

SNS においてフェイクニュースを拡散するユーザーの特徴抽出

岩橋 瑠伊[†] 矢吹 太郎

千葉工業大学 社会システム科学部 プロジェクトマネジメント学科[‡]

1 序論

近年 Facebook や Twitter を始めとしたマイクロブログがスマートフォンなどの普及と共に急激に普及している。Twitter は 140 字以内で記事 (Tweet) を投稿するシンプルな構造であり、リアルタイム性と情報の強い伝播力をもつ。中でも、リツイートは他のユーザーの Tweet を引用し、情報を伝播させるものであり、Twitter 特有の重要な機能である [1]。

Twitter は 2011 年 3 月 11 日に発生した東日本大震災時に、携帯電話がつかない状況下での有用な連絡手段として活躍した。しかし、その有用性はデマや誤情報も大量に拡散させる手助けとなりえる。実際に東日本大震災時に、数十種類のデマや誤情報が情報として拡散されてしまい、日本中を混乱させた。震災時のように連絡手段が限られた状況はこれからも発生する可能性は十分にあり、対策が必要である [2]。

本研究では、デマが拡散されることを防ぐためにデマツイートをリツイートしているユーザーの特徴抽出を行う。デマツイートがリツイートされる原因として、デマをデマと見抜けないユーザー、面白半分でリツイートしているユーザーの 2 種類がいると考えた。この 2 種類のユーザーと、それ以外のユーザーには Twitter の使い方に違いがあるのではないかと考えた。

2 目的

デマが拡散されることを防ぐために、デマツイートをリツイートしているユーザーの特徴抽出を行う。そのために、現時点でデマだとわかっているツイートを過去に拡散したユーザーの活動履歴を収集・分析する。分析結果をランダムサンプリングしたデータと比較することで、デマを拡散するようなユーザーに共通する特徴を抽出し、その結果を報告する。

3 手法

デマツイートをリツイートするユーザーとそれ以外のユーザーの違いを見つけ、その違いが偶然生じたものではないことを示すために以下の手法で研究する。

1. TwitterAPI を用いて日本人ユーザー 50 人をランダムサンプリングする。
2. TwitterAPI を用いてデマツイートをリツイートしたユーザー 50 人を取得する。
3. TwitterAPI を用いて集めた各ユーザーの最新 100 ツイートに含まれるリツイートの数を調べて、平均を計算する。
4. 日本人ユーザー 50 人とデマツイートをリツイートしたユーザー 50 人の最新 100 ツイートに含まれるリツイートの数の平均の差が、偶然的な誤差の範囲にあるものかどうかを判断する為に 2 標本 T 検定 [3] を行う。

4 結果

TwitterAPI を用いて日本人ユーザー 50 人をランダムサンプリングした。4 つのデマツイートからそれぞれ 50 人のリツイートユーザーを取得した。ランダムサンプリングした日本人ユーザー 50 人の平均リツイート数は 20.04 人、デマツイート 1 の平均リツイート数は 56.68 人、デマツイート 2 の平均リツイート数は 62.64 人、デマツイート 3 の平均リツイート数は 58.46 人となった、デマツイート 4 の平均リツイート数は 57.92 人となった。デマツイート 1 と日本人ユーザー、デマツイート 2 と日本人ユーザー、デマツイート 3 と日本人ユーザー、デマツイート 4 と日本人ユーザーのそれぞれ 3 組には対応がないデータなので、F 検定を行い分散が等しいか等しくないかを確かめる。等分散の場合の 2 標本 T 検定と不等分散の場合の T 検定を F 検定の結果に沿って行った結果、全ての組み合わせで有意差が確認できた。以下にデマツイート 1 から 4 のヒストグラムを示す。

Feature extraction of users spreading fake news in SNS.

[†] Rui Iwahashi (MAIL)

[‡] Department of Project Management, Faculty of Social Systems Science, Chiba Institute of Technology.

