

SNSにおける拡散行動パターンの可視化

岩科 智彩[†] 吉田 光男^{††} 伊藤 貴之[†]

[†] お茶の水女子大学大学院 〒 112-8610 東京都文京区大塚 2-1-1

^{††} 豊橋技術科学大学 〒 441-8580 愛知県豊橋市天伯町雲雀ヶ丘 1-1

E-mail: [†]chisae@itolab.is.ocha.ac.jp, ^{††}yoshida@cs.tut.ac.jp, [†]itot@is.ocha.ac.jp

あらまし 情報発信ツールとして利用されている SNS を、団体や個人による情報拡散を目的として利用している例が増加している。SNS の中でも特に Twitter は気軽に拡散（リツイート）できることで知られ、その利用は現在も活発である。芸能人やファッションブランド、企業などがこの機能を利用して、情報告知などを行なっている。ゆえに、Twitter 上では多様なジャンルにおいて影響力のあるユーザが存在する。これらのユーザを本報告では「キーパーソン」と定義する。キーパーソンが存在する大きな要因の一つは情報を拡散（リツイート）するユーザが多数存在しているからである。これらのユーザを本報告では「リツイーター」と定義する。本研究では、キーパーソンのリツイーターに着目し、キーパーソンが発信するツイート群とリツイーター群との関連性をデンドログラムとヒートマップを用いた可視化手法により表現した。情報拡散の性質についてツイート内容とリツイーターの両側面から考察することで、ツイート内容やツイート時刻と関連して反応するリツイーター層やその人数が変化することがわかった。

キーワード SNS, 拡散, 可視化

1. 概要

Twitter や Facebook, Instagram や LINE など多くの SNS が近年普及し、日常生活で欠かせないツールとなっている。SNS は気軽に情報収集手段として、コミュニケーション手段として、また国内・国外への情報発信手段として発達してきた。これらのツールの中でも特に Twitter は、簡素なユーザインタフェースにより短文で即時に情報を発信するのに向いており、かつ一方向的に他者をフォローできるのでフォロワーを増やしやすい。また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。このことから Twitter は、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増えてきた。例えば、災害が起きた時現地の状況を Twitter で発信することで、迅速に物資調達ができる。また、著名人が自分のお気に入りのブランドやお店を紹介することで、知名度が上がることもある。このように Twitter が流行の発信元になることも少なくない。本研究では Twitter の「拡散」という役割に着目して、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。ここで本報告では、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。本報告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。

本論文の構成は以下の通りである。2. 章で関連研究について述べ、3. 章で提案手法について述べる。4. 章で分析結果の一部を、5. 章でまとめと今後の課題について述べる。

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存在する。SNS の中でも Twitter は API が公開されているため、ユーザの属性や友人関係、発言内容や発言時の時刻や位置情報など、多様なデータを収集しやすい。可視化の題材の例として、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。Chae ら [1] の手法では、ツイートの空間的および時間的特性を同時に分析するための単一のビュー表示を提供し、より視認性の高い可視化画面を実現している。

また中国版 Twitter である Weibo のデータを使用した研究も多い [2] [3]。これらの研究では、Weibo 上の重要なユーザや情報の拡散経路、また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

本研究ではリツイート機能に着目しているが、Twitter の機能の一つであるメンション機能に着目し、Tweet の拡散相手を推薦する研究がある。メンション機能とは「@～」の記号で見かけるもので、特定の人に情報発信するときに用いる機能である。Wang ら [4] は、推薦ユーザの Tweet 拡散能力を独自のモデルによって特徴量とし、Tweet の拡散率や拡散する範囲を細かく定義することで、「Tweet の拡散相手」として最適であるユーザを推薦する手法を提案している。

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。また、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。本章では以下の5つの工程にかけて可視化の手順を説明する。

3.1 キーパーソンの選定

本研究では、キーパーソンとはTwitter上で「影響力のある」ユーザを意味する。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、現時点での本研究ではキーパーソンを我々の主観により選定している。本研究では日本または日本人の公式アカウント（公式マークの有無は問わない）を対象として、各ジャンルから1人ずつキーパーソンを選定している。Twitter API [5] 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送信しているユーザを対象とした。

3.2 ツイートの選定

続いて、キーパーソンによる過去のツイートの中で、特に重要と思われる一定数のツイートを選定する。これについてもツイートの重要度を定量化することはある程度可能であるが、現時点での本研究ではツイートを我々の主観により選定している。

