

Relationship between Tourist Attractions and the Baidu Index: A Case Study of National 5A Scenic Spots of Xi'an City During National Day Holiday

Kewei Lei^{a*}, Xiaohui Wang^b and Xiaoning Dou^c

College of Leisure Management, Xi'an Eurasia University, Xi'an, China

^aleikewei@eurasia.edu, ^bwangxiaohui @eurasia.edu, ^cdouxiaoning@eurasia.edu

*The corresponding author

Keywords: Baidu index; Tourist attraction; Xi'an tourism; National Day holiday; 5A scenic spot

Abstract. During Chinese National Day holiday, lots of tourists visit well-known scenic spots leading to heavy overcrowding problems. Consequently, the accurately prediction of tourist numbers for certain small regions/scenic spots within few days is important for tourist attractions management and planning. Most existing methods rely on well-structured statistical data published by the government. However, it usually exists the significant delays, and the statistical data is gathered for long intervals and large regions. These drawbacks lead to an inaccurate prediction of tourist numbers for certain small regions/scenic spots within short periods. In this paper, the popular Baidu Index is employed to predict the tourist flow in national 5A scenic spots of Xi'an city during National Day holidays. A large scale Baidu Index data, from 2012 to 2016 years, is used for overall Xi'an tourism and three national 5A scenic spots, i.e., Terra Cotta Warriors, Huaqing Hot Spring and Dayan Pagoda. The results indicate that the Baidu search index of keyword 'weather' has a weak correlation with tourist volume of Xi'an tourism. While the keywords 'ticket' and 'tourism' has a positive correlation between the increasing Baidu search index and the increasing observed tourist flow.

Introduction

With the rapid development of China's tourism industry, the tourist numbers increase dramatically at some world famous tourism cities, for instance Xi'an city, in May Day and National Day holidays. Accurately prediction of tourist volume has gained more attentions recently in tourism management and planning in order to effectively allocate tourism resources and avoid overflow of carrying capacities. Current prediction methods of tourist volume involves various statistical, econometric, and artificial intelligence techniques [1-2]. However, most of them rely on the well-structured governmental statistical data. It usually takes long time to collect and publish such statistical data, which results in a serious delaying effect in predicting tourist volume. In addition, the prediction model is established on the assumption of consistent pattern and stable economic structure. If there was a dramatic change or unpredictable event, the model may not provide enough accuracy. Especially, these prediction models mainly focus on a long-term scale and large region, such as annually or quarterly reports in certain province or city, instead of monthly or weekly data for certain scenic spot. [3] This aforementioned drawback limits their application in short-term and small region prediction. Consequently, the development of prediction method of tourist numbers for small region and short-term period is important, for instance the tourist number in a scenic spot during several days.

Nowadays, the information technology, for instance Internet, has made a rapid progress. A tourist usually interacts with the Internet through a search engine by a smart mobile phone or computer in order to get some relevant information on destination. The traces of this interaction can be captured and stored, which leads to the so-called big data. Past researches showed that correlations exist between search data and the reality of social behaviour [4]. As a result, the prediction of tourist volume based on online search data has blossomed recently. Tourists use

search engines to find relevant information for all aspects of a trip, for instance ticket price, tourism strategy, introduction of scenic spots, hotel, and so on. It should be noted that tourists visit the search engines in advanced few days. Meanwhile, the destination is comparable small region, say a city or certain scenic spot. Therefore it is possible to overcome the disadvantage of the existing prediction methods for small regions and short-term period [5].

Xi'an is the capital of Shaanxi province in China, one of the birthplaces of the ancient civilization in the Yellow River Basin area. During her 3100 years development, 13 dynasties placed their capitals here. So far, Xi'an enjoys equal fame with Athens, Cairo, and Rome as one of the four major ancient civilization capitals. Xi'an has a semi-moist monsoon climate and there is a clear distinction between the four seasons. Spring and Autumn seasons are quite suitable for travelling. Until now, Xi'an has three national 5A scenic spots, i.e., Terra Cotta Warriors, Huaqing Hot Spring and Dayan Pagoda. Every year, many domestic and foreign tourists visit these world famous scenic spots. Therefore, it is important to accurately predict the tourist numbers of these scenic spots to avoid damage and improve sustainable development.

