New Journal of Physics
The open-access journal for physics

Modeling human dynamics with adaptive interest

Xiao-Pu Han[1], **Tao Zhou**[1,2,4] **and Bing-Hong Wang**[1,3]
[1] Department of Modern Physics, University of Science and Technology of China, Hefei 230026, People's Republic of China
[2] Department of Physics, University of Fribourg, CH-1700, Fribourg, Switzerland
[3] Shanghai Institute for Systemic Sciences, Shanghai 200093, People's Republic of China
E-mail: zhutou@ustc.edu

New Journal of Physics **10** (2008) 073010 (8pp)
Received 27 March 2008
Published 7 July 2008
Online at http://www.njp.org/
doi:10.1088/1367-2630/10/7/073010

Abstract. Increasing recent empirical evidence indicates the extensive existence of heavy tails in the inter-event time distributions of various human behaviors. Based on the queuing theory, the Barabási model and its variations suggest the highest-priority-first protocol to be a potential origin of those heavy tails. However, some human activity patterns, also displaying heavy-tailed temporal statistics, could not be explained by a task-based mechanism. In this paper, different from the mainstream, we propose an interest-based model. Both the simulation and analysis indicate a power-law inter-event time distribution with exponent -1, which is in accordance with some empirical observations in human-initiated systems.

Contents

[4] Author to whom any correspondence should be addressed.

1. Introduction

Human behavior, as an academic issue in science, has a history of about a century according to Watson [1]. As a joint interest of sociology, psychology and economics, human behavior has been extensively investigated during past decades. However, due to the complexity and diversity of our behaviors, the in-depth understanding of human activities is still a long-standing challenge thus far. Actually, in most of the previous works, the individual activity pattern is usually simplified as a completely random point-process, which can be well described by the Poisson process, leading to an exponential inter-event time distribution [2]. That is to say, the time difference between two consecutive events should be almost uniform, and long gaps are hardly observed. However, recently, empirical studies on e-mail [3] and surface mail [4] communication show a far different scenario: those communication patterns follow non-Poisson statistics, characterized by bursts of rapidly occurring events separated by long gaps. Correspondingly, the inter-event time distribution has a much heavier tail than the one predicted by an exponential distribution. The heavy tails have also been observed in many other human behaviors [5, 6], including market transaction [7, 8], web browsing [9], movie watching [10], short message sending [11], and so on. The increasing evidence of non-Poisson statistics of human activity patterns highlights a question: what is the origin of those heavy tails? Based on the queuing theory, Barabási *et al* proposed a simple model [3, 12, 13] where the individual executes the highest-priority task first, and they suggested the highest-priority-first (HPF) protocol, a potential origin of those heavy tails.

The queuing model has great success in explaining the heavy tails in many human-oriented dynamics. However, some other human activity patterns, also displaying the similar heavy-tailed phenomenon, could not be explained by a task-based mechanism. For example, the actions of browsing the web [9], watching on-line movies [10] and playing on-line games [14] are mainly driven by personal interests, which could not be treated as tasks needing to be executed. The in-depth understanding of the non-Poisson statistics in those interest-driven systems requires a new model out of the perspective of the queuing theory. In this paper, different from the mainstream task-based models, we propose an interest-based model. Both the simulation and analysis indicate a power-law inter-event time distribution with exponent -1, which is in accordance with some empirical human-initiated systems.

2. Model

Before introducing the mathematical rules of our model, let us think of the changing process of our interests on web browsing according to our daily experiences. If a person has not browsed the web for a long time, an accidental visit to a browsing outlet may give him a good feeling and arouse his interest in web browsing. Next, during the action, the good feeling continues and the frequency of web browsing may increase. Then, if the frequency is too high, he may worry about it, thus reducing the frequency of browsing. Such similar experiences can be found in the case of many other daily actions, such as playing games, seeing movies, and so on. In a word, we usually adjust the frequency of our daily actions according to our interest: greater interest leads to higher frequency, and vice versa. Some simple assumptions extracted from our daily experiences are as follows: firstly, for a given interest-driven behavior, each action will change the current interest, while the frequency of actions depends on the interest. It is like an active walker [15, 16], whose motion is affected by the energy landscape, while the motion track could

