Computers, Environment and Urban Systems 60 (2016) 79–88

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

Understanding operation behaviors of taxicabs in cities by matrix factorization

Chaogui Kang [a,*], Kun Qin [a,b,**]

[a] School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan, Hubei 430079, China
[b] Collaborative Innovation Center of Geospatial Technology, 129 Luoyu Road, Wuhan, Hubei 430079, China

ARTICLE INFO

Article history:
Received 4 February 2016
Received in revised form 2 August 2016
Accepted 4 August 2016
Available online xxxx

Keywords:
Taxicab mobility
Travel demand-supply
Habitual operation behavior
Digital trace
Matrix factorization

ABSTRACT

Taxicabs play significant roles in public transport systems in large cities. To meet repetitive demands of daily intra-urban travels, cabdrivers acquire self-organized habitual operation behaviors in space and time, largely with assistance of their longitudinal operating experience. Recognizing those collective operation behavior patterns of taxicabs will enable us to better design and implement public transport services and urban development plan. In this paper, we systematically study patterns of the spatial supply of 6000+ taxicabs in Wuhan, China based on a monthly collection of their digital traces and the non-negative matrix factorization method. We successfully identify a set of high-level statistical features of the spatial operation behaviors of taxicabs in Wuhan, providing valuable insights to our knowledge of the demand and supply of taxicabs in (similar) large cities. First, we decouple several spatially cohesive regions with intensive internal taxicab travels (termed as *demand regions*), which intuitively reveal the well-known multi-sectored urban configuration of Wuhan. Second, by applying the non-negative matrix factorization to taxicab's longitudinal traces, we uncover remarkably self-organized operation patterns of cabdrivers in space (termed as *supply regions*) as reactions to the sectored distribution of daily travel behaviors. We find that a large proportion of cabdrivers frequently operate within single specific service area and a small proportion of taxicabs works as shifting tools between different service areas. Last, we focus on performances of taxicabs with distinct spatial operation behaviors and unveil their statistical characteristics in terms of frequency, duration and distance with passenger on board. Our work demonstrates the great potential to understand and improve urban mobility and public transport system from cabdrivers' collective intelligence.

1. Introduction

Taxicabs play significant roles in public transport systems in large cities. It has been intensively reported worldwide that intra-city travels carried by taxicabs can account to >10% of total daily travels in a large city. To meet intensive and repetitive demands of daily intra-urban travels (Huff & Hanson, 1986; Eagle & Pentland, 2009), cabdrivers might acquire self-organized habitual operation behaviors in space and time, with assistance of their longitudinal operating experience and their cognition map of urban built environment. Recognizing those collective and habitual operation behaviors of taxicabs will enable us to rationally design and implement public transport services and urban configuration (Castro, Zhang, Chen, Li, & Pan, 2013). In recent years, the prevalent deployment of vehicle-embedded GPS devices in taxi service and the easy accessibility of relevant data for the public (Farber, 2015) lead to a burst of studies on taxi/passenger mobility, and provide a promising tool for inspecting empirically the spatio-temporal patterns of movement behaviors of a massive volume of taxicabs in various city settings .

In this paper, we systematically study the spatial patterns of labor-supply behaviors of 6000+ taxicabs in Wuhan, China based on a monthly collection of their digital traces and the non-negative matrix factorization method. Our primary research questions are as follows: (1) Whether (and how) do taxicabs' habitual operation behaviors rely on the spatial distribution of intra-urban travel demands? (2) What are the differences between the performances of taxicabs with distinct spatial habitual behaviors? In solving these problems, we identify a set of high-level statistical features of the spatial operation behaviors of taxicabs in the case study city, which provide valuable insights to our knowledge of the demand and supply of taxicabs in (similar) cities. First, we decouple several spatially cohesive regions with intensive internal taxicab travels, which intutively reveals the well-known multi-sectored urban configuration of Wuhan. Second, by applying the non-negative matrix factorization to taxicab's longitudinal digital traces, we uncover that cabdrivers' operation patterns are remarkably self-organized in space as reactions to the sectored distribution of daily intra-urban travel behaviors. We find that a large proportion of

