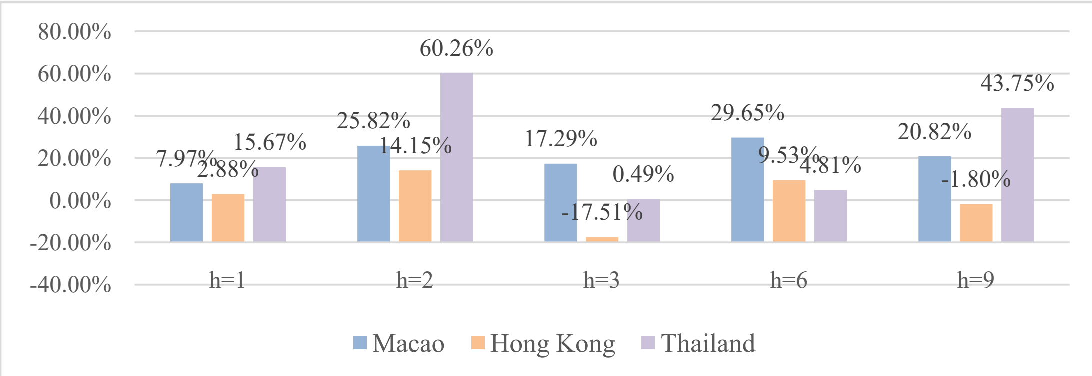

Figure 3. Forecasting effect improvement of PCA-SARIMAX.

It's not hard to see from Figure 3, the IR values of PCS-SARIMAX are all positive when forecasting Macao and Thailand, which indicates that the PCA-SARIMAX model has obvious improvement compared with the SARIMAX model. Meanwhile, for Hong Kong, the IR values are higher than 0 in most cases, on the whole, the forecasting effect improvement of PCA-SARIMAX when forecasting tourism demand from Macao to mainland China is the most obvious, reaching the maximum value of 60.26%. In sum, compared with SARIMAX model, PCA-SARIMAX model shows improved prediction effect, which implies the necessity of Baidu search index synthesis, which again confirmes the above results.

5. Conclusion

This study uses Baidu search engine keywords as the explanatory variable of tourism demand. First, Baidu search keywords related to tourism are collected, and correlation analysis is used for removing some irrelevant Baidu search keywords and reduce dimensions. After that, 10 selected Baidu search keywords are synthetized into an index using principal component analysis, lastly, four models are applied to forecast tourism demand. Research results reveals that introducing the synthetized Baidu search engine data into the PCA-SARIMAX model contributes to improve the prediction accuracy.

There are also shortcomings in this study. For example, in addition to Baidu Index, some other types of big data sources will have an impact on tourism demand in the context of internet big data. Therefore, in the future, more Internet big data sources can be considered to forecast tourism demand. Meanwhile, some emergency events such as political, financial, economic and public health events can be introduced into the forecasting model.

References

Afrianto, M. A., & Wasesa, M. (2022). The impact of tree-based machine learning models, length of training data, and quarantine search query on tourist arrival prediction's accuracy under COVID-19 in Indonesia. *Current Issues in Tourism*, 1-17.

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.

Blunk, S. S., Clark, D. E., & McGibany, J. M. (2006). Evaluating the long-run impacts of the 9/11 terrorist attacks on US domestic airline travel. *Applied Economics, 38*(4), 363-370.

Box, G. E. P., & Jenkins, G. M. (1970). Time series analysis forecasting and control. *Chapman-Hall /crc Boca Raton, 2*(3), 131-133.

Feng, Y., Li, G., Sun, X., & Li, J. (2019). Forecasting the number of inbound tourists with Google Trends. *Procedia Computer Science*, *162*, 628-633.

Goh, C., Law, R., & Mok, H. M. (2008). Analyzing and forecasting tourism demand: A rough sets approach. *Journal of Travel Research, 46*(3), 327-338.

Guizzardi, A., & Stacchini, A. (2015). Real-time forecasting regional tourism with business sentiment surveys. *Tourism Management, 47*, 213-223.

