Impact of habitual behaviors on human dynamics and spreading process

Yu Jiao, Yanheng Liu*, Jian Wang, Jianqi Zhu
1.College of Computer Science and Technology, Jilin University, Changchun 130012, China;
2.Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012, China
Email: mihayee@gmail.com, lyh_lb_lk@yahoo.com.cn, wangjian591@gmail.com, zhujq@jlu.edu.cn
* corresponding author: Yanheng Liu

***Abstract*—Human behaviors are assumed as randomly distribution in current models for human dynamics. But more and more evidence proves that the intervals of human actions follows the power-law distributions with heavy tails. A model considering habit of humans is introduced to explain bursts and heavy tails in human dynamics more exactly, and the simulation results are consistent with the real data from a university emails record and an online movie order web site. Normal distribution is used to simulate intervals of succession of events, and random parameters are set as unexpected events disturbing habit behaviors. Furthermore, a worm propagation model based on the habit model and SI model is presented to investigate the impact of human behavior on virus propagation. The model shows that the consuming time of infecting all nodes in a network increases significantly with the extending of network scale based on the proposed habit model, while the time increases very slowly based on Poisson model.**

I. Introduction

Despite its idiosyncratic complexity, individual and societal human behavior is of enormous practical importance. The scientific attention to this topic is motivated by clear economic and technological purposes since the possibility to monitor and mathematically describe human behavior may have important implications in resource management and service allocation. Besides, the current availability of digital records has made it much easier for researchers to quantitatively investigate various aspects of human behavior. Poisson processes were widely used in the past study to quantify the consequences of human actions, ranging from modeling traffic flow patterns to estimating calls in mobile communication[1]. However, it was reported recently that the probability density for the inter-event time between consecutively activity, such as file downloads, letter correspondence, library usage, broker trades, web browsing, human loco motor activity, and telephone communication, decays asymptotically as $p(\tau) = \tau^{-\alpha}$, with $\alpha \simeq 1$[2]. These observations are in stark contrast with the predictions of a homogeneous Poisson process, describing that the burstiness of humans consists of long periods of inactivity followed by short periods of time in which humans concentrate their actions. This prompts us to search for the mechanisms responsible for its emergence.

Barabási considered that the bursty nature of human behaviour is a consequence of a decision-based queuing process: when individuals execute tasks based on some perceived priority, the timing of the tasks will be heavy tailed[3]. Vázquez studied two classes of queuing models designed to capture human activity patterns and document the existence of two distinct universality classes, one characterized by $\alpha = 3/2$ and the other by $\alpha = 1$[2]. Yet, it cannot exclude the existence of models with some other exponents. Blanchard studied the new dynamic behavior expected when the priority of each incoming task is time-dependent[4]; Bedogne introduced a continuous model of human dynamics instead an infinite queuing list, and found a waiting time distribution explicitly depending on the priority distribution density function[5].

However, there is a missing mechanism in the above models that human as a species with high intelligence is not waiting for 'tasks' to execute. A human cannot be faced with endless 'tasks' which never be all finished as the queuing model assumed. Therefore, researchers investigated models incorporating psychology characters of humans. Vázquez showed the mechanism responsible for these marked nonrandom features is the memory of humans[6]. Humans have some perception of their past activity rate, and based on that they react by accelerating or reducing their activity rate. Han Xiaopu proposed that many actions of humans are mainly driven by personal interests, and introduced an interest-based model[7]. Dean Malmgren demonstrated that the approximate powerlaw scaling of the inter-event time distribution is a consequence of circadian and weekly cycles of human activity, and proposed a cascading non-homogeneous Poisson process which explicitly integrates periodic patterns in activity with an individual tendency to continue participating in an activity[8].

A common feature of all current epidemic models is the assumption that the contact process between individuals follows Poisson statistics[9]. It is proven that the non-Poisson nature of the contact dynamics results in prevalence decay times significantly larger than prediction by the standard Poisson process based models. The decay time of worm propagation is 25 ± 2 days, which is predicted based on a university email dataset capturing the communication pattern between $3,188$ users. In contrast, the Poisson approximation theoretically predicts the value is 0.86 days. Another prediction based on a commercial email dataset is $1-4$ days for Poisson model, and 9 months for human dynamics model[10].

