

ソーシャルネットワークサービスにおける ニュース記事の情報拡散現象に関する統計的分析

藤岡 能史[†] 羽山 徹彩[†]

長岡技術科学大学[†]

1. はじめに

口コミによる情報拡散は、政策やブランドイメージなど社会的に大きな影響を与えてきたが、近年では SNS がその情報拡散の媒体として台頭している。SNS の情報拡散現象については、媒介するユーザ属性(信頼性や専門性など)、ネットワーク構造などが影響することが指摘され、代表的な少数のユーザを対象に検証されてきた[1]。そのため、情報拡散現象の偏向のないデータでの実証研究が望まれているものの、その現象の観測が不可能である点[2]、や拡散されたイベントにだけ注目がいきやすい偏向性[3]が問題視されてきた。それに対し、Bakshy ら[4]は Twitter を対象に、URL を含んだツイートの情報拡散イベントに関する膨大なデータを収集し統計的分析を行った。その知見として、過去に情報拡散に寄与し、一定数以上のフォロワーを持ったユーザが情報拡散に貢献し易いことを明らかにした。これは実証的成果として高く評価されているものの、新たに情報拡散の要因を見出したとはいえない。その原因のひとつとして、無作為に収集したデータを一緒に扱ったため、ユーザの嗜好性による影響を精緻に分析できなかったことが考えられる。

そこで、本研究では Twitter のニュース記事ツイートを対象とし、ニュース記事ジャンルごとの情報拡散現象を実証的に解明することを目的とする。膨大なフォロワー数を持った、特定のニュース記事アカウントに着目することで、ユーザ属性の偏りが少なく、かつ情報拡散の規模に関係ないデータを対象に分析することができる。さらにニュース記事ジャンルごとに分析することで、拡散情報の内容を考慮した拡散現象の特徴について明らかにすることができる。

2. アプローチ

ニュース記事の情報拡散現象を分析するために、一定期間の特定ニュース記事アカウントのツイートとそのリツイートしたユーザのデータを収

集した後、1)記事ツイートごとの情報拡散木の作成(2.1 節参照)、および 2)記事ジャンルごとの情報拡散の統計的な特徴分析(2.2 節参照)、を行う。

2.1 記事ツイートごとの情報拡散木の作成

Twitter のニュースアカウントから記事がツイートされた後に、そのフォロワーが記事ツイートをリツイートすることで伝播していくユーザ関係グラフを情報拡散木と呼び、作成する。しかしながら、ニュースアカウントをリツイートしたユーザとユーザのフォロワー関係は収集データから確認できるものの、ユーザがどのユーザの記事リツイートを参照し、リツイートしたのかは特定できない。

そのため、本研究ではあるユーザにフォローしている複数ユーザが同じ記事をリツイートしている場合に、リツイート時間によってユーザが参照したリツイートのユーザを一意的に特定することで、情報拡散木を作成する。フォローしているユーザのツイートはユーザの Twitter タイムラインに表示されるが、閲覧環境によってユーザが先に参照するタイムライン上の表示位置が異なるため、リツイートのために参照した記事リツイートを発信したユーザも異なってくる。そこで、本研究では図 1 のように、記事をリツイートした、フォローしているユーザのなかで、「最初にリツイートしたユーザを参照」と「リツイートする直前にリツイートしたユーザを参照」の 2 種類を扱う。

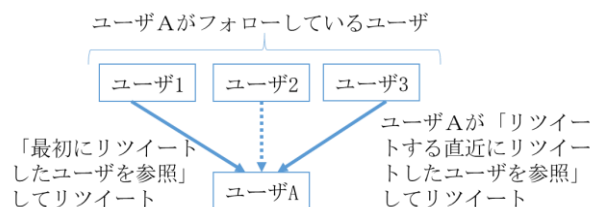


図 1: 2 種類の拡散木の作成方法

2.2 記事ジャンルごとの情報拡散の特徴分析

ニュース記事ごとに作成した情報拡散木に対し、記事ジャンルごとに統計的な特徴分析を行う。

Statistical analysis of information diffusion phenomenon of news articles in social network services

[†]Yoshifumi Fujioka, Tessai Hayama

[†]Nagaoka University of Technology

統計的な特徴分析には回帰木分析によって、情報拡散木の深さと広さに対し分析する。つまり、情報拡散木の深さおよび広さを目的変数とし、情報拡散木に含まれるユーザの属性と記事タイトルをそれぞれの説明変数とする。具体的な説明変数には文献[4]を参考に、拡散木のユーザのフォロワー数の平均と中央値、拡散木のユーザのフォロワー数の平均と中央値、拡散木のユーザの Twitter 登録日数の平均と中央値、および記事タイトルの長さの7項目を用いる。

