Tourism Management 68 (2018) 116–126

Contents lists available at ScienceDirect

Tourism Management

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

Effective tourist volume forecasting supported by PCA and improved BPNN using Baidu index

Shaowen Li [a], Tao Chen [a], Lin Wang [b, *], Chenghan Ming [a]

[a] College of Public Administration, Huazhong University of Science and Technology, Wuhan 430074, China
[b] School of Management, Huazhong University of Science and Technology, Wuhan 430074, China

HIGHLIGHTS

- Proposed PCA-ADE-BPNN is a new combination of dimensionality reduction, optimization algorithm and neural network.
- ADE helps find suitable global initial connection weights and thresholds of BPNN.
- Principal component analysis can decorrelate the input data effectively.
- The effectiveness of PCA-ADE-BPNN is demonstrated by two empirical studies.

ARTICLE INFO

Article history:
Received 13 October 2017
Received in revised form
4 February 2018
Accepted 6 March 2018

Keywords:
Tourist volume forecasting
Principal component analysis
Baidu index
Back-propagation neural network
Adaptive differential evolution

ABSTRACT

The precise forecasting of tourist volume is a very challenging task. This paper aims to propose an effective model named PCA-ADE-BPNN for forecasting tourist volume based on Baidu index. The principal component analysis (PCA), a dimensional reduction, is employed to decorrelate the input data before training a back propagation neural network (BPNN) architecture, and the adaptive differential evolution algorithm (ADE) is for getting global optimization of BP network's weight values and threshold values to enhance the forecasting performance of BPNN. The PCA-ADE-BPNN model is a new combination of a dimensional reduction algorithm, an optimization algorithm, and a neural network. The validity of this model is demonstrated by conducting case studies of Beijing City and Hainan Province, China. The results indicate the proposed PCA-ADE-BPNN always outperforms other models in terms of forecasting accuracies. Therefore, the proposed PCA-ADE-BPNN is a potential candidate for the effective forecasting of tourist volume.

1. Introduction

Developing an accurate tourist volume forecasting method is essential because the tourism industry is perishable (Chandra & Menezes, 2004). Therefore, researchers, practitioners, and policy makers have publicly acknowledged that it's needful to accurately forecast tourism demand (Palmer, Montaño, & Sesé, 2006). Better predictions can contribute to operational, tactical, and strategic decisions made by managers and investors, such as scheduling and staffing, preparing tour brochures, and making hotel investments, respectively. It can also help government agencies design infrastructure for tour like residential site planning and transportation systems designing (Palmer et al., 2006). China's tourism industry has been expanding in recent years, and the number of visitors at famous tourist attractions has also significantly increased. Currently, many destinations cannot afford to hold a large number of tourists from all over the world. On October 2, 2013, more than 2000 visitors stranded in Jiuzhaigou, a tourist attraction in Sichuan Province, China, and the crowd began to evacuate gradually until 9 PM of Beijing time. Reportedly, many of them have detained for more than five hours according to Sohu.com (http://news.sohu.com/20131002/n387584240.shtml). On December 31, 2014, 36 people were killed while 49 were injured in a stampede at Shanghai's bund according to Sina.com (http://news.sina.com.cn/c/z/shwtctsg/). These two events have attracted worldwide attention and have highlighted the necessity of establishing a highly accurate forecasting model and making scientific and reasonable forecasts for tourist flow.

* Corresponding author.
E-mail addresses: lishaowen91@hust.edu.cn (S. Li), chentao15@163.com (T. Chen), wanglin982@gmail.com (L. Wang), mingchenghan@163.com (C. Ming).

https://doi.org/10.1016/j.tourman.2018.03.006
0261-5177/

According to (Huang, Zhang, & Ding, 2017; Wang, Wang, Qu, & Liu, 2018), the data used in traditional forecasting methods are not timely and have a limited amounts because the collection and release of data often lag behind events and the samples are often inadequate. Establishing a model with effective prediction becomes a very challenging work because of the above reasons. However, with the rapid growth of Information and Communication Technology (ICT), the Internet has become the best choice for comprehensive and timely information. The two major search engines, namely, Baidu and Google, have launched the Baidu index and the Google search index, respectively. When people search for information through Internet, the keywords that they enter in will be recorded and used to objectively reflect their concerns, needs, and preferences, so people can learn about how much attention has a keyword received in a particular period with the aid of the search engines. In recent year, search engine data has become a new data source for tourism demand forecasting, and many researches have proved the internet search data can create higher forecast accuracy due to these keywords' real-time performances (Li, Wu, Peng, & Lv, 2016; Zhang, Huang, Li, & Law, 2017). Another factor that warrants consideration is the dimensionality of the search engine data. If all indicators that influence tourist volume are selected, then the computational workload and running speed of prediction model will greatly increase and decrease, respectively. However, if we randomly select several main indicators, then the accuracy rate of forecasting model can be reduced. Therefore, a dimensional reduction algorithm must be imported into the forecasting model. Principal component analysis (PCA) has an advantage to effectively remove redundant information, and its performance and validity have been verified in many previous research (Nastac & Cristea, 2012; Oliverblaskowitz & Helmutherwartz, 2009; Sasmono, Sinisuka, Atmopawiro, & Darwanto, 2013).

