

# 沖縄県内のツイッター情報を用いた異常事態通知システムの開発

小林 未季† 中里 収†

名桜大学国際学群国際学類†

## 1.1 研究の動機

近年では、ソーシャルメディアが普及し、多くの人が様々な状況下で気軽に投稿している。日本では、3500万人(2015年12月現在)がTwitterを利用している[1]。また、2011年の東日本大震災以降は、災害時のライフラインの1つとして認知されるようになった[2]。Twitterでは、即時性や簡便性、双方向性といった特徴から、公衆の注目を集める出来事が発生した時に、ツイート数が平常時と比べて大きく増加する現象が観測される[3]。そのような現象を知ることは、災害や事件など、瞬時に状況確認を必要とする場合に有効である。しかし、人がツイートを常に監視・管理することは、人件費や労働時間を考えると難しい。そこで、本研究では、ツイート数が平時と比べて大きく増える現象を「バースト」と定義し、その発生をリアルタイムに知らせるシステムの開発を行った。

## 1.2 研究の目的

沖縄県内でのツイート数を常に監視し、バーストが発生した際に、音声を用いてリアルタイムに異常を知らせるシステムを開発し、迅速な状況確認を可能にすることを目的とした。また、本システムはバースト通知だけでなく、人によるバースト原因の推測を支援するものとした。

## 2.1 研究の背景

バースト発生時にテキスト内容から原因特定を行った研究として島田らの研究がある[3]。島田らが推測したバースト原因には、台風、金環日食、サッカーワールドカップ、地震がある。沖縄県内におけるバーストの例には、2014年9月4日に発生した石垣島近海の地震がある。その際のツイート数の変化を表したのが図1である。地震が発生した2015年9月5日0時頃に大きなツイート数の増加が見られる。本研究で開発するシステムにより、地震や台風など、緊急性が高い場合の迅速な状況確認を目指す。

Development of the Alert System for Emergency Based on Twitter Information in Okinawa Prefecture  
Miki Kobayashi, Shu Nakazato  
†Meio University Faculty of International studies

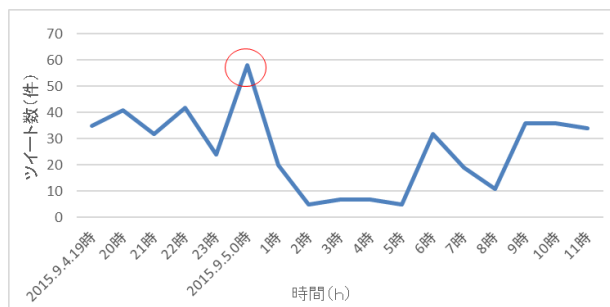


図1 石垣島近海の地震（最大震度3）によるバーストのツイート数の変化

## 2.2 研究の方針

沖縄県内におけるTwitterのテキストデータ及び位置情報GPSデータを利用して、現実起きたイベントと照合し、異常事態を通知するシステムを開発した。

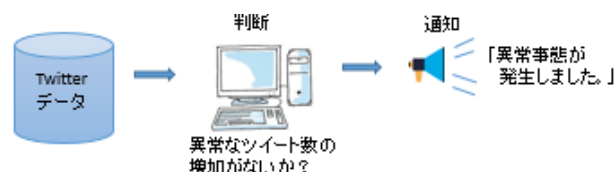


図2 異常事態通知システムの概念図

## 3.1 システム概要

システムの作成手順は以下の通りである。

### ① バーストの閾値設定

沖縄県内における2014年12月から2015年11月までの1年間のTwitterデータから閾値を設定した。閾値は月ごと及び時間ごとに異なる。まず、1時間ごとのツイート数を集計し、月ごとに0時から24時の時間に分けて、平均と分散、標準偏差を求める。閾値は、標準偏差に任意の係数をかけて求めた。システムの利用者が状況に応じて変えられるものとした。これにより、手が空いているときや注意深く監視したいときは閾値を小さくして、少しの増加をバーストとして検出することができる。先行研究[3]においては、閾値を3σとしていた。標準偏差を3倍した値を閾値とすると、図3のようになった。この折れ線グラフから、閾値は、1日を通して時間ごとに大きく変化することが分かった。