In this paper, we used the search query volume data provided by Baidu to analyse the trend of Baidu index during National Day holidays for Xi'an tourism and three national 5A scenic spots. The aim of present work is not trying to focus on the mathematical prediction models while show the evolution of Baidu Index for certain small region and scenic spots within few days. In order to get an overall trend of Baidu index, the search data from September 1st to October 31th was presented although we are focusing on the data in National Day holiday. Moreover, five years historical data (2012-2016) was shown in order to get a statistical result from large samples. The keywords selection was also mentioned in this work.

Baidu Keywords and Data Selection

Baidu is a famous search engine in China, like international tool google. Since 2006, the Baidu Company provided Baidu Index function based on big data mining. The Baidu Index is a free big data analysis service in view of Baidu web searching and news. Someone can get useful information from Baidu Index, which reflects 'user awareness' and 'media attention' during the past period. The tool is very easy to use, the procedures are summarized as below, i) going to the Baidu Index page (<http://index.baidu.com>), ii) entering the keywords in the Baidu Index search textbox, iii) selecting the searched time, range and region, and click on the 'Baidu', vi) the trend line of the keyword for the corresponding period and region is shown.

The keywords selection is the most important work for the Baidu Index-based researches. Until now, there is no unique method to define the search keywords set. Currently, keywords selection methods can be classified into three categories, i.e., technological keyword selection, direct keyword selection and keyword selection scope [6-8]. For the technological keyword selection, the keyword in certain field is automatically raised by computer pattern identification. Direct keyword selection means that the keyword is determined by using subjective assessment of data. For keyword selection scope, a predetermined keyword is firstly used, and then selects text pertinent to keywords. Although the risk of core keyword omission exists for direct keyword selection and keyword selection scope, low workload benefits the proficiency. Therefore the direct keyword selection and keyword selection scope gain more popularities. For the tourism activity, many researches on relationship between tourists' volume and search engine data have been conducted in the past decade. The results indicate that the search keywords, for instance 'ticket', 'ticket price' and 'tourism strategy', are suitable to capture the tourist volume. For simplification, here, we select the search keywords directly by reviewing the existing literatures. In this work, taking Xi'an tourism as the subject of the case study, using the Baidu Index, and analysing the relationship between internet search data and actual tourist numbers through keyword selection, providing an example how to analyse the internet search data[9-10].

Regarding the search keywords, for the overall Xi'an tourism, three keywords 'Xi'an tourism', 'Xi'an tourism strategy' and 'Xi'an weather' are selected. For the national 5A scenic spots, 'Terra

Cotta Warriors ticket’, ‘Huaqing Hot Spring ticket’ and ‘Dayan Pagoda ticket’ are used. In the following sections, the evolution of Baidu Index for each keyword will be presented and discussed.

Results and Discussion

Evolution of Baidu Index for Xi’an Tourism. It can be seen from Fig. 1 that the Baidu Indexes of search keyword “Xi'an tourism” in 2014-2016 are at the same levels, their overall trends are less than that of 2012, while the Baidu index of 2013, user attention, is the largest among five years. For each year, with the National Day holiday’s approaching, the user attention of keyword “Xi'an tourism” shows a sharp increase. In the first day of National Day holiday (October 1st), the user attention achieves the maximum value. Before the National Day holiday, from the September 1st to September 30th, user attentions show a slight oscillation phenomenon, this phenomenon is corresponding to the weekend effect. In the weekend, people have time to travel. Then the tourists to Xi'an are increased. Figure 1 also indicates that the user attentions decrease like a cliff after October 1st to the last day of the National Day holiday.

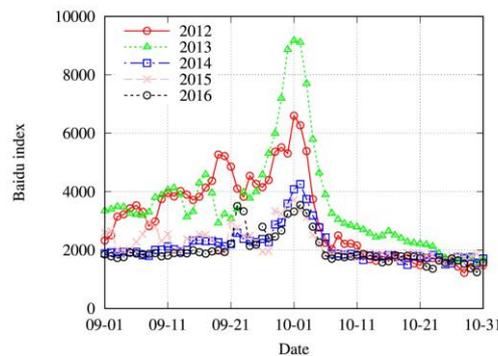


Figure 1. Finite Evolution of Baidu Index for keyword “Xi’ an tourism” in past five years.

