

Twitter のフォロワー数の大小によるツイート拡散の傾向の分析

松香研究室 16L1053U 林 亜紀

1. はじめに

インターネットの普及により、ソーシャルネットワーキングサービス(SNS)は今や情報伝達やコミュニケーションにおける重要なツールとなった。利用者の加齢などの理由により SNS 利用率が年々増加していることに加え、情報収集・発信のためという目的での利用が増加していることから、SNS での情報伝達の重要性はますます拡大しているといえる。

Twitter は、その中でも特に情報のやり取りが盛んな SNS で、全角 140 字もしくは半角 280 字の短文(ツイート)を投稿することができる情報サービスである。特定のユーザーをフォローすることで、そのユーザーのツイートが自分のホームタイムラインに表示されるようになる。Twitter で情報を共有する場合は主にリツイート (RT) という機能を用いる。自分以外のユーザーのツイートや自分が既に投稿済みのツイートを RT することで、自分と自分をフォローしているユーザー(フォロワー)のタイムラインに表示・再表示させることができる。自分のツイートをフォロワーが RT し、そのツイートがフォロワーのフォロワーのタイムラインに表示され、またそれが RT されることで、特定のツイートが広く拡散されていく。Twitter は匿名でアカウントを運用できる上に、一方的にフォローすることが可能なため、ユーザー同士の繋がりが他の SNS よりも緩く、より広範囲への情報の拡散が容易である。

RT による情報の拡散の研究では、ツイートの内容もしくはフォロワー数に注目した研究が多い。RT したユーザーのフォロワー数に関する研究では、フォロワー数が多く影響力のあるユーザーによるツイートからの拡散よりも、平均的もしくは平均以下の影響力を持つユーザーが行う情報拡散に期待する方が効果的であるとされている(Bakshy, Hofman, Mason, & Watts, 2011)。山下・三浦(2018)では、フォロワー数の少ないユーザーが投稿したツイートが拡散された例を研究し、ツイートの拡散範囲と拡散速度を一気に速めるのはフォロワー数の多い強力なインフルエンサーであるが、実際に広範囲に拡散するかどうかはある程度アクティブなユーザーによる RT の積み重ねによると考察した。

本研究では、Twitter の RT による情報拡散について、先行研究からユーザーのフォロワー数に着目し、実際に元ツイートの投稿ユーザーや RT したユーザーのフォロワー数によって RT の様相にどのような傾向があるのか分析した。

2. 分析 1

2.1. 目的

RT 数の多いツイートで、ユーザーのフォロワー数の大小によって、その拡散速度・範囲にどのような傾向があるのか、先行研究と同じような傾向は現れるかどうかを分析した。

2.2. 方法

Twitter の日時別 RT 数ランキングが確認できる外部サイト「TwTimez (<http://www.twtimez.net>)」を利用して、サイト内の 2019 年 6 月 1 日の人気順上位 50 ツイートから、10 ツイートを分析に用いた。フォロワー数の違いの基準として、Twitter 社に公式アカウントと認定されているアカウントによるツイートとそれ以外のアカウントによるツイートを区別し、それぞれ 5 ツイートずつ用いた。API 制限で取得上限が 18,000 ツイートまでという関係上、元ツイートのまで遡って取得することができるツイートのみを分析対象に選択した。

分析対象ツイートの TwitterID (例: @abcde) とツイートの内容をキーワード検索にかけることによってその RT を収集した。RT の収集にはキーワードやユーザー名を用いてツイートやユーザー ID、ツイート日時などが収集できる外部サイト「Web Tweet Crawler (<http://torix.sakura.ne.jp/twitter/>)」を利用した。

収集したデータのツイート日時から 10 分ごとの RT 数とユーザーのフォロワー数を調べた。

2.3. 結果と考察

ツイートの 10 分間毎の RT 回数を可視化し、フォロワー数の多いユーザーによるツイートとフォロワー数の少ないユーザーによるツイートを例として一つずつ挙げた(図 1)。

