Physica A 473 (2017) 313–318

Contents lists available at ScienceDirect

Physica A

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

Double dynamic scaling in human communication dynamics

Shengfeng Wang [a], Xin Feng [b], Ye Wu [a,*], Jinhua Xiao [a]

[a] School of Science, Beijing University of Posts and Telecommunications, Beijing 100876, China

[b] School of Management Science and Engineering, Hebei Geo University, Shijiazhuang, Hebei, 050031, China

HIGHLIGHTS

- In this work, Our contribution is to provide further evidence for the view that the initiator of bursty activity is a Poisson process.
- The paper prove again that even if an inter-event time follows power-law distribution which is guaranteed by KS test, the inherent dynamics are double scaling and the special threshold which is called midrange memory length can be calculated.
- Compare with previous research, we introduce a method to mark the random initial event from the complex dynamics.
- The detection of the random initial event has widely application in topic clustering and topic identification.

ARTICLE INFO

Article history:
Received 11 August 2016
Received in revised form 16 November 2016
Available online 4 January 2017

Keywords:
Communication pattern
Memory length
Heavy tail

ABSTRACT

In the last decades, human behavior has been deeply understanding owing to the huge quantities data of human behavior available for study. The main finding in human dynamics shows that temporal processes consist of high-activity bursty intervals alternating with long low-activity periods. A model, assuming the initiator of bursty follow a Poisson process, is widely used in the modeling of human behavior. Here, we provide further evidence for the hypothesis that different bursty intervals are independent. Furthermore, we introduce a special threshold to quantitatively distinguish the time scales of complex dynamics based on the hypothesis. Our results suggest that human communication behavior is a composite process of double dynamics with midrange memory length. The method for calculating memory length would enhance the performance of many sequence-dependent systems, such as server operation and topic identification.

1. Introduction

There are many complex systems in nature and in human society. Scientists are in pursuit of a deep understanding of such systems because of their value in making predictions. Human communication behavior dynamics has been a popular topic of research in recent years due to the huge quantities of data and server operations available for study. Several impacts of communication patterns have already been discovered. Vazquez et al. [1] first found that the pattern of communication dynamics results in the prevalence of large decay times; their predictions are in accord with detailed time-resolved prevalence data of computer viruses. Ni et al. [2] discussed the impact of inter-active time distribution on viruses' survival times and probability. They found that low, heterogeneous active times lead to a small survival probability, but a large survival time. Takaguchi et al. [3] numerically compared the voter model with long-tailed interaction times on different networks.

* Corresponding author.
E-mail address: wuyeby@163.com (Y. Wu).

http://dx.doi.org/10.1016/j.physa.2017.01.010
0378-4371/

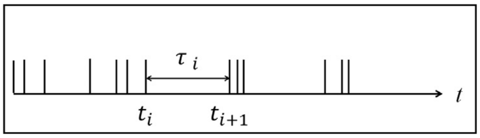

Fig. 1. Schematic diagram of the user time series. The horizontal denotes time and the longitudinal represents a communication event. The IET is calculated using the subscript time stamps.

In the field of human dynamics, a crucial physical quantity is the inter-event time (IET) interval τ. Initially, the dynamics of human communication are widely assumed to be Poisson processes [4], in which an event occurs at a constant rate λ and the IET between two consecutive individual actions follows an exponential distribution $P(\tau) \propto \lambda e^{-\lambda}$. The stationary Poisson processes are memory-less, which means that communication behavior is independent and seemingly unpredictable; such processes are widely used in the quantitative analysis of communication systems.

The current availability of big data has made it much easier for us to quantitatively probe human behavior. In 2005, Barabasi [5] explored letter and email data to determine that IET follows a non-Poisson statistic. The distribution could be described by the power law $P(\tau) \propto \tau^{-\alpha}$. IET interval distributions conforming to the power law distribution were then found in more subjects [6–9]. The power law distribution indicates that the time interval between two successive communications is non-uniform. In contrast, communication alternates between intensive outbreaks over a very short period of time and long periods of silence. In general, the power law distribution represents a long-term correlation, meaning that each individual communication event is not an independent process and that the history of communication events has an impact. So the human behavior can be predicted in long term. Various mechanisms are proposed to explain the heavy tail in human behavior, including priority-queuing processes [10], adaptive interests [8].

In contrast to the above mechanism, Malmgren et al. [11,12] suggested that human correspondence patterns are driven by a cascading, nonhomogeneous Poisson process. The model of cascading process result in that the events within high-activity bursty intervals are independent. A Poisson process as the initiator of localized bursty activity introduced by Wu et al. [13] to explain the bimodal distribution of SM data. For small inter-event time, the distribution follow power law, but the tail of the distribution follow the exponential distribution. The distribution of human communications explained by these model originate a mixture distribution with different time scales.

In this paper, we investigate the widely used hypothesis that the initiator of high-activity bursty intervals are mutual independent. We use a threshold τ_p to distinct different bursty intervals. The initiator of bursty activities are proved to be mutual independent for both bimodal distribution and power law like distribution in empirical data. The behavior beyond the scope of τ_p is independent and inherently impossible to forecast, besides it could be well correlated and forecasted. Thus, the threshold τ_p is a meaningful value called midrange memory length which quantitatively distinguish the time scales of different dynamics. Finally, we apply our method to earthquake data, and find the same pattern. The measurement of the memory length in the communication system can provide the theoretical basis for the design of the topic recognition and operators optimization.

2. Materials and methods

2.1. Data

This paper reanalyzes two datasets. One from a mobile phone operator, contains 65909 users and 130099426 messages over 6 months. Each record comprises a sender mobile phone number, a recipient mobile phone number and a time stamp with a precision of 1 s. The second is the Eckmann et al. e-mail dataset [14], which contains time stamps for about 3×10^5 e-mails between 3188 people over 83 days. A typical time sequence is shown in Fig. 1, in which the horizontal axis denotes time and each vertical line corresponds to a communication event (i.e., a short message or email). Let $\{t_1, t_2, t_3, \ldots, t_n\}$ denote the time sequence of an user, so that the IET is $\tau_i = t_{i+1} - t_i$. First, we plot the IET distribution of users, and find that both bimodal distributions proposed by Wu et al. [13], and a power-law-like distribution (Fig. 2). In the statistics, some real-world data (Fig. 2(b) (c)) can be well-fitted by a power law, and guaranteed by a goodness-of-fit test [15–17]. Marton Karsai et al. [18] indicated that the communications have correlations. Wu et al. [13] studied a modified version and introduced a Poisson process as the initiator of localized bursty activity to explain the bimodal distribution. This raises the question of whether the different bursty intervals of empirical data are mutual dependent or not. Next, we introduce a method to detect the initiator of different bursty intervals to verify the problem.

3. Method

In this section, we discuss the approach to divide the sequences into different bursty activities and investigate whether these initiator of bursty activities are mutual independent or not. In Fig. 3(a), in a user's time sequences, all events can be represented by time stamp t_i $(i = 1, 2, 3, \ldots, k, k+1, \ldots, n)$, and the interval time between two successive events is $\tau_i = t_{i+1} - t_i$. It is assumed that there is a threshold τ_p. When the IET $\tau_k \leq \tau_p$, where $k = 1, 2, 3, \ldots, n$, the two events