

SNS においてフェイクニュースを拡散するユーザの特徴抽出

プロジェクトマネジメントコース 矢吹研究室 1442014 岩橋瑠伊

1. 序論

スマートフォンなどの普及と共に、Twitter を始めとしたマイクロブログが普及している。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。例えば東日本大震災時には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発生する可能性は十分にあり、対策が必要である [1]。

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の特徴抽出を行う。

3. 手法

デマツイートをリツイートするユーザとそれ以外のユーザの違いを見つけ、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

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, 2, 3, 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、1 が 56.68 人、2 が 62.64 人、3 が 58.46 人、4 が 57.92 人となった。

デマツイート 1, 2, 3, 4 と日本人ユーザのそれぞれ 4 組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 %）。

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とデマツイートをリツイートしたユーザで直近 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、デマ拡散ユーザリストにあるユーザと、リツイートの多いユーザ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、デマツイートをリツイートしているユーザ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の特徴として、ツイートに占めるリツイートの割合が高いことが確認できた。このような知識を活用することで、Twitter を閲覧する際に、デマツイートを真に受けて拡散してしまうリスクを下げられることが期待できる。

参考文献

- [1] 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, 2013.