フォロワー数の多いユーザーによるツイートはどれも投稿直後に爆発的に RT されていた。フォロワー数の多いユーザーが深夜に投稿したツイートは朝にも RT が増えており、ユーザーの活動時間帯との関連があると考えられる。その後 RT 速度が急激に上がることはなく、フォロワー数の多いユーザーによるツイートの RT 速度は時間経過に応じて減少していった。

一方で、フォロワー数の少ないユーザーによるツイートで RT に波があり、一回で爆発的に RT されている訳ではなく、じわじわと RT が伸びる傾向があった。また、一旦 RT 数が落ち着いてもフォロワー数の多いユーザーが RT したことで RT が一時的に増える傾向がみられた。RT の増加が一過性のものであるとその時間帯に活動しているユーザーのタイムラインにしかツイートが表示されないため、この RT の波が利用できればより多くのユーザーにツイートを拡散できるのではないだろうか。

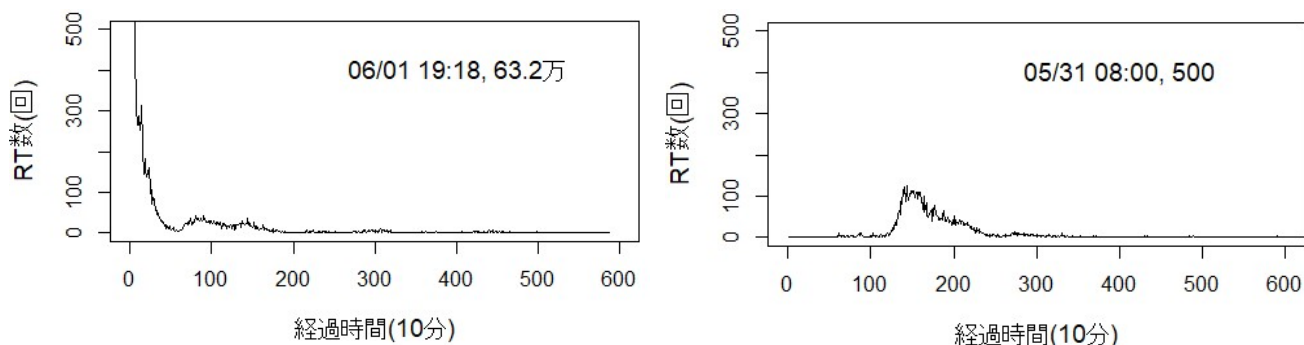


図1 10分毎のRT回数のグラフ
(枠内右上は元ツイートの投稿日時，投稿者のフォロワー数)

2.4. 課題

今回はランキングからツイートを選択したが，この選択方法では無作為とは言えず，結果に偏りが生じてしまった可能性があった。また，フォロワー数の違いを公式アカウント・非公式アカウントで分けるのは適切ではなかったため，ツイートの選択方法を改める必要がある。解析方法についても，複数の変数に対する解析や時系列データを扱うには不十分な解析方法であったので時系列解析の手法を使わなければならない。

3. 分析2

3.1. 目的

前回の分析で生じた課題点の改善と，RTしたユーザーのフォロワー数がRT数にどう影響するかを分析する。

3.2. 方法

3.2.1 ツイートの収集方法

分析対象の選択には，「GET statuses/sample」というTwitterの全タイムラインから無作為に選ばれた1%を取得することができるTwitter Streaming APIを用いた。得られたツイートの中から日本語かつ500RT以上のツイートを取得できるプログラムを作成し，RT収集と分析の際に使用するツイートID，フォロワー数，RT数，ツイート日時を取得した。APIの利用には開発アカウントの申請と認証キーの発行を事前に行った。

取得したツイートを用いてキーワード検索を行い，RTされたツイートを集めた。ツイートの収集にもAPIを利用し，ツイート収集用のプログラムを作成して，RTされたツイートからID，フォロワー数，ユーザー名とRT日時の情報を取得した。RT収集の際に全く別のツイートが混じってしまうなど，うまく収集できなかったものは除き，最終的に13個のツイートとそのRTを10分毎にまとめて分析対象とした。

