

フォロー順序に着目した Twitter ユーザの役割推定手法の提案

武田 悠佑^{1,a)} 佐藤 哲司¹

概要: Twitter は自ら情報を発信するだけでなく、フォローによって他ユーザを情報源とすることができる。一般に、あるユーザをフォローしているユーザはフォロワーと呼ばれ、あるユーザのフォローリストの中でフォローした順序が保存されている。本研究では、複数のユーザが保持するフォローリストを統合することで、フォロワーの大域的な順序関係を生成し、その位置によってフォロワーの役割を推定する手法を提案する。提案法では、試合毎の勝敗から強さの順序を推定する際に用いられる Bradley-Terry モデルを適用してフォローリストを統合している。活発に利用されている実際の Twitter ユーザに提案法を適用した結果、ランキングの順位がフォロワー数や登録されているリスト数と相関があることが確認された。

Estimating Twitter User's Role Based on Sequential Order of Followings

YUSUKE TAKEDA^{1,a)} TETSUJI SATOH¹

1. はじめに

近年、Twitter をはじめとする SNS (Social Networking Service) の利用者が増加している。SNS とは、登録されたユーザがテキストや画像、動画などの投稿を通じてコミュニケーションを行う会員制サービスである。中でも 2006 年にサービスを開始した Twitter は、社会的にも活用が進んでおり、コミュニケーションの場としてのみならず、メディアやマーケティングの場としても注目を集めている。

ユーザが Twitter を利用する目的には、ユーザ間のコミュニケーションや自らの情報発信のほかに情報の獲得がある。この目的を達成するためにユーザは膨大な投稿の中から自分にとって有用な情報を取捨選択する必要がある。Twitter には、ユーザが効率的に閲覧する投稿を選択するための手段として、ユーザ単位で投稿を閲覧するかを選択することができるフォロー、アンフォローと呼ばれる機能が用意されている。フォローとは、継続的にその投稿を閲

覧したいユーザをフォロワーとして登録する機能であり、フォロワーとして登録されたユーザの投稿は、タイムラインと呼ばれる画面上に、時系列に沿って並べて表示されるようになる。アンフォローとはフォロワーとして登録したユーザのフォローを取りやめる機能であり、アンフォローしたユーザの投稿はタイムライン上には表示されなくなる。ユーザはこれらの機能を用いて、自分の興味、関心、属性に応じた投稿がタイムライン上に表示されるよう、フォローリスト (フォロワーの集合) を作成、更新している。フォロワーはユーザにとっての情報源であり、ユーザは自身にとって有用な情報源を求めるため、ユーザの特徴を分析し、フォロワーとして有用なユーザを推薦する研究は数多く行われている。

ここで、本研究ではフォロワーの特徴を推定するにあたり、ユーザがフォロワーを登録する順序に着目する。ユーザがある話題について情報の獲得を目的に複数のフォロワーをフォローする場面を考えると、2 人目以降にフォローされるフォロワーは、その話題についてそれまでにフォローされたフォロワーでは網羅できないよう

¹ 筑波大学
University of Tsukuba

^{a)} ytakeda@ce.slis.tsukuba.ac.jp

な情報を発信するユーザであると考えられる。一方で、同じ場面において他のユーザに先行してフォローされるフォロワーは、その話題に対する興味を継続させ後続のフォロワーを喚起するような情報を発信するユーザであると考えられる。このようにフォロワーがフォローされる順序には、そのフォロワーがある話題において果たしている役割が反映されていると考えられる。

本稿では、フォローの順序性には上記のような性質があるという仮説のもと、フォロー順序を用いてある話題における Twitter ユーザの役割を推定する手法を提案する。ユーザの役割を推定し、情報源としての性質を明らかにすることは、フォロー先ユーザ推薦に応用できると期待される。また、他のユーザの興味のある話題へと引きつける役割を持つフォロワーは、ユーザに対する影響力が強く、効果的な広告対象になると考えられるため、そのようなフォロワーを推定することは、マーケティングにおいても有用であると期待される。

本稿の構成を以下に示す。まず 2 章で本研究に関連する研究として、Twitter ユーザを対象とした、フォロー先ユーザ推薦の研究や、影響力推定の研究を紹介する。3 章で、提案する Twitter ユーザの役割推定手法について詳述する。4 章で、提案手法の有効性を検証するために行った実験について述べ、5 章ではその考察を行う。6 章で今後の課題と展望を述べる。

2. 関連研究

Twitter ユーザを対象にフォロー先ユーザ推薦を行う研究は数多く行われている。久米ら [5] は、コンテンツベースのユーザ推薦手法として、TF-IDF 法を拡張しユーザの興味に応じた重みを用いることで、より興味の近いユーザを推薦することのできる手法を提案している。坪田ら [9] は、相互フォローとなっているアカウントに対するコミュニケーションの偏りや相互フォローを増やす速度を用いてユーザを特徴付け、コミュニケーションの取りやすいフォロー先ユーザの推薦に用いている。北村ら [10] はソーシャルグラフに加えて、ユーザ間のコミュニケーションを特徴として用いた推薦手法を、黒柳ら [6] は、被推薦ユーザとフォローしている著名人ユーザの重複が多いユーザをフォロー先ユーザとして推薦する手法を提案している。また、大村 [7] は、ユーザのツイートから他の語との共起関係を基にユーザの興味語を抽出する手法を提案し、それをユーザ推薦手法において用いることの有効性を確認している。これらの研究に対して本研究は、ユーザ推薦について、被推薦ユーザの興味の対象となる話題を絞った上で、その話題の上で特徴的な役割を果たしているユーザを推薦ユーザ候補とする点において異なる。

