

Twitter 情報の字句探索による安否確認の信頼性向上の試み Trial Improvement of Twitter-based Information Collecting System for Safety Confirmation

原田 龍一^{*1}, 京兼 正和^{*1}, 窪田 英人^{*2}, 堀 幸雄^{*3}, 今井 慈郎^{*2*3}
Ryuichi HARADA^{*1}, Masakazu KYOKANE^{*1}, Hideto KUBOTA^{*2}, Yukio HORI^{*3}, Yoshiro IMAI^{*2*3}

*1 香川大学工学部

*1 Faculty of Engineering, Kagawa University

*2 香川大学大学院工学研究科

*2 Graduate School of Engineering, Kagawa University

*3 香川大学総合情報基盤センター

*3 Information Technology Center, Kagawa University

Email: s11t263@stmail.eng.kagawa-u.ac.jp

あらまし : Twitter 情報をカウントして統計的に処理するだけでは、内容を識別しないで形式的に扱うことで、意味合いを誤認する危険性(初動対応を誤る点)がある。そこで、キーワードをハッシュタグで明示し、ツイート情報から字句探索を行い、統計処理だけでは見落とされるような誤認の危険性を低減することが重要となる。本稿では、Twitter 情報に「助けて」などのキーワードが存在するかどうかを解析することで、安否確認のみならず、支援要求の有無などを走査・調査する、簡易データマイニングを行い、安否確認の信頼性を向上させる試みについて報告する。

キーワード : 字句解析, キーワードの存在確認, 簡易データマイニング, 安否確認の向上

1. はじめに

Twitter などソーシャルメディアと呼ばれる情報を活用することで、緊急時の状況取得を行うことができる。電話やメールだけでは十分に機能しなかった安否確認などを支援し、災害復旧への対策が可能になると報告されている(1)。特に、東日本大震災とソーシャルネットワークとの解析は多方面で行われ、奥村などの解説記事でもその有効性を論じられている(2)。

我々の研究でも、Twitter 情報を発信者の恣意的な行為とみなし、その行動履歴を基に、時系列表示させることで、安否確認に活用する試み(3)を進めている。しかし、単に Twitter 情報をカウントして統計的に処理するだけでは、内容を識別しないで形式的に扱うことにもなりかねず、発信の意図やその意味合いまでを考慮しないと、誤認をおかす危険性も伴う。

そこで、対象を絞ってその字句解析、具体的には、指定されたキーワードの出現頻度を考慮して、発言の意味づけを行い、誤認の危険性を低減することが重要となる。本稿では、Twitter 情報に「助けて」などのキーワードが存在するかどうかを解析することで、安否確認のみならず、支援要求の有無などを調査する、簡易データマイニングを行い、安否確認の信頼性を向上させる試みについて報告する。

2. Twitter 情報解析の有用性と注意すべき問題点

総務省の情報通信白書においても平成23年3月に発生した東日本大震災の影響が克明に記述され、特に、電話やメールといった従来型情報通信手段が混乱した中、ソーシャルネットワーク(メディア)と

呼ばれる新しい情報通信手段が脚光を浴びている。図1は、震災発生時以降に急激に増加する Twitter の利用数を如実に示すグラフである。

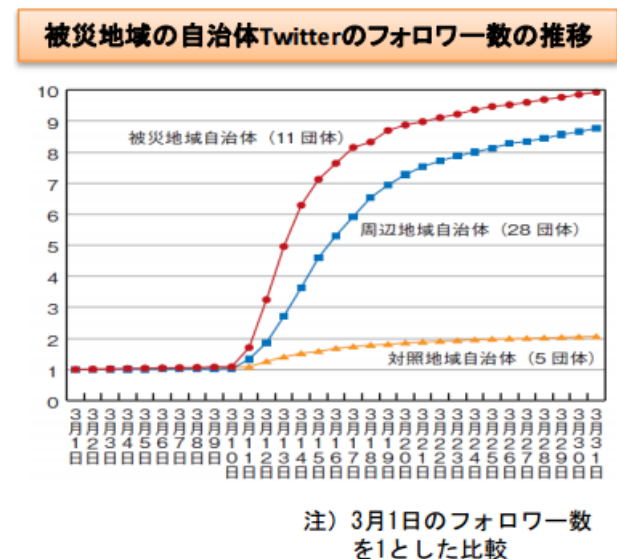


図1 震災発生直後から急上昇する Twitter の利用状況 (情報通信白書 平成23年度版より引用)

このような急激な変化に対しては、人手で情報解析することは難しく、統計処理などよりシステムティックな手法で情報処理することが不可欠となる。実際多くの研究者や技術者が解析していることは、文献(2)にも言及されている。

しかし、問題点もある。これまでデマ情報や詐欺行為など「被害者の傷口に塩を・・・」と指摘されるようなツイートも散見されたとの報告もある。また、

仮に発信者は正しく「助け」を求めているも、発信の時系列変化だけを見て安否確認を行う場合には、発信者は「健在」とのみ報告され、ツイート行為自体を正しく安否確認に反映できない事例も想定される。このような状況への対応がシステム改善の留意点となっている。