Fig. 2 is the Baidu Index of search keyword “Xi'an tourism strategy”, the overall trend of the user attention in 2012 is the minimum. The user attentions of 2013 to 2015 years are almost same, while the overall user attention in the year of 2016 reaches the maximum. The trends of user attentions for keyword “Xi'an tourism” is different with that of ‘Xi’an tourism’. Therefore, different Baidu Index keyword will lead to differences in the overall trend of user attention, so the search keywords should be carefully selected. For certain year, Fig. 2 shows the same trends as shown in Fig. 1, i.e., from September 1st to 30th, the internet user attentions show increased trend overall. There is a local oscillation, the maximum user attention reaches on October 1st. From October 1st to October 7th, internet users attention shows a cliff drop, After the National Day holiday, the overall user attention relatively stable. Meanwhile, there is a slight oscillation which is corresponding to the weekend effect.

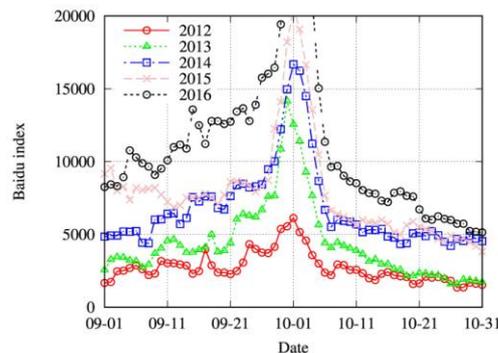


Figure 2. Finite Evolution of Baidu Index for keyword “Xi’ an tourism strategy” in past five years.

Fig. 3 is the internet user attention for the keyword “Xi'an weather” from 2012 to 2016. It can be seen from Fig. 3 that the user attention of 2012 year is almost same as that of 2013, while those of 2014-2016 are comparable higher. However, different with those of Fig. 1 and Fig. 2, the user attention of search keyword ‘Xi'an weather’ does not increase with the approaching of National Day holiday. The possible reasons for this phenomenon may be that people seldom care about the short-term weather in making their traveling plans. When they have made a decision to visit Xi'an, they will not give up their traveling plans even the weather has a slight change. Although the destination climate has a significant impact on tourists traveling behaviour, the impact from short-term weather changes is slight.

Evolution of Baidu Index for Terra Cotta Warriors. The Terra Cotta Warriors is located in Xi'an Lintong District, which is 1.5 kilometres far from east of Mausoleum of the First Qin Emperor. The Terra Cotta Warriors is a part of the Mausoleum of the First Qin Emperor. In the developed three pits, there are a large number of Terra Cotta Warriors in the underneath. In 1987, the Mausoleum of the First Qin Emperor was approved as the protection of the world cultural heritage. More than 200 national leaders had ever visited the Terra Cotta Warriors until now. It becomes a brilliant civilization of ancient Chinese culture. The Terra Cotta Warriors is one of the greatest discoveries in the history of the world.

Fig. 4 is the user attention of keyword “Terracotta Army ticket” from 2012 to 2016 years. The user attention of keyword ‘Terra Cotta Warriors ticket’ in 2016 is the smallest, the overall level of user attention in the year of 2015 is the largest. For the other years, they are in the middle of the years of 2015 and 2016. With the approaching of the National Day holiday, the trend of user attention in September shows a gradual increase. Meanwhile, there is a slight oscillation. Similar as that of Fig. 3, user attention reaches the highest peak on October 2nd, and then there is a cliff drop at the end of the National Day holiday. After that, user attention gets to the steady in the overall trend.

