

Twitter を用いた大規模災害時における情報提供システム A Proposal on Information Service System at the Time of Large-Scale Disaster Using Twitter

六瀬 聡宏†
Toshihiro Rokuse

長島 俊††
Suguru Nagashima

内田 理††
Osamu Uchida

鳥海 不二夫‡
Fujio Toriumi

1. はじめに

大規模災害が発生した際、被害を最小限に食い止めるには、災害発生後の迅速かつ的確な情報収集・伝達が重要である。2011 年 3 月に発生した東日本大震災の際には、140 字以内の短文（ツイート）を投稿できる Twitter[1]というマイクロブログにより速報性の高い情報の発信と収集が行われ、注目を集めた[2][3]。また、東日本大震災発生時、Twitter 利用者の約 6 割が情報収集に役立ったとの調査結果も報告されている[4]。しかし、Twitter で流通する情報は膨大であり、また重複する情報やノイズが多数存在しているため、的確な情報を簡便に得ることは容易ではない。

そのような背景から、我々は大規模災害時に Twitter から情報を収集・整理し、ユーザの属性や状況に応じて適切な情報を提供するシステムの構築を目指している。本稿ではその第一歩として、情報の収集・整理の部分を検討した。具体的には、類似ツイートの集約方法を提案し、実際に東日本大震災時のツイートを利用して検証実験を行った。

2. 提案手法

提案するシステムの概要を図 1 に示す。まずツイート全体からハッシュタグを用いて災害時のツイートを収集する。収集したツイート群からユーザが設定した、もしくはユーザの属性や状況に応じて設定されるキーワードが含まれないツイートを削除し、ツイートの絞り込みを行う。絞り込んだツイートから、ユーザが設定した、もしくはユーザの属性や状況に応じて設定される地域のツイートのみを抽出する。その後、情報をさらに整理するため、類似ツイートの集約を行う。具体的には、絞り込み後のツイート群の中のすべてのツイートを比較し、ツイート間の類似度が設定した閾値を超えた場合、最新のツイートのみを残し、それ以外のツイートは削除する。

2.1 ツイートの収集

Twitter API[5]などを利用してツイートの収集を行う。ツイート収集の際、災害に関係するツイートのみを取得するため、ハッシュタグを利用する。ハッシュタグとは、#記号の後に話題を表す文字列を連結したものである（例えば、“#jisin”など）。東日本大震災の際には、地震関連全般向け（避難含む）のハッシュタグとして“#jisin”や“#earthquaken”、“#saigai”、“#hinan”など、原発関連のハッシュタグとして“#genpatsu”や“#houshasen”など、支援に関連するハッシュタグとして“#bussi”や“#shien”など、電気（節電、停電）関連のハッシュタグとして“#setsuden”や“#yashima”などが利用された。また、各地域を区別するた

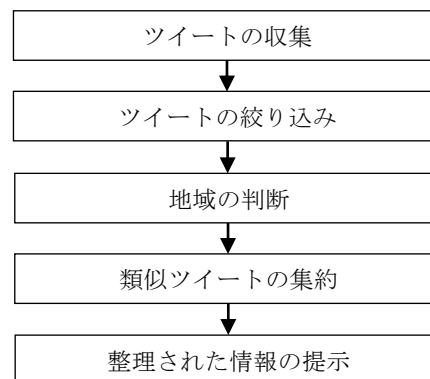


図 1 提案システムの概要

表 1 2012 年 12 月地震発生時に推奨されたハッシュタグ

#地震	地震一般に関する情報
#救助	救助要請
#津波	津波に関する情報
#避難	避難情報
#安否	安否確認

めに、“#iwate”、“#miyagi”、“#fukushima”、“#ibaraki”などが利用され、さらに各地域に関連するより具体的な内容に対応したハッシュタグ、例えば茨城県内のガソリンスタンドに関連するツイート向けの“#gs_ibaraki”なども利用された。

なお、東日本大震災発生時のハッシュタグには“#jisin”と“#jishin”、“#shien”と“#sien”などの表記ゆれが見られ問題となったが、2012 年 12 月 7 日に三陸沖で震度 5 の地震が発生した際には、Twitter 社の公式ブログにより表 1 のハッシュタグの利用が推奨された。

2.2 ツイートの絞り込み

収集したツイート群から不要なツイートを削除しツイートを絞り込む。まずリツイートについては全て削除する。次にユーザが設定した、もしくはユーザの属性や状況に応じて設定されるキーワードが含まれないツイートを削除する。例えば、ガソリン給油に関する情報を収集する場合には、「給油」と「可能」、「ガソリンスタンド」と「営業中」などのキーワードのペアが含まれるツイートは有効であると考えられる。一方、「情報」と「求む」、「情報」と「教えてください」などのキーワードのペアが含まれるツイートは情報提供には有益である可能性が低いと考えられるため、削除することとした。

