

科学研究費助成事業 研究成果報告書

平成 27 年 9 月 25 日現在

機関番号：12102

研究種目：若手研究(B)

研究期間：2013～2014

課題番号：25730184

研究課題名(和文) RT拡散に基づくツイッターバースト時系列の理論的解析とウェブサービスへの応用

研究課題名(英文) Theoretical analysis of Twitter burst time series based on the RT diffusion and its application

研究代表者

岡 瑞起 (OKA, MIZUKI)

筑波大学・システム情報系・助教

研究者番号：10512105

交付決定額(研究期間全体)：(直接経費) 3,200,000円

研究成果の概要(和文)：長い間人間の時間的振る舞いはランダムな過程(ポワソン過程)に従っていると思われていた。しかし、2005年にアルバート・バラバジを中心とする研究グループが、人間の行動は間欠的な振る舞い(バースト)がより顕著であると考え、そうしたバースト的な振る舞いの重要性を説いた。本研究は、ウェブ上のソーシャルメディアからの大量データを分析することで、バーストの詳細なメカニズムを明らかにすることを目的とした。分析の結果、バーストが起こるメカニズムを内因的バーストと外因的バーストに分類し、これら2つのバーストを分ける臨界的な揺らぎの閾値が出現することを明らかにした。

研究成果の概要(英文)：For a long time, the temporal behavior of human beings had appeared to follow the random process (Poisson process). However, in 2005, Albert Barabasi's research group considered human behavior as intermittent behavior (burst) and preached the importance of such a burst-like behavior. The purpose of this study is to clarify the detailed mechanism of the burst using a large amount of data from the social media. Results of the analysis revealed two classes of mechanisms causing burst; one is the endogenous burst and the other is exogenous burst. We found that there exists a critical fluctuation threshold that separates these two bursts.

研究分野：ウェブ工学

キーワード：ソーシャルメディア分析 Twitter exogeneity endogeneity 興奮性媒体

1. 研究開始当初の背景

長い間人間の時間的振る舞いはランダムな過程(ポワソン過程)に従っていると思われていた。しかし、2005年にアルバート・バラバシらを中心とする研究グループが、人間の行動は間欠的な振る舞い(バースト)がより顕著であると考え、そうしたバースト的な振る舞いの重要性を説いた。このようなバースト現象の重要性は以前より指摘されているが、それがウェブ上のさまざまオンラインシステムに関し、大量なデータの上で実証されるようになったことが研究を促進している。たとえば、ツイッターの時系列やグーグルのクエリ-の時系列にバーストするパターンが見つかるが、その詳細なメカニズムは明らかになっていない。

2. 研究の目的

本研究では、このバーストのメカニズムをツイッターに注目して新たに提唱する。それは他のシステムにはない、2種類のバースト・メカニズムを持つと考えられるからである。バーストモデルの妥当性を実際にウェブから収集した大量のデータを解析することで検証し、その有効性を確かめる。このような研究は、バーストのメカニズム解明にとどまるだけでなく、人間の行為予測に関する新しい知見を得ることで、オンラインコンテンツの収益化に寄与するなど重要な技術である。

3. 研究の方法

本提案では、ツイッターの時系列データ(表1)

を用いて、注目する単語を含んだ時系列が「バースト」するメカニズムを解明するモデルを提案し、解析した。まず、これまで行なってきたツイートデータ収集を引き続き行い、形態素解析などの前処理を通じて単語時系列データを取得した(図1)。同時にバーストパターン、RTされたツイート、時間分解能を変えたデータを時系列として抽出し、バースト的な時系列の特徴付けによる分類を行った。次に、バーストモデルの提案と実証を行った。

Total number of tweets	297,792,366
Total number of users	12,677,098
Total number of keywords	1,550,770
Total number of tweets in the top 3000 keywords	162,358,768
Total number of users in the top 3000 keywords	12,037,771

表1: 収集したツイートデータの概要

4. 研究成果

ツイッターのデータを単語ごとにその出現数の時系列を観察すると、A) 周期性の強いパターン、B) 間欠性の強いパターン、C) 平均値の周りで揺らぐパターンなど、さまざまな変動のパターンが見つかる。さらに細かく見ると、そうした周期性や間欠性の破れのような現象も観察できる。そこで、総出現数の大きな上位3,000単語の頻度時系列に関し、バースト現象(ツイートの量の急激な増加)に注目し、各単語の時系列をバースト区間とそれに先行する非バースト(揺らぎ)区間に分類した。その結果の一例を図2に示す。赤い線は時系列を示し、グレーのボックスはバ

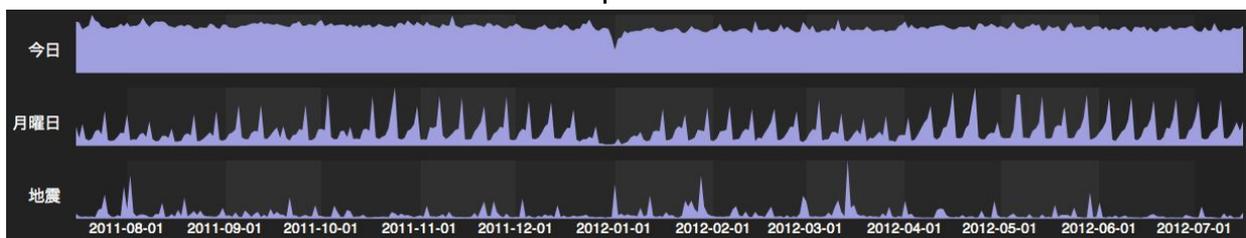


図1: ツイッター単語時系列種類(収集データより抜粋)

