Research on Tourism Demand Forecasting and Tourist Search Behavior Based on Internet Search Index

Ke Ma

Hubei Three Gorges Polytechnic; Yichang, Hubei, 443000, China

Email: ptm033945@163.com

Abstract. This study explores the application of internet search indexes in tourism demand forecasting and analyzes the characteristics and influencing factors of tourist search behavior. Through time series analysis, regression analysis, and clustering analysis, a tourism demand forecasting model is constructed, revealing the significant impacts of seasonal factors, holiday effects, and unexpected events on tourist search behavior. The results indicate that the LSTM model excels in capturing tourism search trends, and tourist search behavior exhibits clear seasonal variations and trend fluctuations. Clustering analysis identifies three distinct groups of tourists with different search behavior characteristics, providing a basis for market segmentation and precise marketing. Specific recommendations for optimizing marketing strategies, real-time monitoring of market demand, and responding to unexpected events are proposed. Future research should further explore multi-dimensional data integration and more predictive models to enhance forecasting accuracy and application scope.

Keywords: Internet Search Index, Tourism Demand Forecasting, Tourist Search Behavior, Big Data Analysis.

1 Introduction

With the rapid development of internet and big data technologies, internet search indexes have garnered widespread attention as a real-time data source that reflects user interests and behaviors[1]. The tourism industry, being an information-intensive sector, is significantly influenced by advancements in information technology. Internet search indexes can analyze user search behaviors to timely capture market dynamics, providing new data support for tourism market forecasting and analysis. Specific application scenarios include tourism route planning, product recommendations.

The research aims to explore the application of internet search indexes in tourism demand forecasting and analyze the characteristics of tourist search behavior, thereby providing theoretical support and practical guidance for tourism market forecasting and analysis. By establishing predictive models, the study attempts to uncover the changing patterns of tourism search trends and analyze the main factors influencing tourist search behavior.

L. Chang et al. (eds.), *Proceedings of the 2024 8th International Seminar on Education, Management and Social Sciences (ISEMSS 2024)*, Advances in Social Science, Education and Humanities Research 867,
https://doi.org/10.2991/978-2-38476-297-2_93

2 Literature Review

2.1 Application of Internet Search Indexes

As a novel source of big data, internet search indexes have been widely applied across various fields[2]. Scholars have applied the Google search index to economic forecasting, discovering its significant predictive power for consumer behavior, unemployment rates, and other economic indicators[3]. Utilizing the Google search index to monitor flu outbreaks has demonstrated the potential of search data in public health surveillance[4]. Additionally, the search index has been used in financial market forecasting, where changes in search volume show a significant correlation with stock market fluctuations[5]. Internet search indexes can effectively capture public behavior patterns.

2.2 Related Research on Tourism Market Forecasting

In the realm of tourism market forecasting, numerous studies have explored different data sources and predictive models[6]. In recent years, with the advancement of big data technology, more studies have focused on the application of big data in tourism market forecasting[7]. For instance, using Weibo data can predict tourist flows to destinations. Analysis based on online tourism reviews can assess tourists' emotional tendencies, thereby predicting tourism demand.

2.3 Analysis of Tourist Search Behavior

Tourist search behavior, as a direct reflection of user demand, has always been a crucial aspect of tourism research[8]. Scholars have investigated the role of online searches in obtaining tourism information, finding that tourists tend to use search engines when planning trips. Seasonal factors, holidays, and unexpected events are significant factors affecting tourist search behavior[9]. Other studies have revealed differences in search behavior among various tourist groups, such as the significant differences in search frequency and keyword selection between leisure tourists and business travelers[10]. Additionally, by analyzing search engine log data, researchers have found that search behavior shows clear temporal patterns, with search volumes significantly increasing on weekends and holidays.

Despite the extensive research on the application of internet search indexes in economic forecasting, public health monitoring, and financial markets, their application in tourism market forecasting remains relatively under-explored. Existing studies have mainly focused on the use of social media data and online review data, with insufficient utilization of search indexes. Furthermore, although some research has analyzed certain characteristics of tourist search behavior, there is a lack of systematic studies combining search indexes with tourism market forecasting.

3 Research Methodology and Data Sources

3.1 Data Sources

(1) Selection of Search Engine This study selects Baidu Search Index as the data source. Baidu, being the largest search engine in China, provides a comprehensive reflection of user search behavior and interest dynamics.

(2) Acquisition of Search Index Data Search index data for keywords such as "tourism," "scenic spots," and "hotels" were obtained through the Baidu Index platform. The data spans from 2015 to 2023, sampled on a daily basis.

3.2 Research Methods

(1) Time Series Analysis. Time series analysis is used to predict tourism search trends. Specific methods include ARIMA, SARIMA, and LSTM models.

