

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

SNS 上でフェイクニュースを拡散するユーザーの特徴について調査する。SNS などのウェブ上のメディアで、フェイクニュースが問題視されている [1]。

例えば、2011 年 3 月 11 日に発生した東日本大震災時に、携帯電話が繋がら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. ユーザ ID を乱数で指定し、日本人ユーザ 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 50 人を取得する。
3. TwitterAPI を用いて集めた各ユーザの最新 100 ツイートに含まれるリツイートの数を調べる。
4. 日本人ユーザ 50 人とデマツイートをリツイートしたユーザ 50 人の直近 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 検定を行う。

4 結果

ユーザ ID を乱数で指定し、日本人ユーザ 50 人をランダムサンプリングした。

以下の 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を取得した。

- デマツイート 1 として、ロンドンのテロ対策で有志のスパルタ兵が巡回しているというデマツイートがある。実際は映画の宣伝である。(TweetID:872255950131822596)
- デマツイート 2 として、国内のアリがヒアリに対抗出来るというデマツイートがある。実際は誤情報である。(TweetID:883170290242527232)
- デマツイート 3 として、PS4 に PS2 互換機能対応するというデマツイートがある。実際は誤情報である。(TweetID:882139486968205312)
- デマツイート 4 として、路上で乾燥海産物を売る人がいるが、それは麻酔薬で匂いを嗅いだら意識を失うというデマツイートがある。実際は昔韓国で流行したデマの内容と同じものである。(TweetID:923151745923948545)

ランダムサンプリングした日本人ユーザ 50 人の直近 100 ツイート中の平均リツイート数は 20.04 人、デマツイート 1 をリツイートしたユーザ 50 人の直近 100 ツイート中の平均リツイート数は 56.68 人、デマツイート

Feature extraction of users spreading fake news in SNS.

[†] Rui Iwahashi (s1442014cr@s.chibakoudai.jp)

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

2 をリツイートしたユーザ 50 人の直近 100 ツイート中の平均リツイート数は 62.64 人、デマツイート 3 をリツイートしたユーザ 50 人の直近 100 ツイート中の 58.46 人となった、デマツイート 4 をリツイートしたユーザ 50 人の直近 100 ツイート中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、デマツイート 2 と日本人ユーザ、デマツイート 3 と日本人ユーザ、デマツイート 4 と日本人ユーザのそれぞれ 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。(有意水準は 5 %)

以下にデマツイート 1~4 をリツイートしたユーザの 100 ツイートあたりのリツイート数のヒストグラムを示す。

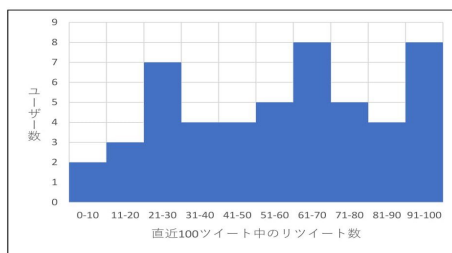


図 1 デマツイート 1 についての結果

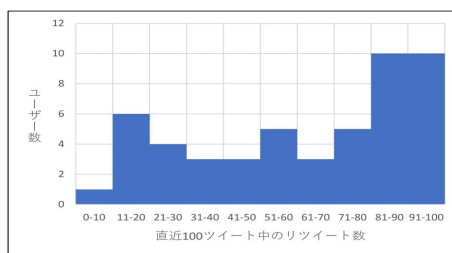


図 2 デマツイート 2 についての結果

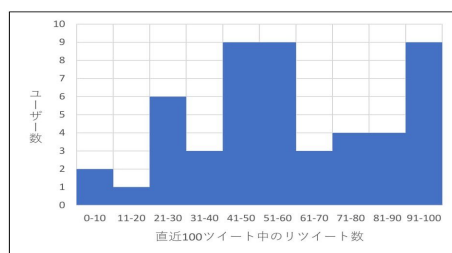


図 3 デマツイート 3 についての結果

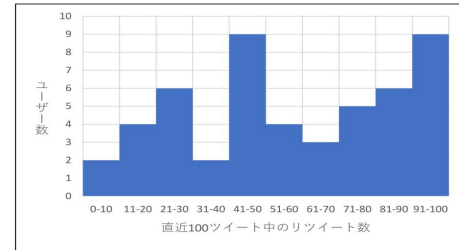


図 4 デマツイート 4 についての結果

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が確認できた。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の将来性がある。

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, 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 (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, 2013.