

Twitter利用者の投稿活動に基づく定量化手法の 災害情報支援システムに向けての展望

Current status and prospects of Twitter users' quantification method based on their posting activity for constructing disaster information support system

*1松本 慎平 *2川口 大貴 *3鳥海 不二夫
Shimpei Matsumoto Hiroki Kawaguchi Fujio Toriumi

*1広島工業大学情報学部

Faculty of Applied Information Science, Hiroshima Institute of Technology

*2広島工業大学大学院工学系研究科

Graduate School of Science and Technology, Hiroshima Institute of Technology

*3東京大学大学院工学系研究科

Graduate School of Engineering, The University of Tokyo

At the time of the Great East Japan Earthquake, many Tweets of the disaster had posted and Twitter had been effectively-utilized as an infrastructure for sharing disaster information and confirming safety. From now the authors have been addressed the researches to utilize Twitter for disaster under the importance of user classification. Concretely by focusing on Twitter user's tweeting, replying, and retweeting activities which is assumed to be the source of Twitter's real time feature, and by numerically-expressing each Twitter user's activities with a quantification method based on entropy, the Twitter users' tendency under the disaster and the possibility for user filtering have been examined. This paper shows the summary of our research results previously reported, and expresses the prospects of the quantification method for constructing disaster information support system.

1. はじめに

東日本大震災発生時、Twitterは貴重な情報交換手段として積極的に活用され重要な役割を果たし、また利用者の情報提供や収集に貢献した[鳥海 14]。災害当日、多くの人々が安否確認や情報収集などの目的で様々な通信手段の利用を試みている。通信回線の被害や停電、トラフィック増や通信規制のために電話やメールが繋がりにくい状況となった一方で、Web閲覧のポケット通信は比較的利用できる状況であったため、多くの人がTwitterを利用し災害に関する大量な情報を即時収集することができた。また、回線に大きな負担をかけることなく情報拡散が可能であった。多くは家族や知人の安否の確認、被災地域の医師による医療相談、行政機関による情報発信などであり、マスメディアに頼らない情報拡散を目的として、被災地内外の地方自治体や個人によって活用されていた[佐々木 11]。

著者らはこれまで、災害時におけるTwitterの有効な活用法の構築を目指して研究を進めている。災害時における利用者分類の重要性を背景に、Twitterのリアルタイム性の源泉であると考えられる利用者の投稿・返信・引用の活動に着目して、災害時のTwitterの使われ方の分析や利用者の属性に応じた特徴の調査を進めている。具体的には、まず、情報量の考え方に基づいて過去の投稿活動を定量化することにより[Ghosh 11]各利用者を多変量定量化し、Twitter利用者の活動を定量的に解釈するための手法の構築を行った。次に、東日本大震災前後にTwitter上に実際に流通した投稿を対象に実験を行い、利用者属性に応じた特徴を分析すると共に、利用者の自動判別に向けての可能性を検証した。本稿では、これまで著者らが進めてきた取組の概要を述べると共に、災害情報システム構築に向けての今後の展望を述べる。

2. ソーシャルメディアの災害時活用に向けての周辺動向

東日本大震災時Twitterでは被災地の状況を知る人間の投稿などマスメディアが報じない貴重な情報が数多く流通していたが、災害に関する情報を効率的に獲得する技術が必要であると考えられる。大規模災害時におけるSNSによる緊急通報の可能性に関する検討会では、災害時において音声通報が途絶えた場合にSNSなどの情報を活用した緊急通報の可能性について報告し、技術的あるいは運用上の課題、SNSの活用方法の在り方を検討しまとめている[総務省消防庁 13]。Twitter上には様々な種類の膨大な量の投稿が常に流通しているため、災害時においてTwitterを円滑に利用できるようにするためには、目的に応じて投稿を的確に自動判別する仕組みが必要であると述べている。また、発信者の位置と災害の適合性、情報の信頼性向上を課題としてあげている。同様に、鳥海らの取り組みでは[鳥海 14]、情報処理に関する課題として、情報の時間的・空間的な精度劣化への対応、情報の信頼性保証、情報の不足・欠落への対応、大規模処理をあげている。この中で信頼性を確保するための技術課題として、ノイズ除去、情報統合、デマの発見と防止、シミュレーションによる精度の検証が示されているが、災害時におけるソーシャルメディアを対象とした多くの研究は、以上で示されたフレームワークのもとで遂行されている。

3. 著者らの取り組みの位置付け

災害時において重要となる情報は、災害に強く関係する利用者や被災者自身によって発信される場合が多い。これら利用者の訴えを的確に抽出するためには、人が発信した情報の識別や利用者の属性に応じた選別、あるいは自動投稿プログラム(bot)により発信される投稿の検知が重要な課題であると考えられる。また、誤情報の拡散は、普段のTwitterには見られ

連絡先: 松本慎平, 広島工業大学情報学部知的情報システム学科, 〒731-5193 広島市佐伯区三宅 2-1-1,
E-Mail: s.matsumoto.gk@cc.it-hiroshima.ac.jp