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、API内部において、リツイート元ツイート本文の冒頭に「RT」の記号が付与されたものとして処理される。本研究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」を付けた上で検索にかける。また、APIには1回の問い合わせにつき100件のリツイートデータしか収集できないため、日時を細かく設定し、時系列順に全てのリツイートデータが収集できるようにする。収集データ項目は、図1に示すようにTweet id・Text・Created at（リツイート実行時間）・User idなどである。

```
tweet_id:881863500963782656
text:フレンドバーク見てくれてありがとうございます！
我々カナナさん！チーム。私たちに頑張りました！是非ともドラマも見てください！
2枚目は中岡さんと合間時間があったので顔入れ替えてみた。 https://t.co/1XHjWZ4JF
created_at:Mon Jul 03 13:13:37 +0000 2017
user_id:113657279
user_desc:29歳女の子。158センチ。42キロ。PUNYUSというS~6Lサイズまでのお洋服
ブランドをプロデュースしております。J-WAVE 81.3 FM RADIO AVALON 毎週月~木
22:00~生放送 私は木曜日だお是非聞いてね
screen_name:watanabe_naomi
user_name:渡辺直美
```

図1 収集データ形式例

なお、図2のようにTwitter APIで収集できるリツイートデータからは、データ拡散の経路を復元することは不可能である。例えばユーザAが元ツイート発信者であるとする、そのツイートを直接閲覧してリツイートを実行するユーザBと、ユーザBのリツイートを閲覧してリツイートを実行するユーザC—Dの存在が考えられる。しかしTwitter APIで収集できるデータ上では、どちらのリツイートも同様に、「元ツイート発信者であるユーザAの発言をユーザB—C—Dがリツイートした」という事実しか記述されない。そのため、本研究ではリツイーターは元ツイートをリツイートしたユーザとし、誰のリツイートからのリツイートであるかという点は考慮しない。

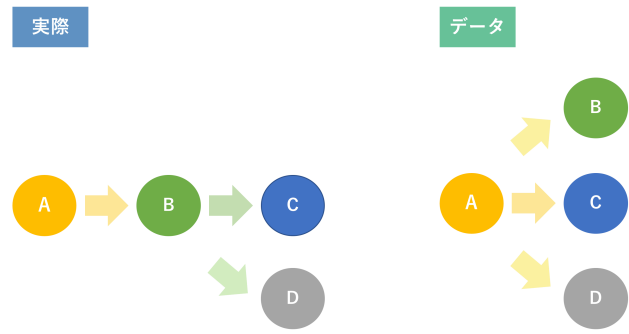


図2 リツイートデータイメージ図

3.4 データ抽出・リツイーターのサンプリング

前工程で収集したデータに記録されているリツイート履歴から、User idとCreated at（リツイーターのIDと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1キーパーソンの各ツイートに対して実施する。同一リツイーターが同一ツイートを複数回リツイートしている場合には、これらのリツイートの中から最も時刻差が小さいものだけを残した。このような現象が生じる理由として、リツイートはワンクリックで解除が可能であることから、リツイートボタンを押すたびにリツイート履歴が記録されたものと考えられる。

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。

$$v_{ji} = \begin{cases} d_{max} - d_{ij} & (\text{リツイート実行があるとき}) \\ -D & (\text{リツイート実行がないとき}) \end{cases} \quad (1)$$

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ので、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）

本研究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 [6]を使用して、図3のようにヒートマップとデンドログラムを一度に可視化した。

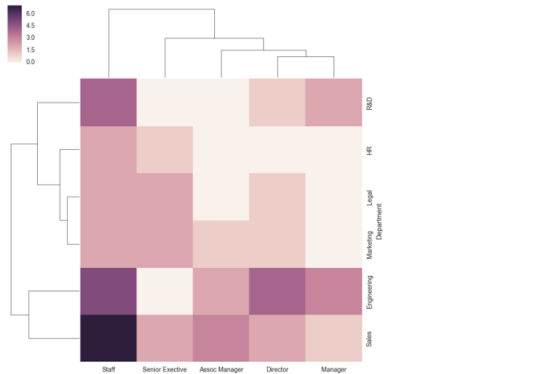


図3 可視化イメージ図

4. 分析例

今回の分析では、以下の5つのジャンルからそれぞれ1人ずつキーパーソンを選定した。

- 俳優
- タレント
- 情報配信
- スポーツ選手
- 政治家

ツイートの選定に関しては、リツイート数・引用リツイート数・画像や動画の有無などを考慮の上、手作業で1キーパーソンにつき5個のツイートを選定した。ただし選定するツイートはデータ収集時点から1ヶ月前までを対象とした。これはTwitter APIの制約によるものである。

リツイーターのサンプリングに関しては、5個のツイートのうち2,3,4個のツイートに反応したリツイーターを抽出することにした。この結果として抽出されたリツイーターが500人を超える場合、さらにランダムで500人を抽出して代表リツイーターとし、これを可視化の対象とした。

