

平成30年6月18日現在

機関番号：12102

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K00420

研究課題名(和文) ソーシャルメディアにおけるバースト現象の理解と予測に関する研究

研究課題名(英文) Understanding and predicting burst phenomena in social media

研究代表者

岡 瑞起 (Oka, Mizuki)

筑波大学・システム情報系・准教授

研究者番号：10512105

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：Facebook、Twitter、Instagramといったソーシャルメディアは、情報共有ツールとして広く一般に用いられている。こうしたソーシャルメディアは、集団行動や集団注意の創発ダイナミクスを理解するために研究プラットフォームとしても注目を集めている。本研究では、特に「タグ」に注目し、新しいタグの生み出され方やタグの使われ方に注目した数理モデルの開発と解析を行った。

研究成果の概要(英文)：Social media such as Facebook, Twitter, Instagram etc. are widely used as information sharing tool. These social media are drawing attention as a research platform to understand the emergent dynamics of collective behavior and group attention. In this research, we focused on "tags" and developed a mathematical model on how new tags are created and used and conducted detailed analysis to understand their dynamics.

研究分野：ウェブサイエンス、人工生命、複雑系

キーワード：ソーシャルメディア ソーシャルタギング Yule-Simon過程

1. 研究開始当初の背景

Facebook, Twitter, Google Plus といったソーシャルメディアは、情報共有ツールとしてその役割を確立している。これらのソーシャルメディアによってさまざまな事象に人々がどのように反応しているかを観察することが可能となった。実際、ソーシャルメディアを使って集団注意 (collective attention) の創発ダイナミクスを研究するためのプラットフォームとして盛んに用いられている。ある出来事や事象が集団として注意を獲得すると、それは「バースト」として観測される。例えば、地震が発生したときに「地震」という言葉を含むツイートがある一定期間に急激に増える、それがバースト現象である。バースト現象は、物理や化学反応系で研究されてきたが、昨今、人の振る舞いにも共通する現象として注目を集めている。特に、人々の行動に関して、大量なデータがウェブ上のオンラインコンテンツなどでの解析可能になったことが研究を促進し、さまざまな応用例が提案されている。例えば、株価が下がる前には Google での株式に関連する検索クエリーがバーストする、という研究報告がある。これは不安な事柄があると、それについて調べるために普段より多く人々が検索するためバーストする。したがってこのバーストは株価下落の早期警報として利用可能である。他にも、映画の興行収入は、映画が公開される前のバーストの大きさに比例することが知られ、映画ヒット予測の指標として使われている。しかし、このようなバースト減少がどのように引き起こされるのか、その詳細なメカニズムはいまだ不明な点が多い。

2. 研究の目的

本研究では、RoomClip という写真共有サイト RoomClip を分析対象とし、数理モデルとの比較や、新しいモデルの提案を行う。

3. 研究の方法

RoomClip データから、タグの時系列データを抽出し、タギングの新規生成と既存タグの選択モデルとして、まず Yule-Simon 過程を用いる。

Yule-Simon 過程は、出現回数に比例した確率でランダムに既存の単語が選択される優先的選択性をもつ。Yule-Simon 過程では、新しいタグを追加することにより、新たな種類の単語を新規タグ生成確率 α で生成し、これまでに使われた単語の中から確率 $1-\alpha$ で選択する。ここで t は時刻である。時刻 t におけるアノテーション数 $N(t)$ 、タグ i が出現している回数を $n_i(t)$ とすると、時刻 t にタグ i が選択される確率 $P_i(t)$ は、

$$P_i(t) = \frac{(1-\alpha)n_i(t)}{N(t)}$$

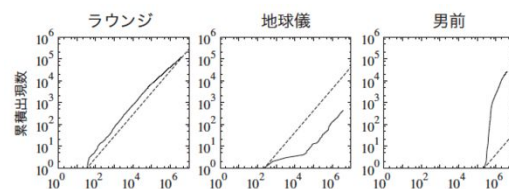
で与えられる。Yule-Simon 過程では以下のような 2 つの性質を過程している。