CHINACOM 2010, August 25-27, Beijing, China

DOI 10.4108/chinacom.2010.87

Here, we demonstrate that the approximate power law scaling and heavy tail of the inter-event time distribution is a consequence of habit property of human activities. We subsequently propose a mechanistic model that incorporates human habitual behaviors and the queuing model. Furthermore, we compare the decay time of email worm based on Poisson process with the proposed habit model. Finally, we discuss the implications of our findings on modeling human activity patterns.

II. Empirical patterns and model rules

Habit is a key human attribute which makes up our everyday lives. Almost everything we do involves the use of habitual behaviors. Scientists agree that at least 95% behaviors are habitual, such as sleeping every night, receiving and sending emails everyday, watching movies every weekend. Experiments show that the average person needs at least 3 weeks to form a habit, and 3 months for a steady habit. It can be seen that habitual behavior always represents remarkable regularity that the same kind of action occurs periodically, though the period differs for different behaviors. Practically, a certain kind of behavior may include several tasks which will be executed in some order.

In order to gain some intuition about habitual behavior patterns, let us consider a fictitious person, Jason. Jason arrives at his office 9:00 am as usual. Before starting his work, he habitually opens outlook express to check emails, deletes spams, responses necessary, and sends several. Then he keeps his outlook email online and begins to work. He receives another email at 10:00 and responses immediately. He deals with all emails received until 3:00 pm, then he logs off outlook and gets to be busy summarizing the work of the whole day. At weekends, Jason likes sleeping until 10:00. After waking up, he still opens his email box first to check emails and then spends hours watching movies online. The behavior of Jason is an typical example of white-collar workers, whose activities are both periodic and cascading accounting for his sleep and work patterns.

It can be concluded from the behavior described above that habitual behavior occurs regularly. Each kind of behavior can be treated as a queue consisting of several tasks according to the queuing model. So dealing with emails and watching movies are two independent queues containing emails to check and movies to watch, respectively. After a queue is selected due to habit of individuals, tasks in the queue are completed as a certain order, such like First-In-First-Out, or the highest priority first, etc..

Habitual behaviors make humans repeat an action every certain time, which results in the intervals of consecutive events fluctuating around a period. The steadier the habit is, the narrower the scope of fluctuation is, and vice versa. But even quite steady habitual behavior can not occur exactly according to the period. For example, an everyday habit fluctuating around 22 to 26 hours can be considered as steady. While, the unsteady habitual behavior could fluctuate around only a few hours or hundreds of hours. The distribution function of the interval time of consecutive events is not fixed but updated with completed tasks. The more times an event repeats, the steadier the habit is, and the narrower the scope of fluctuation of the interval time is. It can not be ignored that even a quite steady habit can be broken up by lack of resources or unscheduled vacations, but it can be back to its original states after the break.

The previous researches assumed that tasks arrive at a constant rate, and we still resume the assumption. Then tasks wait to be executed in a queue in their arriving order. In fact, some tasks like emails will never be responded or just be deleted right away. The ignoring or deleting can also be treated as a kind of executing. It should be noticed that a queue with all tasks completed may still wait at the service table. For example, some e-mail client softwares can notify arriving emails to the users, so emails can be checked at any moment. Similarly, the technology of RSS leads users to focus the latest news at any moment. The duration of events is also affected by habit. Besides, another kind of events driven by human interest which can not be seen as 'tasks', like watching movies or playing games, are completely determined by habitual behaviors. For example, people like to recreate themselves at weekends, so the recreation events last 2 or 3 days.

Based on the above ideas, the rules of our model are as follows:

1) To keep periods of habitual behavior steady in most cases, our model assumes that the distribution function of intervals of consecutive events in the same queue follows the normal distribution:

$$p(\tau,t)=\frac{\exp(-\frac{(\tau-T)^2}{2(\exp(\frac{1}{\xi}\sigma(r(t),f(t))))^2})}{\sqrt{2\pi}\exp(\frac{1}{\xi})\sigma(r(t),f(t))} \quad (1)$$

where $p(\tau,t)$ is distribution function of event occurring at time t. The parameter σ is a function of $r(t)$ and $f(t)$, where $r(t)$ is repeat times of events until time t, and $f(t)$ is the average interval time of happened events. So σ is changed with these two variables.