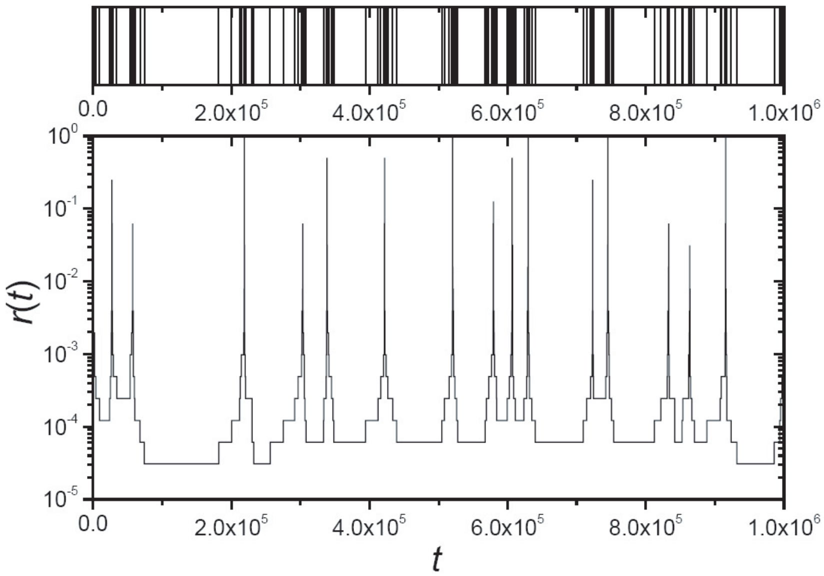

Figure 1. Upper panel: the succession of events predicted by the present model. The total number of events shown here is 375 during 10^6 time steps. Lower panel: the corresponding changes of $r(t)$. The data points are obtained with the parameters $a_0 = 0.5$ and $T_2 = 10^4$.

simultaneously change the landscape. Secondly, the inter-event time τ has two thresholds: when τ is too small (i.e. events happen too frequently), the interest will be depressed, thus the inter-event time will increase; whereas if the time gap is too long, we will increase the interest to mimic its resuscitation induced by a casual action.

According to these assumptions, we propose an interest-based model which is as follows: (i) the time is discrete and labeled by $t = 0, 1, 2, \ldots$, the occurring probability of an event at time step t is denoted by $r(t)$. The time interval between two consecutive events is called the inter-event time and is denoted by τ. (ii) If the $(i+1)$th event occurred at time step t, the value of r is updated as $r(t+1) = a(t)r(t)$, where

$$a(t) = \begin{cases} a_0, & \tau_i \leqslant T_1, \\ a_0^{-1}, & \tau_i \geqslant T_2, \\ a(t-1), & T_1 < \tau_i < T_2. \end{cases} \tag{1}$$

If no event occurred at time step t, we set $a(t) = a(t-1)$, namely, $a(t)$ remains unchanged. In this definition, T_1 and T_2 are two thresholds satisfying $T_1 \ll T_2$, τ_i is the time interval between the $(i+1)$th and the ith events, and a_0 is a parameter controlling the changing rate of occurrence probability ($0 < a_0 < 1$). If no event occurs, r will not change. Clearly, simultaneously enlarging (by the same multiple) T_1, T_2 and the minimal perceptible time, will not change the statistics of this system. Therefore, without loss of generality, we set $T_1 = 1$.

3. Simulation and analysis

In the simulations, the initial value of r is set at $r_0 = r(t = 0) = 1.0$, which is also the possibly maximal value of $r(t)$ in the whole simulation process. As shown in figure 1, the succession

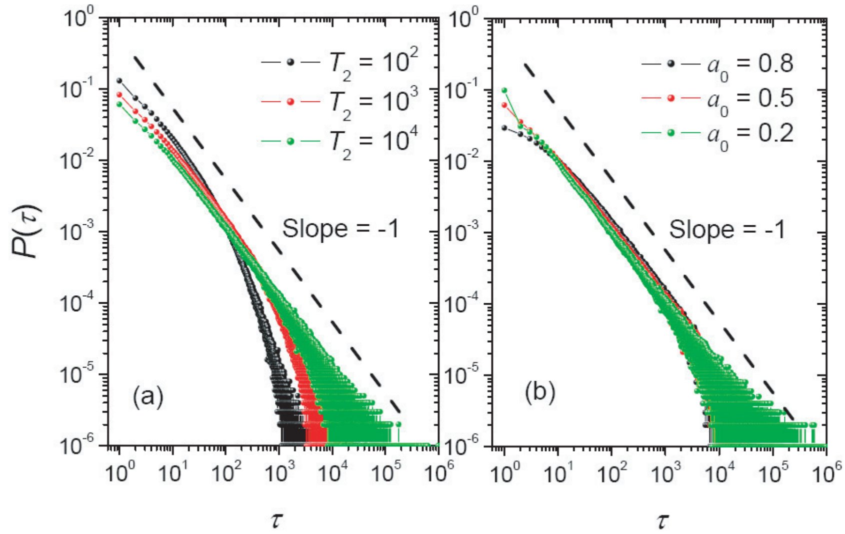

Figure 2. Inter-event time distributions in log–log plots. (a) Given that $a_0 = 0.5$, $P(\tau)$ for different T_2, where the black, red and green curves denote the cases of $T_2 = 10^2$, 10^3 and 10^4, respectively. (b) Given that $T_2 = 10^4$, $P(\tau)$ for different a_0, where the black, red and green curves denote the cases of $a_0 = 0.8$, 0.5 and 0.2, respectively. The black dashed lines in both (a) and (b) have a slope -1. Each distribution contains 10^6 events.

