Physica A 456 (2016) 222–227

Contents lists available at ScienceDirect

Physica A

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

An understanding of human dynamics in urban subway traffic from the Maximum Entropy Principle

Nuo Yong [a,b], Shunjiang Ni [a,b,*], Shifei Shen [a,b,*], Xuewei Ji [c]

[a] *Institute of Public Safety Research, Tsinghua University, Beijing, 100084, China*
[b] *Department of Engineering Physics, Tsinghua University, Beijing, 100084, China*
[c] *Beijing Academy of Safety Science and Technology, Beijing, 100070, China*

HIGHLIGHTS

- Entry time interval obeys power-law distribution with an exponential cutoff.
- Exponential cutoff appears when power exponent is less than 1.0.
- Traffic per unit time influences the parameters of distribution function.

ARTICLE INFO

Article history:
Received 27 October 2015
Received in revised form 3 February 2016
Available online 2 April 2016

Keywords:
Human dynamics
Maximum Entropy Principle
Power-law distribution
Subway traffic

ABSTRACT

We studied the distribution of entry time interval in Beijing subway traffic by analyzing the smart card transaction data, and then deduced the probability distribution function of entry time interval based on the Maximum Entropy Principle. Both theoretical derivation and data statistics indicated that the entry time interval obeys power-law distribution with an exponential cutoff. In addition, we pointed out the constraint conditions for the distribution form and discussed how the constraints affect the distribution function. It is speculated that for bursts and heavy tails in human dynamics, when the fitted power exponent is less than 1.0, it cannot be a pure power-law distribution, but with an exponential cutoff, which may be ignored in the previous studies.

1. Introduction

The study of the timing of human activities has been paid much attention in recent years. Studies show that the timing of many human activities follows non-Poisson statistics, characterized by heavy tailed or power-law distribution of time interval τ as $P(\tau) \sim \tau^{-\alpha}$. Understanding the timing of human activities is extremely important to investigate human related activities. For example, Vazquez et al. [1] investigated the impact of non-Poisson human activity patterns on the email worm's spread and proved that non-Poisson patterns lead to larger decay time than standard Poisson process. Zhang et al. [2] investigated the impact of heavy-tailed human behaviors on the information spreading and proved that for information with different time period of being interested in, heavy-tailed human behaviors may enhance (for small time period) or suppress (for large time period) the spreading.

Most of the explanations of the origin of bursts and heavy tails in human dynamics are based on the specific process of human activities. Barabasi et al. [3] proposed a simple model where the individual executes the highest-priority task

* Corresponding authors at: Institute of Public Safety Research, Tsinghua University, Beijing, 100084, China.
E-mail addresses: sjni@tsinghua.edu.cn (S. Ni), shensf@tsinghua.edu.cn (S. Shen).

http://dx.doi.org/10.1016/j.physa.2016.03.071
0378-4371/

first and suggested that the burst nature of human behavior is a consequence of a decision-based queuing process. Based on that, Saichev and Sornette [4] developed a simple general framework to model the waiting-time distribution between the triggering factor and the actual realization of a job. Stochastic queuing-theory [5,6] was also proposed to explain the burst nature of human dynamics. Furthermore, Muchnik L. [7] performed a causal inference analysis and suggested that the heavy-tailed degree distribution is causally determined by similarly skewed distribution of human activity.

There is still no consensus about the value of the power exponent α. For email communication and view times versus lectures on YouTube [8], it approximates to 1.0, and for letter correspondence, it is 1.5. For the creation time of purchase orders to an individual vendor [9], it approximates to 2.0. Further investigation showed that the power exponent may not be constant [10,11] and the interval distribution of arrival-time may be power-law with an exponential cutoff [12] or the combination of power-law and Poisson [13,14].

To further understand the bursts and heavy tails of human behavior, we deduced the probability distribution function of entry time interval based on the Maximum Entropy Principle. Further analysis of smart card transaction records in Beijing subway proved our theoretical derivation. Both of them indicated that the entry time interval obeys power-law distribution with an exponential cutoff. In addition, we pointed out the constraint conditions for the distribution form and discussed how the constraints affect the distribution function. It is speculated that for bursts and heavy tails in human dynamics, when the fitted power exponent is less than 1.0, it cannot be a pure power-law distribution, but with an exponential cutoff, which may be ignored in the previous studies.

2. Data analysis

We obtained the entry records of the smart card transactions from the Auto Fare Collection (AFC) system of Beijing Subway. In order to study the interval time distribution, we extracted the following information from the entry records. The card ID, which is typically corresponding with a passenger. The entry line and station number, which means the codes of entry subway line and station that the passenger gets into. The exact time that the passenger swipes the card when getting into the station through a certain device, and the ID or tail number of the device which the passenger swipes through.

Here, we analyzed the records of fourteen continuous weekdays in October 2014, and mainly focused on two subway station named YuanMingYuan (abbreviated as YMY) and HaiDianHuangZhuang (abbreviated as HDHZ).

Fig. 1(a) presents the entry traffic of a certain device whose tail number is 762 in YMY station on five continuous weekdays. Fig. 1(c) shows the entry interval time distribution of the traffic data in Fig. 1(a), especially in log–log coordinate. Fig. 1(b) and (d) present the same ones of another device whose tail number is 345 in a HDHZ station.

The typical image of entry time interval is shown in Fig. 1(c) and (d). The image feature can be divided into two parts. For the front part, the image goes up for the limited swiping efficiency of a single device. When getting into the station, passengers have to wait in line and swipe the card one by one, so the probability of short time interval (e.g. 1 or 2 s) is lower. If the swiping efficiency of a single device can be improved, the front part will disappear, which can be confirmed in Fig. 4. For the remaining parts of the image, it decays with a heavy tail. When fitting the image, we mainly focus on the latter part.

It is obvious that there are similar trends of traffic on weekdays for the same device of a certain subway station. And the distribution of the entry time interval is heavy-tailed, which indicates that there exists bursts in the entry behavior of passengers. For different station with different location and surroundings, the busy period and shape feature of entry traffic images are diverse. However, the shape feature of entry time interval distribution is similar. It indicates that there exists common mechanism which leads to similar distributions.

3. Distribution derivation

Jaynes [16] proposed Maximum Entropy Principle (MEP) in 1957, which indicates that the distribution of probabilities is determined by the requirement of maximum of the entropy under additional conditions. For example, when the mean and variance are fixed, the distribution of a series of continuous variables is Gaussian distribution. In 1980, Zhang [17] also introduced the MEP into probability statistics to determine the conditions of common probability density functions. Especially, he proved that when the geometric mean is fixed, the distribution of a series of continuous variables is Power-Law distribution. Based on these precious studies, we deduced the distribution function of entry time interval.

Take the entry time interval as a series of discrete random variables. Note that within one second, at most one passenger is able to get in through a single device. The value of entry time interval sequence is denoted as $\tau = 1, 2, 3, \ldots, M$, and the number of passengers with interval τ is denoted as n_τ. The additional conditions are:

(i) The daily traffic of a single device is finite, denoted by N:

$$\sum_{\tau=1}^{M} n_\tau = N. \tag{1}$$

(ii) The daily working hour of a single device is finite, denoted by L:

$$\sum_{\tau=1}^{M} n_\tau \cdot \tau = L. \tag{2}$$