



Forecasting Chinese tourist volume with search engine data



Xin Yang^{a, b, *}, Bing Pan^{c, 1}, James A. Evans^{b, 2}, Benfu Lv^{d, 3}

^a Management School, University of Chinese Academy of Sciences, NO.80 Zhongguancun East Road Haidian District, Beijing 100190, China

^b Department of Sociology, University of Chicago, 1126 East 59th Street, Chicago, IL 60637, USA

^c Department of Hospitality and Tourism Management, School of Business, College of Charleston, 66 George Street, Charleston, SC 29424, USA

^d Management School, University of Chinese Academy of Sciences, NO.80 Zhongguancun East Road Haidian District, Beijing 100190, China

HIGHLIGHTS

- Web search data help to improve visitor volume forecasting model accuracy.
- Co-integration relationship between search data and visitor volume is verified.
- Baidu data performs better than Google for predicting tourist activities in China.
- Process to select key search queries for visitor volume prediction is proposed.

ARTICLE INFO

Article history:

Received 23 March 2014

Accepted 30 July 2014

Available online 20 August 2014

Keywords:

Search engine data

Google Trends

Baidu Index

Chinese tourism market

Visitor prediction

Tourist volume forecast

ABSTRACT

The queries entered into search engines register hundreds of millions of different searches by tourists, not only reflecting the trends of the searchers' preferences for travel products, but also offering a prediction of their future travel behavior. This study used web search query volume to predict visitor numbers for a popular tourist destination in China, and compared the predictive power of the search data of two different search engines, Google and Baidu. The study verified the co-integration relationship between search engine query data and visitor volumes to Hainan Province. Compared to the corresponding auto-regression moving average (ARMA) models, both types of search engine data helped to significantly decrease forecasting errors. However, Baidu data performed better due to its larger market share in China. The study demonstrated the value of search engine data, proposed a method for selecting predictive queries, and showed the locality of the data for forecasting tourism demand.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Tourism practitioners need accurate forecasts of tourist volume in order to effectively allocate resources and formulate pricing strategies (Song & Li, 2008). This has become especially important in China in recent years, due to the tremendous growth in tourism demand which has accompanied its economic growth. In China, the tourism industry is touted as one of the main sources of non-trade

foreign exchange earnings. In 2013, the tourism income reached 42.9 billion RMB (approximately \$7.15 billion USD), an increase of 13% compared to the previous year (China National Tourism Administration, 2014). This growth can be attributed to both the increased disposable income of Chinese citizens, as well as the government's policy of encouraging the consumption of travel products. However, the crowdedness during certain holidays has also become a major problem. For example, in one of China's National Scenic Areas, Jiuzhaigou, during one of the major holiday periods, the National Day on October 1, 2013, 4000 tourists were stuck at the entrance for five hours due to overcrowding. The administration of the scenic area had to issue a formal apology on news media (Qiu, 2013). This demonstrates the urgency for accurate forecasting, especially near-term forecasting, for anticipating and managing influxes of tourists.

Current tourist volume forecasting techniques include various statistical, econometric, and artificial intelligence methods (Song & Li, 2008). However, most of them rely on historical data to predict

* Corresponding author. Management School, University of Chinese Academy of Sciences, NO.80 Zhongguancun East Road Haidian District, Beijing 100190, China. Tel.: +86 010 8268 2865.

E-mail addresses: qwzhxyangxin@gmail.com, qwzhxyangxin@sina.com (X. Yang), bingpan@gmail.com (B. Pan), jevans@uchicago.edu (J.A. Evans), luf@ucas.ac.cn (B. Lv).

¹ Tel.: +1 843 953 2025; fax: +1 843 953 5697.

² Tel.: +1 773 834 3612; fax: +1 773 702 4849.

³ Tel.: +86 010 8268 0670.

future tourist activities, assuming a consistent pattern and stable economic structure. During certain periods of dramatic change, or at certain one-off events, however, these methods may not provide accurate predictions. In addition, these forecasting techniques mainly focus on a long-term scale, such as annually or quarterly, instead of monthly or weekly, which limits their application in short-term forecasting.

The development of information technology, especially the Internet, has generated another type of data for forecasting. Every time a tourist interacts with the Internet, be it through a search engine, a website, a mobile phone, or a social media platform, the traces of the interaction can be captured, stored, and analyzed. As a result, a new area of forecasting with online data has blossomed. Researchers have used online data such as search engine query volumes, amount and types of tweets, website traffic, and social media posts to forecast contagious disease outbreaks (Carneiro & Mylonakis, 2009), consumer consumption (Vosen & Schmidt, 2011), popularity of songs and movies (Goel, Hofman, Lahaie, Pennock, & Watts, 2010), unemployment rates (Askatas & Zimmermann, 2009), and hotel room demand (Pan, Wu, & Song, 2012; Yang, Pan, & Song, 2014). Online data are useful as predictors not only in developed countries, but also in developing countries where the Internet adoption rate is low (Carrière-Swallow & Labbé, 2011).