2.3 地名の判断

ツイート文を形態素解析し、地名と判断された単語を抽出する。その地名が、ユーザが設定した、もしくはユーザの属性や状況に応じて設定される地域に属するかの判定を、住所データベース（例えば住所基盤データベース[6]など）を検索することなどにより行う。

† 東海大学大学院工学研究科情報理工学専攻

†† 東海大学情報理工学部情報科学科

‡ 東京大学大学院工学系研究科システム創成学専攻

2.4 類似ツイートの集約

類似したツイートを集約するため、N-gram 法を用いて類似度を計算する。N-gram 法とは、テキストを N 文字の文字列片に分解する方法である。2 つのテキストに対し、3-gram を用いて分解する例を以下に示す。

- ・テキスト1：東海大学が避難場所
3-gram による分割結果
東海大・海大学・大学が・学が避・が避難・避難場・難場所
- ・テキスト2：避難場所は東海大学
3-gram による分割結果
避難場・難場所・場所は・所は東・は東海・東海大・海大学

類似度の計算には、文字列片の一致割合を用いる。例えば、テキスト1とテキスト2の類似度を計算する場合、両テキストの文字列片7個中4個が一致している。そのため、この2つのテキスト間の類似度は57.1%となる。

本研究では、予備実験の結果をもとに 3-gram を使用することとし、類似度の閾値は 30%と設定した。全てのツイートを比較し、ツイート間の類似度が閾値を超えた場合、類似したツイートであるとし、時系列的に古いツイートを削除する。また、より正確に文章間の類似度を求めるため、前処理として、ハッシュタグとユーザ名は削除しておくこととした。

2.5 情報の提示

情報提示画面の例を図2に示す。キーワードと場所を入力することにより、関連するツイートが出力される。将来的には、キーワードや場所が、ユーザの属性や状況に応じて自動的に入力できるような仕組みを検討・導入する予定である。今回、ユーザが利用する媒体としてはスマートフォンを想定しており、画面の大きさなどを考慮し一度に表示されるツイート数は4とした。

3. 評価実験

3.1 実験方法

提案手法の有用性を検証するため、東日本大震災時のツイートを利用して、類似ツイート集約実験を行った。実験には、“東京都内の避難所”に関するツイート、“宮城県内の給水”に関するツイート、及び“茨城県内のガソリン”に関するツイートをを用いた。評価に用いる正解データ（集約後のツイート群）は人手で作成した。正解データの作成は、3人が独立して行い、2人以上が正解としたツイートを正解データとした。提案手法によって集約されたツイートと比較し、適合率（式(1)）と再現率（式(2)）を求めることで有用性を検証した。

$$\text{適合率} = \frac{\text{正解データと一致した ツイート数}}{\text{提案手法によって集約されたツイート数}} \quad (1)$$

$$\text{再現率} = \frac{\text{正解データと一致した ツイート数}}{\text{正解データのツイート数}} \quad (2)$$

3.2 実験結果

本システムにより要約されたツイート数、正解ツイート数、適合率、再現率を表2に示す。この結果より、提案手法の有用性が検証できたと考えられる。

図2 情報提示画面の例

表2 実験結果

	避難所	ガソリン	給水
集約前のツイート数	258	35	44
提案手法による 集約後のツイート数	108	30	20
正解データ群の ツイート数	154	30	22
適合率[%]	97.2	100	100
再現率[%]	68.2	100	91.0

4. まとめ

我々のグループでは大規模震災時に Twitter から情報を収集・整理し、ユーザの属性や状況に応じて適切な情報を提供するシステムの構築を目指している。本稿では、主に類似ツイートの集約方法を提案し、実際に東日本大震災時のツイートを利用して検証実験を行った。その結果、提案手法の有用性が検証された。

今後の課題として、ユーザの属性や状況を考慮して情報を厳選する方法の検討やデマツイートへの対策、ユーザインターフェースの検討やシステムの実装などが挙げられる。

謝辞

本研究を行なうにあたり、ツイートデータの収集に協力していただいたクックパッド株式会社の兼山元太氏に感謝する。

参考文献

- [1] Twitter, <https://twitter.com/>
- [2] 風間一洋, “Twitter における情報伝播”, 人工知能学会誌, Vol.27, No.1, pp.35-42, 2012.
- [3] 鳥海不二夫, “大震災・そのときソーシャルメディアは動いた・・・のか?”, 第19回社会情報システム学シンポジウム, pp.1-6, 2013.
- [4] MMD 研究所, “東日本大震災後のソーシャルメディア利用に関する実態調査”, 2011.
- [5] Twitter API, <https://dev.twitter.com/>
- [6] 住所基盤データベース, http://www.japan-database.jp/zip_address.html