* Corresponding author.
** Correspondence to: K. Qin, Collaborative Innovation Center of Geospatial Technology, 129 Luoyu Road, Wuhan, Hubei 430079, China.
E-mail addresses: cgkang@whu.edu.cn (C. Kang), qink@whu.edu.cn (K. Qin).

http://dx.doi.org/10.1016/j.compenvurbsys.2016.08.002
0198-9715/

cabdrivers frequently operate within a single specific service area (i.e., supply region) and a small proportion of taxicabs works as shifting travel tools between (two) different supply regions. Last, we focus on performances of taxicabs with distinct spatial operation behaviors and reveal their statistical characteristics in terms of frequency, duration and distance with passenger on board.

The major contribution of our work lies in the following: (1) We unveil the relationship between longitudinal intra-urban travel demands and taxicabs' self-organized habitual behaviors; (2) We develop a matrix factorization based analytical framework to quantify, visualize and evaluate taxicabs' operation behavior patterns in space. Our work also demonstrates the great potential to understand and improve urban mobility and public transport system from cabdrivers' collective intelligence.

2. Related work

The continued development of urban computing and human mobility studies has been accumulating a rich body of evidences of the descriptive characteristics of intra-urban travel activities. Taxi driving trajectories, as a prevailing data source of related works, contain at least three major facets of human travel behaviors as being summarized as: (1) General mobility patterns of city dwellers represented by passengers' daily usages of taxicabs (Section 2.1); (2) Operation behavior patterns of cabdrivers represented by hunting, waiting and shifting for profit-maximization (Section 2.2); (3) Structural properties of the intra-urban spatial interaction network derived from taxi trip origins and destinations (Section 2.3).

2.1. Passengers' mobility patterns

Patterns of city dwellers' travel behaviors have longstanding importance in time geography, transportation and urban studies. As early as in 1970's, Hägerstrand proposed his initial space-time framework of representing individuals' travel behaviors (Hägerstrand, 1970). Since then, it has turned out to be a widely-reaching research community with concentration on individuals' travel activities in space and time. In the last decade, the rapid development of human dynamics and statistical physics further refine and enrich relevant research topics and methodologies in this field (González, Hidalgo, & Barabasi, 2008). Along with cellphone and social media data, taxicabs' digital traces have contributed substantively to our understanding of individual and collective human mobility patterns. Existing literature, in some ways, regard travel behavior patterns of taxicab passengers as a proxy of the general intra-urban human mobility patterns.

From this starting point, a set of valuable features of taxi passengers' mobility patterns has been discovered. Research focus in this strand lies in the following: (1) The spatio-temporal distributions of trip origins and destinations. An important research question is where and to what extent taxi pick-ups (i.e., trip origins) and drop-offs (i.e., trip destinations) concentrate in spatial and temporal dimensions (Chen et al., 2013; Hu, Miller, & Li, 2014b). Further investigations can involve the correlation between the temporal variations of taxi origin-destination intensity and the underlying land use characteristics at specific locations (Liu, Wang, Xiao, & Gao, 2012b) as well as the temporal variations of taxi trip volume between each single pair of originating and destinating locations (Liu, Kang, Gong, & Liu, 2016); (2) The distance decay effects of collective travel activities. Statistical tools and models for quantifying the distance decay effect of travel behaviors are thoroughly testified. However, debate yet continues between those in favor of the exponential law (Liang, Zheng, Lv, Zhu, & Xu, 2012) and those in favor of the power law (i.e., Lévy flight) (Jiang, Yin, & Zhao, 2009). To uncover underlying mechanism of observed patterns, a couple of further studies focus on the topological structures of road networks (Gallotti, Bazzani, Rambaldi, & Barthelemy, 2015) and the spatial attractions of land units (Liu, Kang, Gao, Xiao, & Tian, 2012a) that could constrain and boost city dwellers' traveling; (3) The regularity of activity-venue and travel-route selections. It implements statistical quantities, such as cyclical rhythm (Liu et al., 2012b) and spatio-temporal entropy (Peng, Jin, Wong, Shi, & Liò, 2012), to describe the high predictability of determining taxi trip origins and destinations in cities. Strong evidences exist to demonstrate that the number of frequent visitation venues and routes of individual dwellers' traveling is astonishingly small, e.g. with an average of four venues and two routes. These studies shed bright light on the repetitious daily routines and the highly limited activity budgets of city dwellers.

