复杂系统中的阵发性和记忆性

Kwang-Il Goh、Albert-László Barabási

摘要

在包括邮件到地震的各种各样的真实系统的动力学中,经常有阵发性的规律存在,表现为短时间的剧烈活动然后是长时间的停止或减弱。对这类现象的理解不足,是因为缺乏可以把不同系统在同一个框架下比较的统一工具。我们引入两种度量来区别阵发性、记忆性两个机制。我们发现尽管阵发性源于时间间隔的分布和记忆性,但由于人类动力学中的记忆性很弱,所以阵发性是由时间间隔分布引起的。最好,我们表明现在的模型还不足以再现实际观察到的规律,所为未来的工作打开了大门。

复杂系统的很多动力学过程是由彼此之间松散连接的小部分引发的,比如社会和细胞。尽管我们这背后的网络的研究中有很多进展 [1], 但我们对动力学的理解进展很慢。随着现在我们有越来越强的能力监控具有时间分辨度的系统,比如电子邮件 [2, 3, 4]、上网 [5] 以及基因表达 [6],我们现在有机会问一个重要的问题: 复杂系统的动力学是由一个通用的原理支配着,还是每个系统有各自的特点?尽管要确定地回答这个问题很困难,但有越来越多的证据表明不同系统都有一个共同的特点:系统活动的阵发性。

阵发性,粗略地说,就是短时间的活跃程度增强然后是长时间的沉寂。阵发性在很多系统里都有发现,包括邮件 [3]、地震 [7, 8] 和基因表达 [6]。但是很多时候阵发性没有确切的度量,而且起源有争议。在人类动力学中,阵发性被认为是源于响应时间的胖尾分布 [3, 4]。然而,在地震和气候现象中,记忆性被认为扮演者重要角色 [9, 10]。阵发性的存在可以影响病毒传播 [3] 和资源分配 [11]。并且,规律的、非阵发的心跳是疾病的征兆 [12]. 出现阵发性的系统如此之多,有必要将其量化。我们的这篇文章是这个方向的第一步,我们制定了度量来帮助量化阵发性的强度和原因。

我们假设系统的活动已经被映射成离散的信号。如果事件的几率不随时间变化,那么活动是个泊松过程。这样相邻事件的时间间隔 (τ) 服从指数分布,$P_{\mathrm{P}}(\tau) \sim \exp(-\tau/\tau_0)$ (图 1a). 一个明显阵发(或者非阵发)的信号可以当 $P(\tau)$ 不同于指数的时候产生, 比如图 1b 的阵发信号和图 1c 的非阵发信号。但是改动间隔时间分布不是造成阵发的唯一办法。比如图 1d 有和图 1a 一样的 $P(\tau)$ 但是阵发性不一样。这是通过改变记忆性做到的:图 1d 中短的时间间隔倾向于跟随短间隔,从而看起来是阵发的。但在图 1e 中,非阵发性是由于另一个方向的记忆性:短的倾向于跟随长的。因此,阵发性是由两个机制引起的:间隔时间分布的改变和记忆性。为区别,我们引入阵发性参数 Δ 和记忆性参数 μ, 来分别量化两者的作用。

