

バースト解析を用いたツイートデータからの 個人の興味・関心の探索手法の提案

佐藤 健太[†] 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科 ^{††}群馬工業高等専門学校電子情報工学科

1. はじめに

Twitter を代表とするマイクロブログサービスの特徴の1つとして、各ユーザがその時々感じたことや考えたことを即座に投稿するリアルタイム性が挙げられる。そのため、それらの投稿を分析することによりユーザの興味や関心の推移を追跡できると考えられる。

本研究では、時系列テキストデータの中で特定のキーワードを含むテキストが急激に増加する「バースト」に注目し、個人のツイートデータに対してバースト解析を行うことで個人の興味・関心の探索を行う。また、個人のバーストと社会全体でのバーストの比較による個人の興味と社会との関係の分析を試みる。

このような方向性の研究として、社会全体のツイートに対しバースト解析を用いて Twitter 上のトレンドの変化を分析している研究や社会全体でバーストしている単語のクラスタを用いて話題の解析などを行う研究がある[1][2]。それらに対し、本研究では個人のツイートに焦点を当ててバースト解析を行う。

2. 提案手法

本研究では、ユーザの興味が何かに向いたとき、その事象についてのツイートがバーストを起こすと仮定し、バースト解析を用いてユーザの興味の推移を探索する。また、その結果を社会全体のツイートデータのバーストと比較する。提案手法の流れを図1に示す。

3. 実験と考察

提案手法の有効性を検証するために、ある特定のユーザの約1カ月間のツイートを個人のツイートとし、同じ期間に Twitter 上に投稿された全ツイートからランダムに取得したツイートを社会のツイートとして解析と比較を行った。

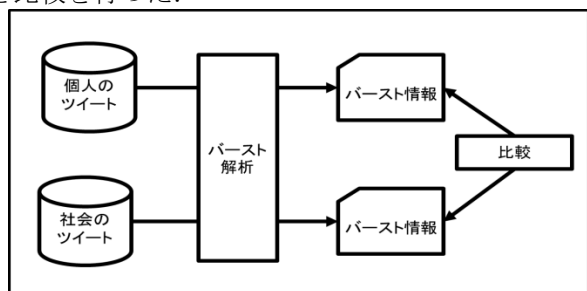


図1：提案手法の流れ

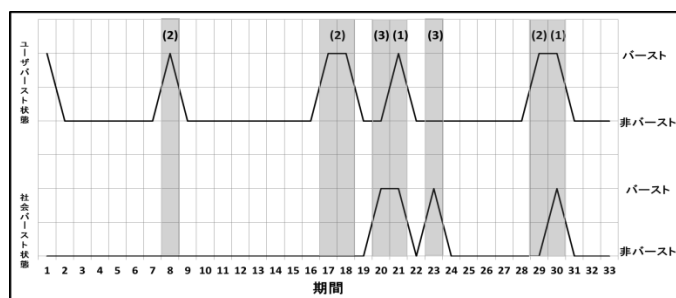


図2：個人と社会全体のバースト比較

解析結果のグラフを図2に示す。図2より(1)個人と社会の両方がバーストしている期間、(2)個人のみバーストしている期間、(3)社会のみバーストしている期間の3種類が存在することがわかる。

(1)の期間に関しては、社会全体が注目している事象に個人ユーザも関心を向けている状況であると考えられるが、偶然バースト期間が一致したという可能性も考えられる。そこで当該期間内のそれぞれのツイートデータを調べたところ、いずれも「高校サッカー選手権」に関するツイートを含んでいることが確認できた。そのため、これらのバースト期間は「偶然の一致」ではなく、実際に社会が注目した事象に対してユーザも関心を持っていたために発生したものだと考えられる。

一方(2)の期間は社会全体とは関わりがなく、ユーザが個人的に関心を持っていた期間、(3)の期間は社会が注目する事象に対し、何らかの理由でユーザが興味を持たなかった、もしくは興味があったとしてもツイートをしなかった期間と考えられるが、当該期間のデータ中にそれを裏付けるツイートは見つからなかった。

4. 今後の予定

提案手法のバースト解析結果の妥当性、および実験結果で明らかになった各バースト期間に対する推測の正当性について検証を進める。

参考文献

[1] M. Mathioudakis, N. Koudas. : “Twittermonitor: trend detection over the twitter stream.” Proceedings of the 2010 ACM SIGMOD International Conference on Management of data p. 1155-1158.

[2] 藤川智英, 鍛冶伸裕, 吉永直樹, 喜連川優 : “マイクロブログ上の話題抽出とユーザの態度の分類に基づく流言検出支援システム” DEIM 2012, F7-1