of events predicted by the present model exhibits very long inactive periods that separate the bursts of rapidly occurring events, and the corresponding $r(t)$ shows a clearly seasonal property (quasi-periodic behavior). Actually, in a period, the maximal and minimal values of $r(t)$ are respectively determined by T_1 and T_2 as $r_{\max} \sim T_1^{-1}$ and $r_{\min} \sim T_2^{-1}$. This quasi-periodic property will be applied in the further analysis. Note that, in a specific quasi-period, $r_{\max}$ can be smaller than T_1^{-1} and $r_{\min}$ can be smaller than T_2^{-1}. This is because $\tau \leqslant T_1$ could result as a consequence of $r(t) < T_1^{-1}$ and $\tau \leqslant T_2$, when $r(t) \leqslant T_2^{-1}$.

Figure 2 reports the simulation results with tunable T_2 and a_0. Given that $a_0 = 0.5$, if $T_2 \gg T_1$, the inter-event time distribution generated by the present model displays a clear power law with the exponent -1; while if T_2 is not sufficiently large, the distribution $P(\tau)$ exhibits a departure from a power-law form with a cut-off in its tail. Correspondingly, given a sufficiently large T_2, the effect of a_0 is very slight, thus can be ignored.

Taking into account the quasi-periodic property of $r(t)$, we raise two approximated assumptions before analytical derivation: (i) the statistical property of $P(\tau)$ is the same as that in a single period; (ii) within one period, the statistical property of $P(\tau)$ in the r-increasing half is the same as that in the r-decreasing half. In the reducing process, $r(t) = r_m a_0^i$, where $i = 0, 1, 2, \ldots, I$. The integer I denotes the number of events in the reducing process (also the number of different values of $r(t)$), whose value is about

$$I \approx -\log_{a_0}(T_2/T_1), \tag{2}$$

since $r_{\max} \sim T_1^{-1}$ and $r_{\min} \sim T_2^{-1}$. The variable r_m is the initial value (it is also the maximum value) of $r(t)$ in a reducing process. Note that, for different reducing processes, the values of r_m are not always the same. Though r_m has the same order of magnitude as $T_1^{-1} = 1.0$, its value can be less than T_1^{-1} in a specific process. The average value of r_m will be calculated later in this paper.

If the current occurring probability is $r(t) = r_m a_0^i$, the probability that the next event will happen at the time $t + \tau$ is

$$Q(\tau) = (1 - r_m a_0^i)^{\tau-1} r_m a_0^i. \tag{3}$$

Considering every value of $r(t)$ in the reducing process, the inter-event time distribution of the reducing process is

$$P(\tau) = I^{-1} \sum_{i=0}^{I} (1 - r_m a_0^i)^{\tau-1} r_m a_0^i. \tag{4}$$

According to the approximated assumptions above, the inter-event time distribution of all the successions can also be expressed by equation (4), which can be approximately rewritten in a continuous form, as

$$P(\tau) \approx I^{-1} \int_0^I (1 - r_m a_0^x)^{\tau-1} r_m a_0^x \, dx. \tag{5}$$

Therefore, $P(\tau)$ can be further expressed as

$$P(\tau) \approx -[(1 - r_m a_0^I)^\tau - (1 - r_m)^\tau](\ln a_0)^{-1} I^{-1} \tau^{-1}. \tag{6}$$

From equation (6), for a fixed r_m, when I is large enough (equivalent to the condition $T_2 \gg T_1$), $P(\tau)$ has a power-law tail with exponent -1. In addition, this analytical result also provides an explanation about the departure from a power law when T_2 is not sufficiently large.