ない特別な使い方をした利用者や、他のリソースを自動引用する bot が原因のひとつであったと考えられている。以上より、bot の検出のみならず利用者の活動形態の把握は、誤情報へ対応するための重要な課題であると考えられる。災害時の Twitter 利用者の特徴を計算可能な形で容易に把握できれば、bot の検出・除外、また、事実を知っている被災者の声の取得に活用可能であると考えられる。そしてその結果として、誤情報の訂正・誤情報拡散の抑制に向けての活用や、被災した地域の人達の生の声の取得が期待できる。利用者の選別、とりわけ bot 検出に着目すれば、その技術として、投稿内容に基づいた機械的フィルタリングが考えられる。しかしながら、投稿内容に基づいた情報選別は従来静的な情報を対象にしており、さらに、これらは現在技術的に成熟の域にある。ソーシャルメディアの実時間性に適合した情報選別手法を構築するためには、質的手法以外からの観点に基づいた計算手順が必要であると考えられている。以上背景のもとで、著者らは利用者の投稿活動に基づいた定量化手法に着眼し、多変量によって表現された災害時における利用者の定量解釈とその分析、利用者分類に向けての可能性検証に取り組んできた [川口 13a].

4. これまでの成果

Twitter 利用者の定量化手法は、処理日を起点として前数日を解析対象期間とし、期間内の利用者の通常投稿、Retweet, Reply の活動パターンからエントロピーを算出する。まず、本研究の手法の基礎となった Ghosh らの手法との比較を行った。Ghosh らの手法では、周囲の Retweet 活動を分析対象としている。一方、本研究では災害時の活用が目的であり、事例に対する即時性を保つことの重要性を踏まえ、エントロピーの算出にかかる計算量を考慮して利用者が投稿した全投稿のみを対象とした。実験の結果、対象データの質が大きく異なるにもかかわらず、両手法とも数値の差違はあるものの利用者の特徴付けが概ね可能であることが確認された。また、震災時における国内の Twitter 情報においても Ghosh らが示した特徴と同様の傾向があり、定量化による分類が可能であることからエントロピーによる定量化手法は普遍性を持つことが示唆された。

次に、解析対象期間の長さがエントロピー時系列に与える影響を分析した。その結果、解析対象期間を最も短く設定した場合に突発的事象の影響が最も顕著に表れていること、一方で、分析期間が広いほどエントロピーの値の変動の反応やボラティリティが小さくなることを明らかにした。突発的事象の検出と利用者属性の分類は相反の関係にあることを踏まえて、突発的事象の検出と利用者属性の分類を同時に行うための手法を提案した。ここでは他に、エントロピーの時間間隔を様々な観点から与えることにより、従来法の特徴を失うことなく次元を拡張することに成功した。

以上の取り組みの中では提案手法の自動判別への応用可能性を検証しており、実験の結果からは、多変量定量化により利用者属性に応じた差違を表現できたばかりでなく、自動判別の応用可能性を示すことができた。

5. まとめと今後の展望

本稿では、これまで著者らが進めてきた取組の概要を述べた。災害情報システム構築に向けての今後の展望として、多変量で表現された利用者の値から利用者同士の距離を算出し、この値を活用した Web システムの開発と応用を考えている。具体的には、投稿活動と投稿履歴の関係性を同時に把握するため、エントロピーの観点から投稿活動の近い複数利用者の投稿



図 1: タグクラウドシステムの一例

内容をタグクラウドで表現するシステムの構築を検討している (図 1 参照)。まず、多様な投稿内容の利用者を判別するため、予め用意された災害キーワードの重複率を Jaccard 係数を用いて利用者ごとに特徴付ける。似た投稿活動の利用者同士はある程度の共起性を持つが明確に投稿内容が重複はしていない点を踏まえ、周囲の距離の短い利用者から収集したテキストデータを用いてタグクラウドを生成する。タグは周辺の利用者の投稿履歴から品詞情報が名詞であるものを対象とする。以上の詳細は当日発表で示す。

参考文献

[Ghosh 11] Ghosh et al.: Entropy-based Classification of ‘Retweeting’ Activity on Twitter, Proc. of KDD workshop on Social Network Analysis (SNA-KDD), <http://arxiv.org/pdf/1106.0346.pdf> (2011).

[川口 13a] 川口 他: Twitter 利用者の定量化手法の災害時活用に向けての有効性評価, 第 15 回 IEEE 広島支部学生シンポジウム論文集, pp.474-477 (2013).

[佐々木 11] 佐々木: 拡大を続ける Twitter の震災における活躍と今後の課題, AD STUDIES, Vol.36, pp.20-24 (2011).

[総務省消防庁 13] 総務省消防庁: 大規模災害時におけるソーシャル・ネットワーク・サービスによる緊急通報の活用可能性に関する検討会報告書 (2013), http://www.fdma.go.jp/neuter/topics/houdou/h25/2503/250327_1houdou/02_houkokusho.pdf, 2013/12/10 参照

[鳥海 14] 鳥海 他, 異種協調型災害情報支援システム実現に向けた基盤技術の構築, 人工知能学会論文誌, Vol.29, No.1, pp.113-119 (2014).