Physica A 398 (2014) 145–151

Contents lists available at ScienceDirect

Physica A

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

Modeling correlated human dynamics with temporal preference

Peng Wang [a,b,c,*], Tao Zhou [d], Xiao-Pu Han [a], Bing-Hong Wang [e,f]

[a] Institute for Information Economy and Alibaba Business College, HangZhou Normal University, Zhejiang 310036, China
[b] Alibaba Group Research Center, Zhejiang 310036, China
[c] School of Business, Renmin University of China, Beijing 100872, China
[d] Web Sciences Center, University of Electronic Science and Technology of China, Chengdu 610054, China
[e] Department of Modern Physics, University of Science and Technology of China, Hefei 230026, China
[f] The Research Center for Complex System Science, University of Shanghai for Science and Technology and Shanghai Academy of System Science, Shanghai 200093, China

HIGHLIGHTS

- We found strong memory in the interevent time distribution of blog-posting.
- We found the decay curve of it have two regimes: it decays as a power law for the short-term part and exponentially for the long-term part.
- We propose a simple model based on temporal preference which reproduces all features we obtained.

ARTICLE INFO

Article history:
Received 10 August 2012
Received in revised form 22 October 2013
Available online 23 December 2013

Keywords:
Interevent time distribution
Power-law
Human dynamics

ABSTRACT

We empirically study the activity pattern of individual blog-posting and observe the interevent time distributions decay as power-laws at both individual and population level. As different from previous studies, we find significant short-term memory in it. Moreover, the memory coefficient first decays in a power law and then turns to an exponential form. Our findings produce evidence for the strong short-term memory in human dynamics and challenge previous models. Accordingly, we propose a simple model based on temporal preference, which can well reproduce both the heavy-tailed nature and the strong memory effects. This work helps in understanding the temporal regularities of online human behaviors.

1. Introduction

Human actions underly many social, technological and economic phenomena, and thus the quantitative understanding of human behavior is very significant [1–3]. Thanks to the development of the information techniques, more and more electronic records available from the Internet may provide us with insights into the pattern of human behavior [4,5]. In recent years, examples of empirically studied human activities include communication patterns of electronic mails [6–8] and surface mails [8–10], web surfing [11,12], short messages [13], movie ratings [14], online games [15,16]. The main result, arising from all these studies, concerns the heavy-tailed nature of human activities: the interevent and/or response times follow a power-law-like distribution at the level of both population and individual.

* Corresponding author at: Institute for Information Economy and Alibaba Business College, HangZhou Normal University, Zhejiang 310036, China. Tel.: +86 18858158273.
E-mail addresses: wangpenge@gmail.com (P. Wang), bhwang@ustc.edu.cn (B.-H. Wang).

0378-4371/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physa.2013.12.014

Traditional models of human activity are based on Poisson processes and predict that the interevent time follows an exponential distribution. Actually, there are crucial differences between exponential and power-law distribution: in a exponential one, the time intervals are relatively regular and cannot be very long; but heavy-tailed processes allow for not only very long periods of inactivity but also bursts of intensive activity.

In order to understand this phenomenon, various mechanisms were suggested to explain the origin of these heavy-tails. An important one is the priority-queue model [6,17], in which human behavior is primarily driven by rational decision making. Another explanation comes from the cascading nonhomogeneous Poisson process in which circadian and weekly cycles play a key role [18,7,8] in the origin of heavy-tails. Other mechanisms include the adaptive interesting model [19], the memory model [20,21] and the interaction model [22]. All these models also can yield the heavy-tailed interevent time (or waiting time) distribution. But which one is the real one of human behavior? Or are there various patterns of human behavior? In order to answer this question, it is not enough only to compare the distribution of interevent time and we need further empirical studies.

