

バースト解析を用いた pixiv のトレンド分析 -Twitter 上のバーストとの比較によるバースト要因分析-

佐藤 紗都[†] 荒川 達也^{††}

[†] 群馬工業高等専門学校専攻科生産システム工学専攻

^{††} 群馬工業高等専門学校電子情報工学科

1. はじめに

我々はテキストマイニングの分野で注目されている技術であるバースト解析[1][2]を用いて イラスト投稿 SNS である pixiv のトレンドを分析する方法を提案している[3]. [3]では特定のタグ(イラストの題材などを示す付加情報)がつけられたイラストの急増を「pixiv のバースト」と定義し、2007 年から 2015 年の期間で 54 種類のタグについて pixiv の投稿データからバーストを検知できたことを報告した. 一方、菊井らの論文[4]では、検索キーワードのバーストと Twitter のバーストを比較することにより、検索キーワードのバースト要因を推定する方法が提案されている. そこで、本稿では [4]の方法を参考に pixiv のバーストと Twitter のバーストを比較することにより pixiv のバースト要因を推定することができるか検討する.

2. 提案手法

以下に提案手法の概要を示す.

(1) pixiv 投稿データのバースト解析

[3]と同様の方法により pixiv のバーストを解析する.

(2) Twitter 投稿のバースト解析

(1)でバーストを検知した pixiv のタグをキーワードとして Twitter のバースト解析を行う.

(3) pixiv バースト要因の推定

(2)で検知した Twitter のバーストに対し、ツイート本文の内容を調べることにより pixiv バーストの要因を推定する. 今回はこの処理を目視で行う. 将来的にはある程度の自動化を考えている.

3. 評価実験

提案手法の有効性を調べるため 2007 年から 2015 年の期間の pixiv 投稿データ及び、2015 年の Twitter 投稿データを用いて評価実験を行っている. 現在までに、2015 年に pixiv のバーストを検知した 14 種類の pixiv のタグについて実験を行った. その結果、pixiv のバーストの直前に Twitter のバーストが起こっているケースは全体の約 93%, そのうち Twitter の本文から pixiv のバーストの原因が特定できたケースは全体の約 78%である.

4. 実行例

実行例として「がっこうぐらし!」に関するタグについてバースト解析とその要因分析を行った結果を図 1 に示す.

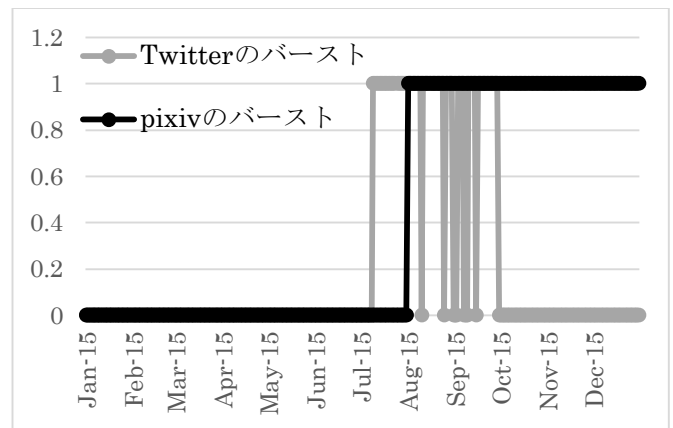


図1. 「がっこうぐらし!」に関するバースト

「がっこうぐらし!」とは 2012 年に芳文社から刊行された千葉サドルによる漫画であり、2015 年 7-9 月に TV アニメが放送された. 図1より、2015 年 7 月頃に Twitter でバーストが起こり、その後 1 ヶ月ほど遅れて、2015 年 8 月頃に pixiv 上でバーストが起こっていることが分かる. さらに Twitter のバースト期間中のツイート本文を調査したところ、その時期に放送されたアニメが話題になっていた. そのため、pixiv のバーストの要因はアニメ放送による作品の人気の高まりではないかと考えられる.

5. まとめ

特定のタグに関する Twitter 上のバーストとの比較による pixiv のバースト要因分析方法を提案した. 現在までに、14 種類のタグで実験を行い、かなり高い割合でバースト要因が特定できた. 今後はより多くの作品を対象に検証を続けたい.

参考文献

- [1] Kleinberg, J. : Bursty and Hierarchical Structure in Streams, Proc. 8th SIGKDD, pp. 91-101, 2002.
- [2]高橋佑介 他:ニュースにおけるバーストキーワードの話題への集約, 第 3 回データ工学と情報マネジメントに関するフォーラム-DEIM フォーラム- 論文集, 2011.
- [3]佐藤 紗都, 荒川 達也: バースト解析を用いた pixiv 投稿のトレンド分析, 電子情報通信学会 東京支部学生会 第 21 回研究発表会, 2016.
- [4]菊井玄一郎 他:検索ホットワードとブログ系テキストの関係を探る, 信学技報, 2012.