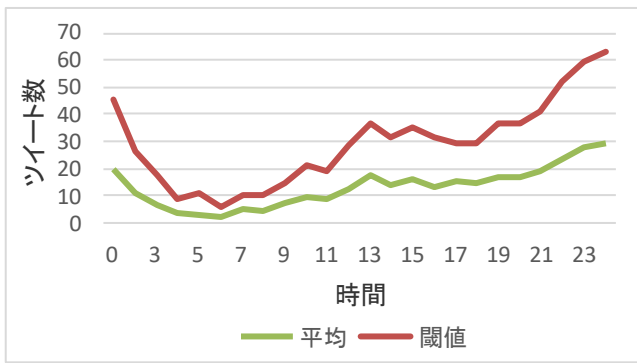


図3 2015年3月の閾値の変化

### ② 閾値に基づいたバーストの検出

ある5分間のツイート数が閾値よりも大きな値であった時は、バーストが発生したものとした。バースト発生時には、ツイート数を知らせ、ツイート本文を10個まで読み上げる。連続でバーストしている場合は、何回連続であるかを伝える。また、MeCab[4]を用いて形態素解析し、得られた結果の名詞のみを表示するものとした。品詞の中でも名詞は、バースト原因に関わりがあるものが多いと考えたからである。

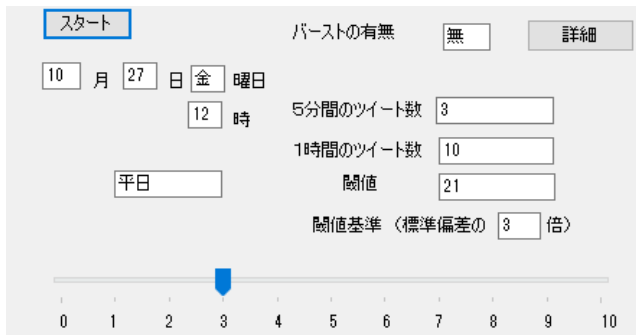


図4 異常事態通知システムの画面

### 3.2 動作テスト

過去にバーストが起きたデータを入力し、本システムが正しく通知できることを確認した。

求めた閾値を使ってバースト判定すると、原因と直接関わりがある内容がつぶやかれる場合は少ないと分かった。サッカーワールドカップを原因とするバーストでは、本文に「サッカー」や「ゴール」という言葉はなく、サッカー選手の名前や「入れたー」「来たー」など原因に対して間接的な言葉が多くつぶやかれていた。

また、マラソンイベントを原因とするバーストでも、本文には「エイド」や「レース」などの原因を間接的に表す言葉が多かった。

### 4.1 考察

今回作成したシステムにより、必要に応じた割合でバーストを検出し、通知することができ

るようになった。このシステムの課題は、バーストの原因とその緊急性がどの程度であるかについて人の確認が必要な点である。

当初は、人によるバースト原因の推測を支援するために、本文の単語を使用したユーザが多い順に音声で知らせることを予定していた。しかし、本文には原因に対して間接的な言葉が多いため、多数出現する単語がバースト原因推測において重要とは言えない。よって、まず、過去に起きた地震・台風を原因とするバーストにおいて良くつぶやかれる単語が何か調べ、それらをまとめたデータベースを作成し、バーストが起きた際にシステムが照合するようにした。これにより、地震・台風に限り、システムによる本文からの原因推測を行うことができた。

### 5.1 おわりに

今回の研究では、バーストを音声を用いてリアルタイムに知らせるシステムを開発した。

このシステムを使ってバーストを検知してみると、バーストには緊急性の高いものから低いものまで様々であり、システムの利用目的に合った緊急性の高さのバーストを示すことが必要ということが明らかになった。災害時の迅速な対応を目的としてシステムを使うのであれば、緊急性が高いバーストを通知すべきである。また、どのようなイベントやテレビ番組が盛り上がっているかを調査したいのであれば、比較的緊急性の低いバーストを通知すべきである。

今後の課題は、緊急性の高さを判定し表示することと、人によるバースト原因の推測をより支援できるシステムへの改良である。

### 引用文献

- [1] THE HUFFINGTON POST, “Twitter が国内ユーザ数を初公表「増加率は世界一」” [http://www.huffingtonpost.jp/2016/02/18/twitter-japan\\_n\\_9260630.html](http://www.huffingtonpost.jp/2016/02/18/twitter-japan_n_9260630.html), 閲覧日: 2016年12月1日.
- [2] 村上圭子, “「震災ビックデータ」をどう生かすか”, <https://www.nhk.or.jp/bunken/summary/150.html>, 閲覧日: 2016年12月1日.
- [3] 水沼 友宏, 池内 淳, 山本 修平, 山口 裕太郎, 佐藤 哲司, 島田 諭, “Twitter におけるバーストの生起要因と類型化に関する研究”, <http://infosocio.org/vol17no2-03.pdf>, 閲覧日: 2016年12月1日.
- [4] 日本語形態素解析システム MeCab, <https://www.mlab.im.dendai.ac.jp/~yamada/ir/MorphologicalAnalyzer/MeCab.html>, 閲覧日: 2016年12月1日.