Computers, Environment and Urban Systems 71 (2018) 88–97

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

Spatial distribution pattern of the customer count and satisfaction of commercial facilities based on social network review data in Beijing, China

Teng Wang[a], Yandong Wang[a,b,*], Xiaoming Zhao[a], Xiaokang Fu[a]

[a] State key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, PR China
[b] Collaborative Innovation Center for Geospatial Technology, Wuhan, PR China

ARTICLE INFO

Keywords:
Network-constrained
Spatial distribution pattern
Social network review data
Commercial facilities
Customer count
Customer satisfaction

ABSTRACT

In cities, commercial facilities play a very important role in economic growth and urban development. Current studies often discuss the spatial distribution of commercial facilities. The spatial distribution and relevant influencing factors of the customer count and satisfaction of commercial facilities, however, has rarely been considered. In this paper, a Weighted Network-constrained Kernel Density Estimation is applied to social network review data to analyze the spatial distribution of customer count and satisfaction of commercial facilities. We found differences in the spatial distribution of customer count, satisfaction, and the location of commercial facilities. To analyze these spatial differences, we present a new method for quantitative analysis using the Network-constrained Local Getis-Ord's General G* as an indicator. Road segments with high-value spatial clustering or low-value spatial clustering were detected, reflecting the spatial distribution pattern of the customer count and satisfaction of commercial facilities. The Network-constrained K-Function was used to explore the spatial clustering pattern of commercial facilities as well as the correlation between the spatial distribution of commercial facilities and other POI data, such as subway stations or business centers. The results of these analyses provide a quantitative reference when deciding locations for commercial facilities, and can help us to identify problems in commercial facility services to improve the quality of life among urban residents.

1. Introduction

In cities, commercial facilities play a very important role in economic growth and urban development. A reasonable layout of commercial facilities is conducive to the economic development of cities, and affects the quality of life of urban residents. Understanding the spatial distribution pattern of urban commercial facilities makes possible the rational allocation of urban resources and commercial facilities, and thus aids the healthy development of the urban economy.

Numerous studies have examined the spatial distribution of urban commercial facilities. Rui, Huang, Lu, Wang, and Wang (2016) applied a geographically weighted Poisson regression model to explore the influence of demographic, economic, and geographic factors on the spatial distributions of KFC and McDonald's outlets in China. With the aim of considering spatial nonstationarity in the model parameters, Suárez-Vega et al. (2014) used a locally calibrated Huff model to analyze the location of retail sites and their implications on the consumer behavior. Roig-Tierno, Baviera-Puig, Buitrago-Vera, and Mas-Verdu (2013) analyzed the commercial distribution sector of frequently purchased products in Murcia (Spain), and a methodology that combines Geographic Information Systems (GIS) and the Analytical Hierarchy Process (AHP) was proposed for the process of selecting a retail site location. Based on the data set of spatial coordinates of seven facilities in 37 major cities, Wu, Chen, and Zhao (2014) used the descriptive statistical analysis to explore the characteristics of the distribution of urban facilities.

In fact, commercial facilities associated with human activities in real life are constrained by a road network. Okabe, Satoh, Furuta, Suzuki, and Okano (2007) show an empirical finding that Euclidean distance is significantly different from the shortest path distance in an urbanized area if the distance is < 500 m. Therefore, the application of the traditional pattern analysis methods for these network-constrained point events will cause false cluster patterns (Lu & Chen, 2007).

In recent years, a large number of scholars have explored and developed spatial point pattern analysis to analyze spatial point events constrained by road networks. These include Network-constrained Kernel Density Estimation (Borruso, 2008; Tang et al., 2015; Yu, Ai, & Shao, 2015), Network-constrained Voronoi Diagram (Tan, Zhao, & Wang, 2012), Network-constrained K-Function (Lamb, Downs & Lee, 2016; Xu & Peng, 2011; Yamada & Thill, 2004), and Network-constrained Cluster Analysis (Mahrsi & Rossi, 2013; Steenberghen, TD, &

* Corresponding author at: State key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, PR China.
E-mail address: ydwang@whu.edu.cn (Y. Wang).

https://doi.org/10.1016/j.compenvurbsys.2018.04.005
Received 9 June 2017; Received in revised form 25 March 2018; Accepted 27 April 2018
0198-9715/

Flahaut, 2004; Yamada & Thill, 2007). Okabe, Satoh, and Sugihara (2009) reveal that using the kernel density estimation to estimate the density distribution of the network-constrained spatial point events may lead to biased conclusions; a network-constrained Kernel Density Estimation method was proposed and applied to analyze traffic accidents. A novel integrated access measure, obtained by means of both Network Kernel Density Estimation and Network Linear Density Estimation, was proposed to compute accessibility to Points of Interest (POIs) constrained by road networks (Li, Zhang, Wang, & Zeng, 2011). Ai, Yu, and He (2015) proposed a constrained network Voronoi Diagram using stream flowing ideas to estimate the service range of the facility in real-time with consideration of traffic restrictions. A new approach, a weighted network Voronoi Diagram, was proposed to model spatial patterns of geographic events on street networks whose street segments can be weighted based on their roles in the events, and the weights were calculated using the Kernel Density Estimation and the local Moran's I methods (She et al., 2015). The similarities and differences was compared in results using planar and network K-functions to analyze the localization of tertiary firms in the CBD of metropolitan areas, finding that the network K-function is a more appropriate method for measuring agglomeration patterns, scale, and intensity at the intra-urban level (Garrocho-Rangel, Álvarez-Lobato, & Chávez, 2013). Taking into account the type of commercial facilities and the grade of road network, Rui et al. (2015) analyzed the spatial distribution characteristics of Nanjing Suguo Supermarket by using the network-constrained Kernel Density Estimation and the network-constrained K-function, exposing the competitive relationship between the Suguo Supermarket and Foreign Supermarkets.