また、SNS ユーザを対象としたユーザの影響力についての研究も数多く行われている。Cha ら [1] は、被フォロー

数、被 RT 数、被 Reply 数を用いてユーザの影響力の分析を行っている。Weng ら [4] は、PageRank ライクな手法を用いてユーザの影響力の推定を行っており、その有効性を確認している。一方で、Fredrik ら [2] は、投稿に対するコメントの共起に着目し、Association rule learning によるユーザの影響力の推定を行っており、精度と計算速度の観点から評価を行っている。これらの研究に対して本研究は、ユーザの影響力を直接的に推定するのではなく、ユーザがある話題において果たしている役割を推定することで、間接的に影響力の高いユーザの抽出を試みる点で異なる。

3. ユーザの役割推定手法

本章では、提案するフォロー順序を用いたある話題における Twitter ユーザの役割の推定手法について説明する。図 1 に手法の概要図を示す。

ある話題を選択した上で、まずその話題に興味を持つユーザ集合を取得し、そのフォローリストを収集する。得られたフォローリスト集合から頻出するフォロワーを抽出した上で、Bradley-Terry モデル [3] を用いてフォローリストごとに存在するフォロワー間の被フォロー順序関係から、被フォロー順序についての 1 つのランキングを作成する。

本稿では、より早期にフォローされるフォロワーほど、フォロワー数が多く、幅広いトピックについての情報を発信しつつその話題へと興味を誘導するようなメディアエーターとしての役割が強いフォロワーであると考えられる。一方で、より晚期にフォローされるフォロワーほど、フォロワー数が少なく、よりその話題に特化した情報を発信するような専門家としての役割が強いフォロワーであると考えられる。また、ユーザはフォローとアンフォローを繰り返してフォローリストを更新すると考えられるため、より早期にフォローされるフォロワーほど、アカウント作成日が早く、より晚期にフォローされるフォロワーほど、アカウント作成日が新しいというような傾向は、強くは見られないと考える。以上の仮説に基づき、提案手法により作成されるランキングにて、上位に登場するユーザほど前者の傾向が強く、下位に登場するユーザほど後者の傾向が強いとして、その役割を推定する。

3.1 フォローリストの収集

フォローリストの収集では、まず、ある話題 a が与えられた上で、その話題に興味を持つユーザ集合 U_a を求める。その後、得られたユーザ $u_a \in U_a$ についてそのフォローリスト $followlist(u_a)$ を取得し、フォローリスト集合 $FollowList_a$ を求める。

本稿では簡便のため U_a を、話題 a を代表する代表ユーザ r_a のフォロワーとする。 r_a は人手により決定する。

Twitter では、Twitter REST API によって、ユーザ

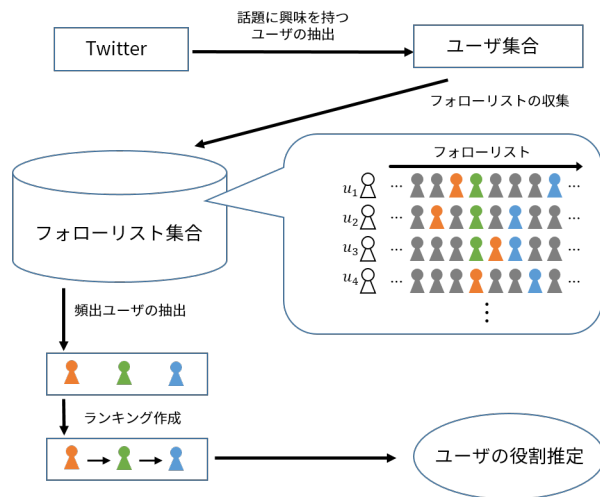


図 1 提案手法概要図

がフォロワーをフォローした順序が保存されたフォロリストを取得することが可能であり、これを用いて $followlist(u_a)$ を取得する。

3.2 ランキングの作成

本節では、3.1 節で取得したフォロリスト集合 $FollowList_a$ から被フォロー順序のランキングを作成する手法について述べる。ランキングを作成するにあたり、まず、 $FollowList_a$ から頻出フォロワーを抽出し、ランキングの対象とするフォロワー集合 $Followee_a$ を求める。本稿では、 $FollowList_a$ 中に 0.25 以上の確率で登場するフォロワーを頻出フォロワーとする。

