Tourism Big Data and Its Application

Meiwei Tang[1*], Xue Ren[2]

[1]Nanjing University of Finance and Economics, Nanjing, China

[2] Chongqing Technology and Business University, Chongqing, China

*Corresponding author: Meiwei Tang

Keywords: Big data, Mnos, Tourist flow

Abstract: This study sketched the basic structure, characteristics, and applications of tourism big data for mobile network operators (MNOs). We discuss the specific impact of big data on tourism statistics and take Nanjing of China as examples to illustrate the application of it in tourism statistics. The innovation of this paper lies in the following two aspects. We theoretically expound the impact of big data from the perspective of tourism statistics. And in terms of application, we put forward the methods of calculating tourist amount by using MNOs. Findings clearly indicate that applying big data to tourism statistics can promote tourism data more timeliness, accuracy and complete.

1. Introduction

First-hand tourism data have been collected from tourists and enterprises through sample and special surveys and statistical reports, as “person-times” and “revenue.” However, statistical results based on sample surveys inevitably cause problems of small coverage, poor timelines, and inconsistent standards. Other issues relate to the survey of accommodation units not covering tourists staying in unregistered bed and breakfast outfits and tents, insufficient identification of attributes of passing tours, lack of reliable and detailed classification statistics on tourism demand, low frequency of publication of statistics, and lack of professionalism of survey staff. These problems seriously affect the quality of tourism statistics, which are one of the main reasons for their being “incomparable horizontally and not addable vertically.”

Applying big data in tourism statistics has theoretical and practical basis. Most big data sources maintain human activity patterns and transaction behavior. Tourism statistics monitor the flow characteristics of those with tourism characteristics (tourists) and their monetary flow. Though the process of collecting information for big data is similar tourism statistics, it is not surprising that the change in tourism data sources and statistical methods driven by big data arises in the latter. Big data has an important impact on tourism statistics in the aspects of tourist identification, tourism consumption calculation, tourism consumption-supply balance, tourism industry cooperation, high-frequency tourism data and sub regional tourism data.

The second part of this study introduces literature review for available big data and its applications; the third presents case study and put forward the basic methods of calculating tourist amount by using mobile signal data; the fourth part contains the discussion and conclusion.

2. Literature Review

MNOs (mobile network operators) are the most widely used big data source for measuring tourism (Eurostat, 2017). Therefore, this section focuses on it in detail. The outstanding feature of MNOs is real-time tracking of the user’s location, which provides information for real-time monitoring of tourism statistics. MNOs record communication activities of users including call detail records (CDR) and tracking data detail records (DDR). Specifically, each call or SMS activity is called an event. Each event in data source should comprise at least the following four attributes: the user ID, which is usually replaced by a number and is different for each user; the country code

  DOI: 10.25236/iiiece.2022.019

to which the user belongs; the time of the event; and the geographic location of the event.

Internationally relevant data on mobile phone positioning has been applied to a certain extent in tourism statistics. In 2012, Eurostat used cell phone positioning data to monitor statistical tourism indicators such as inbound, domestic, and outbound tourism. Statistics Netherlands conducted a pilot study in 2013 to explore whether MNOs data could be applied to tourism statistics. The findings suggested that it had great application value in tourism statistics, especially in studying inbound tourism. Both the number of foreign cell phone roaming and the number of roaming call records were consistent with the trend of the number of inbound tourists in accommodation statistics. The extrapolation of trends in MNOs data to predict the number of inbound tourists deviated from the real data by only 5%.

Several scholars have validated the rationality of using MNOs data for tourism statistics and achieved preliminary results. Ahas et al. (2007) concluded that MNOs data had considerable potential in tourism research and monitoring and used the spatiotemporal MNOs data to systematically study the seasonal change pattern of tourists. Taking inbound tourism in Estonia as an example, Ahas et al. (2008) determined the travel mode, attraction visits, and tourist source composition based on the geographic location of tourists during their travel and associated basic information. The correlation coefficient between the accommodation statistics and MNOs data from inbound tourists from 15 countries were also calculated. In eight countries the correlation coefficient was greater than 0.9; the lowest was 0.68. Nilbe et al. (2014) used MNOs data to study the distance traveled by tourists. Raun et al. (2016) calculated the destination tourism trips by their geographic location and time combination parameters.

The essence of statistics regarding tourist trips is the dynamic monitoring of those who are in a mobile state. MNOs data source has obvious advantages in sensing the distance, frequency, and travel time of tourists (Vij, Shankari, 2015). However, to ensure the accounting accuracy of MNOs data in tourism statistics, the following issues about sample selectivity bias need to be explored: first, the relationship between the number of users and of tourism trips. It is likely that only one cell phone is used in a family trip or one person uses multiple cell phones in a business trip; thus, the estimation of tourist person-times by the number of cell phone users needs to be combined with certain auxiliary surveys to determine the proportional relationship. Second, the inconsistency between the user's cell phone registration place and habitual residence. The phenomenon that nonlocal owners are not necessarily nonlocal tourists is common, and effective algorithms are needed to systematically identify the usual environment of tourists. Third, the accuracy of information coverage. There may be tourists who turn off their mobile phones during the tour or purchase new SIM cards in the locality, causing the number of tourists to be underestimated.

3. Case Study -the Application of Mnos

Consider the availability of data. This section explores the application of MNOs data by taking Nanjing as an example.

MNOs[1] uses cell phone positioning data to monitor the flow of visitors through information such as cell phone origin place, location information, and length of stay. In terms of tourist judgment rules, all mobile users who have voice or traffic roaming behavior outside China (including Hong Kong, Macao and Taiwan) are inbound tourists. Starting from the date when the first voice or roaming record is generated, we take 7 days as a time series cycle. When the first time series cycle starts, one person time will be counted if the records are continuously generated (the second 7 days, the third 7 days...), and two person time will be counted if a time series cycle is disconnected in the middle. In terms of tourists' location attributes, the specific location is judged according to the base station cell and scenic spots. For example, the base station code of Fuzi Temple is 20627, and there are 627 base station cells. Based on the above information, the flow of tourist is carried out.

3.1 Monitoring Tourist Flow

[1] The data comes from the desensitization data of China Mobile.

We use MNOs data to count the daily tourist flow in Nanjing. Compared with the annual data, high-frequency data could provide monitoring and early warning services for tourism suppliers, and finally improve tourism satisfaction.

As can be seen from Figure 1, the tourist flow of Nanjing city has the characteristics of “peak forest structure”. There are four obvious “main peaks” on the figure, which are Spring Festival holiday, Tomb Sweeping Day holiday, International Labour Day and National Day holiday. The zigzag “peaks and valleys” in the Figure 1 are formed by the increase of tourist on weekends compared with working days.

The peak tourist season in Nanjing is February, April to Mid June and early October. During the Spring Festival, Tomb Sweeping Day, Labor Day and National Day, the tourist flow increases sharply. Especially during the Labor Day holiday in 2019, the number of tourists in Nanjing has exceeded 2.84 million; The oral tourism season is from late February to March, June, and November. The suitable weather and student holidays are the main reasons for the increase of tourists in Nanjing; The rest of the time is off-season for tourist. Nanjing, known as the “Stove”, is extremely hot in summer and too wet and cold in winter, which is an important reason for the decrease of tourists.

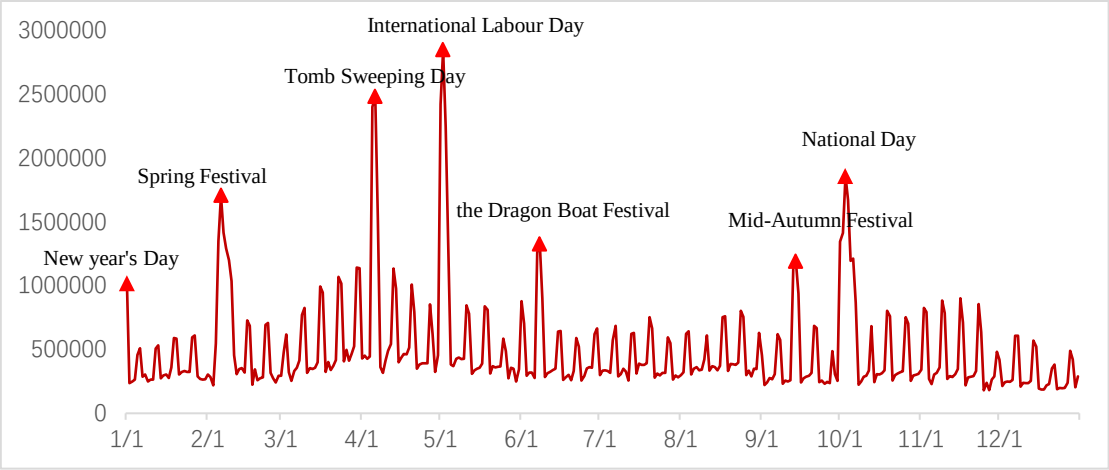

Fig.1 Tourist Flow in Nanjing 2019

3.2 Scenic Spot Combinations Analysis

Figure 2 shows the top ten routes preferred by tourists. The preferred scenic spot combination is from Fuzi Temple to Zhongshan Mountain National Park to the Presidential Palace, with 1.28 million visitors who chose this route from January to December 2019. It was followed by the route from Fuzi Temple to Zhongshan Mountain National Park to Xuanwu Lake (888,229 tourists). Based on tourists’ preferences, relevant sectors can develop traffic routes to facilitate their travel.

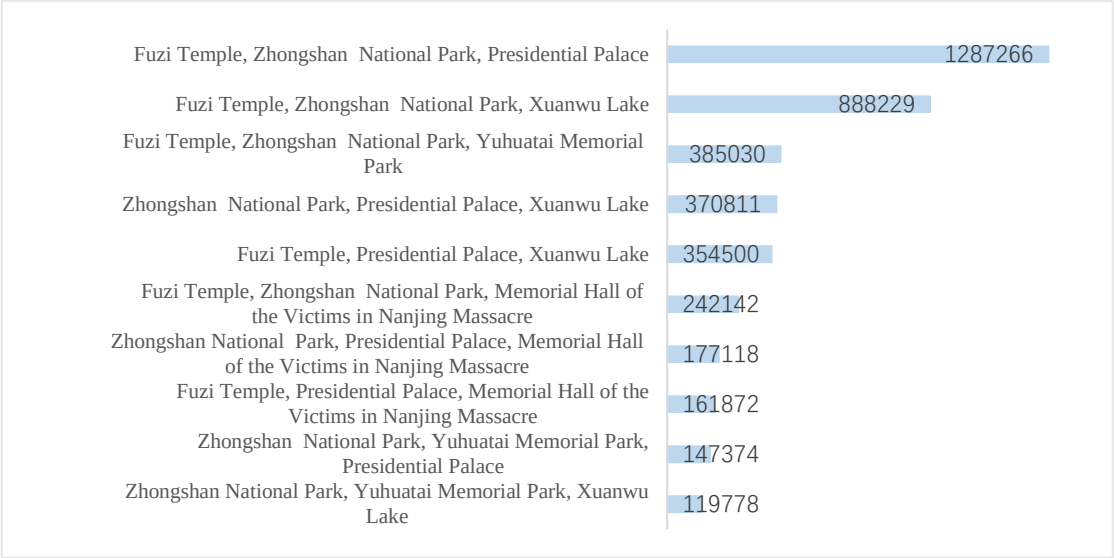

Fig.2 Scenic Spot Combinations in Nanjing (Person-Times)

3.3 Tourist Portrait Analysis

Through MNOs data, we could do in-depth analysis on tourists characteristics, including gender, age, network behavior, tourist origin and other information. Therefore, tourism enterprises may provide more targeted tourism services. As we see in Figure 3, we find that in 2019, male tourists in Nanjing accounted for 55.7%, female tourists accounted for 44.3%, and tourists aged 30-40 were the most, It accounts for 26.8% of the total number of tourists, followed by tourists aged 19-30 and 41-50, accounting for 24.2% and 22.9% respectively. It can be seen from Figure 4 that Nanjing's online tourism supply market shows a dominant situation, with Xiecheng accounting for 45.5%, Qunar accounting for 15.5%, and other apps accounting for less. We observe the domestic and foreign tourist sources in Figure 5 and Figure 6 (places with dark colors represent a large number of tourists). The top five tourist sources outside the province are (excluding Jiangsu): Anhui 22.3%, Zhejiang 11.8%, Shanghai 10.22%, Guangdong 6.41%, Shandong 5.59%. The top five foreign tourist sources are Singapore 12.6%, Japan 8.82%, Thailand 8.09%, UK 6.87%, Korea 5.51%. Foreign tourists are mainly concentrated in Asia. In short, compared with traditional data, tourism big data can provide richer tourist information.

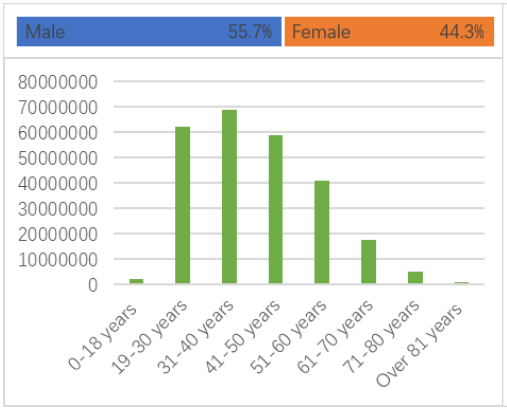

Figure3 Gender and age of tourists

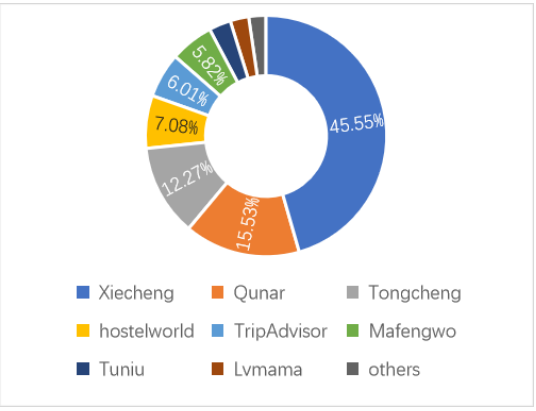

Figure4 Tourist network behavior

Figure5 Main sources of domestic tourists

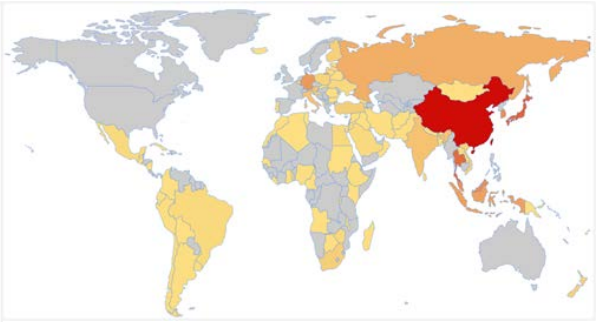

Figure6 Main sources of inbound tourists

4. Discussion and Conclusion

In this paper, we take Nanjing as an example to introduce the application of MNOs. We use MNOs data to calculate tourist flow, which could help us better analyze tourists' characteristics, including daily data of tourist flow, analysis of tourist scenic spot combination and analysis of tourist portrait. These data can make up for the disadvantages of low frequency and insufficient information of current tourism data. The result could provide scientific basis for tourism enterprises to do more targeted tourism services.

Big data has important impacts on tourism statistics and even the official statistics. In this study, based on data availability considerations, we only take MNOS data as an example to study the

application of tourism big data. In the future, in order to make better use of tourism big data, an urgent issue to be resolved is the trade-off between tourism data application and data confidentiality. In fact, most of the big data sources are non-publicly used and contain a large amount of sensitive information such as personal privacy and commercial secrets. Ensuring that the use of these data does not endanger such sensitivities is the key to guarantee continuous access to data use. Therefore, a standard development and application process of micro-big data source and corresponding confidentiality specifications should be laid at the national level as early as possible (Xu et al., 2018), this is the only way to ensure the use of big data under legal and compliant conditions and improve the trust of data providers.

Acknowledgments

The authors would like to acknowledge the financial support from the The National Office for Philosophy and Social Science “Research on Tourism Statistical Monitoring Methods and China Tourism Satellite Account in the Context of Big Data” under grant No. 19CTJ011.

References

[1] Eurostat. Tourism Statistics: Early Adopters of Big Data[R]. Statistical Working Papers, 2017.

[2] Ahas R, Aasa A, Roose A, et al. Evaluating Passive Mobile Positioning Data for Tourism Surveys: An Estonian Case Study[J]. Tourism Management, 2008, 29 (3):469–486.

[3] Ahas R, Aasa A, Ular M, et al. Seasonal Tourism Spaces in Estonia: Case Study with Mobile Positioning Data[J]. Tourism Management, 2007, 28 (3):898–910.

[4] Nilbe K, Ahas R, Silm S. Evaluating the Travel Distances of Events Visitors and Regular Visitors Using Mobile Positioning Data: The Case of Estonia[J]. Journal of Urban Technology, 2014, 21 (2):91–107.

[5] Raun J, Ahas R, Tiru M. Measuring Destinations Using Mobile Tracking Data[J]. Tourism Management, 2016, 57:202–212.

[6] Vij A, Shankari K. When is Big Data Big Enough? Implications of Using GPS-based Surveys for Travel Demand Analysis[J]. Transportation Research, 2015, 56:446–462.

[7] United Nations, et al. (2008). Tourism satellite account: Recommended methodological framework 2008.

[8] United Nations, et al. (2012). National accounts system 2008, translated by Department of National Accounts, the National Bureau of Statistics. Beijing: China Statistics Press.