

SNS ユーザの情報拡散行動パターンの可視化

理学専攻 情報科学コース 1740648 岩科 智彩 (指導教員: 伊藤 貴之)

1. はじめに

近年多くの SNS (Social Networking Service) が普及し、人々にとって日常生活で欠かせないツールとなっている。特に Twitter は、リツイートという機能によりワンクリックで他者の発言を再配信できるため、ユーザが即時性の高い情報を気軽に発信・取得・拡散することができる手段として、世界中で多くのユーザが利用している。これらの特徴を活かし、情報拡散を目的として企業などの団体や個人が利用する例も多い。

本論文では Twitter の「拡散」という役割に着目し、拡散されるツイートと拡散するユーザの行動パターンを可視化する。ここで、Twitter 上で影響力のある著名人や広告アカウントを「キーパーソン」と呼び、キーパーソンのツイートをリツイート機能によって拡散するユーザ群を「リツイーター」と呼ぶ。本論文ではキーパーソンとリツイーターの行動パターンとの関連性を可視化した結果を示し、キーパーソンごとの特徴を議論する。

2. 関連研究

2.1 SNS データの可視化

SNS 上でのコミュニケーション行動の可視化に関する研究は数多く報告されている。Chae ら [1] のように、ハリケーンや竜巻などの自然災害の前後に SNS から抽出されたデータの反応を分析することで、自然災害時の人の動きを可視化することが可能である。

Twitter と同等の機能を有し、主に中国で利用されている Weibo のデータを使用した研究も多く [2]、Weibo 上の重要なユーザや情報の拡散経路、コミュニティ間の相互作用を特定し、それらの可視化画面を一括表示させている。

2.2 SNS の拡散

Kwak ら [3] はリツイートの経過時間を分析し、1 時間以内にリツイートの半分、75% がその日に発生、リツイートの約 10% が 1 ヶ月後に発生することを示した。

3. 提案手法

我々は多様な業界からキーパーソンを選出し、各キーパーソンに対するリツイーターの拡散行動を収集し、その傾向の可視化を試みた。さらに、キーパーソンによってリツイーターの特性がどのように異なるか考察した。

3.1 キーパーソンの選定

Twitter 上での影響力を定量的にはかることはある程度可能であるが、本論文ではキーパーソンを我々の主観により選定している。Twitter API 呼び出しによるデータ収集上の制約を考慮して、ある程度の時間的間隔においてツイートを発信しているユーザを対象とした。

3.2 ツイートの選定

キーパーソンによる過去のツイートの中で、ある一定期間の中で発信されたツイートを、複数の条件をもとに選定する。

3.3 リツイートデータの収集

Twitter API を利用して選定したツイートのリツイートデータを収集する。リツイートされたツイートデータは本文の前に「RT」の記号が付与される。本論文ではこの特性を活かし、収集したい元ツイートの本文に「RT」を付けた上で検索する。収集データ項目は、Tweet id・Text・Created at (リツイート実行時間)・User id などである。

3.4 データ抽出と整理

収集したリツイートデータに記録されているリツイート履歴から、User id (リツイーターの ID) と Created at (リツイート実行時刻) を抽出する。そして、元ツイートの発信時刻とリツイート実行時刻との差を求め、算出された時間差を各リツイートに付与する。これらの作業をキーパーソンの各ツイートに対して実施する。

続いて、1 人のキーパーソンについてあらかじめ選出された各ツイートに対するリツイーター群を統合し、ツイートとリツイーターを行および列とする行列 (以下リツイート行列と称する) を形成する。そして、 i 番目のツイートに対して j 番目のリツイーターによるリツイート実行の有無および時間差 d_{ij} を検索し、リツイート行列中の j 行 i 列の値 v_{ji} に以下の値を代入する。

$$v_{ji} = \begin{cases} d_{max} - d_{ij} & (A \text{ retweet execution}) \\ -D & (Not \text{ a retweet execution}) \end{cases}$$

なお $-D$ は負の定数 (現時点での実装では -1) であり、 d_{max} は全リツイーターの時間差の最大値である。

以上の統合処理により非常に多くのリツイーターが抽出されるため、最後にリツイーターをサンプリングする。

3.5 可視化 (リツイーター間の差分の選出)

本論文ではリツイーターの行動を分類するため、およびツイートの種類に応じてどのように反応が変化するかを観察するために、上述のリツイート行列にデンドログラムを適用した。Python のライブラリである seaborn を使用して、ヒートマップとデンドログラムを一度に可視化した。

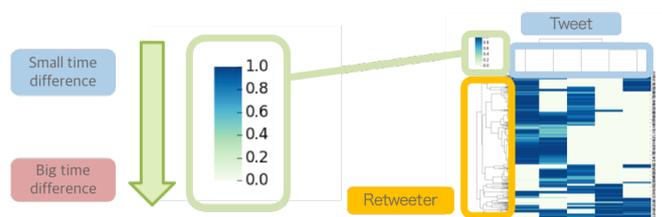


図 1: 可視化結果の読み方

4. 分析結果

本章では、実際の Twitter データセットを使った実験について紹介する。本論文では 5 つのジャンル (俳優, タレント, プロモーションアカウント, スポーツ選手, 政治家)