2)$\sigma(r(t),f(t)) = a\times\exp(\frac{1}{r(t)}) + b\times\ln(|f(t)-T|+1)$ where a and b are variables. The two parts control affection of repeating times and average interval time to distribution respectively. The happened times of events are initialized as 1, $r(t)$ as 1, and $\exp(\frac{1}{r(t)})$ as e. The variable a multiplied with $\exp(\frac{1}{r(t)})$ is to enlarge the scope of fluctuations of initial interval time of consecutive events. With repeat of events, $r(t)$ increases, $\exp(\frac{1}{r(t)})$ decreases, and distribution of intervals and habit of individual tend to be stable. At the limit condition, $r(t)$ tends to infinite when the distribution is most stable. The latter part of the formula, $b\times\ln(|f(t)-T|+1)$, is to control affection of average interval time of happened events to the distribution. When the difference between $f(t)$ and the standard period T decreases, $\ln(|f(t)-T|+1)$ and σ decrease, and the intervals tend to be stable. Under an ideal condition, $f(t)$ equals to T, which means individual behaves periodically exactly, and $\ln(|f(t)-T|+1)$ is 0, the interval time is most stable. Under a limited condition when $f(t)$ tends to be infinity, and so do $\ln(|f(t)-T|+1)$ and σ, the fluctuation is violent, the habitual behavior can hardly exist. The variable b is to magnify or diminish value of $\ln(|f(t)-T|+1)$ properly to control the affection of average interval time and the variable σ. It can be concluded from this rule that the stable degree of habit determines the scope of fluctuation of the interval time of consecutive events.

3) The variable ξ is a random number in $[0, 0.1]$ to simulate random interruptions in reality. The range is so narrow that violent fluctuations occur at a very low probability.

4) It is assumed in the queuing model that all tasks are independent and arrive at a constant rate[12]. Our model assumes that the distribution of the interval time of tasks arriving follows the normal distribution, so that the arriving speed is stable:

$$p(t_{arv}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(t_{arv}-T_{arv})^2}{2}\right) \tag{2}$$

5) Every selected queue keeps waiting at the service table for a while t_{last} which follows the normal distribution:

$$p(t_{last}) = \frac{1}{\sqrt{2\pi} \times \sigma_{last}} \exp\left(-\frac{(t_{last}-T_{last})^2}{2\sigma_{last}^2}\right) \tag{3}$$

This formula is to simulate continuous attention of individuals to an event.

6) The queuing model considers service process as the Poisson process, and tasks are executed by First-in-First-out, random selection, or the highest priority first[12]. Our model simply chooses the First-in-First-out, which can also be considered as the highest priority first where priority is assigned according as arriving order. The consuming time of tasks follows Poisson distribution:

$$p(t_{proc} = k) = \frac{\lambda_{proc}^k}{k!} \mu e^{-\lambda_{proc}} \tag{4}$$

The variable μ is computed as follows:

$$\mu = \begin{cases} 1, & \omega \geq \Omega \\ \frac{1}{\omega}, & \omega \leq \Omega \end{cases} \tag{5}$$

where ω is a random number in (0,1), and Ω is an appointed threshold to express the sporadic events with very long executing time.

III. Analysis and Discussion

To verify our model, we analyze the dataset from Kiel University which records sending and receiving emails[11], and another dataset from Netflix (http://www.netflixprize.com) recording movies ordering online.

