

Twitter における頻出語句ランキング表示システムの開発による 災害検知の検討

吉岡優一 (理工学研究科環境共生系専攻) 瀧本浩一 (理工学研究科環境共生系専攻)

Study on Development of Frequent Phrase Ranking Display System by using Twitter for Disaster Detection

Yuichi YOSHIOKA (Environmental Science and Engineering, Graduate School of Science and Engineering)

Koichi TAKIMOTO (Environmental Science and Engineering, Graduate School of Science and Engineering)

Abstract: Japan, our country, is a disaster-prone area and many people have been killed by earthquakes or powerful storms and large-scale water flood. For example, many casualties caused on March 11, 2013 by the 2013 off the Pacific coast of Tohoku Earthquake because of the tsunami attack. In recent years, microblog service called Twitter been spreads explosively and the opportunity for people to disseminate information freely on Web is increasing. Twitter was used as one of the communication tools because lifeline system such as telephone stopped at the Tohoku Earthquake. Therefore, the author analyzed the utilization situation on Twitter as a useful communication tool during disasters and developed real-time display system by using frequent phrase ranking. Furthermore, the author tried to examine the possibility of disaster detection based on Twitter.

Key Words: Disaster detection, Twitter, Microblog, Disaster prevention

1. 概要

わが国は、「災害大国」と呼ばれるほど災害の多い地域であり、地震や風水害による被害が後を絶たない。2013年3月11日には、東日本大震災により、多くの死傷者が出た。

近年、Twitter¹⁾というマイクロブログサービスが爆発的に普及し、Web上で人々が気軽に情報を発信する機会が増加している。東日本大震災においては、電話回線などの通信手段が途絶された中、コミュニケーションツールの1つとしてTwitterが使われた²⁾。

そこで本研究では、Twitterのつぶやきにおけるリアルタイムな頻出語句ランキング表示システムの開発を行い、風水害などの事前対応に用いるための災害検知の可能性を検討する。

2. システム概要

研究で開発した本システムは、Twitterのつぶやきをリアルタイムに取得し、過去1時間に取得したつぶやきにおける頻出語句のランキングを表示するシステムである。Figure 1に表示画面の一例を示す。

Twitterのつぶやきの取得にはTwitter社の提供するStreaming APIを用い、形態素解析器としてMeCab³⁾を用いた。

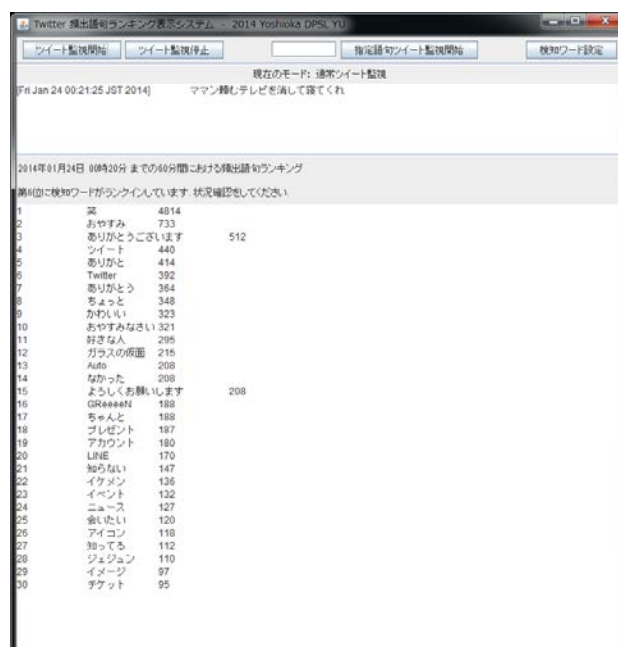


Figure 1. System screen

2-1. システム構成

本システムは、以下の4つの機能から構成されている。

- リアルタイムなつぶやきの取得・表示
- リアルタイムな頻出語句ランキングの表示
- 設定語句のランクイン検知表示

- 指定語句を含むつぶやきのリアルタイム表示
これらの機能について、以下にそれぞれ説明する。

2-1-1. リアルタイムなつぶやきの取得・表示

「ツイート監視開始」ボタンを押すと開始される本機能は、リアルタイムなつぶやきを取得し、表示し続ける。なお、つぶやきの取得は、日本語の含まれるつぶやきに対してのみ行われる。表示される内容は、日付及びつぶやき本文である。

