Computers, Environment and Urban Systems 65 (2017) 66–78

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

# Extraction and analysis of city's tourism districts based on social media data

Hu Shao [a], Yi Zhang [b,*], Wenwen Li [a]

[a] School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ 86287-5302, USA
[b] Institution of Remote Sensing and Geographical Information Systems, School of Earth and Space Science, Peking University, Beijing, China

ARTICLE INFO

*Article history:*
Received 14 April 2015
Received in revised form 24 April 2017
Accepted 24 April 2017
Available online xxxx

*Keywords:*
Tourism district
Social media
Spatial-temporal behavior

ABSTRACT

Through the perspective of tourism, a city as a tourist destination usually consists of multiple tourist attractions such as natural or cultural scenic spots. These attractions scatter in city spaces following some specific forms: clustered in some regions and dispersed in others. It is known that users organize their tours in a city not only according to the distance between different attractions but also according to other factors such as time constraints, expenses, interests, and the similarities between different attractions. Hence, users' travel tours can help us gain a better understanding about the relationships among different attractions at the city scale. In this paper, a methodological framework is developed to detect tourists' spatial-temporal behaviors from social media data, and then such information is used to extract and analyze city's tourism districts. We believe that this city space division will make significant contributions to the fields of urban planning, tourism facility providing, and scenery area constructing. A typical tourism city in China—Huangshan—is selected as our study area for experiments.



## 1. Introduction

Studying the spatial structure of city destinations has long been considered an important topic of tourism study (Pearce, 1998) and the key content of urban tourism planning (Dredge, 1999). According to the tourism-spatial system theory (Gunn and Var, 2002), *Tourism Destination* (as a city) consists of *Tourism Attractions* and *Related Infrastructures*. *Tourism Attractions* are the main inducement of tourist visits, which refers to physical or cultural features of a particular place that individual travelers or tourists perceive as capable of meeting one or more of their specific leisure-related needs. Such features may be ambient in nature (e.g., climate, culture, vegetation, or scenery) or they may be specific to a location (e.g., theater performance, a museum, or ceremony events). *Related Infrastructures* supply a variety of services to tourists, including transportation, accommodation, dining, entertainment, and so forth. Tourism attractions together with its surrounding-related infrastructures are called *Tourism District* (Dredge, 1999; Pearce, 2001), which can be viewed as tourists' main activity area.

As we know, tourists' entire process of traveling usually contains continuous several days. In most situations, each day's route starts and ends with dwelling places (Shoval, McKercher, Ng, & Birenboim, 2011), while the intermediary nodes usually consist of restaurants, tourist sites, gas stations, shops, and so forth to meet their essential and entertainment needs (David A. Fennell, 1996; Bob McKercher, Wong, & Lau, 2006; Rebollo and Baidal, 2003). What is more, restraints such as time and transportation (Lau and McKercher, 2006; Page and Connell, 2014) together with tourists' individual preferences (Bob McKercher et al., 2006) and limitations of knowledge to tourist destinations will also affect their daily tour schedule. All these factors together make tourists' daily travel route relatively comprehensive and concentrated. According to the definition, a tourism district is a concentrated area in city space consisting of multiple nodes, which together are "likely to be able to fulfill a variety of tourist needs and expectations" (Dredge, 1999; Pearce, 2001). Consequently, we believe that tourists' daily travel route should mainly overlap with a single tourism district and can help recognizing the tourism districts in city space.

The lack of appropriate tourist activity data is the main restriction of tourism study (Pearce, 1979, 2001; Lew and McKercher, 2002). However, after entering the digital age (Shoval and Isaacson, 2007), the popularity of location-aware devices and the development of social network service (SNS) have made it possible to access user-centered individual spatiotemporal behaviors and their context information, which are long time series, massive, and highly precise. This information could help geographers understand users' spatiotemporal behavior patterns inside urban space and their interactions with the urban environment, which may be able to fill the gap of data unavailability for the tourism study (Cranshaw, Schwartz, Hong, & Sadeh, 2012; Wood et al., 2013; Liu, Liu, Gao, Gong, Kang, Zhi, Chi & Shi, 2015).

* Corresponding author.
*E-mail addresses:* hu.shao@asu.edu (H. Shao), zy@pku.edu.cn (Y. Zhang), wenwen@asu.edu (W. Li).

http://dx.doi.org/10.1016/j.compenvurbsys.2017.04.010
0198-9715/

In this study, we introduce a novel data-driven method to extract and analyze a city's tourism district structure through tourists' aggregate travel behaviors. The Huangshan City in China is selected as our study area because of its popularity in the tourism market. The rest of this paper is organized as follows: Section 2 reviews related work about tourists' spatial-temporal behavior study and social media study. Section 3 provides the detailed explanation of the data and methods used in this paper. Section 4 provides the description of how our methodologies are applied in the study area of Huangshan City. Section 5 discusses about our work and what we need to do in the future.

## 2. Related works

Pearce summarized in his article that the spatial analysis on urban tourism can be conducted at three scales to develop a comprehensive picture of urban tourism (Pearce, 2001).

On the level of *City Space*, a city is treated as the "focus or unit of analysis" (Pearce, 2001), usually socioeconomic data and aggregate data have been used to study some specific tourism aspects of a single city (Lew, 1992; Shoval et al., 2011; Baležentis et al., 2012; Pons, Salamanca, & Murray, 2014) or the relationships among multiple cities like tourism flow (Oppermann, 1994, 1995; Yan, 2004; White and White, 2007; Xing-zhu and Qun, 2014; McKercher et al., 2006).

