Forecasting London Museum Visitors Using Google Trends Data

Ekaterina Volchek[a], Haiyan Song[a],
Rob Law[a] and Dimitrios Buhalis[b]

[a] The Hong Kong Polytechnic University/ School of Hotel and Tourism Management, Hong Kong
katerina.volchek@connect.polyu.hk, haiyan.song@polyu.edu.hk, rob.law@polyu.edu.hk

[b]Bournemouth University/ Department of Tourism and Hospitality, UK dbuhalis@bournemouth.ac.uk

Abstract

Information search is an indicator of tourist interest in a specific service and potential purchase decision. User online search patterns are a well-known tool for forecasting pre-trip consumer behaviour, such as hotel demand and international tourist arrivals. However, the potential of search engine data for estimating the demand for tourist attractions, which is created both before and during a trip, remains underexplored. This research note investigates the relationships between Google search queries for the most popular London museums and actual visits to these attractions. Preliminary findings indicate high correlation between monthly series data. Search query data is expected to generate reliable forecasts of visits to London museums.

Keywords: Attractions, museum, forecast, search engine, information search, Google Trends.

1 Introduction

Forecasting has been the focus of research for decades because it allows estimation of the main tourism indicators, including number of visitors, and corresponding adjustment of business strategies. Traditional methods are based on statistical, econometric, and artificial intelligence tools (Song, Witt, & Li, 2008) and historical data (Pan, Wu, & Song, 2012). The availability of real-time, high-volume, and high-frequency data has revolutionised the way in which tourist behaviour is monitored and forecasting reliability is achieved (Yang, Pan, & Song, 2014). Information search patterns have been proven to be indicators of future behaviour (Park, Lee, & Song, 2017). Data from search engines are widely applied in the tourism domain to forecast destination-level indicators, such as hotel room demand (e.g. Pan et al., 2012; Yang et al., 2014) and international tourist and visitor arrivals (e.g. Li, Pan, Law, & Huang, 2017; Önder, 2017; Park et al., 2017). Individual attractions traditionally receive less attention from academicians. Although a trend of shortening time lag has been observed between the travel planning stage and the actual trip, decisions regarding a destination and a hotel are frequently made during pre-trip phase. In comparison to them, the complexity of attraction visits forecast increases because the choice for a particular point of interest can be made immediately before and during a trip (Xiang et al., 2015). Tourist decision-making is influenced by a wide range of internal and external factors; hence, the importance of “nowcasting” as a capability to quickly and accurately identify current changes in consumer behaviour (Choi & Varian, 2012) increases. However, the capability of information search patterns data in predicting attraction demand remains underexplored (Huang, Zhang, & Ding, 2017). The current study aims to investigate the opportunity to forecast the number of museum visits based on tourist online search behaviour.

2 Literature Review

A decision-making process in tourism comprises three consequent stages: information search, travel planning, and trip arrangements. The main objectives of information search activities are to increase the overall quality of a future trip and to decrease a perceived high risk, associated with future travel (Jacobsen & Munar, 2012). Tourist attitude towards a service is determined by information, thereby triggering purchase decision (Yang et al., 2014). Therefore, data on information search behaviour are an important research tool that can identify tourist needs, motives, and intentions. Proliferation of ICTs and mobile computing devices have made the Internet the main source of travel information (Park et al., 2017). Search engines, such as Google and Baidu, are the primary sources of tourism information, which shape tourist perceptions on attraction image (Marine-Roig, 2017), and facilitate travel decision (Önder, 2017). When driven by different needs for information, users may switch back and forth between general sources and specialised domains, such as travel portals, online travel agencies, online reservation systems, and social media and news websites (Xiang et al., 2015). Although search engines are perceived to be less useful sources of travel information compared to destination and accommodation websites (Fesenmaier, Xiang, Pan, & Law, 2011), they remain initial tools for attraction search (Xiang & Pan, 2011). Tourists intensively use search engines to support a wide range of activities, such as choosing destinations and hotels, and planning routes to take and attractions to visit. The search for attractions can be motivated by a general idea to plan a visit and by specific needs (Fesenmaier et al., 2011). The search strategy and objectives of a user determine the application of either a general or specific set of keywords. Most search queries (SQs) consist of two to three keywords (Xiang & Pan, 2011) and frequently include the name of the city as the geographical entity. The usage of the exact name of a museum as a SQ indicates tourist awareness of this attraction and implies potential activities to arrange the visit (Yang et al., 2014). Accordingly, applied keywords are a valuable source of data about tourist interests and intentions (Li et al., 2017). Search engines accumulate high volume of individual-level data, which may contain applied keywords, user location, time of search, and other details. Google Trends is a free online tool that provides data on users SQs that is aggregated on monthly or weekly basis (Pan et al., 2012). The index reflects popularity of a SQ at a given moment (i.e., month or week, and for the recent search history, days and hours) (Google, 2017). The data represent the scaled and normalized sample-based index, which is calculated based on the total search performed within a certain geographic area and period since 2004 (Artola et al., 2015). Therefore, the index varies daily to a few percent (Choi & Varian, 2012). The advantages of the data provided by Google Trends are its high volume, high frequency, and immediate sensitivity to changes in user behaviour. Compared to historical data, Google Trends index gives more accurate and less expensive forecasts available for businesses (Pan et al., 2012). Traditional methods mainly focus on the long-term perspective; by contrast, Google Trends can provide real-time data, which enable “nowcasting” (Choi & Varian, 2012). Google Trends data create limitation for research since they are based on the sample mode rather than on the absolute volume of search data (Yang, Pan, Evans, & Lv, 2015), along with the approximation methods, which are not revealed by the publisher, thereby potentially leading to poor representation of the explored population. Consequently, Google Trends data cannot be used to fully substitute traditional analysis. Together, however, they create a powerful tool for policy makers (Artola et al., 2015).