また、本機能を開始すると、自動的に 2-1-2 項の頻出語句ランキング表示機能も開始される。

「ツイート監視停止」ボタンを押すと、本機能は停止し、2-1-2 項の頻出語句ランキング表示機能も停止する。

2-1-2. リアルタイムな頻出語句ランキングの表示

2-1-1 項の機能を開始すると自動的に開始される本機能は、10分毎に過去1時間に取得したつぶやきにおける頻出語句ランキングを表示する。頻出語句ランキングが表示されるタイミングは、毎時00分、10分、20分、30分、40分、50分の6回である。表示される内容は、順位、語句、出現回数である。また同時に、テキストファイルへもランキング情報を出力する。

なお、本機能の開始以降6回の頻出語句ランキング表示を経過するまでは、過去1時間のつぶやきではなく、2-1-1 項の機能開始からのツイートにおける頻出語句ランキングの表示となる。6回目の表示以降では、10分ごとに常に過去1時間のつぶやきにおける頻出語句ランキングが表示される。

本機能は、2-1-1 項の機能を停止すると、自動的に停止する。

2-1-3. 設定語句のランクイン検知表示

本機能は、あらかじめ設定した語句(以下、「検知ワード」)が頻出語句ランキングの上位に入ると、ランクインを知らせるメッセージを表示する。検知ワードは、ユーザーが任意で設定することができる。「検知ワード設定」ボタンを押すと、検知ワードが書かれたテキストファイルがテキストエディタにて開かれる。改行区切りで語句を記述することで、複数の語句を検知ワードとして設定することができる。

表示する内容は、ランクインしている検知ワードの順位である。検知ワードがランクインしていない状態では、「異常なし」と表示される。

本機能は、2-1-2 項の機能と同時に動作する。

2-1-4. 指定語句を含むつぶやきのリアルタイム表示

「指定語句ツイート監視開始」ボタンの左にあるテキストボックスに語句を入力し、「指定語句ツイート監視開始」ボタンを押すと、リアルタイムに、入力した語句を含むつぶやきのみを取得し、表示し続ける。

「ツイート監視停止」ボタンを押すと、本機能は停

止される。

なお、本機能では、2-1-2 項の機能は動作しない。

2-2. MeCab 辞書のカスタマイズ

形態素解析器 MeCab には、標準で辞書が備えられている。しかし、標準の辞書を用いて形態素解析を行うと、固有名詞が個々の単語に別れて判別されてしまうことが多い。例えば、「特別警報」という単語を、標準の辞書を用いて形態素解析すると、「特別」と「警報」の2単語に分割されてしまう。つぶやきを分析する上で、固有名詞はひとまとまりに認識されるべきだと考えられる。そこで、固有名詞が分かれて判別される現象を改善するために、Wikipedia の記事タイトルを MeCab の辞書に追加し、辞書のカスタマイズを行った。この理由としては、Twitter 上でのつぶやきに使われる言葉は、Wikipedia に踏襲されていると考えたためである。Wikipedia のすべての記事タイトルのデータベースが公開されており、本研究では 2013 年 11 月 7 日に公開されたものを使用した⁴⁾。今後、Wikipedia の記事タイトルに代わるものがあれば対応する予定である。

3. 台風発生時のつぶやきにおける頻出語句の分析

先のシステムを用いて、平成 25 年台風第 18 号(以下、T1318)の発生時のつぶやきにおける頻出語句ランキングを得た。また、比較対象として、平常時のつぶやきにおける頻出語句ランキングも得た。

3-1. 台風の概要と調査対象期間

T1318 は、2013 年 9 月 13 日 3 時に小笠原諸島の近海で発生した大型の台風である。しばらく西寄りに進んだ台風は、父島近海を通過し、大型の台風となった。その後、台風は次第に北寄りに進路を変え、上陸直前まで発達し、9 月 16 日 6 時に 965hPa の気圧を記録し、同日 8 時前に愛知県豊橋市付近に上陸した。その後、山梨県や福島県などを通過し、三陸沖へと抜けた。この台風によって、特別警報が 8 月 30 日の運用開始後初めて適用された。