On the level of *Tourist Sites*, research of tourists' spatiotemporal activity within or among different tourist sites in city space is relatively easy to be conducted because of its small scale (Shoval and Isaacson, 2007). Data collecting methods such as questionnaire surveys, camera records, global positioning system (GPS), and land-based tracking systems have all been used to gather the information for analyzing, explaining, and simulating users' movement pattern (Hartmann, 1988; Keul and Kuhberger, 1997; Itami et al., 2003; O′Connor et al., 2005; Lau and McKercher, 2006; Lew and McKercher, 2006; Edwards and Griffin, 2013).

On the level of *Tourism Districts*, tourism does not occur uniformly; instead, it is concentrated in particular areas of urban space. Detailed analyses of urban tourism need to focus into the substructure to develop a comprehensive understanding of patterns, processes, and interrelationships of different parts of a city. This study falls in the scope of tourism district partitioning (Pearce, 2001). Several qualitative studies on this level have been conducted (Teo and Huang, 1995; Savage, Huang, & Chang, 2004; Pearce, 1998) to recognize and explain tourism districts according to their functions. Studies on the scale of tourism district, which is also the scale of our work, can help us gain profound insight on cities' subtourism-region functionality and hopefully bridge the gap between studies of city space and tourist sites. However, the difficulty lies in that tourist districts study requires tourists' spatiotemporal activity data to be both detailed and abundant. Unfortunately, limitations of conventional data acquisition methods such as telephone surveys or onsite investigations on tourist spatiotemporal behaviors can be time and energy consuming (Shoval and Isaacson, 2007). This has restricted the study of urban space at a regional scale. After we entered the digital age, location acquisition techniques have developed rapidly and generated voluminous yet cheap and easy-to-acquire data (Lu and Liu, 2012) such as cellular phone records, GPS data, consumption records, SNS data, and VGI (Volunteered Geographic Information) (Goodchild, 2007). These data depict humans' spatiotemporal behavior in detail, providing us great opportunities to accelerate tourism study.

First, the digital data utilities have been tested or verified in several studies on tourist behaviors. Cheng, Caverlee, Lee, and Sui (2011) explored a lot of potential applications of social media to depict users' spatial-temporal activities: through a huge volume of geo-tagged twitter data, they calculated individuals' patterns of moving length, radius of gyration, detected individuals' home locations, and discussed the factors which might affect users' mobility. Wood, Guerry, Silver, and Lacayo (2013) used Flickr data as their data source to approximate visitation rates of 836 recreational sites around the world, after comparing with official statistics from each site, they concluded that "the crowd-sourced information can indeed serve as a reliable proxy for empirical visitation rates." Ahas, Aasa, Roose, Mark, and Silm (2008) compared the passive mobile positioning data from Estonia with conventional accommodation statistics from the same time; they found that the correlation of these two datasets reached 0.99, proving that mobile positioning data have high precision for depicting users' aggregate spatiotemporal behaviors. Hawelka et al. (2014) performed a similar work on international travelers using twitter data; they validated "geo-located twitter as a proxy of global mobility behavior to a certain extent."

Several works studying city structure and tourist behaviors have also been conducted using these digital data in recent years. By applying cluster analysis on mobile positioning data, Asakura and Iryo (2007) found some topological characteristics of tourists' behavior. Donaire, Camprubí, and Galí (2014) detected different types of photographers' according to their travel photography from Flickr data. Zhai et al. (2015) revealed the popularity of restaurants in cities from social media data. Liu et al. improved the classification of land use in city space by introducing the spatial interaction pattern analysis of mobile phone users from different places. On the basis of point of interest (POI) and social media check-in data, Cranshaw et al. (2012) implemented the spatial partition of urban areas to study the "social dynamics" of cities. Using the same type of data, Yuan, Zheng, and Xie (2012) explored the main functions of a city's different regions. Hollenstein and Purves (2015) explored the "core area" and "border" of cities across the USA using user-generated Flickr data, while Hu et al. (2015) extracted urban areas of interest which attract people's attention in a city space. Yin, Cao, Han, Luo, and Huang (2011) extracted common tour trajectories of Flickr users in different cities. Hot spots of tourism destination inside the city space have also been detected through social media data (Crandall, Backstrom, Huttenlocher, & Kleinberg, 2009; Liu, Sui, Kang & Gao, 2014; García-Palomares, Gutiérrez, & Mínguez, 2015; Zhou, Xu, & Kimmons, 2015).

Differing from the related works, our methodology takes users' geo-weibo sequences and their spatial-temporal characters instead of single pieces of geo-weibos for tourism districts detection. More information will be revealed from our results.

## 3. Data and methods

### 3.1. Study area

In this paper, we selected Huangshan City as our study area, which is a typical tourism city belonging to Anhui Province in the Southeast of China (Fig. 1, a). The main income resource of Huangshan City is its tourism product. Huangshan City's agreeable weather, four distinct seasons, and innumerable nature and culture landscapes help it attract 30 million tourists every year, which is almost 20 times the local population (1.47 million, 2013).[1]

Huangshan City is named by its famous scenic spot Mountain Huangshan (Mt. Huangshan), or literally translated as Yellow Mountain, and ranks among the Great Wall and the famous Terracotta Army as one of China's most luring tourist attractions (Fig. 1, b). In the northwest of Xiuning County, there lies another mountain named Qiyun (literally "Cloud-High Mountain"). Not as famous as Mt. Huangshan, Mt. Qiyuan mainly attracts visitors from the local area. The largest man-made lake in Anhui province, Taiping Lake, is also located north of Huangshan City.

[1] http://ah.anhuinews.com/system/2012/01/13/004705855.shtml
http://www.newshs.com/a/20140329/00104.htm
http://whc.unesco.org/en/list/547
http://www.huangshantour.com/english/
http://www.china-mike.com/china-tourist-attractions/huangshan/
http://www.chinahighlights.com/huangshan/tours.htm