- ・新規タグ生成確率は一定でランダムに新規タグが生成される
- ・出現回数に比例した確率でランダムに既存タグが選択される (優先的選択性)

Yule-Simon 過程では、定義からタグの種類増加はアノテーション数のべきに比例 (Heaps 則) し、タグ出現数の累積確率分布は指数 -1 のべき分布 (Zipf 則) に従う。

4. 研究成果

実データにおけるタグの種類増加や、タグ出現数の累積分布は Yule-Simon 過程に従うが、個々のタグの出現回数が Yule-Simon 過程における予測値から、実データと比較したところ、モデルから逸脱するタグが観測された。図 1 は、RoomClipn におけるタグの累積出現数の増加曲線の一部で、実践は実データで破線が Yule-Simon 過程が予測する出現数である。たとえば、「ラウンジ」というタグではモデルの予測値と実データの振る舞いがほぼ一致している。一方、「地球儀」ではモデルの予測値に比べて増加が遅い。逆に、「男前」では予測値を大きく上回っている。



この実測値とモデルの予測値の間のズレの大きさ x とするとタグ i が初めて使われた時刻 t_i から時刻 s_i までの累積出現数の予測値と実測値の比によって次のように記述できる。

$$x = \frac{n_i(s_i)}{n_i^*(s_i)}$$

ただし、 $n_i^*(s_i)$ は時刻 s_i における予測値で次のように定義される。

$$n_i^*(s_i) = \left(\frac{s_i}{t_i}\right)^{(1-\alpha)}$$

タグの成長が Yule-Simon 過程に従う場合、 x の確率密度分布は、指数関数的減衰を示すことが示されている。たとえば、予測値と実測値の比は、およそ 10 倍程度に収まる。しかし、実データにおける逸脱の大きさの分布を測定すると、Yule-Simon 過程が予測する成

長よりも 100 倍、1,000 倍の大きさになるタグが多く存在することが分析からわかっている (図 2)。

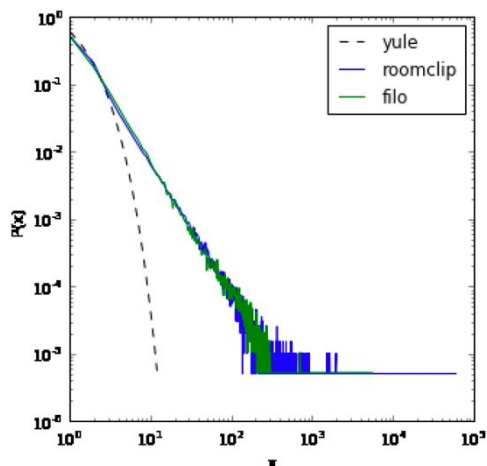


図 1 に示したような予測値からの大きなズレを再現するために Yule-Simon 過程のタグの選択確率について修正を加える。実データでは Yule-Simon 過程と同様に Zipf 則が成り立っており、モデルの修正もこの性質を維持したまま行いたい。そこで Yule-Simon 過程のタグの選択確率を以下のように変形する。

ここで出現回数と同じタグを Class と呼ぶこととする。Class 1 とは累積出現回数が 1 回のタグで、クラス N とは累積出現回数が N 回のタグのことを指す。 $k_i(t)$ とは時刻 t にタグ i が属する Class のタグ総数を表す。上式右辺第 2 項はどの Class が選択されるかを表して、第 3 項は Class の中でどのタグが選ばれるかを表している。Zipf 則が保たれるにはクラスの選択確率の形が必要条件であり、クラスの中でのタグの選択確率の形は問わない。そこで本研究では式の後半にバイアスを与えた First-In, Last-Out (FILO) モデルを提案する。これはある Class の中で最初に Class アップしたタグが必ず最後に選ばれるというモデルである。これによって最近使われたタグが優先的に選択され、より多く使われるタグが出現することが考えられる。

$$\begin{aligned}
 P_i(t) &= (1 - \alpha) \frac{n_i(t)}{N(t)} \\
 &= (1 - \alpha) \frac{n_i(t)k_i(t)}{N(t)} \frac{1}{k_i(t)}
 \end{aligned}$$

