Technological Forecasting & Social Change xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

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

Big Data analytics for forecasting tourism destination arrivals with the applied Vector Autoregression model

Yuan-Yuan Liu[a,b], Fang-Mei Tseng[c,*], Yi-Heng Tseng[c]

[a] School of Management, Guizhou University (The Northern Campus), Huaxi District, 550025 Guiyang City, Guizhou, China
[b] Faculty of Economics and Business Administration, Vilnius University, Sauletekio al. 9, II Building, 10222 Vilnius, Lithuania
[c] College of Management, Yuan Ze University, Taiwan

A R T I C L E I N F O

Keywords:
Big Data analytics
Vector Autoregression model
Granger causality
Destination Management and Marketing

A B S T R A C T

The prediction of tourist numbers is important for Destination Management and Marketing. While most existing methods rely on well-structured statistical data, using web search queries of the destination to forecast its tourist arrivals is a new way to apply Big Data analytics. However, there are no studies exploring correlation of weather, temperatures, weekends and public holidays with tourism destination arrivals and web search queries of the destination, respectively. This study uses the Vector Autoregressive modeling to examine the Granger causality between actual arrivals of the studied cultural tourism destination and its web search queries, and to explore the correlation mentioned above. The striking result is that weather has no correlation either with actual arrivals of the studied cultural tourism destination, or with its web search queries. Meanwhile, unlike previous researchers who discuss the predictive power of web queries on actual tourism flows, this study emphasizes their reciprocal predictive powers upon each other. The originality of this study is exemplifying the utilization of Big Data analytics in the tourism domain with Big Data datasets, data capture techniques, analytical tools, and analysis results. This study further digs possible reasons for an identified short time lag length ($p = 2$), to provide insights for Destination Management and Marketing.

1. Introduction

We are in the era of Big Data (Miller, 2008). Big Data is a term that primarily describes datasets that are so large, unstructured, and complex that require advanced and unique technologies to store, manage, analyze, and visualize (Chen et al., 2012). Since the advent of the World Wide Web, major parts of tourism information processing and transactions are handled electronically (Buhalis, 2006), so numerous travel-related electronic traces are left. These electronic traces include a large variety and a huge number of tourism information, such as, pre-trip planning and information searching, reservation and booking, post-trip experiences sharing and recommendation, as well as photo uploading and other social media interacting activities, and these large, unstructured, and complex electronic traces become tourism Big Data which need to be captured and analyzed so as to uncover hidden patterns, correlation and other insights in the tourism domain.

Recent researches on Big Data in tourism have sprung up. Utilizing Big Data analysis, Liu et al. (2017) reveal determinants of hotel customer satisfaction in language groups; Xiang et al. (2015) prove the inherent relationship between hotel guest experience and satisfaction, and Zhang et al. (2017) show the development of tourism as a discipline in China by using meta-analysis. Big Data analysis has also been employed to forecast tourist flows. Gunter and Önder (2016) test the ability of 10 traffic indicators of Google Analytics of the Viennese Destination Management Organization (DMO) to predict the tourist arrivals to Vienna by applying big data shrinking methods with VAR modeling. Li et al. (2017) adopt data from search engine query to forecast tourism demand for Beijing. Meanwhile, Big Data analytics is found in works related to Destination Management and Marketing (DMM). Fuchs et al. (2014) illustrate how Big Data analytics is applied at a Swedish mountain tourism destination for its knowledge infrastructure to create and apply knowledge for the destination as a learning organization. Marine-Roig and Anton Clavé (2015) highlight the application of Big Data analytics in supporting Barcelona as a smart destination by examining its online image on thousands of travel blogs and online travel reviews (OTRs). Miah et al. (2017) deploy Big Data analytics for tourist behavior study. Besides, Big Data techniques in tourism research are developed, such as text mining in tourism (Godnov and Redek, 2016), opinion mining in tourism (Bucur, 2015; Marrese-Taylor et al., 2014), cross media and multi-agent Big Data collection for

* Corresponding author.
E-mail address: fmtseng@saturn.yzu.edu.tw (F.-M. Tseng).

https://doi.org/10.1016/j.techfore.2018.01.018
Received 27 February 2017; Received in revised form 21 December 2017; Accepted 14 January 2018
0040-1625/

tourism perception research (Guan and Du, 2015), and hardware for tourism organizations to download huge amounts of data (d'Amore et al., 2015).

In the area of tourism research, weather and temperatures are crucial topics, which have led to the widespread attentions and interests of researchers. Travel motivations, satisfactions, tourism demands, and destination images are the most concerned topics in relation to destination weather or temperatures. Weather has been identified as both a motivator and a disincentive factor for travels in the theory of 'push and pull' (Crompton, 1979). Some researchers also find that climate and weather conditions may motivate holiday travels to holiday destinations (Becken, 2013; De Freitas, 2003; Gómez Martín, 2005; Falk, 2014; Gössling et al., 2012). Weather may impact profitability of the tourism industry in a way of reducing or increasing customer satisfaction and loyalty for destinations (Becken and Hay, 2012). Bad weather experiences more negatively impact tourist satisfaction than good weather experiences positively do so (Coghlan and Prideaux, 2009). Weather patterns and tourism demands are widely discussed (Becken, 2013; Falk, 2014; Gössling et al., 2012; Kaján and Saarinen, 2013). Weather is an 'intangible asset' of destinations as it is one part of destination images (Baloglu and Mangaloglu, 2001; Echtner and Ritchie, 1991; Gallarza et al., 2002; Pike, 2010; Tasci and Gartner, 2007). Weather and temperatures are components of tourism demands and destination image (Day et al., 2013). In addition, there is also a research about impacts of weekends and weekdays on tourism motivations, for example, differences in motivations from weekends and weekdays impacting on Japanese spa tourism (Kamata and Misui, 2015). However, there are few research studies on the impacts of weather, temperatures, and weekends together with public holidays in a single research to explore their impacts on tourism destination arrivals, or on web queries about the destination.