(2) Regression Analysis Regression analysis is used to identify factors influencing tourist search behavior. Variables considered include seasonal factors, holiday effects, and unexpected events.

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \ldots + \beta_n \cdot x_n + \varepsilon \quad (1)$$

In the formula: y represents the search index, x_i represents the influencing factors, β_i are the regression coefficients, and ε is the error term.

(3)Clustering Analysis Clustering analysis identifies different types of tourists and their search behavior characteristics. The mathematical formula is:

$$K = \min \sum_{i=1}^{k} \sum_{x \in C_i} \|x - u_i\|^2 \quad (2)$$

In the formula: C_i represents the i-th cluster, and U_i is the center point of the i-th cluster.

3.3 Data Processing

The data cleaning process includes handling missing values, detecting outliers, and standardizing data. Missing values are filled using interpolation methods, and outliers are detected and handled using box plot methods. Data standardization employs the Z-score normalization method, converting data to a standard normal distribution with a mean of 0 and a standard deviation of 1. Data preprocessing includes time series decomposition, extraction of trend and seasonal components, and feature engineering. Time series decomposition uses the moving average method to separate data into trend, seasonal, and random components.

4 Construction of Tourism Search Forecasting Models

4.1 Model Parameter Setting and Optimization

The setting and optimization of model parameters are critical steps to ensure the predictive accuracy of the models. For ARIMA and SARIMA models, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are used to select the optimal parameter combinations. For the LSTM model, parameter optimization mainly includes selecting the number of neurons, the number of layers, the learning rate, and other hyperparameters. This is achieved through cross-validation and grid search methods to iteratively adjust the model parameters for optimal predictive performance.

4.2 Model Parameter Setting and Optimization

After determining the model parameters, proceed with model training and validation. For ARIMA and SARIMA models, use the first 80% of the time series data for training and the remaining 20% for validation. Evaluate the predictive accuracy of the models using error metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

The training process for the LSTM model is similar but requires more computational resources, hence GPU acceleration is used. During training, early stopping and learning rate decay techniques are employed to prevent overfitting. The validation process involves using the same error metrics to evaluate the model's performance and comparing the results with those of the ARIMA and SARIMA models. By comparing the predictive results of different models, the accuracy and reliability of each model can be analyzed. This comparison provides insights into the most effective model for tourism search forecasting based on the specific characteristics of the data.

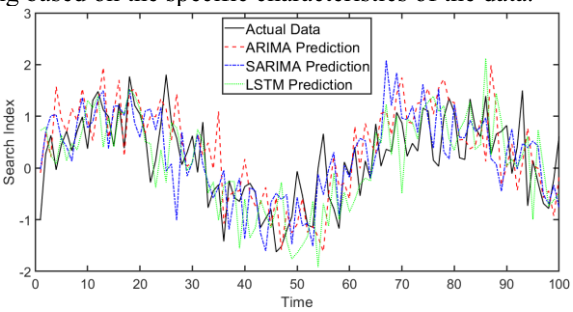

Fig. 1. Tourist Search Frequency Time Distribution.

Figure 1 presents the time series plot of tourism search index predictions by different models. The plot reveals that the LSTM model excels in capturing nonlinear changes and long-term dependencies, while the SARIMA model demonstrates superior performance in predicting seasonal data. Considering the predictive accuracy and practical

application scenarios of each model, this study ultimately selects the LSTM model as the primary forecasting tool.

5 Analysis of Tourist Search Behavior

5.1 Characteristics of Tourist Search Behavior

Tourist search behavior is a crucial pathway to understanding tourism demand. By analyzing the search frequency of tourists over different time periods, the characteristics and patterns of their search behavior can be revealed. This section utilizes Baidu Search Index data to analyze the time distribution of tourist search frequencies, exploring their characteristics and providing result analysis and discussion. Firstly, daily search frequency data for the keyword "tourism" from 2015 to 2023 is obtained. Data processing includes filling missing values and removing outliers. To facilitate analysis, the moving average method is employed to smooth the data, extracting its trend and seasonal components. Figure 2 shows the time distribution of tourist search frequencies.

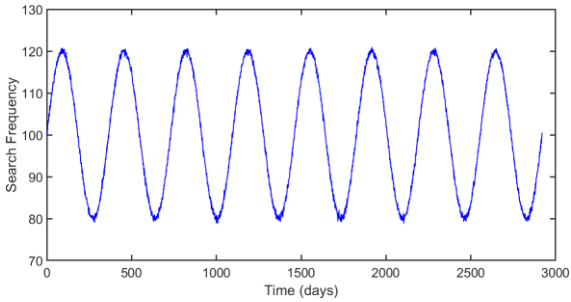

Fig. 2. Tourist Search Frequency Time Distribution.

