

SNS においてデマを拡散するユーザーの特徴抽出

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. 背景

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。特に Twitter は手軽にリアルタイムな情報を多くのユーザーに伝搬出来るため社会に大きな影響を与えている。例えば 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。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も発生する可能性は十分にあり、対策が必要である [1]。

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。デマツイートを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 Twitter の使い方に違いを見つける。違いからデマツイートをリツイートするユーザーなのかを判別できるようにする。

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証明する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 検定を行う。

4. 想定される成果物

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 Twitter の使い方に違いが見つかる。例として 2 標本 T 検定の結果、日本人ユーザー群とデマツイートをリツイートしたユーザー群の最新 100 ツイートに含まれるリツイートの平均値に有意差が確認できる。

5. 進捗状況

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。3 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。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 は 56.68 人、デマツイート 2 は 62.64 人、デマツイート 3 は 58.46 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 と日本人ユーザーのそれぞれ 3 組で 2 標本 T 検定を行った結果、全て有意差が確認できた。

6. 今後の計画

データ量が少ないのでデータ量をさらに増やすために新しいデマツイートを発見し分析を行う。

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, aug 2013.