本報告では俳優・タレント・情報配信アカウントの3つのジャンルに関する可視化結果を紹介し議論する。なお、可視化の前処理としてリツイート行列に標準化（スケールを0.0-1.0の間に調整）を実施している。この値が1.0に近いほどリツイートの時間差が小さく、可視化画面には濃い色で表示される。

可視化結果であるヒートマップには、縦軸にリツイーターが、横軸にツイートが並んでいる。どちらの並び順もデンドログラムを構築した結果による。図3を例に見てみる。リツイーターに着目してみると、可視化結果の下部の方に配置されたリツイーターは、全体的に濃い色で表示されており、どのツイートに対しても反応時間が比較的速く常にこのキーパーソンに注目しているユーザー群であることが分かる。また、左端の列が全体的に濃い色で表示されている。これにより左端に配置されたツイートは、多くのリツイーターにおいて反応時間が速いと言える。

4.1 俳優

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4に示す。

表1 俳優Aのツイートの特徴

no.	特徴	リツイート数(約)	ツイート時間
t1	画像, 他の芸能人	5,000	3:36
t2	画像, 他の芸能人が複数	13,000	21:19
t3	引用リツイート	2,000	22:41
t4	引用リツイート	3,000	19:17
t5	画像, 他の芸能人が複数	12,000	23:32

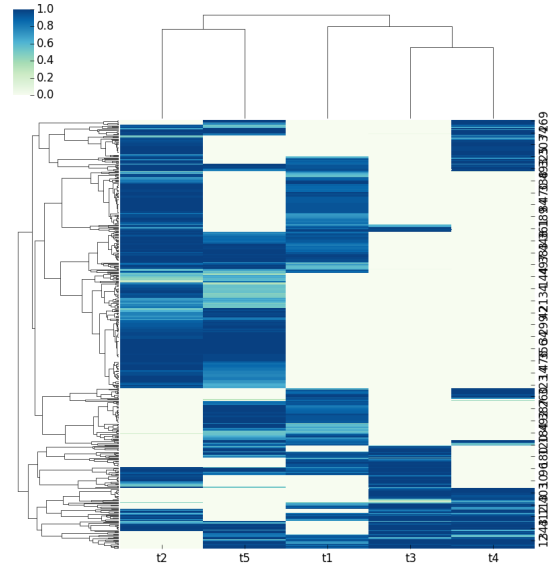


図4 俳優Aのリツイーター

全体的にリツイートのリピーターが多く、5つ全てのツイートに反応しているユーザーも多く見られた。また、この可視化画像は2から4つのツイート反応ユーザーに絞り、さらに500ユーザーをランダムに抽出しているが、あるツイートに対して反応時間が速いユーザーは他のツイートに対しても速いことがわかった。また、横軸に沿って可視化結果を眺めると、青い帯が横に続いていることがわかる。縦軸に沿って眺めると、t2とt5が同じクラスターで、可視化結果からも同等な時間差を有するリツイーター群が存在する。これら2つのツイートの特徴を見てみると、画像付きであり、キーパーソン以外にも芸能人が写っているものだった。キーパーソンを取り巻くリツイーターだけでなく、画像に写っている他の芸能人を取り巻くリツイーターも反応したため、時間差が類似していてリツイート数が多かったのだと考えられる。逆にt3とt4はどちらも引用リツイートだったため、全体として反応するユーザーが少なかったと考えられる。全体を通して時間差が小さく反応回数が多い傾向にあるが、その理由としてこのキーパーソンのリツイーターには熱心なファンが多く、キーパーソン目的でツイートを追っていることが想像される。

4.2 タレント

各ツイートに関する集計結果を表2に示す。また、可視化結果を図5に示す。

表 2 タレント B のツイートの特徴

no.	特徴	リツイート数 (約)	ツイート時間
t1	動画	160	17:50
t2	引用リツイート	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

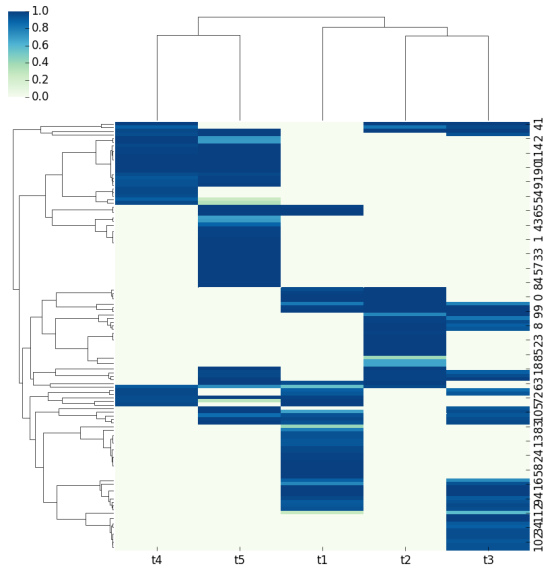


図 5 タレント B のリツイーター

俳優 A のリツイーターほどではないが、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。

4.3 情報配信

各ツイートに関する集計結果を表 3 に示す。また、可視化結果を図 6 に示す。

表 3 情報配信 C のツイートの特徴

no.	特徴	リツイート数 (約)	ツイート時間
t1	画像, グルメ	1,300	0:55
t2	画像, グルメ	600	22:45
t3	画像, 美容	450	20:30
t4	画像, ファッション	100	7:45
t5	画像, ゲーム	280	8:15