3 Methodology

This research investigated the relationships between the SQs for London museums (worldwide) (Google, 2017) and the total visits to these museums (Delaney, 2017). It explored the cases of 10 London museums, which are among the 40 most visited attractions in the UK. The search engine with a higher share provides better prediction of tourist behaviour (Yang et al., 2015). The share of Google search engine in the UK and the UK top international tourism markets exceeds 80% (Statista, 2017), thereby allowing Google to be considered the only source of SQ data. Application of search engine data may lead to significantly overestimated forecasts (Artola et al., 2015), thereby, leading to the challenge of the keywords choice (Park et al., 2017). The relevant keywords should reflect individual word associations between a query and an explored phenomenon (Li et al., 2017), and would exclude single words or word combinations, which may be used to describe another object or to satisfy another information need. To eliminate the items with the same and similar names from the SQs, this study applied one combination of words, recognised by Google as the query for the specific attraction of London, for each museum. To refine that SQs are submitted by tourists, the research considered the keywords that are specifically classified by Google to be under the travel category (Pan et al., 2012). The models used in the forecasting exercise is the autoregressive integrated moving average (ARIMA) and ARIMA with explanatory variable (ARIMAX) models. In the following, y_t denotes the log-transformed visitor data. The monthly data from January 2012 to September 2016 are used to train the models, whereas the data from the last 12 months, i.e., from October 2016 to September 2017, will be used in forecasting. ARIMA models are frequently estimated by following the Box–Jenkins approach. An automatic order selection process for ARIMA is proposed in Hyndman and Khandakar (2008). First, the order of seasonal differencing D is selected based on the OCSB test (Osborn, Chui, Smith, & Birchenhall, 1988). Second, the order of non-seasonal differencing d is chosen using the KPSS unit root test (Kwiatkowski, Phillips, Schmidt, & Shin, 1992). Third, a stepwise procedure is used to traverse the model space to choose the order of autoregressive and moving average terms, p, q, P, and Q. After this step, the model with the lowest Akaike information criterion is selected. ARIMAX is constructed in a similar way to ARIMA, with the exception that a linear regression of y_t on lagged SQ data is first conducted, then the automatic order selection process is used for the errors from the linear regression.

4 Preliminary Findings

The preliminary data and correlation analysis results indicate that consistent relationships exist between the SQ series and the actual visits to the London museums under investigation (Table 1). The correlation coefficient is relatively high for the primary and secondary attractions, which are popular among international and domestic visitors. Predominance of both short-haul and long-haul tourists at the first two iconic attractions (London & Partners, 2015), and the fact that iconic attractions drive tourists to destinations (Leiper, 1990), triggering information search on pre-trip search, cause aggregation of heterogeneous information search patterns in these sample. Therefore, a significant but lower correlation is confirmed for the British Museum and the National Gallery, where the number of international tourists exceeds domestic visits (London & Partners, 2015). Nevertheless, the occurrence of consistent relationships without a time lag between information search and visit confirms that tourists search for in-destination activities and attractions immediately before or during a trip (Xiang et al., 2015), thereby allowing the use of online search data for museums to estimate their immediate demand.