からそれぞれ1人のキーパーソンを選出した。本章では、タレントのジャンルに属するキーパーソンのツイートデータセットの可視化結果を紹介し、議論する。

4.1 小規模データによる可視化

リツイーター数、引用ツイート数、画像や動画の有無に基づいて、キーパーソンごとに5つのツイートを手動で選択した。ただし Twitter API の制限により、データ収集対象期間を1か月間としている。

各ツイートに関する集計結果を表1に示す。また、可視化結果を図2に示す。

no.	特徴	リツイート数 (約)	ツイート時間
t1	動画	160	17:50
t2	引用ツイート	110	21:57
t3	特になし	130	16:20
t4	画像, 他の芸能人が複数	100	19:40
t5	画像, 共演者	530	22:04

表 1: タレントのツイートの特徴

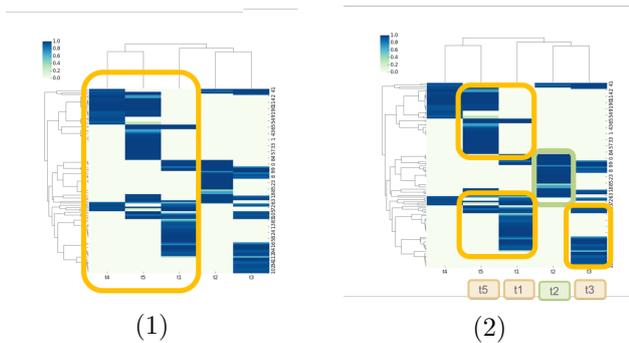


図 2: タレントのリツイーター可視化の詳細

複数のツイートに反応しているユーザは一定数存在している。リツイーターの多くは時間差が全体的に小さい印象がある。また、動画や画像が付いているツイートの方がツイート時刻に関わらずリツイート数が多い傾向にある。収集時期に放送されていたキーパーソンが主演のドラマに関するツイートは反応人数は多く、時間差が小さかった。このドラマに関するツイートは t1・t3・t5 であり、これらの3つのツイートと t2 の両方をリツイートしているユーザが少ない。これはドラマファンのリツイーターとそれ以外の目的でのリツイーターとで注目しているツイート内容が異なっているのではないかと考えられる。

4.2 大規模データによる可視化

本章では、各キーパーソンの1年間のツイートを全て収集して可視化した。収集条件を以下の2点とした。

- retweeted_status->user->id が各キーパーソン ID
- created_at が 2017 年のもの (JST 基準)

リツイートについては時間差が1週間以内のものを抽出した。その後のリツイート行列の生成手順は前章と同様である。

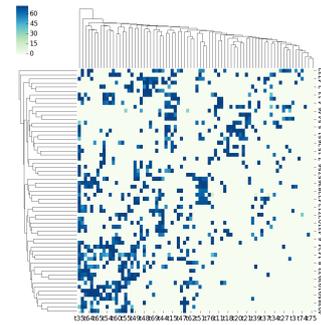


図 3: タレントの1年間のツイートにおけるリツイーターの行動パターン

可視化結果を図3に示す。個々のツイートに対するリツイート件数はさほど大きくはないが、その中で共通したリツイート行動を伴うリツイーター群を観察することができた。個々のリツイーターの拡散行動パターンを観察してみると、複数のリツイートに対して同様な行動を起こしていることが分かった。具体的に、あるツイートにすばやく反応しているリツイーターは、他のツイートに対しても同じ行動をとる傾向が見られた。また、リツイーターの人数が多いツイートには、文字だけでなく画像や動画を含んだツイートが多いことが分かった。

5. まとめと今後の課題

本論文では Twitter 上で影響力のあるキーパーソンとそのリツイーターに着目し、キーパーソンによるツイート群とそのリツイーター群を可視化した。本論文の可視化により、キーパーソンを取り巻くリツイーターの全体像を観察することができたため、このような考察が可能となった。

今後は、リツイーターの興味分野なども考慮しながら、さらにキーパーソンとリツイーターとの関連性について考察を深めたい。

参考文献

- [1] J. Chae, D. Thom, Y. Jang, S. Kim, T. Ertl, and D. S. Ebert, "Public behavior response analysis in disaster events utilizing visual analytics of microblog data," *Computers & Graphics*, vol. 38, pp. 51-60, 2014.
- [2] S. Chen, S. Chen, L. Lin, X. Yuan, J. Liang and X. Zhang, "E-Map: A Visual Analytics Approach for Exploring Significant Event Evolutions in Social Media," in *IEEE Conference on Visual Analytics Science and Technology*, 2017.
- [3] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a Social Network or a News Media?," in *Proceedings of the 19th International Conference on World Wide Web*, pp. 591-600, 2010.