Applied Geography 73 (2016) 26–37

Contents lists available at ScienceDirect

Applied Geography

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

Characterizing geographical preferences of international tourists and the local influential factors in China using geo-tagged photos on social media

Shiliang Su [a, b, c, d], Chen Wan [a], Yixuan Hu [a], Zhongliang Cai [a, c, d, *]

[a] School of Resource and Environmental Sciences, Wuhan University, Wuhan, China
[b] Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan, China
[c] Key Laboratory of Geographical Information Systems, Ministry of Education, Wuhan University, Wuhan, China
[d] Key Laboratory of Digital Mapping and Land Information Application Engineering, National Administration of Surveying, Mapping and Geoinformation, Wuhan University, Wuhan, China

ARTICLE INFO

Article history:
Received 30 January 2016
Received in revised form
6 May 2016
Accepted 16 June 2016
Available online 23 June 2016

Keywords:
Geo-tagged photos
International tourism
Destination preferences
Place attractiveness
Photo sharing
Flickr
Spatial analysis
Big data

ABSTRACT

Understanding the geographical preferences of international tourists is critical for the tourism planning and marketing. However, it is not an easy endeavor to gather the corresponding information, given the absence of city-level tourism statistical data and high costs of participant survey. This paper characterizes the geographical preferences of international tourists using geo-tagged photos on social media (the Flickr in particular). Data are harvested for 333 prefecture-level cities in China from 2008 to 2013, and the intensity of photo sharing (IPS: the ratio between total number of uploaded photos and total area of the city/region) is used an indicator of tourist geographical preferences. IPS visualization shows the geography that tourism hotpots generally concentrate in regional capital cities and economically developed megaregions. More specifically, the East Asia and Oceania tourists exhibit more preferences towards the eastern coastal cities, while the Europe and North America visitors show increasing interest in exploring the western and northern places. Spatial regression is employed to quantify the local influential factors of tourists' geographical preferences. It is found that international tourists usually consider the local economy, accessibility, infrastructure and cultural attractions when they choose their destinations in China. East Asia and Oceania visitors particularly appreciate the local economy and cultural attractions. Europe and North America tourists especially value the cultural attractions and local openness. The demonstrated methodological framework is not restricted to Flickr data, and it can be applicable to social media offer geo-tagging service. This paper is therefore believed to advance the applications of social media into geographical research.

1. Introduction

Today, millions of people travel away from their own country to explore and experience the outside world (World Travel and Tourism Council, 2012). It is expected that the total number of international tourists will reach 1.6 billion worldwide by 2020 and the tourism revenues will exceed 2 trillion US dollar (World Travel and Tourism Council, 2012). Tourism without doubt becomes the most profitable industry in the world (Candela & Figini, 2012). Developing tourism has been given high priorities in government's agenda, as it is regarded as an effective means to promote economic growth (Li, Wu, & Cai, 2008). Hence, understanding the geographical preferences of tourists is critical for the national tourism administration and local travel agencies to formulate themed planning and deploy promotion strategy. However, it is not an easy endeavor to gather the corresponding information given the absence of city-level tourism statistical data, especially in developing countries. Planners, authorities and researchers cannot therefore effectively gain the knowledge of travelers' geographical preferences at national level. Some studies attempt to collect the information through participant survey, but the small sample size and biased interviewee choices greatly impede the analysis at broad level (Xu & Zhang, 2015). Under the circumstances, some

* Corresponding author. No. 129 Luoyu Rd, Wuhan, Hubei Province, China.
E-mail addresses: shiliangsu@163.com, shiliangsu@whu.edu.cn (S. Su), zlcai@whu.edu.cn (Z. Cai).

http://dx.doi.org/10.1016/j.apgeog.2016.06.001
0143-6228/

alternative data sources and more sophisticated techniques should be taken advantage to characterize geographical preferences of international tourists at broad level.

Thanks to the advent of Web 2.0 and location aware technologies, social media enables people to add extra timestamps and geocoordinates as tags, making uploaded files more easily to be shared, searched, indexed and interpreted based on the time-space information (Kou, Hou, Yang, & Gong, 2015; Luo, Joshi, Yu, & Gallagher, 2011; Majid, Chen, Mirza, Hussain, & Chen, 2015). Following this tendency, most popular photo sharing websites (e.g., Flickr, Panoramio, Instagram, Imgur and Picasa) provide the geo-tagging interface, either extracting the geocoordinates automatically based on photos' EXIF, or letting the users themselves provide the location information (Zhou, Xu, & Kimmons, 2015). For example, over 40 million photos on Flickr and more than 4 million photos on Panoramio are geo-tagged. These openly accessible photos not only contain a wealth of geographic information, but also convey the emotions and perspectives of people with various demographic backgrounds (Li & Zhang, 2011). The geo-tagged photos on social networks therefore provide new opportunities for geographical science of different applications (García-Palomares, Gutierrez, & Mínguez, 2015; Paldino, Bojic, Sobolevsky, Ratti, & González, 2015). Scholars have attempted to employ the big data of geo-tagged photos on social media into the tourism geographical issues (Brabyn & Mark, 2011). Some studies adopt the information of the geo-tagged photos for personalized location or routing recommendation (Hu et al., 2015; Jiang, Yin, Wang, & Yu, 2013; Memon et al., 2015; Sun, Fan, Bakillah, & Zipf, 2015; Zhou et al., 2015). Other cases map the photo sharing intensity to identify the hotspots within cities (García-Palomares et al., 2015). Paldino et al. (2015) measured the attractiveness of ten global cities using the geo-tagged photographs intensity.