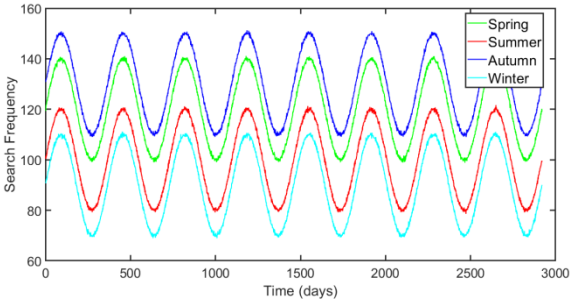

Fig. 3. Impact of Seasonal Factors on Search Behavior.

From figure 2, it can be observed that tourist search behavior exhibits clear seasonal variations and trend fluctuations. Tourist search frequency shows significant peaks and troughs at specific times of the year. Indicating a significant increase in tourism demand during these holidays. Higher search frequencies in summer and winter may be

related to travel demand during the summer and winter vacations. These seasonal variations reflect the cyclical nature of tourism demand, influenced mainly by holidays, climate changes, and school vacations. For instance, the peak in searches during the Spring Festival is closely related to the family travel tradition of this Chinese holiday, while the peaks in summer and winter reflect the concentrated travel periods for students and families.

5.2 Factors Influencing Tourist Search Behavior

Tourist search behavior is influenced by various factors, including seasonal factors, holiday effects, and unexpected events. Analyzing these factors can provide a deeper understanding of their impact and mechanisms on search behavior, offering valuable references for tourism market forecasting and management.

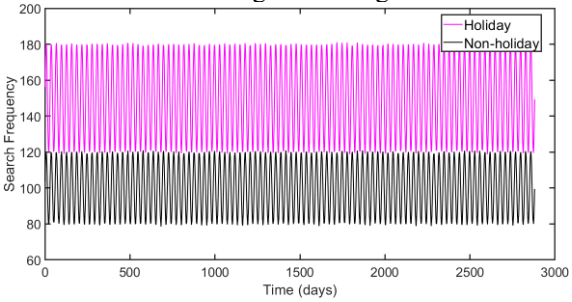

Fig. 4. Impact of Holidays on Search Behavior.

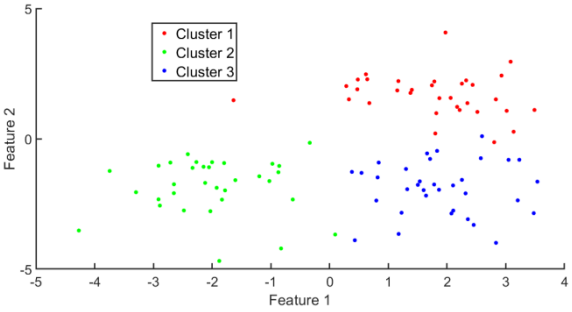

Fig. 5. Clustering Results of Tourist Search Behavior.

(1) Seasonal Factors Seasonal factors are one of the primary influences on tourist search behavior. Climate conditions, the availability of tourism resources, and festive activities in different seasons directly affect tourists' willingness to travel and their search behavior. Generally, the pleasant weather in spring and autumn makes these seasons ideal for tourism activities, leading to higher search frequencies. In contrast, travel demand in summer and winter is closely related to vacation schedules, such as summer and winter holidays. Figure 3 illustrates the impact of seasonal factors on search behavior, showing the variations in search frequency across different seasons.

Analysis of Figure 3 indicates that search frequencies in spring and autumn are significantly higher than in summer and winter, suggesting robust tourism demand during these seasons.

(2) Holiday Effects Holiday effects are another significant influence. During major holidays such as the Spring Festival and National Day, tourist search behavior increases markedly. This is because holidays provide the public with more leisure time, thereby enhancing the possibility and demand for travel. Figure 4 depicts the impact of holidays on search behavior. The analysis shows that search frequencies during holidays are significantly higher than on regular days, indicating that holidays have a substantial promoting effect on tourist search behavior.

(3) Impact of Unexpected Events Unexpected events, such as natural disasters and pandemics, also significantly affect tourist search behavior. During such events, tourism search frequency typically drops sharply, reflecting a decline in the public's willingness to travel. For example, during the outbreak of the COVID-19 pandemic, tourism search frequency decreased dramatically, indicating a sudden drop in tourism demand.

5.3 Clustering Analysis of Tourist Behavior

To better understand the search behavior characteristics of different types of tourists, this study employs clustering analysis methods to categorize tourist search behavior data into several groups and analyze the features of each group. Clustering analysis can identify tourist groups with similar search behaviors, providing a basis for market segmentation and precise marketing.

First, daily search data for the keyword "tourism" from 2015 to 2023 was extracted from the Baidu Search Index. Through data preprocessing, features such as search frequency, search keywords, and search time were extracted and organized. Principal Component Analysis (PCA) was used to reduce the dimensionality of the high-dimensional feature data. The characteristics of each group of tourists are shown in Figure 5.