Except for data, the method can greatly affect the prediction accuracy. According to (Ren et al., 2014; Sun & Wang, 2018; Wang, Zeng, & Chen, 2015; Zeng, Zeng, Choi, & Wang, 2017), Back Propagation Neural Network (BPNN) has became one of the most popular prediction techniques in many fields, especially in solving complicated nonlinear prediction problems. However, several important parameters are easy to fall into local optimal values and then must be optimized by intelligent algorithms for the training of a BPNN. In recent years, a few studies have combined search engine data with an improved BPNN to predict tourist volume, however, no study thus far has combined PCA, Adaptive differential evolution (ADE) with BPNN for tourist volume forecasting. Therefore, the objectives of this study are (a) to propose a model (PCA-ADE-BPNN) that contains the hybrid algorithm for tourist volume forecasting, (b) to apply this model to new data represented by the Baidu index, and (c) to demonstrate the validity and the stability of this model by taking Beijing City and Hainan province as examples. Moreover, vector autoregression model (VAR), PCA-VAR, and autoregressive integrated moving average model (ARIMA) are also used as comparative models in this study. As shown in Fig. 1, PCA is adopted to obtain several irrelevant principal components and ADE is for getting global optimization of BP network's weight values and threshold values. The PCA-ADE-BPNN model, which is a novel combination of dimensionality reduction algorithm, optimization algorithm and neural network, provides new insights that this study aims to explain, and serves as a new model for predicting tourist volume.

The rest of this paper is organized as follows. Section 2 introduces the methods for forecasting tourist volume, and reviews the literature related-to the search engine data and PCA-ADE-BPNN model. Section 3 briefly describes the methodological approach, while Section 4 presents an empirical analysis. Section 5 concludes the paper and outlines its future prospects.

2. Literature review

2.1. Forecasting with search engine data

Adopting highly powerful predictors is a good way to increase forecast accuracy. The users' online behavior data, obtained from search engines, provides a great potential for improving forecast accuracy. Previous studies show a correlation between large-scale search engine data and the reality of social behavior, Ginsberg et al. (2009), who first applied the relevance of such correlation to epidemiological surveillance in their study, demonstrated how flu epidemics can be predicted by using Google search queries as a data source. The application of search engine data then spread from the study of infectious diseases to economics and social behavior research, like predicting unemployment rate (Askitas & Zimmermann, 2009), cinema admissions (Hand & Judge, 2012), labor and housing markets (McLaren & Shanbhogue, 2011), consumer consumption (Vosen & Schmidt, 2011), and stock market

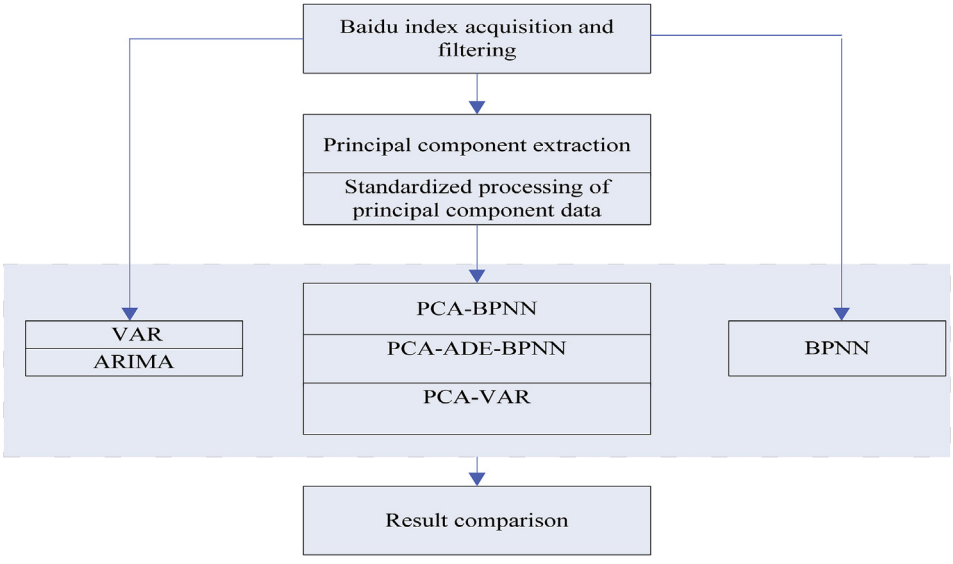

Fig. 1. Flowchart for tourist volume forecasting used in this study.