

# Twitterを用いた大規模災害時におけるシステム制御に有用な情報の抽出

丸 千尋<sup>†</sup>

榎 美紀<sup>‡</sup>

小口 正人<sup>†</sup>

<sup>†</sup>お茶の水女子大学

<sup>‡</sup>IBM Research - Tokyo

## 1. はじめに

地震や台風などの大規模災害が発生すると、電話やインターネットが繋がりにくくなるといった問題が生じる。現在、人間が見て判断可能な単純なネットワーク制御は行われているが、緊急災害時のようにデータの種類や量が多くなると、人間が障害を検知するのは困難である。

近年、Twitterなどのマイクロブログサービスの利用が人々の生活の一部として定着しており、Twitterから得ることのできるソーシャル情報は実世界の動きを反映していると言えることができる。例えば、緊急災害時などに特定の地域で大きな出来事が起きていると、その地域でのツイートが同じ時刻に急激に増加する [1]

本研究ではこの点に着目し、大規模災害時における通信状況等に言及するツイートの時間変化と位置情報を解析し、ネットワークのトラフィック障害などを検知する手法の提案を行う。

## 2. 提案システム

Twitterから検出したネットワークトラフィック障害に関する情報を出力することで、直接システム制御を行うという目的を達成するために、本研究では以下のシステムを提案する。提案システムの概要を図1に示す。

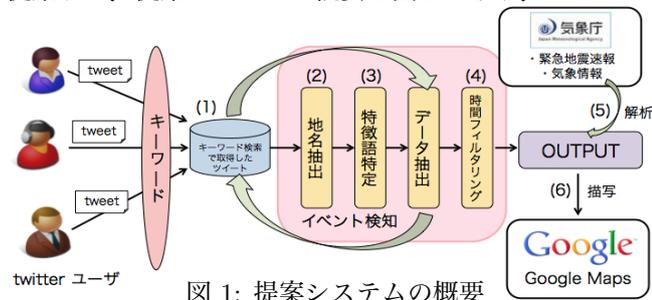


図1: 提案システムの概要

- (1) 特定のキーワードを設定して、キーワードを含むツイートを取得する。
- (2) キーワード検索で取得したツイートを同じ地名名詞が含まれるツイートごとにまとめる。
- (3) 地名名詞ごとにまとめたツイートの中から特徴語を特定し、特徴語を含み地名名詞を含まないツイートを、キーワード検索で取得したツイートの中から抽出して加える。
- (4) 関係のないツイートを排除するために、ツイートされた時間を考慮し、時間フィルタリングを行う。
- (5) 緊急地震速報や気象情報などを取得し解析を行う。
- (6) ツイートから検知した障害と外部情報を照らし合わせ、Googleマップ上に表示する。

## 3. システム詳細

### 3.1 地名抽出

キーワード検索で取得したツイートの本文とユーザのプロファイルとGEOタグをMeCab[2]を使ってそれぞれ品詞分解し、地名名詞を抽出する。GEOタグの緯度・経度は、Yahoo!社が提供する「Yahoo!リバーズジオコードAPI」を使って市区町村に変換する。抽出した地名名詞の出現回数をカウントし、5回以上出現した地名名詞ごとに、キーワード検索で取得したツイートをまとめる。

### 3.2 特徴語特定

地名名詞ごとにまとめたツイートだけでは、同じ障害について言及したツイートであるが地名名詞を含んでいないツイートを全て捨ててしまうことになる。そのため、地名名詞ごとにまとめたツイートの中に出現する特徴的な単語を抽出し、抽出した特徴語を含み地名名詞を含まないツイートを、キーワード検索で取得したツイートの中から加えることを考える。

特徴語は、地名名詞ごとにまとめたツイートの中から、MeCabを使って名詞のみを抽出することで見つける。抽出した名詞それぞれに対してtf-idf値を求める。そして、tf-idf値が0.2以上となった名詞を特徴語と定義する。

2014年7月8日にキーワードを「電話 and 繋がらない」としたときに取得したデータに対して、特徴語の特定を行った。tf-idf値が0.2以上になったのは、地震、心配、番号、iPhone、大丈夫、震源、ビックリといった名詞であった。この結果から、地震に関係のある名詞を特徴語として抽出できたことが分かる。地震が発生したときに「地震」「震源」などの名詞が特徴語として抽出されるのは予想することができるが、「心配」「大丈夫」「ビックリ」といった感情的・感覚的な名詞も抽出することができた。このことから、特徴語特定の有効性を示すことができた。

### 3.3 時間フィルタリング

地名名詞でまとめたツイートの中には、内容の異なるツイートが多く混じっている。そのため、ツイートされた時刻を考慮して関係のないツイートを排除することを考える。大規模災害が発生した際には、Twitter上の複数のユーザが特定の時間に似たような内容のツイートをする。本研究では、その点に着目して、ツイートを排除する時間の閾値を決定する。閾値を決定するために、地震に言及するツイート数の時間変化を調べ、一般化した。2014年11月22日に長野県北部で発生した地震の際の通信状況等に言及するツイート数の時間変化を図2の緑の棒グラフに示す。