T1318 発生時と平常時それぞれの、調査開始日時、調査終了日時、取得つぶやき数を、Table 1、Table 2 に示す。

Table 1. Period during T1318 occurrence

開始日時	2013 年 9 月 16 日 00:00:00
終了日時	2013 年 9 月 16 日 23:59:59
取得つぶやき数	924,340

Table 2. Period of time of normal

開始日時	2013 年 11 月 4 日 00:00:00
終了日時	2013 年 11 月 4 日 23:59:59
取得つぶやき数	764,776

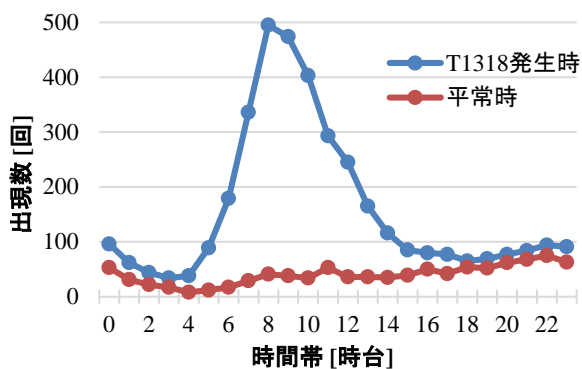


Figure 2. “気をつけ” number of occurrences transition

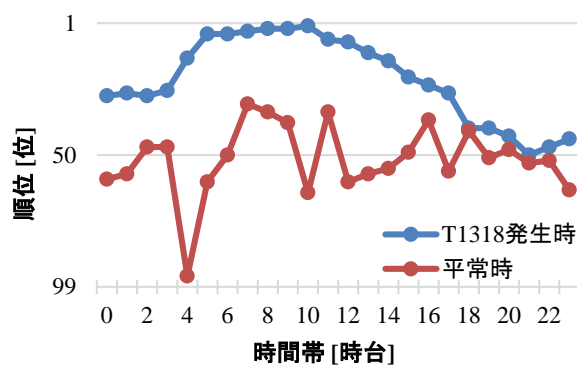


Figure 3. “気をつけ” ranking transition

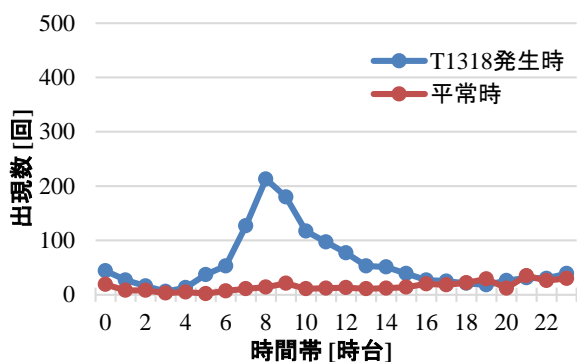


Figure 4. “気を付け” number of occurrences transition

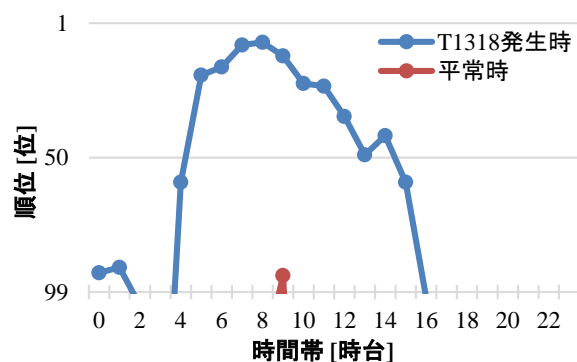


Figure 5. “気を付け” ranking transition

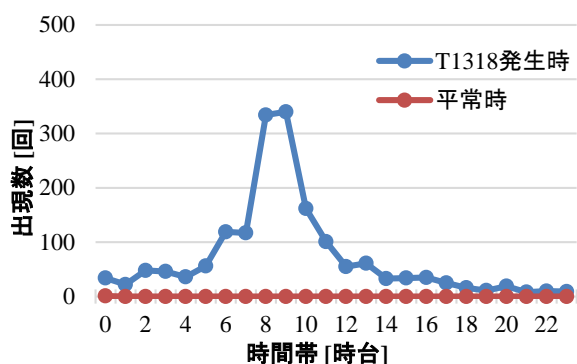


Figure 6. “避難勧告” number of occurrences transition

