Vol.2016-MPS-111 No.17
2016/12/12

SNS における流言拡散の時系列データ分析の一考察

牛込龍太郎[†1]　松田健[†2]　園田道夫[†3]　趙晋輝[†1]

概要：ユーザー自身が様々な情報を発信することができる SNS は，マーケティングや医療情報連携，災害時利用など多岐に渡ってその活用方法が検討されている．しかしながら，インターネットや SNS に投稿される内容には事実とは異なるデマや誤りも拡散されている可能性があり，特に災害時のような緊急の場合に，その信憑性をできる限り早く確認できる手法の開発は必要性が高いと言える．しかしながら，大規模災害時にはおいては投稿内容の単語頻度解析は困難であることが指摘されており，様々な情報を統合して投稿内容の信憑性を判断する必要があると考えられる．本研究では，流言が拡散されていく状況を時系列的に解析し，他者の投稿や他の関連する情報と付き合わせることで，その信憑性の度合いの変化をモデル化することで，投稿された内容がデマである場合にそれがどのように収束していくかを考察する．

キーワード：SNS，デマ

1. はじめに

SNS(Social Network Service)はユーザー同士のサイバー空間上でのコミュニケーションツールとしての機能を主な役割とする Web サービスである．SNS サービスはその手軽な情報発信能力という長所を，大衆の情報発信手段だけでなく企業におけるマーケティングや教育分野などにおいても活用されている一方，発信される情報の正当性が保証されているとは限らず，事実と異なる情報が多くの人に伝わってしまう可能性がある．特に大規模な災害の発生時に悪質な流言が拡散した場合，内容次第では現場での情報の混乱を招き，支援活動の妨げとなる可能性も考えられる．このような状況の現場では情報の真偽の判定に人手を割くことは好ましくなく，機械的に情報が流言かどうか判断できることが望まれる．その機械的な判断を可能にするために，流言拡散のプロセスを分析することは重要である．本稿では SNS サービスの一つである Twitter[1]に投稿された流言と通常の書き込みを収集した．また投稿された流言の信憑性の度合いの数値化を行うと共に，流言のリツイートに対して形態素解析を用いてリツイート(RT における感情を数値化し，得られた結果について考察を行った．

2. 関連研究

既存研究[2]では頻繁にリツイートされているツイートを検出するとともに，各ツイートの形態素解析の結果からツイートの感情を数値化している．またその数値とリツイートの反復の程度（リツイートの深さ）に対して SVM を用いてデマツイートかどうかを判別している．また別の既存研究[3]では，デマの内容が多種多様であるため，デマツイート自体を直接抽出することは困難であるとしているが，一方デマであると指摘するツイートは特徴的な語句を含むことが多いため，これらの抽出を行うシステムを提案している．前者の既存研究ではリツイートを通じた情報の広まりについての言及がなされていない．また後者の既存研究はデマを指摘するツイートの抽出条件を追加していくことで特徴量が増加することが懸念される．

3. 提案手法

提案手法を述べるにあたり，Twitter のリツイート機能の概要を記す．リツイートとは他のアカウントがツイートした内容を複製し，自アカウントでもツイートする機能である．複製時に自分の文を加えてツイートすることもできる．本稿では区別のため，後者を「引用リツイート」と呼ぶ．

データの収集は主に Twitter の Web ページから行った．収集の対象は流言のツイート（デマツイート）およびそのリツイートである．デマツイートは 5 件，それらのリツイートは合わせて 412 件集めた．ここでデマツイートとは，事実と異なる情報を発信するツイートを指す．その後，リツイートデータに対してユーザーがデマツイートに対してどの程度信頼を置いているかを 1, 0.5, 0 の三段階で評価した．各値の意味とツイートに付す基準は以下の通りである．本研究においてはこれらの基準は著者の主観で定めた．

信頼度=1：信頼している

- デマツイートをリツイートしたもの
- 感嘆詞やそれに準ずる表現を含む引用リツイート
- （犯罪，事件などのデマツイート事象に対して）危惧を含む引用リツイート

信頼度 0.5：信頼も疑いもしない

- デマツイートの後に，デマツイートの事象と関連のない文をつぶやいている引用リツイート

信頼度 0：疑っている

- デマツイートに対し否定語句を含む引用リツイート

また，リツイートデータのうち引用リツイートに該当するものに対して日本語形態素解析システム juman[4]を用いた形態素解析を行い，得られた結果に対して日本語評価極性辞書[5]でツイートの感情に関するスコア付けを行った．スコア付けの方法として極性辞書において感情的にポジティブ、ネガティブであると評価されている語句にそれぞれ+1, -1 を加算し，加算の結果を形態素の数で除した．