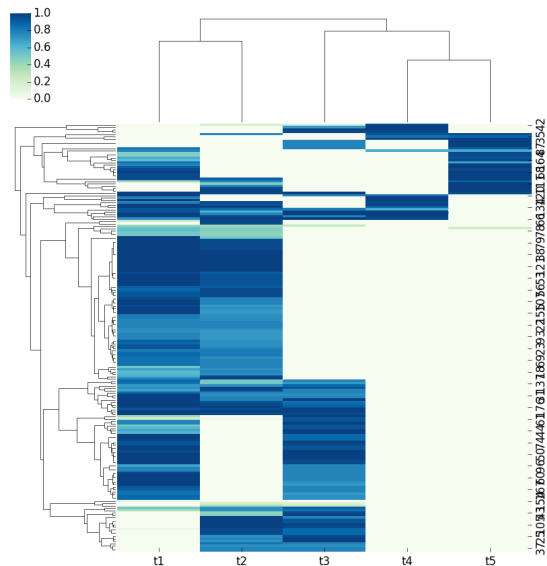


図 6 情報配信 C のリツイーター

横軸に沿って可視化結果を眺めると、各ツイートともに類似した色が見られる。このことから総じて、リツイーター個人の行動パターンはどのツイートに対しても同じ傾向にあると考えられる。t1・t2 はリツイーターの多くが共通している。これはツイートの内容が 2 つともグルメであることから、同じテーマに興味を持っているリツイーター群であることが可視化画像から推察される。また、t1・t2 にはツイート時刻が遅いが、薄い色（時間差が大きいリツイート）が多く、遅い時刻のツイートに対して遅く反応するリツイーターの存在が伺える。一方で、グルメに関する t1・t2 と美容に関する t3 ではリツイーターの層が大きく異なり、グルメと美容の両方に興味のあるリツイーターが少ないと考えられる。t4・t5 に関してはリツイーターの数が極端に少ない。これら 2 つのツイートの共通点としてツイート時刻が早朝であったことから、リツイーターの数が少なかった一番の要因はツイート時刻ではないかと考えられる。

以上の可視化結果から推察される点として、情報配信アカウントの多くは特定のユーザを対象にすることよりも幅広いユーザに情報を届けることを目的としているため、そのぶん内容も多岐にわたり、近い内容の話題でない限りは多くのツイートに反応するユーザは少ない傾向にあると考えられる。

5. まとめと今後の課題

本報告では 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。これらの可視化結果から、ツイート内容やツイート時刻と関連して反応するリツイーター層やその人数が変化することがわかった。デンドログラムやヒートマップを用いた可視化により、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。

今後はデンドログラム構築結果を詳しく分析し、リツイーターの興味分野なども考慮しながら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。また、キーパーソンとリツイーターの関係をさらに効果的に表現するために、必要に応じて新たな可視化手法を開発したい。

謝 辞

本研究の一部は、日本学術振興会科学研究費補助金の助成に関するものです。

文 献

- [1] J. Chae, D. Thom, Y. Jang, S. Kim, T. Ertl, D. S. Ebert, "Public behavior response analysis in disaster events utilizing visual analytics of microblog data", *Computers & Graphics Volume 38*, pp. 51-60, 2014.
- [2] S. Chen, S. Chen, Z. Wang, J. Liang, X. Yuan, N. Cao, Y. Wu, "D-Map: Visual Analysis of Ego-centric Information Diffusion Patterns in Social Media", *IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2016.
- [3] S. Chen, S. Chen, L. Lin, X. Yuan, J. Liang, X. Zhang, "E-Map: A Visual Analytics Approach for Exploring Significant Event Evolutions in Social Media", *IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2017.
- [4] B. Wang, C. Wang, J. Bu, C. Chen, W. V. Zhang, D. Cai, X. He, "Whom to Mention: Expand the Diffusion of Tweets by @ Recommendation on Micro-blogging Systems", *the International World Wide Web Conference Committee (IW3C2)*, 2013.
- [5] "Twitter API," <https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets.html>
- [6] "seaborn," <http://seaborn.pydata.org>