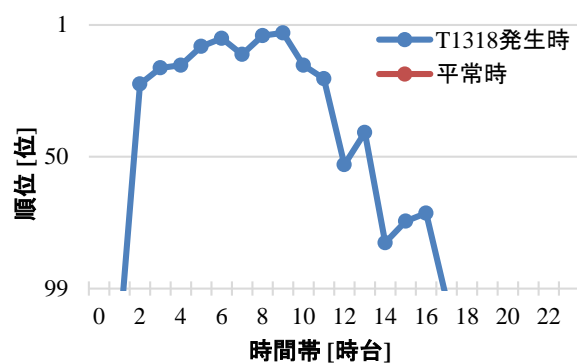


Figure 7. “避難勧告” ranking transition

3-2. 分析の方法

先のシステムを用いて得た頻出語句ランキングにおける、特定の語句の1時間ごとの出現数と出現数順位を分析した。ここで、分析対象の語句は、「気をつけ」、「気を付け」、「避難勧告」、「避難指示」、「暴風警報」、「特別警報」の6個とした。

3-3. 分析結果

以上の方法より得られた結果について述べる。

台風が日本に接近したのは2013年9月16日の午前中であるが、調査対象の語句は、その時間帯において出現数、出現数順位共に高い数値を記録している。

「気をつけ」の出現数の推移と出現数順位の推移を、それぞれ Figure 2、Figure 3 に、「気を付け」の出現数

の推移と出現数順位の推移を、それぞれ Figure 4、Figure 5 に示す。「気をつけ」と「気を付け」は、平常時もある程度出現している語句であることが分かる。台風接近日においては、平常時よりも多く出現し、7時から11時までの台風接近時間帯には特に多く出現している。

次に、「避難勧告」の出現数の推移と出現数順位の推移を、それぞれ Figure 6、Figure 7 に、「避難指示」の出現数の推移と出現数順位の推移を、それぞれ Figure 8、Figure 9 に、「暴風警報」の出現数の推移と出現数順位の推移を、それぞれ Figure 10、Figure 11 に示す。「避難勧告」、「避難指示」及び「暴風警報」は、平常時にはほとんど出現しない語句であることが分かる。台風接近日においても、前述の2個の語句と同様に台風接