†1 中央大学理工学部情報工学科
†2 長崎県立大学情報システム学部情報セキュリティ学科
†3 サイバー大学

4. 分析と考察

5 件のデマツイートを hoax1～hoax5 とし，リツイートに対して提案手法のスコア付けを行った．得られた結果を表 1, 2 に示す．表 1 は各デマツイートのリツイートに対する信頼度ごとの件数をあらわしたものである．ここで信頼度が 0.5 の列名はその他の分類としてある．表 2 はリツイートの positive/negative のスコアが左列より 0 より大きい(ツイート内容がポジティブ)，0 より小さい（内容がネガティブ)，0 に等しいものの件数と割合をあらわしたものである．表 1 より hoax1,3,4 に対するリツイートのおよそ 75%以上がデマの内容を信頼していることが確認できる．これらのデマは事件・事故に関連する内容であることから，この類の内容のツイートはユーザーに信用されやすいと考えられる．また表 2 では positive/negative スコアが 0 に等しいものが hoax1~5 全てにおいて多数を占めていることが分かる．これはリツイートしたユーザー独自のつぶやきの長さが短く形態素の数が少なくなったためと考えられる．また感情を表現する上で本来重要視するべきと考えられる顔文字がただの記号として解析されてしまったことも要因であると考えられる．顔文字は SNS では頻繁に用いられ，今回のデータにも無視できない件数のツイートに顔文字が含まれている．これについては極性辞書と juman において顔文字を独自に定義することで回避できるものと推測される．その一方で，hoax5 以外では positive よりも negative の割合の方が高いことも見て取れることから，デマツイートに対して信用するユーザーよりも疑念を抱くユーザーの方が多いと分かる．また，リツイートされ続けた時間が最も短い hoax1 と最も長い hoax4 の信頼度と positive/negative スコアを時系列順にグラフ化した（図 1～4）．各グラフの横軸はデマツイートが投稿されてからの経過時間を，縦軸は信頼度もしくは positive/negative スコアを表す．2 つの信頼度のグラフ図 1,2 から，デマツイートが投稿されてから早い段階ではデマツイートの内容が信頼されていることが分かる，一方，図 2 より時間が経過すると疑念を抱くリツイートが増加していることが分かる．このことからデマツイートの信頼度と時間経過の間には何らかの関係があることが予想される．

5. まとめと今後の課題

本稿では，デマツイートとそのリツイートのデータセットから，デマに対する信頼度と感情極性を評価した．評価の結果，デマが投稿されてからの時間によってデマに対する信頼度が異なること，またデマに対してユーザーはネガティブな印象をもつ傾向が見て取れた．今後は，より多くのデータの収集や極性辞書の語句の追加，positive/negative 評価の見直し，信頼度についてのより客観的な基準作成などを行い，より精度の高いモデルの作成を行う予定である．

表 1 デマツイートに対する信頼度

	信頼	不信	その他
hoax1	20(100%)	0(0%)	0(0%)
hoax2	77(43%)	86(48%)	15(8%)
hoax3	29(94%)	1(3%)	1(3%)
hoax4	111(77%)	29(20%)	5(3%)
hoax5	16(42%)	16(42%)	6(16%)

表 2 デマツイートに対する Positive/Negative スコア

	>0	<0	=0
hoax1	1(5%)	1(5%)	18(90%)
hoax2	40(4%)	6(29%)	91(66%)
hoax3	4(12%)	3(16%)	18(72%)
hoax4	17(13%)	24(18%)	89(68%)
hoax5	9(26%)	2(6%)	24(69%)

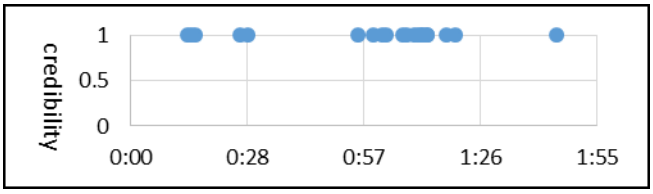

図 1 hoax1 信頼度

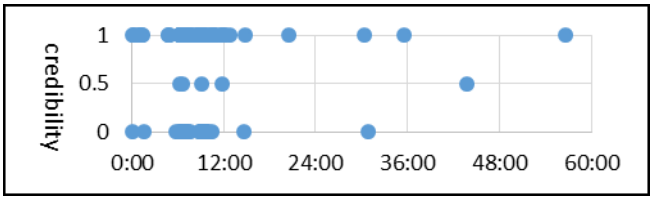

図 2 hoax4 信頼度

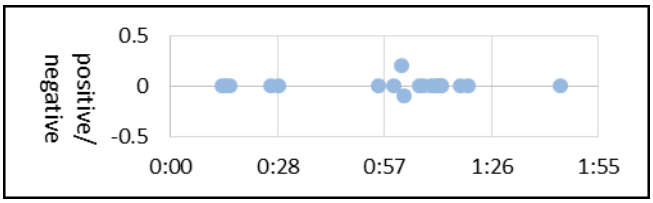

図 3 hoax1 positive/negative スコア

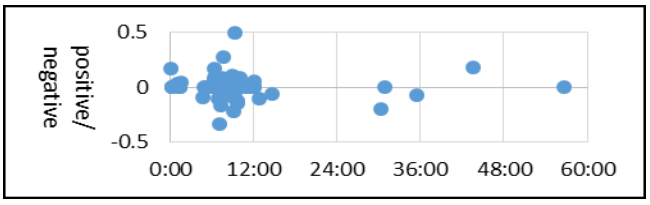

図 4 hoax4 positive/negative スコア

参考文献

[1] “Twitter”. https://twitter.com/ (参照 2016-11-13)
[2]須田 剛裕，小嶋 和徳，伊藤 慶明，石亀 昌明，鳥海不二夫，震災時におけるツイッターのトレンドワードと拡散情報を利用したデマ推定の一考察，第 75 回全国大会講演論文集, pp.99-100, 2013
[3]渡邊 建太，山田 剛一，絹川 博之，訂正投稿の傾向を利用したデマ訂正ツイートの抽出，情報科学技術フォーラム講演論文集
[4] “JUMAN - KUROHASHI-KAWAHARA LAB”. http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN (参照 2016-11-13)
[5]日本語評価極性辞書(名詞編) 東山昌彦，乾健太郎，松本裕治，述語の選択選好性に着目した名詞評価極性の獲得，言語処理学会，