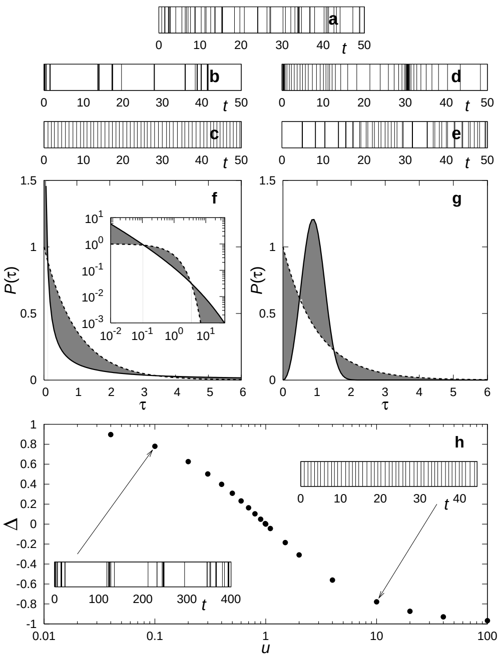

图 1: (a) 泊松分布的信号. (b,c) 改变间隔时间: 阵发信号 (b) 由幂律 $P(\tau) \sim \tau^{-1}$ 产生, 非阵发信号 (c) 由 $m = 1$、$\sigma = 0.1$ 的高斯分布产生. 阵发信号 (d) 通过打乱 (a) 以增加记忆性产生. 非阵发、负记忆性的信号 (e) 也是打乱产生的.(a), (d) 和 (e) 有相同的间隔时间分布。

阵发性参数 Δ 定义为

$$\Delta \equiv \frac{\operatorname{sgn}(\sigma_\tau - m_\tau)}{2} \int_0^\infty |P(\tau) - P_{\mathrm{P}}(\tau)| d\tau \ , \qquad (1)$$

其中 m_τ 和 σ_τ 是均值和标准差 [13]. Fig. 1f–h 解释了 Δ, 我们比较了具有泊松间隔时间分布的阵发、非阵发的 $P(\tau)$。当长、短的时间间隔比随机序列出现得更多时,序列看上去是阵发的 (图 1f)。如果平均长度的间隔更多、长短间隔更少,看起来就不是阵发的 (图 1g)。$\Delta \in (-1, 1)$, 其大小对应于阵发性: $\Delta = 1$ 意味着阵发性最强,$\Delta = 0$ 意味着中性(泊松), $\Delta = -1$ 对应于规律(周期)信号。

很多复杂系统具有异质性：一些部分很活跃,一些不是。比如有的人每天很多邮件,有的人只有一两封。我们可以把活动程度相似的放在一起比较 $P(\tau)$ 。如图 2 所示，曲线被平移了．如果我们把 $\tau_0 P(\tau)$ 作为 τ/τ_0 的函数画图，其中 τ_0 为平均间隔时间，数据会落在曲线 $\mathcal{F}(x)$ 上 (图 2)，表明间隔时间服从 $P(\tau) = (1/\tau_0)\mathcal{F}(\tau/\tau_0)$，其中 $\mathcal{F}(x)$ 与平均活动强度无关，表示系统的一个特征 [8, 15]。这样提出一个重要的问题:Δ 依赖于 τ_0 吗? Δ 随着时间放大缩小不变,随着 $\tilde{\tau} \equiv \tau/\tau_0$ and $\tilde{P}(\tilde{\tau}) \equiv \tau_0 P(\tau)$，由于 $\tilde{\Delta} \equiv \int_0^\infty |\tilde{P}(\tilde{\tau}) - \tilde{P}_0(\tilde{\tau})| d\tilde{\tau} = \int_0^\infty |\tau_0 P(\tau) - \tau_0 P_0(\tau)| d(\tau/\tau_0) = \int_0^\infty |P(\tau) - P_0(\tau)| d\tau \equiv \Delta$，也就是说其刻画了 $\mathcal{F}(x)$.

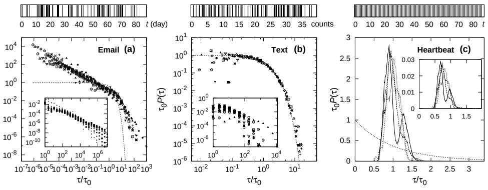

图 2: 真实信号的间隔时间分布 $P(\tau)$。(a) 一个大学电子邮件的 $P(\tau)$ [2]. τ 对应于一个用户两次发邮件的间隔。(b) 相邻两个字母在 *David Copperfield* 中出现的间隔 [16]. (c) 心律的间隔时间 [21]. 我们还画出了指数间隔分布作为参考 (点)。

