Transportation Research Part C 96 (2018) 22–37

Contents lists available at ScienceDirect

Transportation Research Part C

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

A universal distribution law of network detour ratios

Hai Yang[a], Jintao Ke[a,*], Jieping Ye[b,c]

[a] Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China
[b] Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, United States
[c] AI Labs, Didi Chuxing, Beijing, China

ARTICLE INFO

Keywords:
Network detour ratios
Euclidean distance
Road distance
Horn-shaped distribution law

ABSTRACT

Using trajectory data of normal taxis and ride-sourcing vehicles for 10 cities with various sizes in China, we analyze trip distance characteristics by examining the distribution of network detour ratios. The detour ratio for a specific ride is the ratio of the actual driving distance to the corresponding Euclidean (straight-line) distance. We find that, in spite of their different sizes and geographical features, the various cities exhibit an amazingly similar distribution law of network detour ratios: the mean of the detour ratios is inversely proportional to the Euclidean distance with an intercept. We further verify our findings with extensive simulation experiments for a hypothetical circular city with a directional grid street network. Our finding of this universal distribution law of network detour ratios contrasts sharply with the traditional wisdom of modeling throughout the past 50 years that have typically assumed a constant road detour ratio or factor within the range of 1.25–1.41. Our finding in the urban context also has far-reaching implications for fundamental research in many fields such as human mobility, human geography, facility location problems, logistic distribution networks and urban transportation planning.

1. Introduction

The distance between any two points in a city is a key parameter in the analysis of human activities in many fields. Throughout the past 50 years, researchers have typically assumed that the actual road distance is proportional to the Euclidean (straight-line) distance ever since Cole and King (1968) introduced a road detour factor to characterize this relationship. Historically, the Euclidean distance has been widely used due to its ease of calculation and data availability. However, the actual road distance is usually more relevant, although it is an expensive and labor-intensive parameter to determine (Boscoe et al., 2012). Understanding the relationship between the two distance metrics is undoubtedly important for the analysis of human activities in geographical space. For example, analysis of human mobility patterns typically employs a Euclidean distance using the data for circulation of banknotes (Brockmann et al., 2006), mobile phone traces (Gonzalez et al., 2008; Song et al., 2010a, b; Simini et al, 2012), check-in records of social networks (Yan et al., 2017) and taxi GPS data (Liang et al., 2012; Ren et al., 2014; Zheng et al., 2016). Various distributions such as the power-law distribution, the exponential distribution and the log-normal distribution have been found to fit the various datasets. Many geographical studies such as equity and disparity analysis of public service facilities (e.g., medical services) use distance as an accessibility measure (Haggett, 1970). A distance decay function, such as an inverse power or a negative exponential function, is often used to describe the effect of distance on spatial interactions (Tobler's 'first law of geography' (Tobler, 1970)). Urban and transportation planning, such as public transit planning, involves identifying the accessibility of facilities and traffic forecasting, both

---

[*] Corresponding author.
E-mail address: jke@connect.ust.hk (J. Ke).

https://doi.org/10.1016/j.trc.2018.09.012
Received 9 August 2018; Received in revised form 10 September 2018; Accepted 12 September 2018

of which are based on the Euclidean distance or on the calculation of distances along a street network (Gutiérrez, and García-Palomares, 2008). Freight distribution in urban areas usually occurs in two steps, first from large-scale logistics centers to small warehouses and then from the small warehouses to customers (Savelsbergh and Van Woensel, 2016). The detour factor may arise in other fields ranging from ecology to computer science where space and topology are relevant and there is a cost associated with the length of edges (Cardillo et al., 2006; Buhl et al., 2009; Zhang et al., 2015; Barthélemy, 2011).