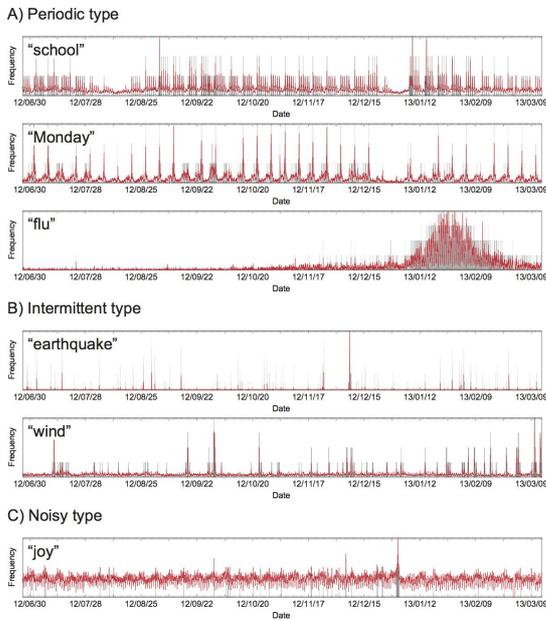


図 2 : 1 年分の Twitter の時系列データ。単語ごとに抽出。

バーストとして検知された区間である。

さらに、バーストが起こるメカニズムを内因的バーストか外因的バーストに分類するために、各バーストのサイズ S に対するピーク p の割合 p/S を計算する。バーストのサイズとは各バースト区間内の総ツイート量、バーストのピークとは各バースト区間で最もツイートの多かった時刻のツイート量である。これまでの先行研究から、が院生バーストの特徴は、 p/S 比が高いことが報告されている。すなわち、外因性バーストはパルス的なものだといえる。つまり外因的バースト

は S に依存しない傾向を持つ。一方、内因的バーストは p がほぼ一定であるため $1/S$ となる。これらの結果を図 3 に示す。

機械的に外因性と内因性バーストを判定するために $y = ax^b$ でフィットし、このフィットしたラインが $b = -1$ よりも上であれば外因性バースト（赤）、下であれば内因性バースト（青）とラベル貼りする。どの単語も純粹に外因的や内因的という分類ができるわけではなく、その度合いは変化する。しかし単語の出現頻度によりその割合は大きく異なっている。

定義上、ベースラインの揺らぎとバースト応答の期間は交互に出現する。そこでベースラインの揺らぎの区間の相対標準偏差の値

σ/E と、直後のバースト区間の相対的なピークの値 p/E を比較してみる。そこには時系列からは予想できないパターンが存在していた。図 4 に示したように、揺らぎ σ/E とバースト応答 p/E がほぼ比例する場合（左）、ある揺らぎの大きさを境にバースト応答が大きく分散する場合（中）、揺らぎの幅が狭くバースト応答が大きく分散する場合（右）が観察された。また、上記で判定した外因性バーストと内因性バーストの結果を重ね、赤青で色付けを行うと、揺らぎの閾値（臨界的な c ）はバーストの内因性 / 外因性起源の境界になっていることがわかる。

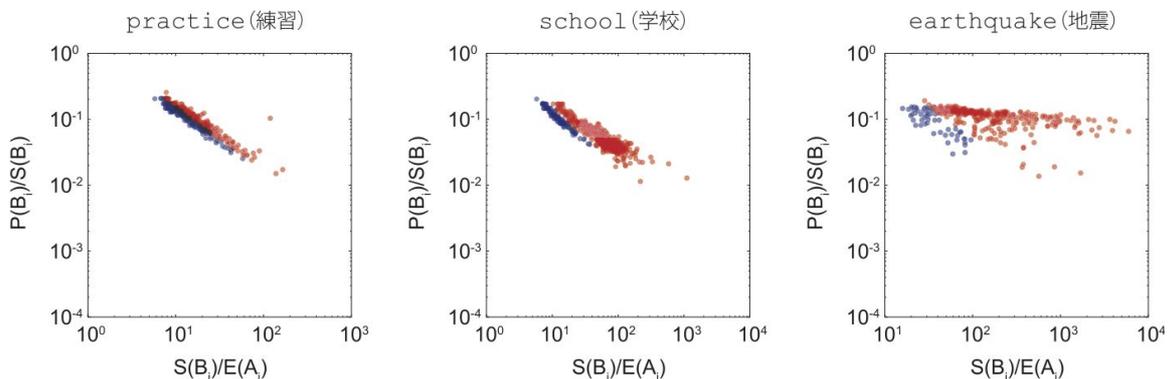


図 3 : バーストのピークとサイズの比 (p/S) とバーストサイズを直前の揺らぎ区間の平均ツイート値 E で割った量 (S/E) の関係。 E で規格化するのは全体的なトレンドによるバイアスを取り除くためである。