信号的记忆性系数 μ 被定义为相继两个序列的关联系数。即,给定所有相继的时间间隔对 $(\tau_{k,i}, \tau_{k,i+1})$, $\{k = 1, \cdots, N\}$,

$$\mu \equiv \frac{1}{N} \sum_{k=1}^{N} \sum_{i=1}^{n_k - 1} \frac{(\tau_i - m_{k1})(\tau_{i+1} - m_{k2})}{\sigma_{k1}\sigma_{k2}} , \qquad (2)$$

其中 N 是系统组成部分的数量,n_k 是该部分的事件数目,$m_{k1}(m_{k2})$ 与 $\sigma_{k1}(\sigma_{k2})$ 分别是 $\tau_{k,i}$ $(\tau_{k,i+1})$, 的均值和标准差。当长间隔倾向与跟随长间隔时,记忆性系数是正的;反之是负的。μ 与 τ_0 无关。

将复杂系统映射到 (μ, Δ) 空间 — 我们已经指导阵发性有两个可能的原因，对他们可以用 μ 和 Δ 两个参数来刻画，画在 (μ, Δ) 相图中 (图 3)。作为第一个例子,我们测量了不同文本中一个字母相继出现的间隔 [16]. 对于这些信号，我们发现 $\Delta \approx 0$，即间隔时间分别接近于指数分布 (图 2b),同时 $\mu \approx 0.01$，缺乏记忆性。在相图原点附近的信号 (图 3) 可以用泊松过程描述。相反，自然现象，比如地震 [17] 和天气 [18] 在对角线的边上，表明 $P(\tau)$ 和记忆性对于阵发性的贡献差不多。然而，对于人类活动,例如电子邮件、电话、上网 [2, 5, 4, 20]，情况就不同了。对于这些现象我们发现大的 Δ 和可以忽略的小 μ，表明这些系统的记忆性来自 $P(\tau)$，记忆性对于异质性的贡献很小。记忆性小是理所应当的,因为它表明它们相对于自然现象的可预测性小,而自然现象的强记忆性使得其可以被预测。最后,对于心率 (图 2c) [21]，我们发现对于健康人 $\Delta_{\text{cardiac, healthy}} = -0.73(4)$，对于心脏衰竭的人 $\Delta_{\text{cardiac, CHF}} = -0.82(6)$ 两者都很规律。因此 Δ 表明心脏衰竭的人比健康的人心跳更规律 [12]。进一步,我们发现 $\mu \approx 0.97$,,表明记忆性在心律的规律性中有很大作用。

各种系统在 (μ, Δ) 相图上有规律的分布着:人类行为集中在 Δ 高 μ 低的区域，自然现象在对角线附近，心律在 μ 高 Δ 为负的区域,语言文字靠近 $\Delta = \mu = 0$，表明这些系统的时间行为由不同种类的动力学机制支配着。

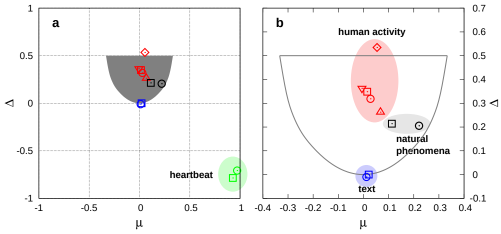

图 3: (a) (μ, Δ) 相图。人类行为 (红色) 包括电子邮件 (□) [2]、图书馆借阅 (○) [4]、大学里的打印 (◇) [19] 、一个银行的通话 (△) [20] 以及一个公司的拨打电话行为 (▽)。自然现象的数据包括日本的地震 (○) [17] 和新墨西哥州的每日降水量 (□) [18]。语言文字数据 [16] 包括英文版的 *David Copperfield* (○) 和匈牙利文版的 *Isten Rabjai* (□). 生理现象的数据 (绿色) 包括正常人 (○) 和心脏衰竭者 (□) 的心律 [21]. 灰色是 2-状态模型的区域 [23]。