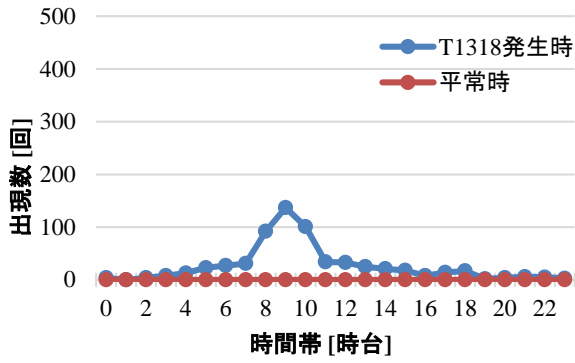


Figure 8. “避難指示” number of occurrences transition

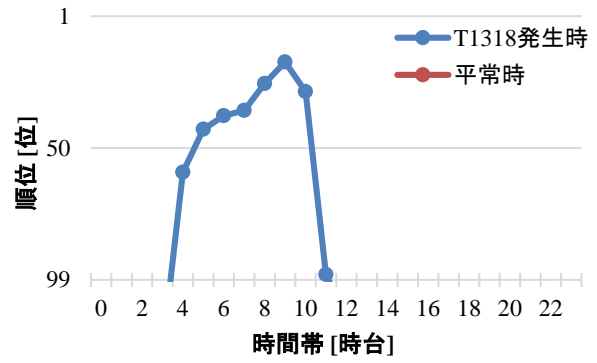


Figure 9. “避難指示” ranking transition

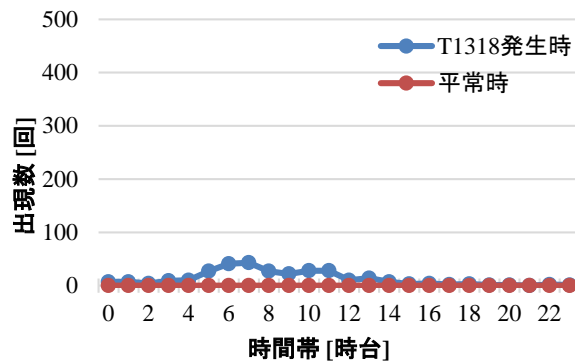


Figure 10. “暴風警報” number of occurrences transition

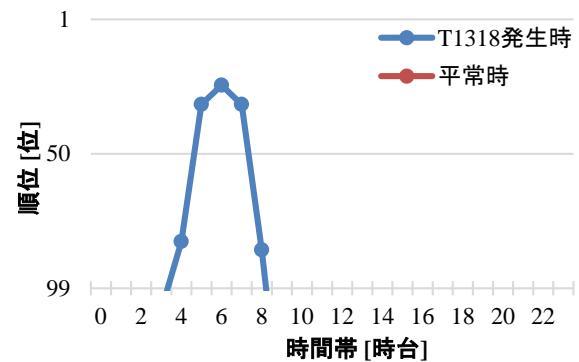


Figure 11. “暴風警報” ranking transition

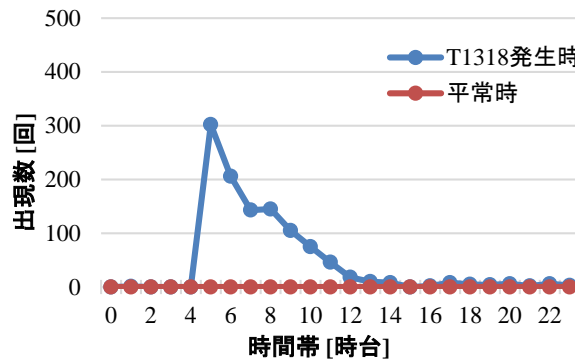


Figure 12. “特別警報” number of occurrences transition

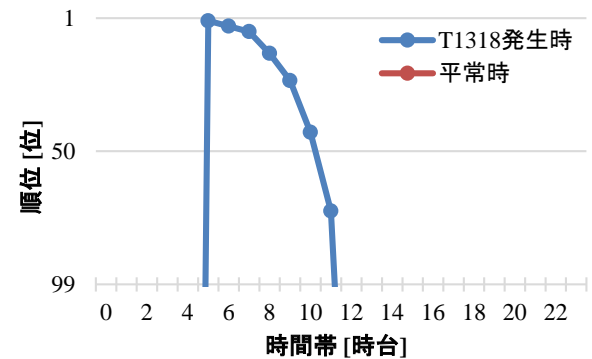


Figure 13. “特別警報” ranking transition

近時間帯に出現数が増加しているが、その時間帯以外ではほとんど出現していない。

また、「特別警報」の出現数の推移と出現数順位の推移を、それぞれ Figure 12、Figure 13 に示す。「特別警報」という語句は、平常時には1日を通して1度も出現していない。また、台風接近日においては、4 時台までは1度も出現していないが、5 時台には302 回の出現が確認された。これは、5 時 5 分に京都府、福井県、滋賀県に大雨特別警報が発表されたことによるものと考えられ、人々の関心がつぶやきに顕著に現れることを示している。

これらの結果より、災害の発生時には関連する特定の語句が増加すると考えられる。今回は、特定の6 個の語句に注目して出現数と出現数順位の時間推移を分

析したが、他にも災害時に増加する語句はあるものと考えられる。

3-4. 本システムを用いた災害検知の検討

本システムを用いた、災害検知の可能性を検討する。先の分析では、災害の発生時には関連する特定の語句が急激に増加するということが判明した。ここで注目した特定の語句以外にも、災害時に増加する語句はあるものと考えられる。どのような災害の発生時にどのような語句が増加するかがあらかじめ分かれば、その語句を検知ワードとして登録しておくことで、本システムを用いた災害検知ができると考えられる。

また、「気をつけ」という語句については、人々が身の危険を感じたときにつぶやかれるつぶやきに多く含

まれる語句であると考えられ、検知ワードとしては有効であると考えられる。しかし、別の意味で使われている可能性もあるため、他の危険性のある言葉と相互的に分析し、判断する必要がある。加えて、平常時と災害の発生時のしきい値を探るためにも、今後様々な災害のケースでつぶやきデータを取得し、分析する必要がある。