FILO モデルにおけるゆらぎの大きさ x の分布を計算し、カルバック・ライブラー情報量 (Kullback-Leibler Divergence, KLD) によ

って実データの分布の類似度を計測した。 x の分布を図 2 に示す。図 2 をみると FILO モデルはオリジナルの Yule-Simon 過程に比べて実測値により近い振る舞いを再現できていることがわかる。また、KLD による実データとモデルの分布の類似度は Yule-Simon が 0.833, FILO が 0.00390 という結果になった。これらの結果から提案する FILO モデルは Class 内のタグにバイアスをもった選択を与えることによって、大きな逸脱を生じ得ることを示している。

本研究では、Yule-Simon 過程では捉えきれない、出現数の増加曲線に大きなゆらぎをもつタグが実データに存在することを示した。また、Yule-Simon 過程の Class 内の選択にバイアスを加えることによって、タグの出現回数に大きな逸脱を生む効果を導入できることを示した。しかし、実データでは FILO モデルでも捉えきれない大きなゆらぎをもったタグが存在することが図 2 からわかった。本研究で提案した FILO モデルでは、クラス内で最近使われたタグが必ず選ばれる、という 1 つの極端な例を考えた。このようなタグの「鮮度」に依存した使われ方は、インターバルの分布を見ることでも明らかであることが分かっている。今後はこのような人間の注意の時間的な変化にも着目してモデル化を試みる予定である。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 1 件)

Takashi Ikegami, Mizuki Oka, Yasuhiro Hashimoto, Yoichi Mototake, Life as an Emergent Phenomenon: Studies From Large-Scale Boid Simulation and Web Data, Philosophical Transactions of The Royal Society A Mathematical Physical and Engineering Sciences. (査読有) DOI:10.1098/rsta.2016.0351, 15 pages, 2017.

〔学会発表〕(計 6 件)

Yasuhiro Hashimoto, Mizuki Oka, Takashi Ikegami, Measuring Bursty Growth of Words in Social Tagging Ecosystems, ACM WebSci '18 conference, Proc. of the 10th ACM Web Science Conference, p.17, 2018.

Yasuhiro Hashimoto, Mizuki Oka, and Takashi Ikegami, Large fluctuation in social tagging ecosystems, International Conference on Computational Social Science (IC2S2 2017), 2017.

Mizuki Oka, Yasuhiro Hashimoto, and Takashi Ikegami, Understanding Evolutionary Dynamics in Online Social Networks. Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence (SSCI), DOI : 10.1109/SSCI.2017.8280796, 4 pages, 2017.

橋本康弘, 佐藤晃矢, 岡瑞起, 池上高志: ソーシャルタギングシステムにおける語彙の個性, 2017 年度人工知能学会全国大会, 2017.

佐藤晃矢, 岡瑞起, 橋本康弘, 池上高志, 加藤和彦: 新しいタグの出現とソーシャルタギングシステム, 2016 年度人工知能学会全国大会, 2016.

橋本康弘, 佐藤晃矢, 岡瑞起, 池上高志: ソーシャルタギングにおけるタグ出現数のゆらぎ, 2016 年度人工知能学会全国大会, 2016.

〔図書〕(計 0 件)

〔産業財産権〕

○出願状況(計 0 件)

名称:
発明者:
権利者:
種類:
番号:
出願年月日:
国内外の別:

○取得状況(計 0 件)

名称:
発明者:
権利者:
種類:
番号:
取得年月日:
国内外の別:

〔その他〕

ホームページ等

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・准教授
研究者番号: 10512105

(2) 研究分担者

池上高志 (IKEGAMI, Takashi)
東京大学大学院・総合文化研究科・教授
研究者番号: 10211715

橋本康弘 (HASHIMOTO, Yasuhiro)
東京大学大学院・総合文化研究科・研究員
研究者番号: 10376494