3. 提案システム

そこで本研究では京兼らの文献(3)の研究を発展させる形で、Twitter のハッシュタグを用いツイートの内容を安否判定に利用する方法を提案する。

3.1 提案するツイート内容の安否判定手法

まず、Twitter フォローの最新ツイート情報を取得する。その取得情報の文字列に「#help」というハッシュタグ（すなわち、キーワード）が含まれているかどうかを判定し、より発信者の要求に即した状況判断を行う。

発信者の時系列行動の把握（最終ツイートの時刻などの取得）とツイート情報の文字列走査（字句検索）の両方を用いる場合のサーバ内での処理：

- ・災害発生日の確認
- ・対象フォローの最終ツイート確認
- ・災害発生日と最終ツイートの日付から安否を判定
- ・緊急性のある単語が含まれていないか字句検索
- ・検索結果から安否の再判定
- ・判定をもとに一覧表示のためのテーブル作成を、フローチャートで表現したものを図1に示す。

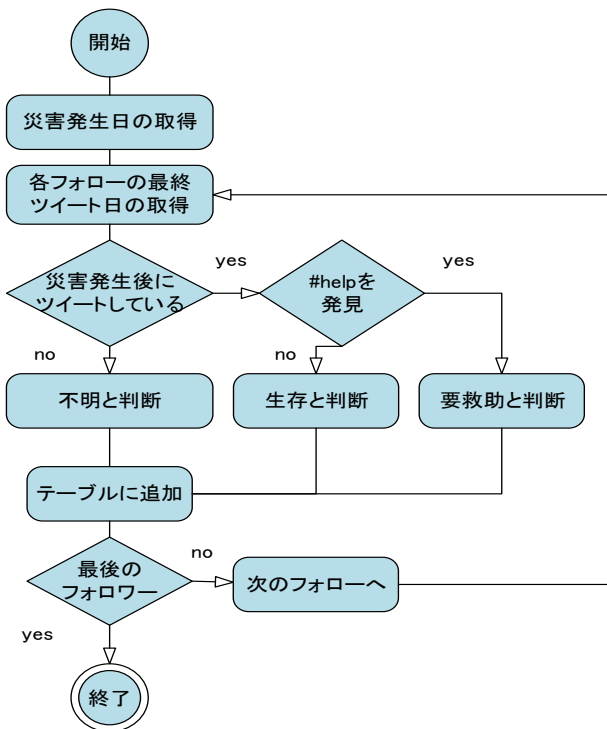


図1 Twitter 情報収集サーバ内の処理フロー

緊急性のある単語の検索のためにハッシュタグを利用し、ハッシュタグ付きの文字のみを検索する。

今回は緊急性のある単語として「#help」を検出した場合にのみ反応するシステムを作成する。

3.2 改善可能な事項（安否確認の信頼性向上）

京兼らのシステムでは、最終ツイートの時刻と災害発生時刻の比較のみによって安否を確認していたため誤検知の可能性も生じる。bot による自動ツイートや、被災して助けを求めるツイートも統計処理の上では、生存情報として分類するので、生存判定のみを返し、救助などの初動対応には寄与しない。

字句検索を用いて被災後に助けを求めるツイートを別途マイニングできれば、安否確認に加え、救助などの初動体制の充填に繋がり、信頼性向上に資すると考えられる。(1)において、Twitter で助けを求めるユーザの存在を平行してマイニングできれば、危険な状態の被災者への安否確認だけではなく、緊急を要する救助も可能になり、機能向上の期待できる。

4. システムの現状と今後の課題

上のような手法で字句検索によって助けを求めるツイートを判定し表示するための仕組みが可能となる。現在、ハッシュタグを付けた場合に当該領域の字句探索に特化して判定する仕様となっている（実装が容易との判断のため）。しかし、ハッシュタグを前提に Twitter データの字句探索を行う仕組みを改良し、ハッシュタグ以外にも「助けて」などの指定単語が文面に出現していた場合にも緊急情報として判定できるよう改造したいと考えている。

状況を想定すれば、それほど冷静には情報発信できない場合も考慮すべき、と考える。具体的には、「助けて」以外のキーワードも探索対象とし、その他の危険を意味する単語も判定できるよう機能拡張（利便性向上）を図りたい。

また、bot による自動ツイートを生存判定にいれてしまう誤検知を取り除くことができていない。何等かの方法で bot のツイートも除外して生存判定を行うようにしたい。

5. おわりに

Twitter 情報から指定の「キーワード」を抽出し、発信者の時系列的行動のみに基づく統計処理だけでは正しく認識されない可能性のある発信者の意図を反映させる仕組みを導入し、統計処理結果への信頼度を向上させる手法について報告した。今後は、システムの完成を目指し、実証的動作確認を目指す。

参考文献

- (1) 総務省. 情報通信白書 平成 23 年版 (概要). <http://www.soumu.go.jp/johotsusintokei/whitepaper/> (アクセス確認：2014-Feb).
- (2) 奥村晴彦：“震災とソーシャルネットワーク”，情報処理 Vol.52 No.9 pp.1072-1073 (2011 年 9 月)
- (3) 京兼正和, 堀 幸雄, 今井慈郎：“安否確認のための Twitter を活用した情報収集システム”，電子情報通信学会技術研究報告 (知能ソフトウェア工学) 113 巻, 277 号, pp. 7-11 (2013)