Table 1. Correlation between museum SQs and actual visits, 2012-2017

Attraction	***R***	***Attraction***	***R***
British Museum	0.483***	Victoria and Albert Museum	0.521***
National Gallery	0.358**	National Portrait Gallery	0.668***
Tate Modern	0.708***	Tate Britain	0.651***
Natural History Museum	0.858***	Imperial War Museum	0.679***
Science Museum (Group)	0.732***	Horniman Museum	0.697***

$^{***}p<0.01$, $^{**}p<0.05$, $^{*}p<0.10$

5 Conclusions

This study is still in progress. The preliminary findings provide statistical evidence that tourists use search engines to plan primary and secondary attraction visits close to or during the trip. The expected outcome of the study is that SQ data will generate reliable forecasts of visits to London museums. The forecasting performance of the ARIMA model will be significantly improved by incorporating SQ data. The major limitation results from the aggregated nature of the Google search index and museum visits, which restricts usage immediate behaviour patterns into the model. The study investigated only the case of London museums; therefore, the findings cannot be generalised to other destinations. Further research based on different types of attractions and additional destinations is required to test the generalisability.

References

Artola, C., Pinto, F., & Pedraza, P. D. (2015). Can internet searches forecast tourism inflows? *International Journal of Manpower*, 36(1): 103-116.

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record, 88*(s1), 2-9.

Delaney, L. (2017). *Monthly Museums and Galleries visits*. Department for Digital, Culture, Media & Sport, UK.

Fesenmaier, D. R., Xiang, Z., Pan, B., & Law, R. (2011). A framework of search engine use for travel planning. *Journal of Travel Research, 50*(6), 587-601.

Google. (2017). Google Trends. Retrieved from https://trends.google.com/

Huang, X., Zhang, L., & Ding, Y. (2017). The Baidu Index: Uses in predicting tourism flows – A case study of the Forbidden City. *Tourism Management, 58*, 301-306.

Jacobsen, J. K. S., & Munar, A. M. (2012). Tourist information search and destination choice in a digital age. *Tourism Management Perspectives, 1*(1), 39-47.

Leiper, N. (1990). Tourist attraction systems. *Annals of Tourism Research, 17*(3), 367-384.

Li, X., Pan, B., Law, R., & Huang, X. (2017). Forecasting tourism demand with composite search index. *Tourism Management, 59*, 57-66.

London & Partners (2015). *London. Tourism Report 2014-2015*.

Marine-Roig, E. (2017). Measuring Destination Image through Travel Reviews in Search Engines. *Sustainability, 9*(8), 1425.

Önder, I. (2017). Forecasting tourism demand with Google trends: Accuracy comparison of countries versus cities. *International Journal of Tourism Research*, doi:10.1002/jtr.2137

Pan, Wu, D. C., & Song, H. (2012). Forecasting hotel room demand using search engine data. *Journal of Hospitality and Tourism Technology, 3*(3), 196-210.

Park, S., Lee, J., & Song, W. (2017). Short-term forecasting of Japanese tourist inflow to South Korea using Google trends data. *Journal of Travel and Tourism Marketing, 34*(3), 357-368.

Song, H., Witt, S. F., & Li, G. (2008). *The advanced econometrics of tourism demand*. NY: Routledge.

Statista. (2017). Market share held by the leading search engines in the United Kingdom (UK) as of July 2017. Retrieved from https://www.statista.com/statistics/280269/market-share-held-by-search-engines-in-the-united-kingdom/

Xiang, Z., & Pan, B. (2011). Travel queries on cities in the United States: Implications for search engine marketing for tourist destinations. *Tourism Management, 32*(1), 88-97.

Xiang, Z., Wang, D., O'Leary, J. T., & Fesenmaier, D. R. (2015). Adapting to the Internet: Trends in Travelers' Use of the Web for Trip Planning. *Journal of Travel Research, 54*(4), 511-527.

Yang, X., Pan, B., Evans, J. A., & Lv, B. (2015). Forecasting Chinese tourist volume with search engine data. *Tourism Management, 46*, 386-397.

Yang, Y., Pan, B., & Song, H. (2014). Predicting hotel demand using destination marketing organization's web traffic data. *Journal of Travel Research, 53*(4), 433-447.