To explore the complex relationships among these variables, we introduce Xijiang Thousand Households Miao Village (the Miao Village, in short) as the studied destination. It is a leading Miao cultural tourism destination in Guizhou Province, China. Moreover, it is the biggest and most complete original Miao Village in China, probably even in the world. It presents tourists with the original Miao lifestyle in a Miao Village locating on mountains, surrounded by rivers and streams. It is a cultural destination with abundant natural resources and about 95% of its tourists are domestic. Thus, the research questions are the following:

1: Can web search queries of the Miao Village be able to predict the actual tourism arrivals to the village, and vice versa?
2: What is the time lag length if prediction in (1) is possible?
3: How will weather, temperatures, weekends and public holidays impact actual tourism arrivals to the village and the web search queries of the village?

The main contribution of the study lies in its uniqueness of involving weather, temperatures, weekends and public holidays together with tourism destination arrivals and web search queries of the destination in one model to explore their complex relationships. Furthermore, being different from previous tourism flows forecasting papers, this paper emphasizes the reciprocal predictive power between tourism destination arrivals and web search queries of the destination. Besides, adopting Big Data analytics for data and analysis needed exemplifies its application in tourism arrivals forecasting and Destination Management and Marketing (DMM).

The remainder of the paper is structured as follows. Section 2 overviews the recent literature relevant to tourism forecasting modeling, providing the rationale for using the chosen methodology. Section 3 presents the data preparation and cleansing procedures with a conceptual framework of the destination's Big Data dataset while Section 4 contains a discussion of the VAR (*p*) modeling methodology. Section 5 presents the results, and Section 6 concludes the study with implications and a discussion.

2. Literature review

2.1. The impacts of weather and temperatures on tourism

Holiday travels are motivated by the climate and weather conditions at holiday destinations (Becken, 2013; De Freitas, 2003; Gómez Martín, 2005; Gössling et al., 2012). Crompton (1979) present a 'push-and-pull' model where push factors are these factors motivating tourists traveling away from home while pull factors are those driving them toward destinations. Falk (2014) find that warm weather is a pull factor for tourists to select a destination, and in addition, temperatures, duration of sunshine, as well as precipitation are important factors in summer seasons. On the contrary, Thapa (2012) find that in long-term tourist expectations of poor weather at a destination may restrain their visitations. However, the impact of weather condition considerably differs according to different destinations and types of touristic activities (Lohmann and Kaim, 1999). Smith (1993) finds that cultural tours and urban breaks are not that dependent on weather conditions. Instead, holidays are reliant on natural resources and outdoor activities. Besides, tourist perceptions of weather condition attractiveness vary according to different destinations. Jeuring and Peters (2013) conclude that for nature-based tourism the mist in the mountains may limit possible visitations while the clouds surrounded valley is more impressive than that with the bright sunny weather condition.

Regression models have been used to estimate the impact of weather on tourism. The results of the first-difference regression models of Falk (2014) show that average sunshine duration and temperatures have significant positive impacts on domestic overnight stays while average precipitation has significant negative effects. In his research, Falk (2014) approve that the positive impact of temperatures on tourism is limited. The relationship between temperature and the number of visitations is non-linear in the form of an inverted u-shape curve (Gössling and Hall, 2006; Rossello-Nadal et al., 2011), which indicates a decline in temperature's effect after a given point (Falk, 2014). Besides, there is a 1-year time lag before temperatures and sunshine duration positively impact foreign overnight stays (Falk, 2014). However, Scott et al. (2007) reveal the impact of monthly temperatures and precipitation on tourist flows to Waterton Lakes National Park of Canada by using monthly data from 1996 to 2003 and remarkably find no relationship between visitations in peak summer (July and August) and weather. Agnew and Palutikof (2006) present a relationship between domestic tourism demand and weather conditions in the UK, and a relationship between outbound tourist flows and weather conditions. They find that in the UK, domestic tourism is more responsive to variations of weather conditions, and in general, the tourism industry of the UK benefits from warm and dry conditions. The time lag in their research is 1-year (or -season) before international tourism being affected by weather conditions. Taylor and Ortiz (2009) explain time lag variations that domestic residents are more spontaneous, whereas international tourists need more time to plan their visitations well in advance.

2.2. Forecasting tourism volumes and Vector Autoregressive (VAR) modeling in the forecasts

Different methodologies can be applied to forecast tourism volumes, but their performances vary in terms of accuracy. Athanasopoulos et al. (2011) evaluate the performance of various methods for forecasting tourism flows. In their research, monthly series, quarterly series, and annual series are included. Uni- and multivariate time series approaches, and econometric models are implemented. Algorithms such as Forecast Pro, ARIMA, exponential smoothing is included. Specific methods, such as the Theta method and damped trend, are employed. Frameworks, such as static and dynamic regression, autoregressive distributed lag models, and time varying parameter models, are incorporated. They find that pure time-series approaches are more