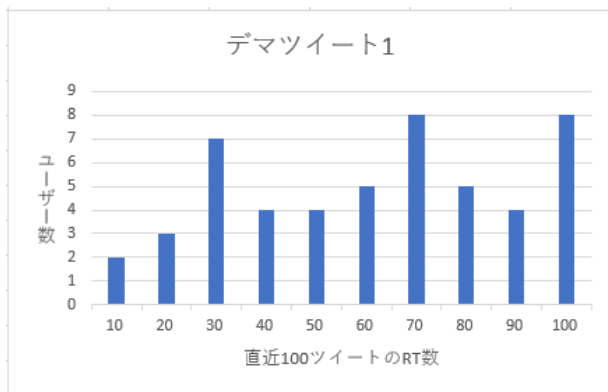


図1 デマツイート1のヒストグラム

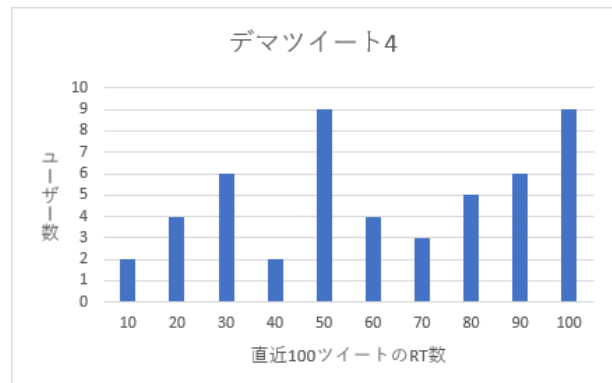


図4 デマツイート4のヒストグラム

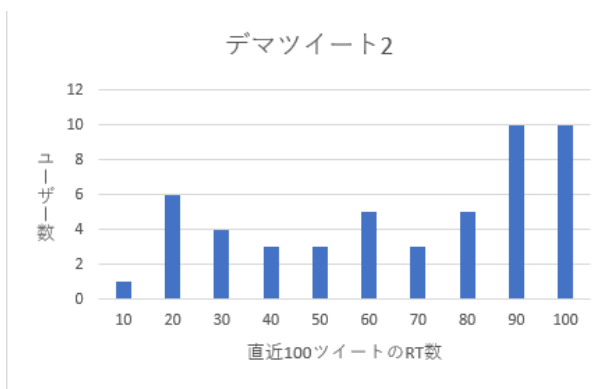


図2 デマツイート2のヒストグラム

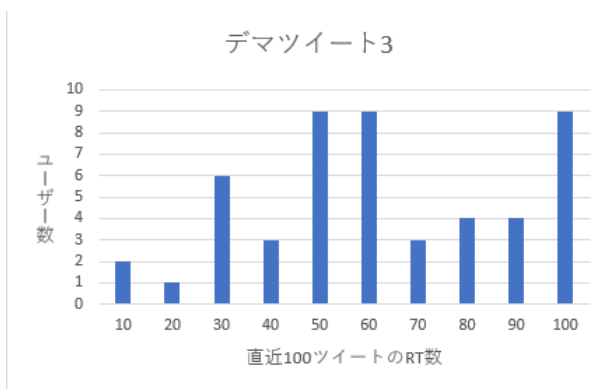


図3 デマツイート3のヒストグラム

5 考察

デマを拡散するようなユーザに共通する特徴として、リツイート数に着目しランダムサンプリングしたユーザーとデマツイートをリツイートしたユーザーで直近100リツイート内のリツイート数を調査した結果、大きく数値が異なった。この結果からデマを拡散するようなユーザーはリツイート機能を多用する傾向にあり、ツ

イート内容の真偽を確かめる前にリツイートをし、デマ拡散者の一員となっていると考えられる。

自分がデマ拡散者にならない為の手段として、デマ拡散ユーザーリストにあるユーザーと、リツイートの多いユーザーを排除することが有効だと考えられる。

6 結論

本研究では、デマが拡散されることを防ぐために、デマツイートをリツイートしているユーザーの特徴抽出としてリツイート数の調査を行った。その結果、デマツイートを拡散するユーザーの特徴としてリツイート数が多いことを証明することができた。データ数が少ないことは否めないが、T検定の結果からデータの整合性は高いものと判断できる。本研究のデータを利用し、リツイート数以外のデータ（ツイート数、フォロー数、フォロー数）などからもデマ拡散者の特徴を抽出し、更なる精度の向上とデータ数の増加に期待される。

参考文献

- [1] 荒川唯, 亀田堯宙, 相澤彰子, 鈴木崇史. Retweet に着目した広がりやすい tweet の特徴分析. 第74回全国大会講演論文集, Vol. 2012, No. 1, pp. 617–618, mar 2012.
- [2] 榎本光, 内田理, 鳥海不二夫. O-054 東日本大震災時のツイート分析によるデマ判別に有用な特徴抽出 (o 分野:情報システム, 一般論文). 情報科学技術フォーラム講演論文集, Vol. 12, No. 4, pp. 649–650, aug 2013.
- [3] 統計 WEB. 2 標本t検定とは. <https://bellcurve.jp/statistics/course/9427.html> (2017.12.20 閲覧) .