

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

丸 千尋 (指導教員：小口 正人)

## 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桁違いに多くなると、少数のネットワーク管理者が多数の障害を手作業で検知するのに限界がある。この問題を解決するために、本研究では、ソーシャルネットワークサービス (SNS) による集合知を利用した障害検知を迅速かつ高精度に行うシステムを提案する。

## 2 提案システムの概要

本研究では以下の「ネットワーク障害検知システム」を提案する。提案システムの概要を図 1 に示す。

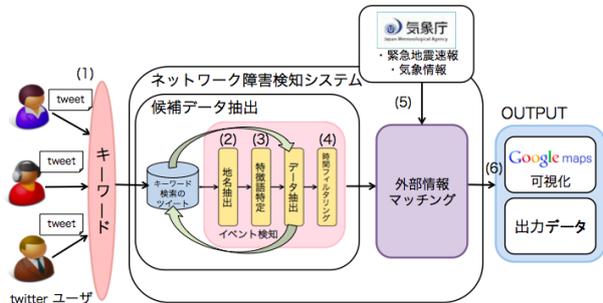


図 1: ネットワーク障害検知システム

- (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刻を考慮し、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の公的機関から発行された外部情報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。

また、Twitter から検出したネットワーク問題に関する情報を出力することで、ネットワーク制御システムに直接連携するという目的を達成するために、ネットワーク障害検知システムの自動化を行った。2014年11月22日22時08分に長野県北部で発生した地震を対象に、Twitter のデータを用いて自動化システムの動作を確認した。自動化にあたり、ツイートのサンプリング間隔を1分とし、現在の時刻から60分遡った時刻までのツイートを扱うことにした。その結果、企業からネットワーク障害に関する第一報が出されたのが、地震が発生してから48分後であったのに対し、提案システムを用いると5分後の22時13分に障害を検出することができた。

## 3 システム詳細

提案システムの候補データ抽出処理の各段階について、以下の各節で説明する。

### 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ーのプロファイルと GEO タグを MeCab を使ってそれぞれ品詞分解し、地名名詞を抽出する。抽出した地名名詞の出現回数をカウントし、一定回数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。ここでは、出現回数の閾値を5回とした。

### 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ま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てく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、有益そうな特徴語を発見することができたと言える。

### 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 上の複数のユーザー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、ネットワーク障害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