As discussed before, for different reducing processes of $r(t)$, the possible values of r_m are not always the same (see also the lower panel of figure 1 for different quasi-periods, the maximum values of $r(t)$ are different). Since the order of magnitude of r_m is comparable with $T_1^{-1} = 1.0$ (it is equal to r_0), the minimum value of $r(t)$, $r_m a_0^I$, has the same order of magnitude as $r_0 a_0^I$. Making the approximated assumption that the minimum value of $r(t)$ is given by $r_0 a_0^I$ in an r-increasing process, and the maximum value of $r(t)$ in the next r-decreasing process is $r_0 a_0^k$ ($r_0 a_0^k$ is also the start point in the next decreasing process), then the probability density of k reads

$$\Omega(k) = r_0 a_0^k \prod_{i=0}^{I-k-1} (1 - r_0 a_0^{I-i}). \tag{7}$$

Therefore, the average value of r_m is

$$\langle r_m \rangle = \sum_{k=0}^{I-1} r_0 a_0^k \Omega(k) = \sum_{k=0}^{I-1} (r_0 a_0^k)^2 \prod_{i=0}^{I-k-1} (1 - r_0 a_0^{I-i}). \tag{8}$$

This average value of r_m calculated by equation (8), as well as the integer part of $-\log_{a_0}(T_2/T_1)$ (as the approximation of I), can be directly used in the approximate calculations of equation (6). Given that $r_0 = 1.0$, $a_0 = 0.5$, $T_2 = 10^4$ and $T_1 = 1$, one obtains $I \approx -\log_{a_0}(T_2/T_1) = 13$, and $\langle r_m \rangle \approx 0.50$ from equation (8). Accordingly, figure 3 reports the comparison of analytical and simulation results, which are well in accordance with each other.

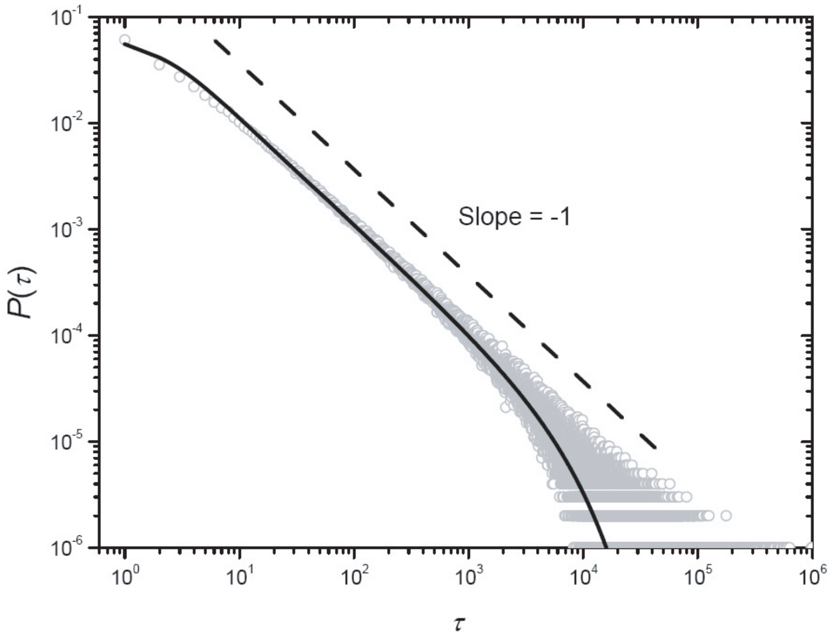

Figure 3. Comparison of the analytical (black solid line) and numerical (gray circles) results of inter-event time distribution. The numerical data are obtained with parameters $r_0 = 1.0$, $a_0 = 0.5$ and $T_2 = 10^4$. The analytical results are calculated by equation (6) with $a_0 = 0.5$, $I = 13$ and $r_m = 0.50$. The black dashed line has slope -1. The numerical results contain 10^6 events.

4. Conclusion and discussion

A novel model of human dynamics is proposed in this paper. Different from the mainstream queuing models, the current model is driven by personal interests. In this model, the frequency of events are determined by the interest, while the interest is simultaneously affected by the occurrence of events. This interplay working mechanism, similar to the active walk [15, 16], is a genetic origin of complexity of many real-life systems. The rules in the current model are extracted from our daily life, and both the analytical and simulation results agree well with empirical observations, such as the activity pattern of web browsing [9]. Our work indicates a simple and universal mechanism in human dynamics, that is, people could adaptively adjust their interest in a specific behavior (e.g. watching TV, browsing the web, playing on-line games, etc), which leads to a quasi-periodic change of interest, and this quasi-periodic property eventually gives rise to the departure from Poisson statistics.

Besides the HPF protocol and the current model, there are also some other mechanisms that can lead to a power-law inter-event time distribution. For example, Hidalgo [17] pointed out that a Poissonian individual with characteristic time varying randomly in time could generate a power-law inter-event time distribution with exponent -2. In addition, Vázquez [18] showed that if the current executing rate is linearly correlated with the average executing rate in an immediate predecessor period, the inter-event time distribution will follow a power-law form.

