Tourism Management 59 (2017) 57–66

Contents lists available at ScienceDirect

Tourism Management

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

Forecasting tourism demand with composite search index

Xin Li [a,*], Bing Pan [b,c], Rob Law [d], Xiankai Huang [e]

[a] Institute of Tourism, Beijing Union University, Beijing, 100101, China
[b] Department of Recreation, Park, and Tourism, College of Health and Human Development, Penn State University, University Park, PA, 16802, USA
[c] School of Tourism and Environmental Sciences, Shaanxi Normal University, Xi'an, China
[d] School of Hotel and Tourism Management, The Hong Kong Polytechnic University, Kowloon, Hong Kong
[e] Beijing Open University, Beijing, 100081, China

HIGHLIGHTS

- We propose a framework to accurately forecast Chinese tourism demand.
- Search engine query data is collected to forecast tourist volumes to Beijing.
- A generalized dynamic factor model is used to create a composite search index.
- Our method improves forecast accuracy better than two benchmark models.

ARTICLE INFO

Article history:
Received 14 January 2016
Received in revised form
8 July 2016
Accepted 11 July 2016

Keywords:
Tourism demand forecast
Big data analytics
Search query data
Generalized dynamic factor model
Composite search index

ABSTRACT

Researchers have adopted online data such as search engine query volumes to forecast tourism demand for a destination, including tourist numbers and hotel occupancy. However, the massive yet highly correlated query data pose challenges when researchers attempt to include them in the forecasting model. We propose a framework and procedure for creating a composite search index adopted in a generalized dynamic factor model (GDFM). This research empirically tests the framework in predicting tourist volumes to Beijing. Findings suggest that the proposed method improves the forecast accuracy better than two benchmark models: a traditional time series model and a model with an index created by principal component analysis. The method demonstrates the validity of the combination of composite search index and a GDFM.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Advances in information technology have given rise to a massive amount of big data generated by users, including search queries, social media mentions, and mobile device locations (Mayer-Schonberger & Cukier, 2013). In particular, search query data provide valuable information about tourists' interests, opinions, and intentions. Tourists use search engines to obtain weather and traffic information, and to plan their routes by searching for hotels, attractions, travel guides, and other tourists' opinions (Fesenmaier, Xiang, Pan, & Law, 2011). Search query data, including its content and volume, can capture tourists' attention to travel destinations and can be useful in accurately forecasting tourist volumes. The abundant search trend data became a favorable source for tourism forecasting in the era of Big Data (Pan, Wu, & Song, 2012; Yang, Pan, & Song, 2014; Yang, Pan, Evans, & Lv, 2015). However, they also bring challenges in the modeling process of tourism forecasting.

In particular, in forecasting tourist volumes with search trend data, one needs to collect tourism-related keywords, obtain their search trend data, select appropriate data series to construct an aggregated index, and construct econometric models. The major challenges are keyword selection and search data aggregation. Keyword selection has received significant attention from researchers. For example, Brynjolfsson, Geva, and Reichman (2016) proposed a crowd-squared method. They prompted individuals through an online interface to produce word associations and the results verified that this method performed efficiently in the keyword selection task. In comparison, the process of index

* Corresponding author.
E-mail addresses: leexin111@163.com (X. Li), bingpan@gmail.com (B. Pan), rob.law@polyu.edu.hk (R. Law), xiankai@buu.edu.cn (X. Huang).

http://dx.doi.org/10.1016/j.tourman.2016.07.005
0261-5177/© 2016 Elsevier Ltd. All rights reserved.

aggregation has received limited attention (Brynjolfsson et al., 2016). This step is generally conducted through three main approaches: (1) incorporating keywords directly into the models; (2) extracting the index using principal component analysis (PCA); and (3) index aggregation from multiple variables (Yang et al., 2015). Although these approaches could predict more accurately than their benchmark models, they are still not optimal. First, high dimensions of variables may incur multicollinearity or overfitting problems (Varian, 2014). In particular, out-of-sample forecasts may fail even when in-sample forecasts perform well; second, a large amount of the original information will be lost if the data series is weighed equally in aggregating an index from multiple keywords. Incomplete information may reduce the forecast accuracy.

This study aims to propose a feasible variable selection method in forecasting tourist volumes with search trend data. In order to generate a universally acceptable framework, the approach should follow two rules: first, it should acquire one representative and meaningful index that reflects the dynamic correlation among all search trend data series; second, the new method should be able to deal with a large number of search data series. As a result, a generalized dynamic factor model (GDFM) is adopted to incorporate many keyword variables. An advantage of GDFM is its ability to process high-dimensional data and to use a composite index (Amstad & Potter, 2009). GDFM is commonly adopted in the analysis of economic or financial cycles, but it is seldom used in tourism forecasting. We applied our proposed methodology to predict tourist volumes in Beijing, one of the most renowned travel destinations in the world. By collecting specific search trend data from Baidu including tourism-related keywords ("dining," "lodging," "trip," "traveling," "shopping," and "recreation"), this study empirically tested the method in the forecasting of weekly Beijing tourist volumes from January 2011 to August 2015. The empirical results demonstrate that our method is superior to the benchmark models of an autoregressive model and a model with PCA as a predictor. This study contributes to existing literature in two aspects: first, it validates the performance of the aggregated index from large search trend datasets; second, empirical results demonstrate that the GDFM model is suitable for accurate tourism demand forecasting.

