Computers, Environment and Urban Systems xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

Estimation of urban crowd flux based on mobile phone location data: A case study of Beijing, China

Zide Fan[a], Tao Pei[a,b,c,*], Ting Ma[a], Yunyan Du[a], Ci Song[a], Zhang Liu[a], Chenghu Zhou[a]

[a] State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, CAS, Beijing 100101, China
[b] University of Chinese Academy of Sciences, Beijing 100049, China
[c] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

ARTICLE INFO

Keywords:
Crowd flux estimation
Mobile phone location data
Trajectory reconstruction
Urban dynamics

ABSTRACT

In previous urban planning research, the fine-grained population was considered a crucial factor. However, this population, which is generated from census data, represents only the number of people who live in a region. This static figure cannot indicate the underlying number of people and their temporal variation. Therefore, any decision-making based on the static population may be separated from reality. To overcome this difficulty, in this paper, the urban crowd flux is initially proposed and defined as the number of individuals flowing into or out of a region per unit time interval. Then, the urban crowd flux is estimated using an approach of human trajectory gridding reconstruction based on mobile phone location data. This approach is divided into three steps. First, the trajectory of each individual is extracted from sparse sampling of mobile phone location data after data cleansing. Then, combining with a road network, the trajectory of each individual is reconstructed by network interpolation based on the shortest path algorithm in regular grids. Third, we use the velocities of each user's trajectory record to estimate the urban crowd flux on spatio-temporal grids at each time slice. Finally, urban crowd flux in Beijing, China were estimated using our method and the spatio-temporal characteristics of flux is analyzed.

1. Introduction

Census data are commonly viewed as the crucial factor in previous urban planning research; however, census data often simply indicate the number of people who live in a region and cannot fully account for the underlying people and also fails to reveal the dynamic changes. In the case of shop location selection, the sales of a shop rely more on the mobile population size (i.e., the number of people passing by) than the static population size in a region. For example, in a transport hub area, the static population size is often low, while the mobile population size is very high. Therefore, it is necessary to develop an index and a method for measuring the dynamic population. With the rapid development of mobile positioning, wireless communication and sensor technologies, the locations of individuals can be recorded through urban big data of location-based services (LBS) such as mobile phone location data, traffic card data, Wi-Fi location data and so on. Therefore, estimation of the fine-grained dynamic population becomes possible using urban LBS big data. In this paper, we address this problem using mobile phone location data to propose a new concept of the dynamic population and its computation method.

Before urban LBS big data became universal, the distribution of urban fine-grained populations mainly relied on census data. Population estimations were obtained by spatial downscaling analysis from large-area irregular administrative region demographic data to regular grids (Wu, Qiu, & Wang, 2005). The spatial downscaling method is divided into the simple area weighting method (Balk & Yetman, 2004; Deichmann, Balk, & Yetman, 2001) and the partition density modeling method (Balk, Deichmann, & Yetman, 2006). The main idea of these two methods is that auxiliary data are used to redistribute the demographic data within the district. Nevertheless, since census data are based on the sampling of the administrative region, the estimation accuracy of downscaling methods is usually lower, especially for fine-grained population estimation. However, with the enrichment of urban LBS big data, their use in the estimation of urban fine-grained populations has become more popular. Krings, Calabrese, and Ratti (2009) propose a linear relationship between the number of mobile phone users and the regional population size. Based on this relationship, the regional population distribution can be approximated

* Corresponding author at: State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, CAS, Beijing 100101, China
E-mail address: peit@lreis.ac.cn (T. Pei).

https://doi.org/10.1016/j.compenvurbsys.2018.01.005
Received 8 June 2017; Received in revised form 10 January 2018; Accepted 10 January 2018
0198-9715/

Please cite this article as: Zide, F., Computers, Environment and Urban Systems (2018), https://doi.org/10.1016/j.compenvurbsys.2018.01.005

by the number of mobile phone users (Kang, Liu, Ma, et al., 2012; Reades, Calabrese, & Ratti, 2009). On this basis, Deville, Linard, and Martin (2014) and Stevens, Gaughan, and Linard (2015) used random forest models to estimate the population distribution of different cities based on mobile phone location data. The main use of this type of research is fine-grained urban population distribution, but in some applications, we need to know more about the flow of people in the city rather than just the static distribution of the population, such as for effectively locating businesses and other applications.

To study human flow in a city, more studies have focused on the patterns of human flow by relying on multiple datasets, such as mobile phone location data, taxi trajectories data and social media check-in data (Gonzalez, Hidalgo, & Barabasi, 2008; Liu, Kang, Gao, et al., 2012; Sagl, Delmelle, & Delmelle, 2014), among others. Human flow patterns have been analyzed in different cities (Isaacman, Becker, & Cáceres, 2012; Secchi, Vantini, & Vitelli, 2015; Xu, Zhang, & Li, 2016) and are widely used in areas such as public health (Wesolowski, Metcalf, & Eagle, 2015), air pollution (Gariazzo, Pelliccioni, & Bolignano, 2016), and post-disaster population mobility assessment (Lu, Bengtsson, & Holme, 2012). At this stage, the estimation of the fine-grained urban crowd flow within a city is one of the hotspots in this research field. Hoang, Zheng, and Singh (2016) and Zhang, Zheng, and Qi (2016) integrated large amounts of data, including traffic, weather and events, to predict the incoming and outgoing flows for various areas in the city for the upcoming hour in order to prevent urban catastrophic incidents (such as trampling). Similarly, Hong, Zheng, Yung, et al. (2015) used the dynamic changes of urban crowd flows to identify the convergence and divergence of crowds. Yang et al. (2016) detected urban black holes and urban volcanoes using urban crowd flow data. Nevertheless, most of the existing studies were devoted to the analysis of human flow patterns in a city, and there is no clear indicator so far to define and measure the flux of urban crowds.