Travelers use search engines to find relevant information for all aspects of a trip, including accommodations, attractions, activities, and dining (Pan, Litvin, & Goldman, 2006). In this study, we used the search query volume data provided by both Google and Baidu, two search engines used in China, to predict tourist volumes to a specific destination. According to China Internet Network Information Center (CNNIC), the number of Chinese citizens using the Internet has reached 591 million, and 80% of them queried search engines in 2013. Travel planning is one major search activity (China Internet Network Information Center, 2013). The main questions we wanted to address were: are search engine data a valid predictor for tourist volumes in China? Which search engine data are more powerful predictor, those from Google or Baidu? How can a researcher select the candidate queries for forecasting tourist volumes among possibly hundreds or thousands of queries?

In this study, we proposed a conceptual model on the role of search engines in the travel planning and travel process of tourists. We then compared the forecasting power of models with actual Google and Baidu data, with their equivalent time series counterparts. In this process, we also proposed a systematic way to obtain relevant search engine queries from search engine query volume tools. Thus, this work contributes to two areas of the literature. First, previous researchers only focused on one type of search query data, and did not discuss a search engine's locality when examining the relationship between search data and visitor numbers. Our study adds to the literature by comparing the goodness of fit and forecasting power of data from two types of search engines, Google and Baidu. Second, relative to the unclear search query selection process in previous studies, we proposed a systemic mechanism to better pick search queries for predicting visitor volumes to aid in improving the reproduction of our study.

2. Literature review

This section reviews relevant studies in the areas of tourism demand and socioeconomic activity forecasting with search engine query data. Specifically, a few recent studies on forecasting tourism demand with search engine data are also discussed and research gaps are identified.

2.1. Forecasting tourism demand

Researchers have adopted two main types of methods for forecasting tourism demand and tourist volumes. The first is based on time series or statistical techniques, such as linear regression, exponential smoothing, and autoregressive models (Song & Witt, 2000). The other type consists of artificial intelligence methods, such as artificial neural network, grey theory, rough set theory, fuzzy theory, genetic algorithm, Monte Carlo simulation, and expert systems (Abratt, Nel, & Nezer, 1995; Andrew, Cranage, & Lee, 1990; Law & Au, 1999; Weatherford & Kimes, 2003). However, recent studies have also demonstrated that no single method outperforms others in forecasting accuracy, and a combination of methods can produce better forecasting results (Chan, Witt, Lee, & Song, 2010; Song & Li, 2008).

The traditional time series model and many of its derivatives are well-established and widely adopted in forecasting tourism demand, and are superior to other methods (Song, Witt, & Li, 2008). In the Asian-Pacific area, Lim and McAleer (2002) compared several exponential smoothing models in order to estimate quarterly tourist volumes from Hong Kong, Malaysia and Singapore to Australia during the period of 1975–1999. The results showed that the Hoyt-Winter additive model and multi-variant seasonal model performed better than the secondary Hoyt-Winter non-seasonal exponential smoothing model. Liu (2008) demonstrated that the seasonal product model was more accurate than the autoregressive and exponential smoothing models in predicting the visitors to Guilin, China. Feng (2008) built an ARMA model that predicted visitors to Yue Temple in China from 1980 to 2007 and achieved a better result than other models. Feng (2008) used Hong Kong travel demands as an example and compared three typical visitor forecast techniques: exponential smoothing, univariate ARIMA, and Artificial Neural Networks. He concluded that Artificial Neural Networks performed better than the other two time series prediction methods. Chaitip and Chaiboonsri (2009) explored two different statistical models, one named X-12-ARIMA seasonal adjustment, and an autoregressive fractionally integrated moving average (ARFIMA), in the application of travel amount prediction and conducted an empirical study on India from 2007 to 2010. They argued that in order to achieve the most optimal forecast performance, one needs to diligently test parameters when using different statistical models.

The development of computer technology has facilitated the adoption of artificial intelligence methods in predicting tourism demand. The neural network method is especially useful when the relationship between predictors and predicted variables is nonlinear. For example, Chen, Chen, Xing, and Fu (2005) built an effective back propagate neural network model of tourism demand to Yunnan Province. Rough set, grey theory, fuzzy theory and other artificial intelligence methods have also been adopted. For instance, Goh and Law (2003) used rough set theory to predict visitor volumes from 10 source markets to Hong Kong with an accuracy of 87.2%. Weng, Zhen, Liu, and Zhang (2008) used the GM-Markov model to analyze and predict the amount of inbound tourists to China and achieved superior accuracy. Kan, Lee, and Chen (2010) adopted six grey models, each with different parameter settings, to predict Taiwan travel demands from 2009 to 2013. The simulation results showed that the highest growth rate would reach about 6% while the lowest growth rate would be around 5%. Alvarez-Diaz, Mateu-Sbert, and Rossello-Nadal (2009) employed a Genetic Program (GP) to forecast monthly visitors from UK and Germany to the Balearic Islands of Spain. Their empirical results showed the GP was more robust, easy-to-use, and allowed for simple ad hoc interpretation compared to other non-parametric methods.

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات