Empirical Analysis of the Impact of High Speed Rail on Tourism Spatiotemporal Behavior—A Case Study in Jiangsu-Zhejiang-Shanghai-Anhui Region

Jing SHI  
Tsinghua University, Department of Civil Engineering  
Haidian District, Beijing, China  
jingshi@tsinghua.edu.cn

Lei XIN  
Tsinghua University, Department of Civil Engineering  
Haidian District, Beijing, China  
xinlei931223@126.com

ABSTRACT

The target of this research is exploring the influence of high speed rail on tourism spatiotemporal behavior and quantitatively analyzing the mechanism of the impact of high speed rail by social media chick-in data from SINA Weibo. The Nanjing-Hangzhou high speed rail, which has put into operation since July 2013, is taken as a case to analyze its effect on tourists’ spatiotemporal behavior of Hangzhou Weibo users during the National Day of 2012 and 2013, the holiday before and after high speed rail operation. A systematic analysis method is established to handle social media big data which is relatively sparse in space. First, destination space boundaries are rearranged to maintain evenness of data in each destination. Second, transfer rate is used to characterize the spatial mobility between destinations. Third, spatial network model is established and analyzed by community discovery algorithm. It is found that, high speed rail would promote the reconstruction of tourism circle which indicates the travel range of tourist, and the integration ability of tourism resources is significantly enhanced for the cities along the high speed rail. In addition, from the prospective of microscopic transfer rate, for the cities with different tourism resources, the mechanism of promotion tourism development by high speed rail is different. These analysis confirms the significance of high speed rail to the spatiotemporal behavior of tourist and the research results can provide reference for the planning of high speed rail.

KEYWORDS: big data, transfer rate, community discovery, travel behavior, tourist distribution.

1 INTRODUCTION

With the rapid development of urbanization in China, in order to meet the needs of intercity communication, transportation facilities have been continuously upgraded and improved. Especially, the intercity high-speed railway has been rapidly promoted in recent years. The convenient transportation reduces the time distance between cities, thus intercity travel and multi-destination travel are become more and more popular.

Since the emergence of high speed rail all over the world, its impact on travel behavior has been widely concerned by researchers. However, with different transportation culture background, the influence of high speed rail on travel behavior is different in different countries. Bazin et al. (2013) finds that French TGV (train à grande vitesse) promotes tourists to well-known tourism destinations, while the tourist number of unknown tourism destinations have rarely increased. Pagliara et la. (2015) argues that high-speed rail is not the decisive factor for international tourists to choose Madrid as
destination, but it successfully helps to attract more tourists to surrounding towns which are connected with Madrid by high speed rail. Through panel data, Albalate and Fageda (2016) find that high speed rail has only weak contribution to attract tourists, and this effect will be gradually decayed. While in most of cases study in densely populated areas, such as Japan (Kurihara and Wu, 2016) and China (Chen and Haynes, 2015), it concludes that high speed rail significantly impacts the spatiotemporal behavior of tourists and promotes the development of regional tourism. The difference of these findings may be related to travel habit which reflect the culture of the counties under study. Most Chinese tourists regard the transportation as a way to realize space transfer, because of the large population and wide range territory of China. However, the foreign tourists pay more attention on landscape along the route, and they regard the transportation as part of their traveling (Lu, 2009).

Most of the quantitative researches obtain the data of tourists through the official statistical yearbook (Albalate and Fageda, 2016; Kurihara and Wu, 2016; Chen and Haynes, 2015 ). Although these data can partly explain the relationship between high speed rail and the number of tourists, there are inherent defects in the data itself. First of all, the annual or monthly data may lack detailed information about the attributes of tourists, and cannot distinguish the different travel pattern between workdays and holidays. Secondly, most of the data in yearbook are collected in national or provincial unites. That means it doesn’t provide accurate data on the scale of cities which are directly connected by high speed rail (Kurihara and Wu, 2016). When assessing the impact of high speed rail on tourism in China, Chen and Haynes (2015) only use the statistical data of international tourists as dependent variable, while the domestic tourist data is difficult to obtain directly in China. In order to overcome the shortcomings of official statistics data, many researchers formulate special questionnaires for their own purpose and investigate in major transportation hubs, scenic spots entrances and hotels to collect data (Wang et al., 2014; Wang et al., 2015). However, limited by finite manpower, the survey result is not sufficient enough for large area analysis.

With the development of mobile Internet and the popularity of smart phones, Location Check-in Aggregator has gradually become one of the most popular function for tourists. In tourism activities, the enhanced interaction between cyber society and real world action will further encourage tourists to use mobile social network to increase their satisfaction (Dickinson et al., 2014). Compared with traditional data, the spatiotemporal data in social media covers wider range in time and space, that means it can overcome the shortcomings of traditional questionnaires.

SINA Weibo is a famous public social media for sharing short real-time status in China. Since July 2011, SINA Weibo has launched its check-in function on mobile application and attracted a large number of users. These check-in data accumulate a large amount of information about tourists’ spatiotemporal behavior, which inspires a new way to quantify the impact of high speed rail. In recent year, the research on microblog check-in data has gradually attracted attention, while the research methods are not mature yet. At present, the analysis of spatiotemporal behavior by social media data mainly focuses on the characteristics of spatiotemporal activities of residents (Wang et al., 2014), the boundaries of spatial activities (Cao et al., 2014), and the spatial interaction between various destinations (Liu et al., 2015; Sui et al., 2013). There are relatively few studies on tourists’ spatiotemporal behavior, and they mainly focus on the following aspects. On the one hand, some researches analyze the spatiotemporal distribution of tourists. Wang et al. (2017) analyze the distribution of tourists in Lanzhou by micro-blog check-in data and reveals the relationship between number of tourists and time. Zhang et al. (2015) discuss the evolution of spatiotemporal distribution of tourists in Zhongshan scenic spots of Nanjing and finds the distributions of tourists with different genders and regions are different. Li et al. (2015) find that the spatiotemporal co-occurrence of tourism activities is weaker compared with that of daily travels, which indicated the tourists’ travel are more random. On the other hand, some researches mainly focus on the analysis of content and location of micro-blog. Xie et al. (2017) construct a topic model for micro-blog with location information, and analyzes the social cultural and tourists’ behavior of different regions. Han and Zhang (2011) summarize the content of micro-blog in tourism areas and compared it with field survey, which confirms the consistency between micro-blog content and actual situation.

Generally speaking, the existing researches on social media data have accumulated some achievements on discovery of Point of Interest, association of nearby spots, community discovery of tourism and so on. However, most existing researches focus on the description of spatiotemporal
distribution by aggregating single location data. Few studies involve the analysis of travel routes of tourists, and research methods in this area need to be further explored.

2 RESEARCH OBJECT SELECTION AND DATA ACQUISITION

The case area of this research is a very famous tourism area in East China which contains four provinces, i.e., Zhejiang, Shanghai, Jiangsu and Anhui. There are plenty of beautiful natural sceneries and cultural historic heritage attractions. It contains more than 1/4 of the number of 5A level scenic spots, the highest tourism spot level in China. Besides, it is also one of the most economic areas in China, and the travel demand of inhabitants in this area is relatively high. Meanwhile, because of the developed transportation network, there are almost no commercial air lines in this area. Thus it is helpful to focus on the influence of high speed rail.

In this research, we focus on the tourists from Hangzhou and their tourism activities on National Day, a main Chinese holiday, in 2012 and 2013. Because a new important high speed rail from Hangzhou to Nanjing started operation between 2012 and 2013(line 7 shown in Figure 1), and the travel time between these cities is reduced from 2.5 hours to 1.25 hours. That means the traveling convenience of Hangzhou residents is significantly improved.

![operation time](Figure 1: The high speed rail map in case area)

The data of SINA Weibo can be obtained through its Application Programming Interface Platform (API Platform). We downloaded the data in the format of JSON and saved the data in Mongo DB, a non-relational database. JSON format records data in the way of key-value pairs. The information contained in each microblog is shown in the table 1.
Table 1: Introduction of key of SINA Weibo check-in data (partial content)

<table>
<thead>
<tr>
<th>The name of key</th>
<th>The meaning of key</th>
</tr>
</thead>
<tbody>
<tr>
<td>“created_at”</td>
<td>Publishing time of microblog</td>
</tr>
<tr>
<td>“test”</td>
<td>Content of microblog</td>
</tr>
<tr>
<td>“mid”</td>
<td>Identification number of microblog</td>
</tr>
<tr>
<td>“poiid”</td>
<td>Identification number of point of interest</td>
</tr>
<tr>
<td>“title”</td>
<td>Name of point of interest</td>
</tr>
<tr>
<td>“lat”</td>
<td>Latitude of check-in location</td>
</tr>
<tr>
<td>“lon”</td>
<td>Longitude of check-in location</td>
</tr>
<tr>
<td>“id”</td>
<td>Identification number of user in Weibo</td>
</tr>
<tr>
<td>“location”</td>
<td>Registered residence city of user</td>
</tr>
<tr>
<td>“gender”</td>
<td>Gender of user</td>
</tr>
<tr>
<td>“followers count”</td>
<td>Fans number of user</td>
</tr>
</tbody>
</table>

According to the point of interest (POI), the label indicating the location types of data in Weibo, the data related to tourism activity are screened out. During the National Day in 2012 and 2013, 27264 and 21197 Weibo users of Hangzhou were detected to tweeted at least one tourism micro-blog in the area of Jiangsu, Zhejiang, Shanghai or Anhui. According to their user ID, the location data of their travel route is obtained.

3 DATA PROCESSING AND ANALYSIS

3.1 Destination recognition

According to previous researches, the longer the distance between origin and destination is, the less visitors would visit this destination (Tang, 2015). Thus, the spatial distribution of Hangzhou tourists in the four provinces is uneven. In order to appropriately describe the relationship between destinations, it is necessary to determine the spatial boundaries of each destination to ensure the data contained is enough to reflect arrival rate.

In this research, destinations are recognized by visit quantity of social media data. If user $n$ has visited destination $D_i$ in $d$ day, $C_{nd}(D_i) = 1$ is set, otherwise it is 0. And thus, $\sum_d \sum_n C_{nd}(D_i)$ is the total number of person-day in destination $D_i$. The spatial boundaries of destination are decided by following rules. If the number of person-day of tourists in a county level distinct is less than 0.1% of the number of person-day in the whole area, then the county is seen as a destination. If not, the personal-day of the cities which are included in the county is considered. And if the number of a city is more than 0.1% of the total number of person-day, this city would be seen as an independent destination. Other cities that do not meet this proportion are merged with the core city of county or set as independent destinations according to whether they are geographically adjacent. The reason for setting the threshold to 0.1% is that the destinations can be distinguished appropriately in this case.
The repartition of destination could avoid the significant difference in data scale of arrival rate between different destinations, so as to better observe the movements of tourists. Figure 2 is the conclusion of spatial boundaries of 98 tourist destinations based on the social media data in 2012 and 2013. There are 58 destinations in Zhejiang, 19 destinations in Jiangsu, 20 destinations in Anhui and 1 destination in Shanghai. As shown in the figure, tourists tend to choose the spots around the origin Hangzhou, so the cities nearby Hangzhou are more like to be considered as independent destinations.

3.2 Transfer probability of destinations

In this research, the transfer probability of destinations is used to express the movement of tourists. The check-in sequences of Weibo users are obtained by Weibo API Platform.

First, the destinations sequence $L_n$ of tourist $n$ can be gotten by analyzing every points in check-in sequence. If the destination of current point is the first destination or is different from the last destination in current sequence, it needs to be inserted into the end of $L_n$.

Next, beginning with the second elements of destination sequence $L_n$, if the current element is $j$ and its former element is $i$, then the value of $T_{nij}$ is equal to 1. Thus $\sum_n T_{nij}$ is the total number of transfer for all tourists from $i$ to $j$.

The destination transfer probability is the ratio of transfer number of a path to the total transfer number of all paths. The transfer probability distribution shows an obvious long tail effect, that means the paths with low probability occupy overwhelming majority of all paths. Therefore, the paths in top 10% cumulative probability are defined as frequent transfer paths to clearly describe the main spatial relationship between destination and avoid overly complex network.

3.3 Identification of tourism circle

Because the transfer probability matrix is relatively large, it is difficult to fully display the overall distribution results, so the tourism circle is introduced in this section. Tourism circle reflects the spatial cooperation of multiple tourism attractions, meanwhile it also indicates the frequent spatiotemporal distribution of tourists. In order to identify the tourism circle, a spatial network based...
on graph theory is established. It is assumed that destinations are nodes in the spatial network, and every two destinations are connected by an undirected edge whose weight is the frequent transfer rate between these two destinations. So, a weighted undirected network reflecting the strength of connection between destinations is established.

The community refers to a subset of nodes in a complex network. The nodes in same community are closely related, while the nodes in different community are relatively independent (Girvan and Newman, 2001). The community discovery algorithms for weighted undirected network mainly include Girvan-Newman algorithm (Girvan and Newman, 2001), Fast Newman algorithm (Newman and Girvan, 2004), and Spectral Clustering algorithm (Ng et al., 2001). The Girvan-Newman algorithm is used in this research. The basic principle of this algorithm is to ignore the weight of edges and calculate the betweenness of each edge. The betweenness of edge is the number of passes in all shortest paths between each two nodes. Then the betweenness is divided by the weight of corresponding edge, and the ratio is defined as the new betweenness of each edge in the weighted network. In this way, the edges with large weight tend not to be removed, which is in line with the goal of community discovery.

In order to evaluate the results of community discovery, Newman and Girvan (2004) propose the concept of modularity and extend it to weighted complex network. The formula of modularity is shown as below.

\[
Q = \frac{1}{2s} \sum_{ij} \left[ w_{ij} \left( \frac{s_i}{2s} \right) \delta(G_i, G_j) \right]
\]  

(1)

In this formula, \(Q\) is modularity; \(s\) is the sum of weights of all edges in the network; \(w_{ij}\) is the weight of the edge which connects node \(i\) and \(j\); \(s_i\) is the sum of the weights of all edges connecting node \(i\); \(G_i\) is the community of node \(i\); if \(G_i\) and \(G_j\) are same, the value of \(\delta(G_i, G_j)\) is equal to 1, otherwise it is 0.

The modularity can be seen as the difference in connection strength between the actual community and randomly selected community. The value of modularity is between 0 and 1. The higher the modularity is, the better the result of community division is. The maximum modularity of actual network is generally between 0.3 and 0.7.

In this research, the weighted network is generated from the transfer probability matrix in section 3.2, and the tourism circle of Hangzhou tourist is obtained by Girvan Newman algorithm. The results of tourism circle in 2012 and 2013 are shown in figure 3 in which destinations in same tourism circle are represented by the same color. The modularity \(Q\) of tourism circle results in 2012 and 2013 are 0.623 and 0.579 which means the results are good.

Figure 3: tourism circle discovery in 2012(left) and 2013(right)
As shown in figure, there are both eight tourism circles for the Hangzhou tourists in Jiangsu-Zhejiang-Shanghai-Anhui area in 2012 and 2013. And the tourism circles are mostly located around the main transportation corridors, especially railways and expressways. That also reflects the important influence of transportation on leisure travels. In the identification results of tourism circle in 2012 and 2013, the spatial coverage of most tourism circles are nearly same, which indicates that the multi-destination travel behavior of tourist is relative stable.

It is noteworthy that there are two tourism circles expanded significantly in 2013, after the operation of Nanjing-Hangzhou high speed rail. The Hefei tourism circle (medium-purple in figure 3) covers more prefecture-level cities in Anhui Province, including Luan, Anqing, Chaohui, Huaibei and Changxing. The Nanjing tourism circle (royal-blue in the figure 3) absorbed the counties along the Nanjing-Hangzhou high speed rail, such as Yixing and Ma’anshan, and extended northward to Bengbu. Among these new destinations, Luan connects Hefei by the Hefei-Wuhan high speed rail, and Yixing and Bengbu is located along the Nanjing-Hangzhou railway and Beijing-Shanghai railway. As the operation of the Nanjing-Hangzhou high speed rail significantly reduces the travel time from Hangzhou to these cities, the tourism distribution ability of the corresponding transportation center of these cities are also enhanced. It proves that the operation of high speed rail has a great impact on the spatiotemporal distribution of tourist.

Table 2: Transfer probability of Hefei and Nanjing

<table>
<thead>
<tr>
<th></th>
<th>Transfer probability in 2012</th>
<th>Transfer probability in 2013</th>
<th>Change rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>All paths connected with Hefei</td>
<td>0.73%</td>
<td>1.44%</td>
<td>+97.53%</td>
</tr>
<tr>
<td>The path connected Hefei and origin</td>
<td>0.23%</td>
<td>0.26%</td>
<td>+14.84%</td>
</tr>
<tr>
<td>The paths connected with Hefei and other cities except origin</td>
<td>0.50%</td>
<td>1.18%</td>
<td>+134.90%</td>
</tr>
<tr>
<td>All paths connected with Nanjing</td>
<td>3.72%</td>
<td>4.87%</td>
<td>+30.98%</td>
</tr>
<tr>
<td>The path connected Nanjing and origin</td>
<td>1.34%</td>
<td>2.20%</td>
<td>+64.63%</td>
</tr>
<tr>
<td>The paths connected with Nanjing and other cities except origin</td>
<td>2.38%</td>
<td>2.67%</td>
<td>+12.07%</td>
</tr>
</tbody>
</table>

From the calculation results of transfer probability in section 3.2, the mechanism of the influence of high speed rail on Hefei and Nanjing tourism circles is further analysed. As shown in table 2, after the operation of railway in 2013, the total transfer probabilities of Nanjing and Hefei both increase compared with 2012, with growth rate of 30.98% and 97.53%. While, when it comes to the reasons for the growth of transfer probability, the mechanism of these two cities is different. The increase of transfer probability in Nanjing mainly comes from the direct transfer between Nanjing and Hangzhou. The increase of direct transfer probability reaches 64.63%. That means the tourism resources of Nanjing are highly attractive to Hangzhou tourists. More and more Hangzhou tourists would like to visit Nanjing by high speed rail. The increase of transfer probability in Hefei mainly comes from the transfer of other destinations except Hangzhou. The increase of transfer probability with other destinations is as high as 134.9%. That means the tourism resources of Hefei don’t appeal to Hangzhou tourists yet. However, because the high speed rail compresses the travel time between Hangzhou and Hefei, Hefei is more likely to become a destination for tourists after visiting their most desirable scenic spot. It also reveals that high-speed rail not only promotes the cities along the route with abundant tourism resources, but also significantly enhances the distribution ability of transit cities along the route.

4 CONCLUSION AND DISCUSSION

As an important development of infrastructure in recent years, high speed rail has great significance in promoting the development of related industries. Quantifying the impact of high speed rail on the tourism development can help relevant government departments manage and plan this type
of infrastructure. This research confirms that the operation of high speed rail is an important factor in tourist spatiotemporal behavior from the prospect of big data with a case study in Jiangsu-Zhejiang-Shanghai-Anhui area. The results are as follows.

The tourism circle, which reflects the frequent travel of tourists, basically expands outward along the main transportation corridors. As an important component of regional transportation network, high-speed rail can bring significant spatiotemporal compression effect, promote the reconstruction of regional tourism circle along the line, enhance the integration ability of surrounding tourism resources, and boost the expansion of the tourism circle centered on the cities along the railway.

From the view of transfer probability, this research quantitatively analyses the mechanism of tourism development of the cities along the railway. For cities with abundant tourism resources, the increase of transfer probability mainly comes from the direct attraction of tourism resources to tourists from origin. However, for the cities along the route which are not rich in tourism resources, the improvement of the transfer probability mainly comes from the enhancement of the city's status as a transit station. It also reveals the transfer probability is an effective method to quantify the impact of high speed rail on spatiotemporal behavior.

The methodology of this research can be used to analyze the spatiotemporal behavior of tourists through location big data. By destination recognition, transfer probability representation, and community discovery, the microscopic location data can be used to analyze the macro-distribution, which has a certain reference value for related research.

5 ACKNOWLEDGEMENT

We acknowledge the financial support of the National Natural Science Foundation of China [Grant number: 51778340].

REFERENCES