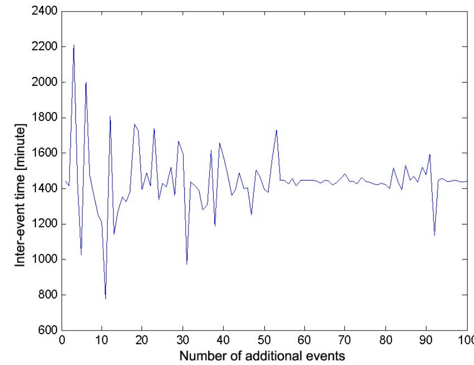
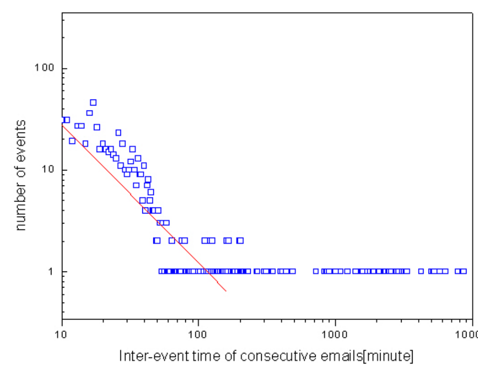
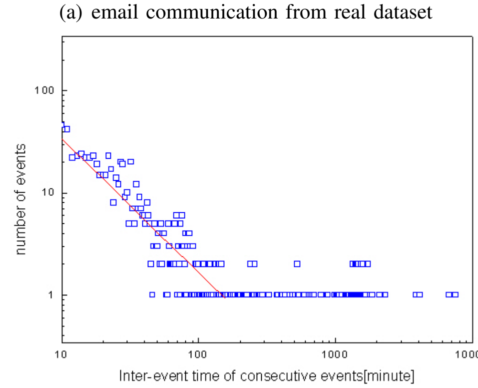


Fig. 1. Distribution of the interval time of consecutive events generated by our model. The interval time fluctuates around $1,440$ minutes, and especially violent at the beginning. With repeat of events, it tends to be stable. The burst at the tail can be seen as an interrupt by occasional break after which the distribution resumes stable.

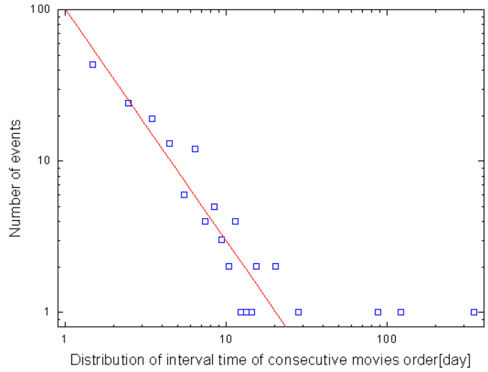


(a) email communication from real dataset

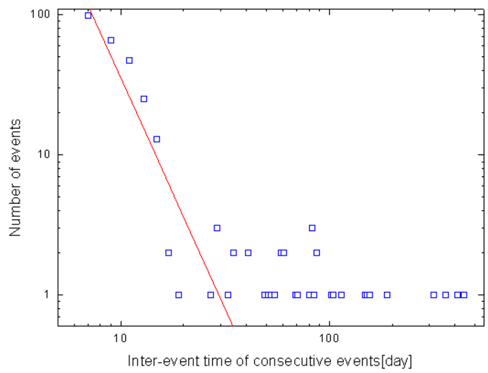


(b) data generated by the habit model

Fig. 2. Comparison of the email recording dataset and data generated by our model. Figure 2a shows the distribution of the interval time of consecutive emails sending by one user in 3 months, which can be fitted by power-law distribution with exponent -0.89 and with heavy tail. Figure 2b shows the distribution of interval time of consecutive events simulated by our model, which can be fitted by power-law distribution with exponents -0.8996, and with heavy tail.

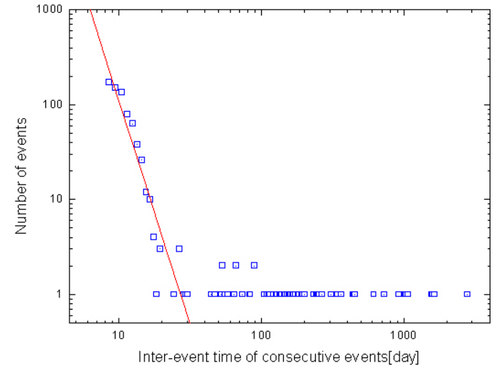


(a) movie order record from real dataset

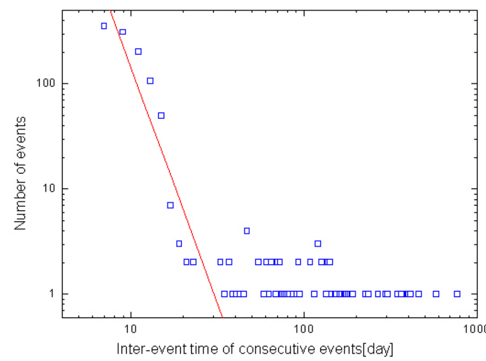


(b) data generated by the habit model

Fig. 3. Comparison of the movies online order dataset and data generated by our model. Figure 3a shows the distribution of interval time of consecutive movies order from the record of Netflix, which can be fitted by power-law distribution with exponent -0.9255 and with heavy tail. Figure 3b shows distribution of the interval time of consecutive events simulated by our model, which can be fitted by power-law distribution with exponents -0.9061, and with heavy tail.