Analysis of Figure 5 reveals the following: Tourists in Category 1 have a high search frequency throughout the year, with significant increases during holidays. These tourists are frequent travelers and travel enthusiasts. Tourists in Category 2 show a significant increase in search frequency during specific seasons (such as summer and winter). Their travel preferences are clearly influenced by seasonal factors, likely including family tourists or student groups. Tourists in Category 3 exhibit relatively stable search behavior, but their search frequency fluctuates significantly during unexpected events (such as pandemics). This group may include more cautious travelers and occasional travelers.

6 Discussion

This study validates the potential application of internet search indexes in tourism market forecasting. Compared to existing research, this study makes breakthroughs in

the selection and optimization of predictive models, particularly with the application of the LSTM model, which significantly improves prediction accuracy. In the analysis of tourist search behavior, this study provides more detailed characteristics and influencing factors of search behavior. The clustering analysis further reveals the diversity of tourist groups and the complexity of their search behaviors.

Despite the achievements of this study, there are some limitations. Firstly, the data source is singular, primarily relying on the Baidu Search Index, which may not fully capture the search behavior of all tourists. Future research can incorporate other data sources, such as social media data and online tourism reviews, to conduct multi-dimensional data integration analysis, enhancing the comprehensiveness and accuracy of the research. Secondly, this study uses ARIMA, SARIMA, and LSTM models for time series analysis, and while good predictive results were obtained, other models such as Prophet and GRU also show potential. Future studies can explore more models for comparison and optimization. Additionally, the clustering analysis in this study mainly relies on features like search frequency and keyword selection. Future research can introduce more behavioral features, such as search paths and click behaviors, to improve the granularity and explanatory power of clustering results. Finally, the impact analysis of unexpected events is primarily based on historical data. The adaptability and responsiveness of the predictive models to future unexpected events need further validation and optimization.

7 Conclusion

This study systematically analyzes internet search indexes, constructs tourism search forecasting models, and deeply explores the characteristics and influencing factors of tourist search behavior. The main research conclusions are as follows:

1. Through time series analysis, the effectiveness of ARIMA, SARIMA, and LSTM models in tourism search forecasting is verified. Particularly, the LSTM model significantly improves prediction accuracy by capturing nonlinear changes and long-term dependencies.
2. Search frequencies are highest in spring and autumn, with significant increases during holidays, while unexpected events (such as pandemics) lead to sharp declines in search frequency.
3. Clustering analysis identifies three types of tourists with different search behavior characteristics, including travel enthusiasts, student groups, and cautious travelers. Each type of tourist exhibits significant differences in search frequency, timing, and keyword selection.

References

1. E. Cebrián, J. Domenech. Addressing Google Trends inconsistencies[J]. Technological Forecasting and Social Change, 2024, 202: 1-7.
2. H. Y. Song, A. Y. Liu, G. Li, etal. Bayesian bootstrap aggregation for tourism demand forecasting[J]. International Journal of Tourism Research, 2021, 23(5): 914-927.
3. Q. C. Jiang. Dynamic multivariate interval forecast in tourism demand[J]. Current Issues in Tourism, 2023, 26(10): 1593-1616.
4. B. R. Zhang, Y. L. Pu, Y. Y. Wang, etal. Forecasting Hotel Accommodation Demand Based on LSTM Model Incorporating Internet Search Index[J]. Sustainability, 2019, 11(17): 357-361.
5. K. J. He, Q. Yang, D. Wu, etal. Tourist arrival forecasting using feed search information[J]. Current Issues in Tourism, 2023: 1-8.
6. B. Huang, H. Hao. A novel two-step procedure for tourism demand forecasting[J]. Current Issues in Tourism, 2021, 24(9): 1199-1210.
7. W. Y. Li, H. Z. Guan, Y. Han, etal. Short-Term Holiday Travel Demand Prediction for Urban Tour Transportation: A Combined Model Based on STC-LSTM Deep Learning Approach[J]. Ksce Journal of Civil Engineering, 2022, 26(9): 4086-4102.
8. X. D. Liang, X. Y. Li, L. L. Shu, etal. Tourism demand forecasting using graph neural network[J]. Current Issues in Tourism, 2024: 573-579.
9. Y. C. Hu, G. Wu, P. Jiang. Tourism Demand Forecasting Using Nonadditive Forecast Combinations[J]. Journal of Hospitality & Tourism Research, 2023, 47(5): 775-799.
10. H. Y. Wang, T. Hu, H. H. Wu. Tourism demand with subtle seasonality: Recognition and forecasting[J]. Tourism Economics, 2023, 29(7): 1865-1889.