2.2. Cabdrivers' mobility patterns

For the sake of improving taxi service efficiency, another strand of studies heavily emphasize on cabdrivers' operation strategy for gaining optimal (e.g., minimal and maximal) profit (Liu, Andris, & Ratti, 2010). For most cabdrivers, there are three key decisions they have to make during working as follows: (1) Where to search for the next passenger by cruising? (i.e., hunting behavior). Previous literature have reported several well-received searching strategies based on hotspot detection (Chang, Tai, & Hsu, 2009), cabdrivers' historical performances (Li et al., 2011), and time-series-based prediction (Moreira-Matias, Gama, Ferreira, & Damas, 2012) of travel demands. The hunting behavior should also take the cost of energy (as a proxy of cruising distance) and time (as an indicator of traffic condition) into consideration (Yuan, Zheng, Zhang, & Xie, 2013); (2) Where to, if not hunt, wait for potential passengers? And how long to wait? (i.e., waiting behavior). This operation behavior is strongly intertwined with the hunting strategy. In most circumstances, cabdrivers would choose to wait at, or nearby to, the previous drop-off location if the current location is a hotspot with massive potential passengers (Li et al., 2012) or the cost of hunting next passenger is too high. It is also noteworthy that certain literature (Li et al., 2011) argues that hunting is a dominant strategy upon waiting for most cabdrivers in cities; (3) When and where to interrupt regular taxi working behaviors? This problem might involve cabdrivers' refueling, resting, shifting and ceasing behaviors. Economists develop several profound theories including income reference dependence (Camerer, Babcock, Loewenstein, & Thaler, 1997) and gain-loss utility (Farber, 2015) to explain such kind of daily labor supply decisions of cabdrivers. Urban researchers focus more on the optimization of urban infrastructures (such as the allocation of gas stations and parking lots) and taxicab dispatch service (Zhang, Yuan, Wilkie, Zheng, & Xie, 2015). To the best of our knowledge, existing studies demonstrate more intensive interests on cabdrivers' passenger hunting and waiting behaviors than other operation behaviors.

2.3. Structures of taxi-trace-based spatial network

From a high-level point of view, geographers and regional scientists usually simplify taxicabs' digital traces into an origin-destination matrix or a set of location visitation sequences. In either transformation mode (i.e., matrix or sequence), they can be regarded as quantities of the spatial connections between different locations within a city. In the notation of graph theory, we can derive (taxi-trace-based) spatial networks from taxicabs' digital traces by treating locations as nodes and taxi passenger travels as edges linking a pair of originating and destinating nodes. The structural properties of the resulting network, such as centrality and community, are closely related to a couple of key topics in geography and transportation, such as traffic distribution (Gao, Wang, Gao, & Liu, 2013), congestion propagation (Wang, Lu, Yuan, Zhang, & Van De Wetering, 2013) and regionalization (Goddard, 1970). Particularly, with advances of complex network science, community detection analysis (Fortunato, 2010) has became a powerful technique for exploring taxi-trace-based spatial networks.

Depending on the mode of transformation as discussed above, two kinds of spatial communities have been repetitiously recognized from