It is well know that memory plays an important role in human behavior [23–25]. But, unexpectedly, recent research on human dynamic shows small memory/auto-correlation between two successive interevent times in several human activities [26]. This result supports those non-memory models such as the priority-queue model [6,17], the cascading nonhomogeneous Poisson process [18,7,8] and rejects the memory-based models such as the adaptive interesting model [19]. However, it is not clear yet whether the lack of memory is common in human behavior and more empirical study is required.

In this paper, we empirically study the activity patterns of individual blog-posting and find not only the heavy-tails in interevent time distribution but also significant memory effects. We propose a simple model based on temporal preference, which can well reproduce both the heavy-tailed nature and the strong memory effect. This paper is organized as follows. In Section 2, we introduce the empirical observations, followed by the model and simulation results in Section 3. We conclude this work in Section 4 with some discussion about the relevance of our work to the real human behavior.

2. Empirical analysis

A blog is a kind of so-called web2.0 application emerging in recent years, in which people post some words, read and comment on each other. For most ordinary bloggers (the user of blog), they post according to their interests and treat it only as an amusement or an optional way of communication with friends. This makes their blogging behavior unstable and the frequency is relatively low. Our data was collected from a campus blog website of Nanjing university [29], most users of which are students or teachers of Nanjing university. The first article is posted at 25/03/2003 when this blog website was established. As of 01/09/2009, there are 1627,697 posts belonging to 20,379 users on it.

Here, the interevent time is the time interval between consecutive posting by the same user in one's blog. So, if one user has n articles, we will get $n - 1$ interevent times. We collect all the 20,379 users' interevent times and get the global distribution of interevent times which is shown in Fig. 1. As we can see, the distribution of the number of the post decays as power law just like previous results. Here, we fit the inter-day distributions with the so-called "shifted power-law" [27,28]:

$$y \sim (x + a)^{-\beta}. \tag{1}$$

The exponent is -1.98 which is very close to the one in movie rating [14] and web activities on AOL and Ebay [12].

For individual behavior, we only consider users whose number of posts is more than 200 to avoid characterizing users who post too little and get 2211 qualified users. Fig. 2 shows the cumulative distributions of interevent times of six users. As we can see, all of the cumulative distributions also decay as a power law asymptotically and the exponents are around 1. Correspondingly, the exponents of distributions should be around 2. Actually, the average of the exponents of all qualified users we obtained is 2.23. It should be noted that this result is close to the one observed on the global distribution.

The interevent times for consecutive events of two users are shown in Fig. 3(a) and (b). As we can see, one of the important features is the clustering of long interevent times which also is called a mountain-valley-structure and can be found in many complex systems [30,31]. This feature inspired us to investigate the memory coefficient of this succession, although the lack of it in human activities was already reported by Goh and Barabási [26]. The definition of it is as follows [26]:

$$M_k = \frac{1}{n_\tau - 1} \sum_{i=1}^{n_\tau - 1} \frac{(\tau_i - m_1)(\tau_{i+k} - m_2)}{\sigma_1 \sigma_2}, \tag{2}$$

where τ_i is the interevent time values and n_τ is the number of interevent times and $m_1(m_2)$ and $\sigma_1(\sigma_1)$ are sample mean and sample standard deviation of τ_i's (τ_{i+k}'s). The two interevent times τ_i' and τ_{i+k} are separated by k events.

Here, we calculate M_k of all these qualified users with k ranging from 1 to 40. Because the number of posts of one single user is still too little, we only study the average M_k of all users which was shown in Fig. 3(c). As we can see, M_1 is 0.211 which shows there is strong memory between the nearest interevent times. Interestingly, there are two regimes in this decay curve: when $k < 10$, it decays asymptotically as a power law: $y = 0.23 * x^{-0.48}$; when $k > 10$, it decreased exponentially: $y = 0.1 * e^{-x/23.76}$. This means that the short-term (when $k < 10$) and long-term memory (when $k > 10$) in this behavior are likely due to different mechanisms. For the part of $k < 10$, it is strong and probably the cause of the mountain-valley-structure above. On the other hand, for the part of $k > 10$, it decreases more quickly and may have something to do with the long-term change of personal interests.