Note that, although in the recent empirical works the power-law form is widely used to fit the inter-event time distribution of human behaviors, there exists a debate about the choice of fitting functions for this distribution in e-mail communication [19, 20]. Actually, a candidate, namely a *log-normal distribution*, has also been suggested [19] for describing the non-Poisson temporal statistics of human activities. The *stretched exponential distribution* [21, 22], interpolating between a power law and an exponential form, serves as another candidate (see, for example, the distribution of inter-event time between two consecutive transactions initiated by a stock broker [13]). A clear understanding of the tails in the inter-event time distribution requires in-depth exploration of empirical data in the future.

The concept and methodologies related to the statistics of the inter-event time can also find applications in some other systems. For example, similar statistical analysis can be carried out on the spacing between the consecutive occurrences of the same letter in written text [5], and the time difference between successive events above a certain threshold (i.e. extreme events) [23].

Finally, we point out some limitations in the current model. Firstly, it can only generate the power-law inter-event time distribution with exponent -1, which does not agree with some real human-initiated systems with different power-law exponents. Secondly, we assume that the changing rate of the occurring probability, a_0, is fixed as a constant in every rising or decaying process. This assumption is very ideal, and we could not find any support from the empirical data. Thirdly, as stated by Kentsis [24], there are countless ingredients affecting the human dynamics, and for most of them, we do not know their impacts. Those ingredients, such as the social content, the semantic content and the periodicity due to circadian and weekly cycles, have not been considered in the present model, neither has the HPF protocol. However, although this model is rough and may contain some artificial assumptions, it provides a starting point for modeling interest-based human dynamics. Human-initiated systems are the most complex systems, and there must be many underlying mechanisms that have not yet been discovered. We believe our model could enlighten readers in this rapidly developing field.

Acknowledgments

We acknowledge useful discussions with Wei Hong and Shuang-Xing Dai. This work was partially supported by the National Natural Science Foundation of China (grant numbers 10472116 and 10635040) and the 973 Program 2006CB705500.

References

[1] Watson J B 1913 *Psychol. Rev.* **20** 158
[2] Haight F A 1967 *Handbook of the Poisson Distribution* (New York: Wiley)
[3] Barabási A-L 2005 *Nature* **435** 207
[4] Oliveira J G and Barabási A-L 2005 *Nature* **437** 1251
[5] Goh K-I and Barabási A-L 2008 *Europhys. Lett.* **81** 48002
[6] Zhou T, Han X-P and Wang B-H 2008 *Preprint* 0801.1389
[7] Plerou V, Gopikrishnan P, Amaral L A N, Gabaix X and Stanley H E 2000 *Phys. Rev.* E **62** 3023
[8] Masoliver J, Montero M and Weiss G H 2003 *Phys. Rev.* E **67** 021112
[9] Dezsö Z, Almaas E, Lukács A, Rácz B, Szakadát I and Barabási A-L 2006 *Phys. Rev.* E **73** 066132
[10] Zhou T, Kiet H A-T, Kim B J, Wang B-H and Holme P 2008 *Europhys. Lett.* **82** 28002
[11] Hong W, Han X-P, Zhou T and Wang B-H 2008 *Preprint* 0802.2577
[12] Vázquez A 2005 *Phys. Rev. Lett.* **95** 248701
[13] Vázquez A, Oliveira J G, Dezsö Z, Goh K-I, Kondor I and Barabási A-L 2006 *Phys. Rev.* E **73** 036127
[14] Henderson T and Nhatti S 2001 *Proc. 9th ACM Int. Conf. on Multimetia* (New York: ACM) p 212
[15] Lam L 2005 *Int. J. Bifurcation Chaos* **15** 2317

[16] Lam L 2006 *Int. J. Bifurcation Chaos* **16** 239
[17] Hidalgo C 2006 *Physica* A **369** 877
[18] Vázquez A 2007 *Physica* A **373** 747
[19] Stouffer D B, Malmgren R D and Amaral L A N 2005 *Preprint* physics/0510216
[20] Barabási A-L, Goh K-I and Vázquez A 2005 *Preprint* physics/0511186
[21] Laherrere J and Sornette D 1998 *Eur. Phys. J.* B **2** 525
[22] Zhang P P *et al* 2006 *Physica* A **360** 599
[23] Bogachev M I, Eichner J F and Bunde A 2007 *Phys. Rev. Lett.* **99** 240601
[24] Kentsis A 2006 *Nature* **441** E5