Transportation Research Part C 85 (2017) 249–267

Contents lists available at ScienceDirect

Transportation Research Part C

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

Open PFLOW: Creation and evaluation of an open dataset for typical people mass movement in urban areas☆

Takehiro Kashiyama[a], Yanbo Pang[a,b,*], Yoshihide Sekimoto[a]

[a] Institute of Industrial Science, University of Tokyo, Ce507, 4-6-1, Komaba, Meguro-ku, Tokyo 153-8505, Japan
[b] Department of Civil Engineering, University of Tokyo, Ce507, 4-6-1, Komaba, Meguro-ku, Tokyo 153-8505, Japan

ARTICLE INFO

Keywords:
Public Participation GIS (PPGIS)
Data visualization
Urban simulation
Open data
Attitude formation

ABSTRACT

Understanding people flow at a citywide level is critical for urban planning and commercial development. Thanks to the ubiquity of human location tracking devices, many studies on people mass movement with mobility logs have been conducted. However, high cost and severe privacy policy constraints still complicate utilization of these data in practice. There is no dataset that anyone can freely access, use, modify, and share for any purpose. To tackle this problem, we propose a novel dataset creation approach (called Open PFLOW) that continuously reports the spatiotemporal positions of all individual's in urban areas based on open data. With fully consideration of the privacy protection, each entity in our dataset does not match the actual movement of any real person, so that the dataset can be totally open to public as part of data infrastructure. Because the result is shown at a disaggregate level, users can freely modify, process, and visualize the dataset for any purpose. We evaluate the accuracy of the dataset by comparing it with commercial datasets and traffic census indicates that it has a high correlation with mesh population and link-based traffic volume.

1. Introduction

1.1. Background

Understanding people flow at a citywide level is critical for urban planning and commercial development. Recently, the emerging technologies for collecting and analysis huge amount mobility data have changed the way society understand human movement. Unlike previous methods that travel data was collected by active self-report surveys via questionnaires, the new ubiquity of human location tracking devices have enabled the collection of a large amount of passive data on human movement in low cost. Comparing to traditional active data, passively collected data shows the ability to observe large amount individuals' travel behavior in long period with finer spatial and temporal resolution. Chen et al. (2016) offers a focus review on ideas and technologies that employ passive collected data to analysis travel behavior patterns recognition and behavioral factor identification. Transportation researchers try to derive urban travel demand from Call Detail Records data and GPS data (Calabrese et al., 2011; Toole et al., 2015; Ge and Fukuda, 2016).

In Japan, the GeoSpatial Information Center, a platform for one-stop showcasing of various types of geospatial information from government and the private sector, provides a platform where everyone can find and purchase all kinds of spatial data easily. Table 1

☆ This article belongs to the Virtual Special Issue on "Agent technology".

* Corresponding author at: Department of Civil Engineering, The University of Tokyo, Ce509, 4-6-1, Komaba, Meguro-ku, Tokyo 153-8505, Japan.
E-mail addresses: ksym@iis.u-tokyo.ac.jp (T. Kashiyama), pybdtc@iis.u-tokyo.ac.jp (Y. Pang), sekimoto@iis.u-tokyo.ac.jp (Y. Sekimoto).

http://dx.doi.org/10.1016/j.trc.2017.09.016
Received 31 January 2017; Received in revised form 15 September 2017; Accepted 18 September 2017

Table 1
Variety of commercial data about people mass movement in urban areas.