本システムを活用できる場面の例を、「誰が」、「いつどのようなときに」、「本システムをどのように活用して」、「どのような効果を得ることができるか」の4つの観点に注目して、以下に2点挙げる。

- ① すべての人が、普段の生活中に、本システムの検知ワードに災害に関連する語句を登録しツイート監視を行うことで、日本での異常をニュースよりも早く知ることができる。
- ② 自治体職員が、通常業務中に管理区域内の異常を監視したいときに、本システムの検知ワードに管理区域内の地名や川の名前などをあらかじめ登録しておきツイート監視を行うことで、管理区域内の異常を早く知ることができる。

4. まとめ

本研究では、Twitter のつぶやきにおける頻出語句ランキング表示システムを開発し、つぶやきのリアルタイム監視と頻出語句ランキングの表示機能、設定語句のランクイン検知機能を実現した。

本システムを用いて、災害発生時における頻出語句の特徴を、平常時と比較、分析した。その結果、災害時には関連する語句の出現数が急激に増加することが判明した。

また、本システムを用いた災害検知の可能性を検討し、Twitter からの災害の検知の可能性を見出した。しかし、災害を検知するという観点では、まだ至らない点が多い。今後の課題について以下に挙げる。

① 災害時に増加する語句の調査

本研究で開発した頻出語句ランキング表示システムでは、あらかじめ設定した特定の語句(検知ワード)の頻出語句ランキングへのランクインを知らせる機能を持っている。検知ワードの適切な設定のためには、どのような災害の発生時にどのような語句が増加するかを調査する必要がある。

② バースト検知

普段出現しない語句の急激な増加(バースト状態)を検知する機能が必要である。これを実現することで、検知ワードとして設定していない語句の急激な増加も検知することができ、システムの活用用途が広がると考えられる。

③ 位置情報の利用の検討

本研究では、つぶやきに付加された位置情報は利用していない。位置情報の付加されたつぶやきの割合などを調査し、利用できるかどうか検討する必要がある。位置情報を利用することで、特定の地域からのつぶやきに特定した、精度の高い災害検知が行えると考えられる。また、位置情報の地図上への可視化を行うことで、ツイートの急増や広がりなどを視覚的に確認でき、視認度の高いシステムができると考えられる。

④ デマの検出

Twitter では、出所の不確かな情報などを鵜呑みにした利用者によって誤った情報が急速に広まる特徴がある。デマのつぶやきの広まり方や、デマを否定するつぶやきによるデマの収束を調査し、デマを検出できる必要がある。

⑤ 言葉のゆれ

つぶやきは口語で書かれることが多く、同じ意味の言葉でも様々に変化する「言葉のゆれ」が発生する。変化した言葉も、同じ言葉だと判別する必要がある。

⑥ つぶやきのカテゴライズ

つぶやきの内容をカテゴリに分類することで、増加しているつぶやきのおおまかな内容などが判明し、システムをより活用できると考えられる。分類方法としては、つぶやき内にカテゴリに関連するキーワードが存在するかどうかの検索などが考えられる。

参考文献

- 1) Twitter, Inc. . “Twitter” . Twitter . 2015 年 . <https://twitter.com/>
- 2) NEC ビッグロブ株式会社 . “ツイッター分析：2011 年 4 月 27 日 東日本大震災におけるツイッターの利用状況について | ツイッターでいま HOT な話題は？- ついっぷるトレンド” . ついっぷるトレンド . 2011 年 4 月 27 日 . <http://tr.twipple.jp/info/bunseki/20110427.html>
- 3) 工藤拓 . “mecab - Japanese morphological analyzer - Google Project Hosting” . mecab : Japanese morphological analyzer . 2013 年 3 月 9 日 . <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>
- 4) ウィキメディア財団 . “Wikipedia:データベースダウンロード - Wikipedia” . Wikipedia:データベースダウンロード . 2014 年 1 月 2 日 . <http://ja.wikipedia.org/wiki/Wikipedia:%E3%83%87%E3%83%BC%E3%82%BF%E3%83%99%E3%83%BC%E3%82%B9%E3%83%80%E3%82%A6%E3%83%B3%E3%83%AD%E3%83%BC%E3%83%89>

(平成27年1月30日受理)