(a) simulation results of $2,000$ executions based on the habit model



(b) dsimulation results of $2,000$ executions based on the habit model

Fig. 4. Distribution of the interval time of consecutive events simulated by our model, which can be fitted by power-law distribution with exponents -0.94431 and -0.92853, and both with heavy tails.

A. Compared with the real dataset of emails

The dataset is from Kiel University including emails of $9,665$ users during three months. Each record consists of sending time and the ID of senders. According to our model based on habit, the distribution of the interval time of consecutive events is shown as Figure 1. It can be seen that with the repeat of events, habit becomes more and more steady. But even a steady habit can be interrupted by occasional break which causes the violent fluctuation at the tail. But when the break is over, the habit can resume to its origin states. Besides, habit is formed after 2 months repeat. Figure 2 describes the data generated by our model and from the email dataset. Comparing with the real data, there are more data generated by our model gathering around 60 minutes, while the real data gather around over ten minutes. This is caused by the assumption of our model that the interval time of consecutive tasks arrive is about 1 hour, and every task has to be executed. But users do not need to response every email in reality, and may not deal with the email received immediately. However, the most important is that our model expresses the very long periods of inactivity separating bursts of intensive activity, which is similar to the reality that users deal with emails during work time and turn to focus other things after work until the next day. Besides, it can be seen from Figure 2a that the distribution of the interval time of consecutive emails sending can be fitted by power-law distribution with a heavy tail which allows long periods of inactivity. The similar phenomenon also happens in Figure 2b. The fitted exponent 0.89 is close to the conclusion that email electrommunication patterns follow the power-law distribution with exponent 1. And our model can generate random interrupt to break habit occasionally, which causes the data showing around $10,000$. This phenomenon also exists in reality.

B. Compared with the dataset of movie order

The dataset is from Netflix, one of the most online movie share websites, which shares a database (only part of its whole database) containing $17,770$ movies, $447,139$ users, and almost 100 million ordering records. The unit used in the records is day, so we ignore multi-order in one day and compute interval days between two consecutive orders of a single user by tracking his order records.

Figure 3 describes the data generated by our model and from the email dataset. As the unit is day, the data quantity is not large enough. But the heavy tail is still obvious. Compared with the real data, there are more data generated by our model gathering around 2 or 3 days. The interval time of ordering movies is about 7 days, and fluctuates around over ten days, which makes the distribution decrease almost linearly. The violent fluctuation caused by interrupt causes a heavy tail.

As the real dataset is too small, we rerun the simulations and specify the times of task execution as $2,000$, two results of which are shown in Figure 4. Because of the period is set as 7 days, the interval time gathers around this value and decreases linearly after that. Actually, the unit forces the distribution fluctuate only around 1 or 2 days to hundreds of days, so the power-law appearing is not satisfied. While the more important is that the linear decrease and the obvious heavy tail can be a reference to searching the potential reason of busty and heavy tail in human dynamics.

IV. Impact of habit on worm propagation

We study the spread of email worms among email users based on the proposed habitual behavior model of human dynamics to investigate the impact of the observed non-Poisson activity patterns on spreading processes. The spreading process can be simply described by the susceptible-infected (SI) model on the email network. It is assumed that only one node is infected at the beginning. Each node communicates with other nodes according as Poisson model or human behavior model. More nodes are infected with the communications between nodes in the network. Figure 5 describes the consuming time of infecting all nodes in networks of different sizes based on Poisson model and the habitual behavior model. It can be concluded that the consuming time increases significantly with extending of network based on the habitual behavior model, while the time increases very slowly based on Poisson model. The consuming time of the two models in network of small size do not differ much, but the time using the habitual behavior model is several times larger than the Poisson. Besides, it takes $33,065$ minutes to infect all the nodes in the network of $3,000$ nodes, which is consistent with the prediction of 25 ± 2 days in [10].