Previous literature evidences the promising applicability of geo-tagged photos into geographical research. However, few studied have characterized the preferences of international tourists for temporal and cross-regional comparisons using geo-tagged photos. Besides, the local influential factors of IPS remain poorly understood. More efforts are required to be made in these unsolved issues. Geographers have good theoretical reasons to believe the feasibility of geo-tagged photos to indicate tourists' preferences, since prior psychological studies conceptualize geo-tagging as the same means of self-impression on social media (Gilbert & Barton, 2013; Kim & Sundar, 2011). In other words, people are prone to share the photos that are preferred by their own. Ames and Naaman (2007) reported that two essential motivations drive the photo sharing on Flickr: functional and social. The geo-tagged places are always personally satisfying, and people upload the information not only for their own memory of the enjoyable destinations but also to gain popularity and build reputation. Nov, Naaman, and Ye (2009) identified that enjoyment is one of the most principle motivations to share the geo-tagged photos. Lindqvist, Cranshaw, Wiese, Hong, and Zimmerman (2011) found that people shared their location via Foursquare with the purpose of expressing their enjoyment of the places and exploring new destinations. It was pointed that check-in intensity was highly correlated with users' personality traits on social media (Wang & Stefanone, 2013; Wang, 2013). It is feasible to assume that the places preferred by the tourists are those that are densely geo-photographed (Girardin, Vaccari, Gerber, Biderman, & Ratti, 2009; Majid et al., 2015; Paldino et al., 2015). Thus, the intensity of photo sharing (IPS: the ratio between total number of uploaded photos and total area of the city/region) can be used to indicate and compare tourist preferences across cities (Paldino et al., 2015).

China is an ideal destination for international tourists given the rich and diverse tourism resources, including numerous historical sites and scenic spots, colorful folk customs and long cultural traditions. Following France, Spain and USA, China ranks the fourth place in global tourism industry (World Travel and Tourism Council, 2012). International tourist arrivals have kept an increasing trend since the middle 1990s. China's national development plan recognizes tourism industry as an essential contributor to economic growth (Yang, Lin, & Han, 2010). The central government calls for improving service facilities and infrastructure to expand the international tourism industry. Given the vast territory and various types of travel spots (Fig. 1), in practice, the authorities are faced with two key questions when establishing tourism promotion strategy: (1) what are the geographical preferences of international tourists? (2) what are the local influential factors for travelers' choice of destinations? However, it is rather difficult to have a whole knowledge of the two issues due to the absence of city level statistical data. The case of China therefore provides a useful example to demonstrate the applicability of geo-tagged photos to characterize the geographical preferences of international tourists and the local influential factors.

Against the above backdrop, this paper harvests the geo-tagged photos from Flickr and applies them into the case of China. We specifically attempt to: (1) characterize the geographical preferences of international tourists by visualizing IPS; (2) analyze how the geographical preferences vary with time and with tourists' origins; and (3) quantify the local influential factors of tourists' destination preferences across time and space.

2. Data and methods

2.1. Data and processing

Instagram, Flickr and Panoramio are the major social media platforms that offer the option of geo-tagging photos. Instagram and Panoramio are completely forbidden and inaccessible to users in China. Flickr is therefore used to conduct the analysis. We use the Flickr's public API (https://www.flickr.com/services/api/) to retrieve the data (including Flickr photos and corresponding metadata) for the entire China from Jan 2008 to Dec 2013. Flickr photos are in standardized JPG format. The metadata records a wealth of information, including photo title, photo ID, taken date, geo-coordinates (latitude and longitude), textual tags, host server ID, owner ID, and some owners' personal information (e.g., the origin). Only the photos whose metadata records location information are retrieved. Only one photo is kept when a set of photos is uploaded by one user at the same place. As tourism can be seen as a short-term mobility, we refer to the 'home location' concept to select the photos taken by international tourists. Bojic, Massaro, Belyi, Sobolevsky, and Ratti (2015) summarized five approaches to home location inferring in the literature: (1) maximal number of photos; (2) maximal number of active days; (3) maximal timespan of photo sharing; (4) maximal number of photos during nighttime; and (5) maximal number of active days during nighttime. Since this paper focuses on the international tourists in China, the 'home location' should be a 'temporary' concept for the Flickr users from foreign countries. We define it as: (1) maximal number of active days do not exceed one month (García-Palomares et al., 2015); and (2) timespan does not exceed three months within one year. When the samples meet the two criteria at the same time, the Flickr users whose origins are foreign countries are regarded as the international tourists in China. In addition, previous studies report that the geo-tagging information can be problematic and even wrong when people are in cities that they are not familiar with (Hecht, Hong, Suh, & Chi, 2011). This problem should be insignificant since we just analyze the IPS for certain cities rather than intra-urban geo-tagging patterns. In total, we obtain a volume of 434,470 data