Although previous studies have made substantial progress, two key problems remain unresolved. The first is how to define the crowd flow. Urban crowd flow not only indicates the inflow and outflow of people per unit area in the city but also is closely related to the speed with which the crowd passes through this region. To solve this key problem, we define the concept of the urban crowd flux and then propose an approach of human trajectory gridding reconstruction based on mobile phone location data to estimate the urban crowd flux. The second problem is how to reconstruct individual trajectories using mobile phone location data of low accuracy. The accuracy of mobile phone location data is relatively low in two aspects: (1) The location point is usually the location of a tower station rather than the actual location of the mobile phone users and (2) because the sampling frequency of mobile phone location data is generally low, all the users' actual movements are often not recorded. The existing trajectory reconstruction method based on mobile phone location data mainly includes an interpolation method (Hoteit, Secci, & Sobolevsky, 2014) and a machine learning method (Qiao, Shen, & Wang, 2015; Qiao, Tang, & Jin, 2010). As the sampling of mobile phone data is sparse, the reconstruction accuracy using mobile phone location data is relatively low and unstable. To overcome this key problem, considering that the flow of people is almost always through the road network, we take the road network into account to reconstruct people's trajectories using the sparse sampling of mobile phone location data.

This paper is organized in six sections. In Section 2, the definition of urban crowd flux is proposed. In Section 3, the proposed estimation method is represented. In Section 4, the study areas and data are introduced. In Section 5, a case study of Beijing datasets is selected to demonstrate the validity of the proposed method. In Section 6, conclusions are described.

2. Definition of urban crowd flux

In this paper, a city is divided into many regular spatio-temporal grids expressed as $Grid(S_p, T_q)$, $p = 1…N$, $q = 1…M$, where S_p is the ID of the spatial grids, T_q is the ID of the time slices, N is the total number of spatial grids and M is the total number of time slices. A personal movement trajectory can be expressed as a time series of spatial sampling points expressed as $Trj_i(p_j, t_j)$, $i = 1…n$, $j = 1…m$, where p_j is each sampling point of a personal movement trajectory and t_j is the timestamp of these sampling points, n is the total number of all users and m is the total number of all sampling points for this personal trajectory. Each personal trajectory is mapped to the spatio-temporal grids to obtain the intersection. The human gridding trajectory is then defined as the time series of the spatial intersections between each personal movement trajectory and the spatial–temporal grids as $Tg_i(S_j, T_j, s_j)$, $i = 1…n$, $j = 1…m$. There are three major parameters: the spatial grid id S_j, the temporal slice T_j and the velocity of a person passing though this spatio-temporal grid, namely, s_q. The velocity s_q is an important parameter for human gridding trajectory, which can be calculated by the distance and time interval between two adjacent records. When the velocity is small, it is helpful for understanding people's local activities. Therefore, we believe when the velocities of crowd movement is slow, people are more likely to stay in this region. Karamshuk, Noulas, Scellato, et al. (2013) and Yu, Zhang, and Yang (2013) proposed that when a region has more people through and stay, this region is more suitable for commercial location.

Next, we define the concepts of urban crowd flux. The urban crowd flux is divided into two types regarding its direction:

Definition 1: in-flux, the number of people entering the target area during a given time interval within a specific speed range. For example, as people from workplaces return to their homes at night, the in-flux in people's living areas will rise. The calculation formula of in-flux can be expressed as follows:

$$Flux_{S_p,T_q}^{In} = \sum_{i=1}^{n}\sum_{j=2}^{m} \#(T_j = T_q \cap S_j \neq S_{j-1} \cap s_j \in s^*) \tag{1}$$

where $Flux_{S_p}$, $T_q{}^{In}$ is the in-flux in grid S_p at time T_q. For a personal trajectory, $\#(x)$ is a binary function. When the following conditions are met, $\#(x)$ is equal to one; otherwise, $\#(x)$ is equal to zero: (1) The time of the jth record T_j is equal to the time of in-flux T_q. (2) The grid of the jth record S_j is not equal to the grid of the previous record S_{j-1}. (3) The speed of the jth record s_j is within a given threshold range s^*.

Definition 2: out-flux, the number of people leaving the target area during a given time interval within a specific speed range. For example, as people leave their homes to go to their workplaces in the morning, the out-flux in people's living area will decline. The calculation formula of out-flux can be expressed as follows:

$$Flux_{S_p,T_q}^{Out} = \sum_{i=1}^{n}\sum_{j=1}^{m-1} \#(T_j = T_q \cap S_j \neq S_{j+1} \cap s_j \in s^*) \tag{2}$$

where $Flux_{S_p}$, $T_q{}^{Out}$ is the out-flux in grid S_p at time T_q. $\#(x)$ is a binary function. When the following conditions are met, $\#(x)$ is equal to one; otherwise, $\#(x)$ is equal to zero: (1) The time of the jth record T_j is equal to the time of out-flux T_q. (2) The grid of the jth record S_j is not equal to the following record S_{j+1}. (3) The speed of the jth point s_j is within a given threshold range s^*.

When the time interval is one hour and the side length of the spatial grid is one kilometer, the unit of the crowd flux is trips/(km^2·h). The chart of the human gridding trajectory and spatio-temporal grids in a region are shown in Fig. 1. The blue cube represents the time slices, and the black grid shows the spatial grids. In this figure, a personal trajectory is shown as a crimson curve and the green points represent the sampling points of this trajectory. The intersection of this human movement trajectory and the spatial grid is shown as the yellow cells, and the red five-pointed star is the intersection of the human movement trajectory and the spatial grid.