实证数据在相图中有不同区块,我们不禁要问：我们现在的模型对于复现这些现象表现得如何？排队模型被用来描述等待时间 [3, 4, 22],画在这里没有意义。阵发性信号可以由 2-状态模型产生 [23]，以概率 p 切换对应于两种速率 $\lambda_0 < \lambda_1$ 的泊松过程。长时间下,当 $p > 0$，Δ 独立于 p ，取值范围为 $0 < \Delta < 0.5$, approaching 0 当 $\lambda_0 \approx \lambda_1$ 时接近 0,当 $\lambda_1 \to \infty$ 且 $\lambda_0 \to 0$ 时接近 0.5。模型的记忆性为 $\mu = A(0.5 - p)$，其中 A 依赖于 λ_0 和 λ_1 的常数。模型在 (μ, Δ) 空间所能表示的范围由图 3a 的灰色部分标出。模型可以通过调整参数覆盖所有观察到的范围。但是,这可能是个假象：比如人类活动的 $P(\tau)$ 是胖尾的,不同于模型。这表明 Δ and μ 只提供了阵发性原因的第一阶近似,而我们需要在比较系统的时候考虑其它的度量，比如 $P(\tau)$ 的函数形态。这也表明现在还没有能阵发性背后机制的细节的模型,未来还有很多工作可以做。

参考文献

[1] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks* (Oxford University Press, Oxford, 2002); M. E. J. Newman, SIAM Rev. **45**, 167 (2003); R. Pastor-Satorras and A. Vespignani, *Structure and evolution of the Internet* (Cambridge University Press, Cambridge, 2003); S. Boccaletti, *et al.*, Phys. Rep. **424**, 175 (2006); M.

E. J. Newman, D. J. Watts, and A.-L. Barabási (eds.), *Structure and Dynamics of Complex Networks* (Princeton University Press, Princeton, 2006).

[2] J. P. Eckmann, E. Moses, and D. Sergi, Proc. Natl. Acad. Sci. U.S.A. **101**, 14333 (2004).

[3] A.-L. Barabási, Nature (London) **207**, 435 (2005); A. Vázquez, Phys. Rev. Lett. **95**, 248701 (2005).

[4] A. Vázquez, *et al.*, Phys. Rev. E **73**, 036127 (2006).

[5] Z. Dezső, *et al.*, Phys. Rev. E **73**, 066132 (2006).

[6] I. Golding, *et al.*, Cell **123**, 1025 (2005); J. R. Chubb, *et al.*, Curr. Biol. **16**, 1018 (2006).

[7] P. Bak, *et al.*, Phys. Rev. Lett. **88**, 178501 (2002).

[8] A. Corral, Phys. Rev. E **68**, 035102(R) (2003).

[9] A. Bunde, *et al.*, Phys. Rev. Lett. **94**, 048701 (2005).

[10] V. N. Livina, S. Havlin, and A. Bunde, Phys. Rev. Lett. **95**, 208501 (2005).

[11] W. E. Leland, *et al.*, IEEE/ACM Trans. Networking **2**, 1 (1994).

[12] S. Thurner, M. C. Feurstein, and M. C. Teich, Phys. Rev. Lett. **80**, 1544 (1998).

[13] As an alternative, σ_τ/m_τ can also be used intead of Δ to meausre burstiness [A. Vázquez, private communications].

[14] J. Laherrère and D. Sornette, Eur. Phys. J. B **2**, 525 (1998).

[15] A. Saichev and D. Sornette, Phys. Rev. Lett. **97**, 078501 (2006).

[16] Project Gutenberg, http://gutenberg.org.

[17] Japan University Network Earthquake Catalog, http://wwweic.eri.u-tokyo.ac.jp/CATALOG/junec/.

[18] National Resources Conservation Service, http://www.nm.nrcs.usda.gov/snow/data/historic.htm.

[19] U. Harder and M. Paczuski, Physica A **361**, 329 (2006).

[20] I. Guedj and A. Mandelbaum, http://iew3.technion.ac.il/serveng/callcenterdata/.

[21] PhysioBank, http:// www.physionet.org/ physiobank/.

[22] J. G. Oliveira and A.-L. Barabási, Nature (London) **437**, 1251 (2005).

[23] J. Kleinberg, Proc. ACM SIGKDD ’02, pp. 91 (2002).