Kinds of data	Contents	Data source	Coverage area	Provider organization
Traffic volume data	Hourly volume of vehicle traffic on each link	Pioneer car navigation user's GPS log	All parts of Japan	Pioneer Corporation
Vehicle travel data	Average travel speed, hourly traffic volume, and transit time on each link	Navitime application user's GPS log	All parts of Japan	Navitime Japan Co., Ltd.
Population flow data	Hourly floating population data in 250 m-mesh or 500 m-mesh or Point format	Agoop smartphone application user's GPS log	All parts of Japan	Agoop Corporation
Congestion degree data	Hourly 250 m-mesh population data	Mobile phone GPS log of NTT DoCoMo users	All parts of Japan	ZENRIN DataCom Co., Ltd.
Mobile spatial statistics	Hourly 500 m-mesh population distribution data	Call Detail Records of NTT DoCoMo users	All parts of Japan	NTT DOCOMO, Inc.

shows a variety of commercial data provided by the GeoSpatial Information Center on people mass movement in urban areas. These datasets provide large-scale hourly population distribution and traffic flow information over a long period. By utilizing these kinds of data, the dynamic population potential for store opening planning, urban design, traffic engineering, and anti-disaster planning can be determined (Song et al., 2013; Sudo et al., 2016; Ge and Fukuda, 2016).

However, these datasets have the same drawbacks in common for researchers:

- The raw data used to estimate population and vehicle volume are garnered from the location data sent from smartphones with GPS function by individuals. Thus, there is a bias against residents who do not possess a smartphone or who do not utilize these services. As a result, the sampling rate of these datasets is approximately 0.1% to 1%—much lower than the common census, which is nearly 2% of the overall population of Japan. This causes a missing-data problem in many areas, especially small cities and countryside.
- To secure users' anonymity, the data collected are processed collectively and statistically (aggregated in mesh or link unit). A popular form of anonymization is population distribution data, in which the number of people in a specific region is recorded at a specific time (Sevtsuk and Ratti, 2010). Furthermore, in some cases, the data in which mesh (or link) contains less than five individuals (or vehicles) are eliminated in order to conceal the private information. Thus, users cannot analyze people movement at a disaggregate level based on such datasets.
- In addition, the datasets are extremely expensive. For example a one month population distribution map of Tokyo costs 900,000 Japanese yen (approximately US$7627). This is a huge undertaking for any researchers or institutions.

In short, to the best of our knowledge, there is no dataset that depict urban mobility at individual level and anyone can freely access, use, modify, and share for any purpose. Thus, as the situation currently stands, studies about macroscopic people mass movement can only be conducted by persons who can afford the commercial data or who can collect raw location GPS data by themselves.

On the other hand, a surge in open data platforms has resulted in new opportunities for the scientific and technical communities: they can now take advantage of a vast amount of data provided by open government to develop urban mobility models that are useful for different decision makers. Many open datasets have already been utilized in traffic field studies. Jäppinen et al. (2013) investigated the potential effect of shared bicycles on public transport travel times based on open data provided by Helsinki Region Transport. Chen et al. (2015) proposed a novel method using heterogeneous urban open data to predict bike trip demand. Further, Caiati et al. (2016) discussed how open data could play a key role in gaining insight into urban mobility when used for transportation modeling purposes. In these studies, surveys of populations and daily activities, traffic census data and public transportation information are effectively applied to estimate people movement in urban areas. In Japan, the Person Trip survey is conducted every 10 years in large metropolitan areas by the Ministry of Land, Infrastructure, Transport and Tourism (MLIT) and results in a variety of datasets that help us to know people's travel behavior. An open version person trip survey data is also available. In general, the information from open government is comprises demographic, economic, transportation, and environment open data. One of the major challenges faced by researchers is the question of how to process these statistical data for research purposes.

In summary, although various types of data on people mass movement are available, high cost and severe privacy policy constraints still complicate utilization of these data in practice. Further, even though the open data have been shown to have considerable value in the field of traffic engineering and human ability research, the research subject and methods applied to date are still rudimentary. In this study, a people mass movement dataset that reports individuals' location change continuously and allows users to freely modify, process, and visualize people mass movement for any purpose was developed.

1.2. Research objective

This work is motivated by existing issues associated with the results of open version household travel and several other surveys on residents' daily activities and vehicle usage. We consider the high cost and privacy policy restrictions associated with using commercial data on people mass movement and develop a method for generating people flow data based on open data. Specifically, our