Using the massive amount of mobility data available from the DiDi Chuxing ride-hailing platform (Didi, 2017), this study investigates the hierarchical and geographical properties of urban road systems—the infrastructure enabling people movement and urban growth and sprawl (Makse et al., 1995; Batty, 2008; Bettencourt, 2013). As one of the largest ride-sharing companies in the world, DiDi serves more than 400 million users across 400 + Chinese cities. DiDi's taxi e-hailing service connects waiting passengers and cruising taxis, while its private car hailing service—in the form of ride-sourcing—efficiently matches a ride request with a nearby, affiliated private car (which is named "express" in DiDi's terminology). We study the distribution of the detour ratios for both trips to pick-up locations and the trips from pick-up locations to the drop-off locations, under two different matching mechanisms applied for taxis and expresses, respectively. More importantly, we find that, for all ten cities examined, the detour ratios of the trips from pick-up locations to the drop-off locations exhibit a striking similar relationship in the sense that the mean of the detour ratios is inversely proportional to the Euclidean distance with an intercept. Moreover, the parameters in the relationship are scale-invariant. This distribution law of network detour ratios implies that the actual road distance of mobility rides can be written as the sum of a distance that is proportional to the Euclidean distance and an extra detour distance that is incurred by the local directed streets and natural/artificial barriers.

The remainder of the paper is organized as follows. Section 2 presents the major observations from the mobility data of taxis and expresses in ten cities, and compares the distribution of detour ratios on trips to pick-up locations and trips to drop-off locations. Section 3 proposes the distribution law of network detour ratios which fits the mobility data in various cities. Section 4 conducts simulation experiments for a hypothetical circular city with a grid street network for further verification of our findings. Section 5 concludes the study and highlights future research directions.

2. Observations from E-hailing trajectory data

Here we use data on a total of 54.4 million taxi ride orders (from 2017/03/01 to 2017/05/31) and 102.5 million express ride orders (from 2017/04/01 to 2017/04/30) in four large cities and six medium and small cities in China. Note that the administrative region of cities in China is large and includes both the urban area and rural area, and we only collect the orders of rides that have origins and destinations within the urban area. The sizes of express and taxi samples collected in 10 cities, and the population and area of these cities are given as Table 1.

As shown in Fig. 1, data for each order include latitude/longitude coordinates and the time when the driver: a) accepts the order (location ①); b) reaches the pick-up location (or somewhere nearby) (location ②); c) picks up the passenger and presses the button 'begin charge'; and d) reaches the drop-off location (location ③). The trip from a) to b) is viewed as the trip to the pick-up location (pickup-trip), while that from c) to d) is regarded as the trip from pick-up location to the drop-off location (ride-trip).

Two different matching mechanisms—the broadcast mode for taxis and the dispatch mode for expresses—are introduced in Fig. 2. The broadcast mode for taxis is akin to an agent-based system, where individual drivers aim to minimize their pick-up time or distance and maximize the ride fare (mainly dependent on the ride distance). The platform receives orders from waiting passengers and then broadcasts these orders to nearby taxi drivers according to a set of rules (for example, it first broadcasts to taxis within a 1 km radius and then to taxis within a 2 km radius if no response is received within a certain period). The order is assigned to the first taxi driver to accept. In contrast, the dispatch mode for expresses matches idle drivers and ride requests by considering the overall system efficiency rather than the utility of individual drivers. The platform simultaneously collects orders and nearby locations of

Table 1
Sample sizes and city characteristics.

City	Taxi samples	Express samples	Population* (million)	Area* (km^2)
Beijing	18,214,456	24,856,371	18.78	12187.00
Shenzhen	4,654,443	18,498,240	11.38	1997.27
Guangzhou	3,376,301	17,145,063	6.15	2099.18
Shanghai	16,916,886	11,675,124	24.15	6340.50
Hangzhou	7,127,155	12,578,019	3.32	1484.96
Quanzhou	673,473	5,290,512	0.99	539.00
Foshan	1,452,660	7,802,565	1.28	763.16
Sanya	270,246	567,824	0.25	188.00
Nanchang	796,788	2,999,187	2.42	358.90
Xuzhou	905,803	1,048,278	1.71	603.25

* Sourced from the China Urban-Rural Construction Statistical Yearbook in 2015 (http://www.mohurd.gov.cn/xytj/tjzljsxytjgb/index.html). Area refers to the urban area defined by the government, which can be a good proxy for measuring the activity range of urban residents. Population refers to the population living in the urban area of the city (those living on the outskirts are not recorded).