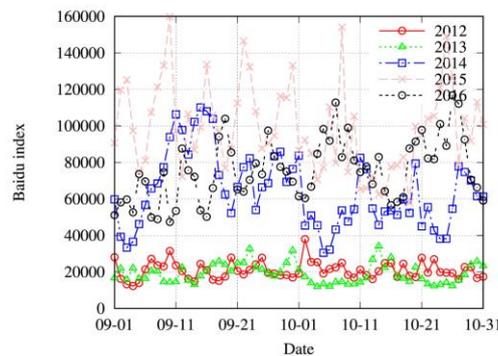


Figure 3. Finite Evolution of Baidu Index for keyword “Xi’ an weather” in past five years.

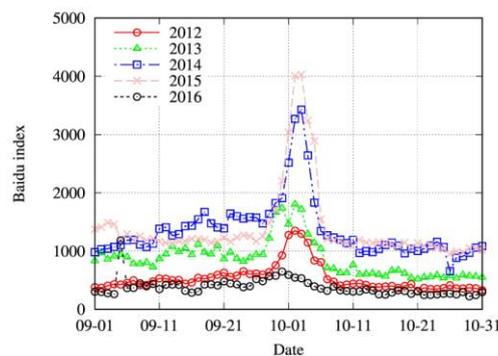


Figure 4. Finite Evolution of Baidu index for keyword “Terra Cotta Warriors ticket” in past five years.

Evolution of Baidu Index for Huaqing Hot Spring. Huaqing Hot Spring, also named as Huaqing Palace, locates in Xi'an Lintong district, in the north of Mount Li, which is 30 kilometres far from west of Xi'an. Huaqing Hot Spring is famous for the hot spring pool, a Chinese ancient palace. Historical documents and archaeological excavations have proved that Huaqing Hot Spring has a history of 6000 years of hot springs, and the history of the Royal Garden in the past 3000 years. In 1982, Huaqing Hot Spring was listed as the first batch of national key scenic spots. On May 8 2007, Huaqing Hot Spring was approved as the national 5A scenic spot.

Since the Baidu index does not create the user attention of search keywords 'Huaqing Hot Spring tourism', 'Huaqing Hot Spring tourism strategy" and "Huaqing Hot Spring weather", the present work will focus on the user attention of keyword 'Huaqing Hot Spring ticket'. It can be seen from Fig. 5 that a similar user attention trend was found as that of keyword 'Terra Cotta Warrior ticket' during the period from September 1st to October 1st. The user attention of keyword "Huaqing Hot Spring ticket" shows a gradually increasing trend, with a local oscillation. In the first two days of National Day holiday, the user attention reaches dramatically. The user attention has a sharp increase on October 2nd, and achieves the maximum value. At the end of October 7th, the user attention also shows a cliff drop. After the National Day holiday, the user attention shows a stable trend. Due to the impact of weekend effect, a local oscillation is observed as well.

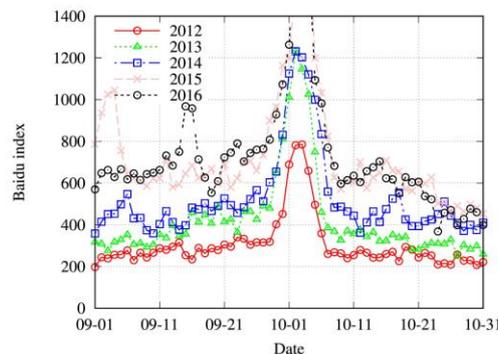


Figure 5. Finite Evolution of Baidu Index for keyword “Huaqing Hot Spring ticket” in past five years.

Evolution of Baidu Index for Dayan Pagoda. The Dayan Pagoda is located in Xi'an Da Ci'en Temple, is considered as a symbol of the ancient Xi'an and Shaanxi province. The Dayan Pagoda is the earliest and largest Tang Dynasty Square Pavilion style pagoda. The pagoda is the Indian style temple building with the spreading of Buddhism into the Central Plains region. It is the embodiment of iconic Chinese people's wisdom. The existing tower has seven layers, 64.5 meters high. On June 22, 2014, the United Nations held in Doha, Qatar on the UNESCO 38 session of the World Heritage Committee. The Dayan Pagoda is successfully included in the list of world heritage.