図2より、地震が発生した直後にツイート数が急速に増えて、その後収束していることが分かるので、地震の場合はツイート数の時間変化が特徴的な形になると考えられる。本研究ではサンプル数が少ないため、累積度数を考える。

累積度数の時間変化を見ると、指数分布の累積分布関数に似た形になっている。そのため、指数分布の累積分布関数を用意し、最小二乗法を使ってフィッティングを行う。指数分布の累積分布関数を以下に示す。

$$f(x) = 1 - e^{-\lambda x}$$

図2に累積度数の実データを赤い点、フィッティングの結果を赤線で示す。

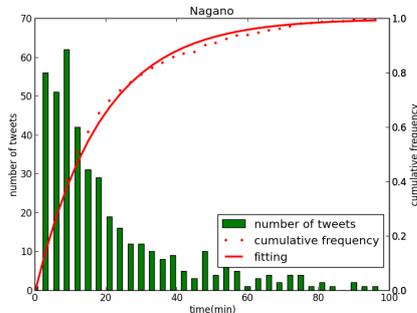


図 2: 通信障害に言及するツイート数の時間変化と累積度数の時間変化の指数分布の累積分布関数へのフィッティング

図2より、指数分布の累積分布関数にフィッティングさせることができた。北海道や関東で発生した地震のケースも同様によくフィットした。このことから、地震の場合のツイート数の時間変化は特徴的であることが分かる。

地震が発生した際に通信障害等に関するツイートが出現してから60分後には、どの実験対象ツイートも累積度数が80%に達しているため、本研究では時間の閾値を60分に設定する。

#### 4. 提案手法の評価

本章では、キーワード検索で取得したツイートに対して、三種類の手法を適用したときのネットワークトラフィック障害検知の効果を検証した。手法は以下の通りである。

- (1) 地名抽出
- (2) 地名抽出 + 関連キーワード抽出
- (3) 地名抽出 + 関連キーワード抽出 + 時間抽出

検証方法としては、実験対象のツイートを正解のツイート、曖昧なツイート、不正解のツイートに分類して手法を適用したときの適合率、再現率、F値を求める。正解のツイートは、地震が原因で通信障害が起きていることが明らかに分かるツイートを人手で集めた。2014年11月22日にキーワードを「電話 and 繋がらない」としたときに取得したデータに対して、三種類の手法を適用したときの結果を表1に示す。

表 1: ネットワークトラフィック障害検知の効果

手法	適合率	再現率	F 値
(1)	0.8071	0.6141	0.6974
(2)	0.6748	0.8913	0.7680
(3)	0.9625	0.8369	0.8953

適合率は特徴語を加えると下がってしまったが、その後ツイートされた時刻を考慮しノイズとなっているツイートを排除することで上げることができた。このことから、ツ

weetされた時刻を考慮することは有用であることが分かる。再現率は特徴語を加えることで大幅に上げることができたが、時間を考慮すると少し下がってしまった。しかし、三種類の手法を全て使うことにより最終的にF値を0.85以上にすることができた。

#### 5. 外部情報の取得

Twitterの情報は即時性があるが、必ずしも正確であるとは言えない。そのため公的機関から発行された外部情報を取得し、信頼性を確保する。本研究ではTwitterで発信される緊急地震速報や気象情報のツイートを外部情報として利用する。

#### 6. 検知した障害と外部情報の可視化

提案手法を使って検知した障害と緊急地震速報から取得した情報を照らし合わせ、Googleマップ上に表示した。



(a) 可視化の結果 (1)

(b) 可視化の結果 (2)

図 3: 長野県北部で発生した地震を可視化した結果

2014年11月22日に長野県北部で発生した地震の結果を図3(a)と図3(b)に示す。ピンの色はツイート数を表しており、多い方から赤桃黄緑青の順である。一方、外部情報から特定した震源地を円の中心に設定し、マグニチュードに応じて円の半径が決まる。ピンをクリックすると、図3(b)のようにピンが刺さっている地域に関するツイートが出力される。図3(a)を見ると、円に近い場所ほど多くツイートが検出されており、震源地付近で大きな障害が起きていることが分かる。この結果から、障害の原因は地震であり、ピンの色と円の位置から提案手法を使って取得した障害と外部情報が一致していることを示せた。

#### 7. まとめと今後の課題

緊急時における通信障害に言及するツイートの時間変化と位置情報に焦点を当て、ネットワークのトラフィック障害を検知する手法の提案と評価を行い、有効性を示した。今後の課題としては、システムの自動化を行いたい。

#### 謝辞

本研究は一部、総務省戦略的情報通信研究開発推進事業(SCOPE)先進的通信アプリケーション開発推進型研究開発によるものである。

#### 参考文献

- [1] T. Sakaki et al., "Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors," Proc. WWW2010, pp.851-860, 2010.
- [2] <http://mecab.sourceforge.net/>