However, at present, most of the research on spatial analysis of commercial facilities focuses on the spatial distribution of commercial facilities. Few studies have explored the spatial distribution and relevant influencing factors of the customer count and satisfaction of commercial facilities, which may be more important for the rational allocation of commercial facilities and management of the service level of the commercial facilities. In this paper, based on review data collected from social networks, Weighted Network-constrained Kernel Density Estimation is used to explore the differences in the spatial distribution between the customer count and the customer satisfaction of commercial facilities. In order to analyze the spatial difference, we present a quantitative analysis method using the Network-constrained Local Getis-Ord's General G* as an indicator. The road segments with high-value spatial clustering or low-value spatial clustering is detected, which reflect the spatial distribution pattern of the customer count and satisfaction of commercial facilities. On the other hand, the Network-constrained K-Function is used to explore the spatial cluster pattern of commercial facilities and the correlation between the spatial distribution of commercial facilities and other POI data, such as subway stations or business centers. The results of these analyses can provide a quantitative reference for the rational allocation of urban resources and the location of commercial facilities.

The remainder of this paper is structured as following: Section 2 introduces the methods used in this article. Section 3 describes our study area and data. Section 4 analyzes the experimental results. We then conclude our discussion with a summary of the findings and directions for further works.

2. Network-constrained point pattern analysis methods

At present, Network-constrained Kernel Density Estimation and Network-constrained K-Function are two of the most widely used methods to analyze spatial point events constrained by road networks. In addition, we expand the traditional planar Local Getis-Ord's General G* to analyze spatial autocorrelation of the network-constrained point events. These methods will be described in detail in the following paragraphs.

2.1. Network-constrained kernel density estimation

The kernel density estimation (KDE) is a non-parametric way to estimate the probability density function of a random variable. Taking into account the constraints of the urban road network, the network-constrained kernel density estimation (N-KDE) is a variant for road network space. The N-KDE uses the kernel function and bandwidth to compute the density function distribution of point events on the network. The estimated function is shown in formula (1):

$$\lambda(s) = \sum_{i=1}^{n} \frac{1}{r} k\left(\frac{d_{is}}{r}\right) \quad (1)$$

where$\lambda(s)$is the density estimate at location$s$,$r$is the bandwidth, and$d_{is}$is the shortest path distance from the estimation point to the observation point event marked as$i$, and$k()$is a kernel function of the ratio between is$d_{is}$and$r$to measure the distance decay effect.

The bandwidth (r) and kernel function ($k()$) are two required parameters. When the bandwidth is large, the KDE is smoother and details will be ignored. When the bandwidth is small, the KDE is more abrupt and difficult to reflect the overall trend. Some studies have shown that 100 m to 300 m are ideal bandwidth values when analyzing commercial facilities related to urban economic activity (Porta, Latora, Wang, et al., 2009; Porter & Reich, 2012). Because 100 m to 300 m is close to the distance of a typical city block, which is an acceptable range for pedestrian walking in a block (Okabe et al., 2009; Rui et al., 2015). Therefore, in this paper, we choose 300 m as the N-KDE bandwidth.

The effect of the choice of kernel function is less than the effect of the choice of the bandwidth (Bíl, Andrášik, & Janoška, 2013). Common kernel functions include Gaussian, Quartic, Epanichnekov and so on. In this paper, we use the Gaussian kernel function, as shown in formula (2):

$$k\left(\frac{d_{is}}{r}\right) = \begin{cases} \sqrt{2\pi}\exp\left(-\frac{d_{is}^2}{2r^2}\right) & 0 < d_{is} < r \\ 0 & d_{is} > r \end{cases} \quad (2)$$

Based on the traditional planar KDE, the N-KDE extended measurement of distance between two points from the Euclidean distance to network distance. In addition, the N-KDE divide the road network into a fixed length as the smallest unit of study.

In this paper, the specific implementation steps are as follows:

(1) Each road segment of road network is divided into a fixed length (10 m (Nie, Wang, Du, Ren, & Tian, 2015)) as a basic unit of study (called lixel (Xie & Yan, 2008)).
(2) Each commercial facility is constrained to the nearest lixel, and the distance from the commercial facility to the current lixel is defined as 0. The shortest path distance from the other lixel to the current lixel is taken as the distance from the other lixel to this commercial facility.
(3) According to the bandwidth (r) and kernel function ($k()$), for each commercial facility, search for the lixel objects within 300 m from the current commercial facility, and the distribution density is calculated according to the kernel function and the distance between the lixel and the current commercial facility.
(4) All the commercial facilities are traversed in turn, and, eventually, the distribution density value of each lixel is obtained by calculation and summation.

2.2. Network-constrained local Getis-Ord's General G*

Getis-Ord's General G* is one of the most commonly used indexes of spatial autocorrelation. It is an index based on distance weight matrix, which can evaluate the spatial aggregation of study objects, such as high-value aggregation and low-value aggregation. The high-value aggregation or low-value aggregation identified can be identified from the