This paper proceeds as follows: Section II briefly reviews the relevant literature. Section III proposes a framework of integrated index construction. Section IV presents our empirical study and research findings. Finally, Section V concludes by discussing the study's contributions and implications for future research.

2. Literature review

In this section, we first review the current studies on tourism demand forecasting. Second, we focus on big data forecasting with search trend data, including the major techniques in keyword and variable selection. We also introduce the generalized dynamic factor models along with their applications. Third, we address the research gap at the end of this section.

2.1. Tourism demand forecasting: data and techniques

Tourism demand forecasting is a well-established research area, and it has been the subject of many studies in the tourism and hospitality field. Song and Li (2008) conducted a detailed literature review on tourist demand forecasting methods and techniques in recent decades. The commonly adopted forecasting techniques are time series, econometric models, artificial intelligence approaches, and hybrid methods.

Time series models predict tourist arrivals based on historical patterns. Many studies used time series models to analyze and forecast tourism demand (Akın, 2015; Athanasopoulos & Hyndman, 2008; Chu, 2008, 2009; Guizzardi & Stacchini, 2015; Gunter & Önder, 2015). The most popular models are autoregressive moving average models (Song & Li, 2008). Econometric models explore the causal relationship between tourist arrivals and influencing factors, which are especially useful when a correlational relationship exists (Song, Romilly, & Liu, 2000; Song & Witt, 2000, 2006; Song, Witt, & Jensen, 2003; Wong, Song, & Chon, 2006; Wong, Song, Witt, & Wu, 2007). Artificial intelligence methods adopt neural networks and support vector machines to model the nonlinear data series (Hadavandi, Ghanbarib, Shahanaghic, & Abbasian-Naghneh, 2011; Pai & Hong, 2005; Pai, Hong, Chang, & Chen, 2006; Palmer, Montano, & Sese, 2006; Palmer et al., 2006). Some studies have proposed a hybrid forecasting by combining econometric and data mining techniques (Pai, Huang, & Lin, 2014; Sun, Wang, Zhang, & Gao, 2016). Researchers also used methods such as meta-analysis and singular spectrum analysis in the modeling and forecasting of tourist arrivals (Hassani, Webstera, Silvaa, & Heravic, 2015; Peng, Song, & Crouch, 2014).

In terms of forecasting accuracy, different models have their own advantages and disadvantages. No single model can consistently outperform others in all situations (Song & Li, 2008). Artificial intelligence techniques can model limited observations. For example, Wang (2004) used fuzzy time series and grey models to predict tourism demand with only 12 data points. Econometric models need a large amount of observations to achieve a higher forecasting accuracy; in comparison, artificial intelligence models lack theoretical foundation for modeling tourism demand (Song & Li, 2008). Researchers are unable to illustrate the detailed influences of each explanative variable on an explained variable. By contrast, econometric models have sound theoretical underpinning, and they can validate the relationship between explained and explanative variables from an economic perspective.

2.2. Big data analytics in tourism research

Big data analytics has become increasingly important in both the academic and the business communities over the past two decades (Chen, Chiang, & Storey, 2012; Xiang, Woeber, & Fesenmaier, 2008). Travelers' decision-making is intrinsically complicated and multidimensional, such as selecting destinations, reserving hotels, planning itineraries, and other activities. The new data sources generated by users based on Internet technology (search engines or social media platforms) have become popular in studying travelers' decision-making and behavior.

Some extant literature has attempted to introduce user-generated content and big data analytics in tourism-related research. With big data sources, tourist arrivals or hotel sales can be forecasted more accurately (Blal & Sturman, 2014). Choi and Varian (2012) investigated the predictive ability of search engine data in travel destination planning. By using keyword search volume data from Google, they increased the prediction accuracy for Hong Kong tourist arrivals from several countries, such as the United States, Canada, Great Britain, and Germany. Yang et al. (2014) predicted hotel demand by combining traditional econometric models with web traffic volumes and demonstrated the use of web volumes in predicting hotel occupancy in a tourist destination.

In addition, these search engine and social media data sources can also help improve customer service, user experience, and satisfaction (Pan, Litvin, & Goldman, 2006). Ye, Law, and Gu (2009) examined the effects of online consumer-generated reviews on hotel room sales. The data were collected from the largest travel website in China. Their research findings indicated a significant relationship between online reviews and the business performance

Download English Version:

https://daneshyari.com/en/article/1011800

Download Persian Version:

https://daneshyari.com/article/1011800

Daneshyari.com