Fig. 6 is the user attentions of search keyword 'pagoda tickets' from 2012 to 2016 years. The overall user attentions is similar as those of keywords 'Huaqing Hot Spring ticket', while is lower than that of keyword 'Terra Cotta Warriors ticket'. The main reason can be explained as that the Terra Cotta Warriors scenic spot is more famous than that of Huaqing Hot Spring and Dayan Pagoda scenic spots in the world. From 2012 to 2016, the user attention in 2012 is the smallest, the largest in 2016, while 2013 to 2015 in the middle. The peak of the user attention appears on October 2nd or October 3rd, which is the same as that of Terra Cotta Warriors and Huaqing Hot Spring scenic spots.

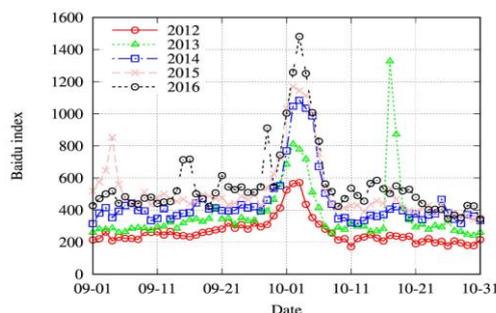


Figure 6. Finite Evolution of Baidu Index for keyword “Dayan pagoda ticket” in past five years.

Conclusion

This paper analyses the user attentions of three national 5A scenic spots in Xi'an city from 2012 to 2016. The results indicate that: (1) with the increase of the year, the overall level of the user attention is increasing. (2) In National Day Holiday, the daily user attention shows a sharp increase with a local oscillation. The daily peak of user attention is on October 2nd or October 3rd, after that, the daily user attention shows a cliff drop until the end of National Day holiday. Subsequently, the overall level of daily user attention tends to a smooth flat. (3) For Xi'an city, the Baidu Index of search keyword ‘Xi'an weather’ has a slight change with the increase of tourist numbers during the National Day holiday. It has a week correlation with tourist numbers. As for a scenic spot, the user attention of search keyword ‘(scenic spots) ticket’ is very important, there is a strong coupling between the actual amounts of tourists and Baidu Index.

Acknowledgements

The authors acknowledge the Scientific research project of Shaanxi Provincial Department of Education: the application and forecast of domestic tourism market demand in Shaanxi Province under the background of big data (Grant: 17JK2076); Xi'an Social Science Planning Project: Study on the spatial structure of tourism resources and market forecast in Xi'an (Grant: 2016EA11)

References

- [1] Huang X K, Zhang L F, Ding Y S, The baidu index: Uses in predicting tourism flows – A case study of the Forbidden City [J]. *Tourism management*, 2017, (58), pp. 301-306.
- [2] Song H, Li G, Tourism demand modelling and forecasting-A review of recent research. *Tourism Management [J]*. 29(2), pp. 239-259.
- [3] Li X, Pan B, Law R, Huang X K, Forecasting tourism demand with composite search index, *Tourism management [J]*. 2017, (59), pp. 57-66.
- [4] Yang X, Pan B, Evans J A, Lv B F, Forecasting Chinese tourist volume with search engine data, *Tourism management [J]*. 2015, (46), pp. 386-397.
- [5] Sun Y, Lv B, A review of researches on the correlation between internet search and economic behaviour. *Management Review [J]*. 23(7), pp. 72-77. (in Chinese)
- [6] Ginsberg J, Mobebbi M H, Patel R S, et al. Detecting influenza epidemics using search engine query data [J]. *Nature*, 2009 (2) : 1012—1014.
- [7] Geoffrey Peter Smith. Google Internet search activity and volatility prediction in the market for foreign currency [J]. *Finance Research Letters*. 2012 (9): 103—110.

- [8] Hyunyoung Choi, Hal Varian. Predicting the Present with Google Trends[J]. *The Economic Record*, 2012, 88: 2—9.
- [9] Y. Fondeur, F. Karamé. Can Google data help predict French youth unemployment[J]. *Economic Modelling*, 2013, 30 : 117—125.
- [10] Getz D. *Event management & event tourism* [M]. New York: Cognizant Communication Corporation, 1997: 4—11.