Hong, W. C., Dong, Y., Chen, L. Y., & Wei, S. Y. (2011). SVR with hybrid chaotic genetic algorithms for tourism demand forecasting. *Applied Soft Computing, 11*(2), 1881-1890.

Hu, M., Li, H., Song, H., Li, X., & Law, R. (2022). Tourism demand forecasting using tourist-generated online review data. *Tourism Management, 90*, 104490.

Hu, Y. C., & Wu, G. (2022). The impact of Google Trends index and encompassing tests on forecast combinations in tourism. *Tourism Review,* (ahead-of-print).

Law, R., & Au, N. (1999). A neural network model to forecast Japanese demand for travel to Hong Kong. *Tourism Management*, *20*(1), 89-97.

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.

Li, X., Shang, W., Wang, S., & Ma, J. (2015). A MIDAS modelling framework for Chinese inflation index forecast incorporating google search data. *Electronic Commerce Research and Applications, 14*(2), 112-125.

Morley, C. L. (1992). A microeconomic theory of international tourism demand. *Annals of Tourism Research, 19*(2), 250-267.

Nadal, J. R., Font, A. R., &Rosselló, A. S. (2004). The economic determinants of seasonal patterns. *Annals of Tourism Research, 31*(3), 697-711.

Önder, I. (2017). Forecasting tourism demand with Google trends: Accuracy comparison of countries versus cities. *International Journal of Tourism Research*, *19*(6), 648-660.

Önder, I., & Gunter, U. (2016). Forecasting tourism demand with Google Trends for a major European city. *Tourism Analysis, 21*(2/3), 203-220.

Pan, B., Litvin, S. W., & Goldman, H. (2006, June). Real users, real trips, and real queries: An analysis of destination search on a search engine. *In Annual Conference of Travel and Tourism Research Association* (TTRA 2006). Ireland, Dublin (pp. 16-18).

Pan, B., Wu, D. C., & Song, H. (2012). Forecasting hotel room demand using search engine data. *Journal of Hospitality and Tourism Technology.*

Park, S., Lee, J., & Song, W. (2017). Short-term forecasting of Japanese tourist inflow to South Korea using Google trends data. *Journal of Travel & Tourism Marketing*, *34*(3), 357-368.

Shen, S., Li, G., & Song, H. (2011). Combination forecasts of international tourism demand. *Annals of Tourism Research, 38*(1), 72-89.

Song H & Li G. (2008). Tourism demand modelling and forecasting—a review of recent research. *Tourism Management 29*(2): 203-220.

Song, H., & Witt, S. F. (2006). Forecasting international tourist flows to Macau. *Tourism Management, 27*, 214-224.

Song, H., Qiu, R. T., & Park, J. (2019). A review of research on tourism demand forecasting: Launching the Annals of Tourism Research Curated Collection on tourism demand forecasting. *Annals of Tourism Research, 75*, 338-362.

Tang, L., Zhang, C., Li, T., & Li, L. (2021). A novel BEMD-based method for forecasting tourist volume with search engine data. *Tourism Economics, 27*(5), 1015-1038.

Tsui, W. H. K., Balli, H. O., Gilbey, A., & Gow, H. (2014). Forecasting of Hong Kong airport's passenger throughput. *Tourism Management*, *42*, 62-76.

Wang, C. H. (2004). Predicting tourism demand using fuzzy time series and hybrid grey theory. *Tourism management, 25*(3), 367-374.

Wen, L., Liu, C., & Song, H. (2019). Forecasting tourism demand using search query data: A hybrid modelling approach. *Tourism Economics*, *25*(3), 309-329.

Wen, L., Liu, C., Song, H., & Liu, H. (2021). Forecasting tourism demand with an improved mixed data sampling model. *Journal of Travel Research*, *60*(2), 336-353.

Yang, X., Pan, B., Evans, J. A., & Lv, B. (2015). Forecasting Chinese tourist volume with search engine data. *Tourism Management, 46*, 386-387.

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, B., Li, N., Law, R., & Liu, H. (2021). A hybrid MIDAS approach for forecasting hotel demand using large panels of search data. *Tourism Economics*, 13548166211015515.