3.2.2 解析方法

解析するデータはフォロワー数と RT 数の 2 変量を扱う時系列データなので、ベクトル自己回帰モデル(VAR モデル)を用いた。自己回帰モデルは過去のデータから現在の値を回帰するモデルのことで、自己回帰モデルを多変量に拡張したものが VAR モデルである。1 次 2 変数の VAR モデルは以下のように表される。

$$\begin{cases} y_{1t} = c_1 + \phi_{11}y_{1,t-1} + \phi_{12}y_{2,t-1} + \epsilon_{1t} \\ y_{2t} = c_2 + \phi_{21}y_{1,t-1} + \phi_{22}y_{2,t-1} + \epsilon_{2t} \end{cases}$$

時系列データ同士の影響を推定する手法としてはグレンジャー因果性検定とインパルス応答関数を用いた。グレンジャー因果性検定は 2 変数の時系列データに因果関係があるかどうかを調べる手法で、インパルス応答関数は 1 変数の変動が他の変数にどう影響を与えるかを調べるものである。インパルス応答関数を求めることでフォロワー数が RT 数にどう影響を与えているかを調べた。

3.3. 結果と考察

10 分毎の RT 回数に加え、10 分間に RT したユーザーのフォロワー数の合計と、その内フォロワーが最も多いユーザーのフォロワー数を可視化した(図 2)。

このツイートは 440 分後あたりから RT 速度が上昇しているが、フォロワー数の多いインフルエンサーがきっかけになって

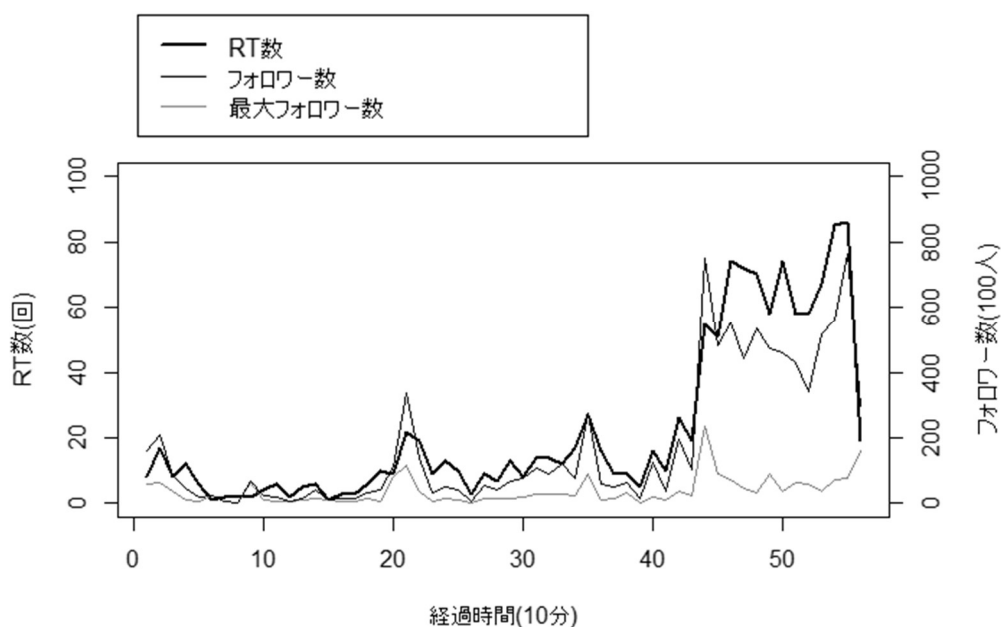


図 2 10 分毎の RT 回数とフォロワー数のグラフ

いるように見える。しかし、このツイートでグレンジャー因果性検定を行ったところ、RT 数と合計フォロワー数、最大フォロワー数ともに因果関係は見られなかった。他のツイートも因果関係がみられず、原因としては、RT 数の範囲が小さすぎてうまく分析できなかった、フォロワー数は RT 数に有意差が出るほど大きな影響を与えていない、もしくは VAR モデルの推定方法がそもそも適切ではない、などが考えられる。