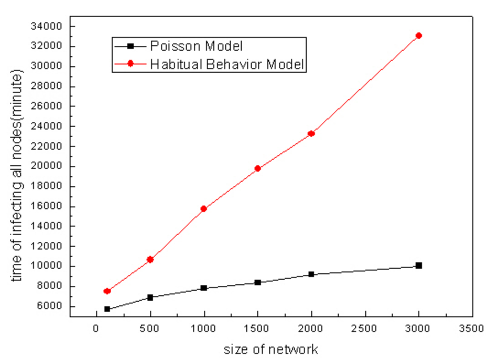


Fig. 5. Consuming time of infecting all nodes in different sizes network

V. Conclusion

In this research, we propose a human dynamics model incorporating the queuing model and psychology factors, imitate habitual behavior with the normal distribution, adjust scope of interval fluctuation according to repeat times and stable degree of behaviors, which causes habit to be stable after numbers of repeat. Besides, we force a queue waiting at the service table for a while, which reflects the situation that humans always pay attention to an event for several hours or days. Actually, most researches on human dynamics were based on the queuing model introduced by Barabási, which considered long waiting time of low priority tasks as response for the bursty and heavy-tailed features. Different from that, we find out human behaviors are so various and complex that the same event can only occur after a long period, and the time spending on it must be relatively ephemeral. Simulations show that our intuition is reasonable, and the model results can fit the real data with obvious power-law and heavy-tailed features. Furthermore, we study the impact of human dynamics on virus spreading, and prove that the consuming time of infecting all nodes in network based on human dynamics model is much more than the prediction based on Poisson model, and the difference grows with extending of network.

Human dynamics with non-Poisson features has caused great concern since it was proposed because the research has important theoretical significance and commercial values, though there is still huge research space in evidence and modeling. The model we propose is to provide a referential framework for recovering the potential mechanism. Our future work will focus on the deep reason for longer decay time caused by human behavior.

Acknowledgment

This work was supported in part by National Nature Science Foundation under Grant 60973136, the China - British Columbia Innovation and Commercialization Strategic Development under Grant 2008DFA12140, Jilin Province Science and Technology Development Plan under Grant 20070708, the ”985” Graduate Student Innovation Fund of Jilin University Grant 20101029 and the Europe Cooperation Project under Grant 155776-EM-1-2009-1-IT-ERAMUNDUS-ECW-L12.

References

[1] Haight Frank A. Handbook of the Poisson distribution. New York: Wiley, 1967

[2] Vázquez A, Dezso Oliveira Jg. Modeling bursts and heavy tails in human dynamics. Physical Review E, 2006, 73(3): 36127

[3] Barabási, AL. The origin of bursts and heavy tails in human dynamics. Nature, 2005, 435(7039): 207-211

[4] Blanchard Ph, Hongler M. -O. Modeling human activity in the spirit of Barabsi’s queueing systems. Physical Review E, 2007, 75(2): 26102

[5] Bedogne C., Rodgers G. J. A continuous model of human dynamics. Physica A, 2007, 385(1): 356-362

[6] Vázquez A. Impact of memory on human dynamics. Physica A-Statistical Mechanics and its Applications, 2007, 373: 747-752

[7] Han XP, Zhou T, Wang BH. Modeling human dynamics with adaptive interest. New Journal of Physics, 2008, 10: 073010

[8] R. Dean Malmgren, Daniel B. Stouffer, Adilson E. Motter, Luis A.N. Amaral. A Poissonian explanation for heavy-tails in e-mail communication, PNAS, 2008, 105(47): 18153-18158

[9] Barthélemy M, Barrat A, Pastor-Satorras R, Vespignani A. Velocity and hierarchical spread of epidemic outbreaks in scale-free networks, Physical Review Letter, 2004, 92: 178701.

[10] Vázquez A, Rácz Balázs, Lukács András, Barabási, AL. Impact of non-Poisson activity patterns on spreading processes, Physical Review Letter, 2007, 98: 158702.

[11] Ebel H, Mielsch L-I, Bornholdt S. Scale-free topology of e-mail networks. Physical Review E, 2002, 66(3): 35103