3. 情報拡散現象の分析

3.1 データセット

Twitter API を使用しニュースアカウント (@YahooNewsTopics) の記事ツイートを 2 週間 (2019/11/13-27) 収集し、5 種類のニュースジャンル (スポーツ, 経済, 政治, 国内外, エンターテイメント) に分けて分析した。分析データの内訳を表 1 に示す。

3.2 結果と考察

ニュースジャンルごとに記事ツイートの拡散木を回帰木分析した。紙面の都合上、「経済」に関する記事ツイートの拡散木の深さと広さの回帰木分析結果だけを、図 2 に示す。なお、これは「最初にリツイートしたユーザを選択した拡散木作成」を適用した。

図 2 から、経済に関するニュース記事の情報拡散の深さではフォローしているユーザ数の平均が 2000 以下、フォロワー数の中央値が 600 以下の場合に深くなる傾向がある。その一方で、経

済に関するニュース記事の情報拡散の広さではフォロワー数の平均が 1600 以下の場合に広く伝わる傾向がある。

それ以外のニュースジャンルの情報拡散の特徴として、スポーツではフォロワー数の平均が 1200 以上かつ Twitter 登録日数が 4 年以上前の場合に拡散木が深くなる傾向がみられた。政治ではフォロワー数とフォロワー数の中央値が 500 以上の場合に深くかつ広くなる情報拡散の傾向がみられた。国内外に関する記事では、Twitter 登録年数が 6 年で、平均のフォロワー数が 1300 以下の場合に深くかつ広くなる情報拡散の傾向がみられた。エンターテイメントに関する記事では、フォロワー数の中央値が 600 以上で、平均が 1900 以下の場合に深くなる傾向がみられ、Twitter 登録年数が 6 年以上の場合に拡がりやすい傾向がみられた。また異なる情報拡散木作成法で行った場合にも、ほぼ同様の傾向がみられた。

以上から、ニュースジャンルごとの情報拡散の分析を行った場合に、フォロワー数以外の異なるユーザ属性も影響していることがわかった。

参考文献

- [1] M. Gladwell: "The Tipping Point: How Little Things Can Make a Big Difference", Little Brown (2000).
- [2] Sun et al.: "Gesundheit! Modeling contagion through facebook news feed", In International Conference on Weblogs and Social Media, AAAI (2009).
- [3] Bakshy et al.: "Social influence and the diffusion of user-created content", In 10th ACM Conference on Electronic Commerce, (2009).
- [4] Bakshy et al.: "Everyone's an Influencer: Quantifying Influence on Twitter", WSDM (2011).

表 1: 分析データの内訳

記事ツイートのジャンル	記事数	総リツイート数	拡散木の深さ		拡散木の広さ	
			平均	標準偏差	平均	標準偏差
スポーツ	107	12,940	3.03	1.49	121	141
経済	47	16,303	3.39	2.09	347	1,422
政治	144	40,040	3.55	1.65	278	577
国内外	194	69,235	3.75	1.75	357	1,106
エンターテイメント	57	57,298	5.51	2.78	1,005	2,483

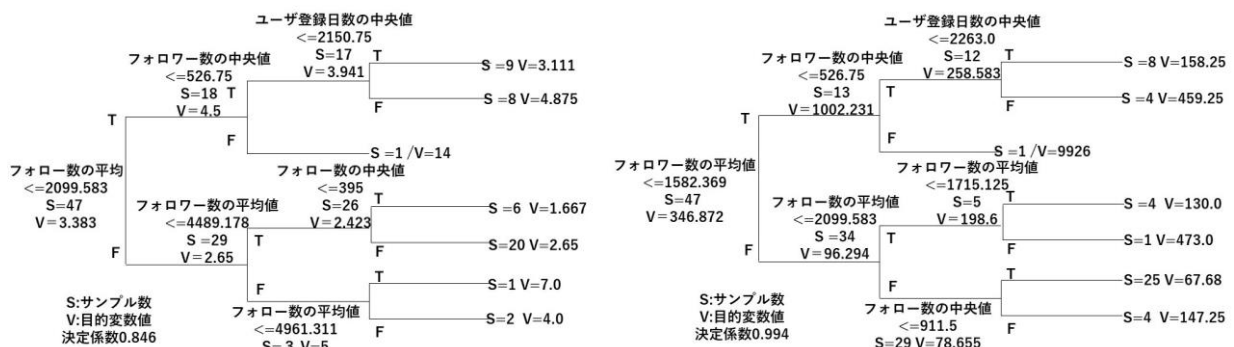


図 2: 経済に関する記事ツイートの拡散木の深さ(左)および広さ(右)の回帰木分析結果