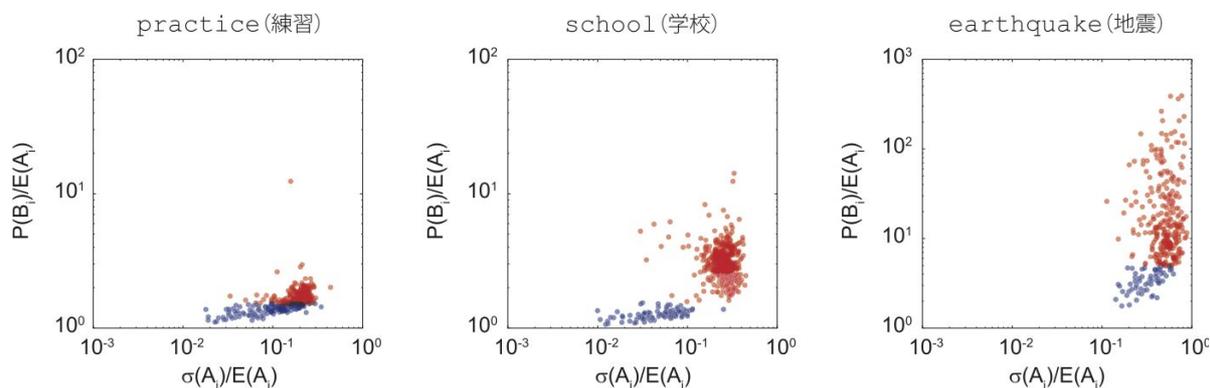


図4：経時する揺らぎの大きさと直後のバースト区間の相対ピーク値の比較．ここで揺らぎ（標準偏差）は $\sigma(A_i)/E(A_i)$ ，ピーク値は $P(B)/E(A)$ で、 E は直前の揺らぎ区間での平均値である．

以上のことから、単語ごとに揺らぎとバースト応答の関係は異なり、多くの単語に内因性 / 外因性バーストを分ける臨界的な揺らぎの値 c が出現することがわかった。つまり、内因性から外因性への切り替わりは大きなバーストへの転移にもなっており、揺らぎの臨界値が対応している。この臨界値の存在は、実際の脳の神経細胞のような2状態（静穏状態と興奮状態）がTwitterにも出現することを示す。つまりTwitterシステムは「興奮性媒質」となっていて、興奮状態への転移がポジティブ・フィードバックにより組織化される。この臨界的な揺らぎ値があるということは、いま測定している揺らぎの大きさが c を超えると、それに続く大きなバーストが予測できるということでもある。そのような予測が実際に可能かどうかは以後詳細に検証してみる価値がある。

さらに、Twitterが興奮性媒質であるならば、「Twitter脳」のような存在も考えうる。なにが「Twitter脳」の記憶であり、意識状態であるのか。そうしたことが本研究の真骨頂であり、それは今後の研究を待たざるをえない。それは現行のインターネット・ウェブをさらに拡張したものになるのは間違いないだろう。

5. 主な発表論文等

〔雑誌論文〕(計 3 件)

1. [Mizuki Oka](#), Hirotake Abe, Takashi Ikegami : Dynamic Homeostasis in Packet Switching Networks, Adaptive Behavior, Vol.23(1), pp.50-63, 2014, doi:10.1177/1059712314556369, 査読有り.
2. [Mizuki Oka](#), Yasuhiro Hashimoto, Takashi Ikegami : Self-organization on social media: endo-exo bursts and baseline fluctuations, PLOS ONE 9(10):e109293, doi:10.1371/journal.pone.0109293, 8 pages, 2014, 査読有り.
3. [Mizuki Oka](#), Takashi Ikegami : Exploring Default Mode and Information Flow on the Web, PLOS ONE 8(4): e60398, doi:10.1371/journal.pone.0060398, 11 pages, 2013, 査読有り.

〔学会発表〕(計 3 件)

1. [Mizuki Oka](#), Yasuhiro Hashimoto, Takashi Ikegami, Fluctuation and Burst Response in Social Media, The

2nd International Web Observatory
Workshop WOW2014 at WWW 2014,
pp. 1037-1042, April 8, 2014, Seoul
(Korea).

2. Takashi Ikegami, Mizuki Oka,
Massive Data Flows:
Self-organization of energy, material,
and information flows, 6th
International Conference on Agents
and Artificial Intelligence ICAART
2014, pp. 237-242, March 6-8, 2014,
Angers (France).
3. Norihiro Maruyama, Mizuki Oka,
Takashi Ikegami: Creating
Space-Time Affordances via an
Autonomous Sensor Network, The
2013 IEEE Symposium on Artificial
Life, pp. 67-73, April 16-19, 2013,
Singapore (Singapore).

6 . 研究組織

(1) 研究代表者

岡 瑞起 (OKA, Mizuki)
筑波大学システム情報系
助教
研究者番号 : 10512105