次に、得られた $Followee_a$ について、Bradley-Terry モデルを用いて、フォロリストごとに存在するフォロワー同士の被フォロー順序関係から、被フォロー順序についての 1 つのランキングを作成する。Bradley-Terry モデルとは、複数の順序関係から 1 つの順序関係を導出するモデルであり、複数のランキングを統合して、1 つのランキングを作ることができることから、スポーツチームの勝敗予測などに用いられる手法である [8]。

Bradley-Terry モデルでは、ある集合における要素のそれぞれについて、各要素間の確率的な順序関係を表現するような、強さとも呼ぶべき π という量が存在すると仮定し、その値を、複数の順序関係から推定する。

本稿では、それぞれのフォロリストを被フォロー順序によって順序付けられるフォロワー間の順序関係とみなして、Bradley-Terry モデルを適用することで、フォロワーについて、より早期におけるフォローのされやすさを推定し、被フォロー順序を表した 1 つのランキングを作成する。

4. 実験

実験として、提案手法により作成されたランキングにお

表 1 データセット概要

	機械学習	ポータブルオーディオ
代表ユーザ	@shima_shima	@kindo3
収集フォロリスト数	3154	2966

けるユーザの順位と、ユーザの個々の特徴量との順位相関係数を算出することにより、提案手法の有効性を検証する。本稿ではユーザの特徴量として、ツイート数、フォロー数、フォロワー数、リスト数、アカウント作成日、トピックの分散を選択し、順位相関係数にはスピアマンの順位相関係数を使用する、リストとは、Twitter においてユーザが話題や属性ごとに作成することができるユーザ集合であり、リスト数とは、そのユーザを含むリストの数である。ツイート数、フォロー数、フォロワー数、リスト数、アカウント作成日の 4 つの特徴量については、Twitter REST API によってユーザごとに取得することが可能である。トピックの分散については求める方法も含めて 4.2 節で説明する。

4.1 データセット

実験は、機械学習 ($a = ml$) とポータブルオーディオ ($a = pa$) という 2 つの話題を対象に行なう。 r_{ml} には @shima_shima を、 r_{pa} には @kindo3 を人手により選択し、 U_{ml} , U_{pa} ならびに $FollowList_{ml}$, $FollowList_{pa}$ を収集した。このとき、5000 人以上をフォローしているユーザ、また非公開アカウントとなっているユーザについては、フォロリストの収集を行わなかった。機械学習の話題については、2016 年 8 月 14 日に、ポータブルオーディオの話題については、2016 年 8 月 17 日にそれぞれのデータの収集を行った。表 1 にデータセットの概要を示す。

4.2 トピックの分散

本稿では、ユーザのツイート集合を結合して作成した 1 つの文書について LDA によって推定されるトピック分布と一様分布との JS ダイバージェンスをトピックの分散として扱う。LDA のモデルは、話題 a ごとに構築を行う。コーパスには、Twitter REST API によって取得できる限り取得した、 $Followee_a$ 内のユーザが行ったツイートを、MeCab^{*1}によって名詞のみの分かち書きにしたものを使用する。モデルの構築には Python のライブラリである gensim^{*2}を用いる。また、トピック数には 15 を指定する。

4.3 実験結果

まず、それぞれの話題において作成されたランキングを表 2 に示す。機械学習の話題では 18 ユーザが、ポータブルオーディオの話題では 13 ユーザが頻出ユーザとして抽出され、ランキングされたことが確認できる。

*1 <http://taku910.github.io/mecab/>

*2 <https://radimrehurek.com/gensim/>

表 2 ランキング概要

順位	機械学習	ポータブルオーディオ
1	@masason	@e_earphone
2	@yukihiro_matz	@ear_ryouta
3	@hyuki	@FUJIYAATIC
4	@mamoruk	@ear_takuya
5	@shima_shima	@iriver_Lyumo
6	@hillbig	@avwatch
7	@ibisml	@fitear
8	@hamadaoichi	@OYAIDE_NEO
9	@preferred_jp	@phileweb
10	@AntiBayesian	@kindo3
11	@sla	@nomurakenji
12	@unnonouno	@KumitateLab
13	@sylvan5	@KumitateK
14	@TJO_datasci	
15	@neubig	
16	@issei_sato	
17	@iwiwi	
18	@beam2d	

表 3 順位相関係数 (機械学習)

	順位相関係数	有意性
ツイート数	-0.119	x
フォロワー数	0.461	x
フォロワー数	-0.616	o
リスト数	-0.656	o
アカウント作成日	0.284	x
トピックの分散	-0.212	x

表 4 順位相関係数 (ポータブルオーディオ)

	順位相関係数	有意性
ツイート数	-0.473	x
フォロワー数	-0.560	x
フォロワー数	-0.725	o
リスト数	-0.687	o
アカウント作成日	0.434	x
トピックの分散	-0.110	x

次に、それぞれの話題のランキングにおけるユーザの順位と各特微量との順位相関係数を表 3, 表 4 に示す。なお無相関検定は有意水準 1%で行った。表 3, 表 4 中で、左の端の列はユーザの特微量の種類を、中央の列はその特微量とランキングにおける順位との順位相関係数を、右端の列は順位相関が検定により有意だと認められたか否かを表している。表 3, 表 4 よりどちらのランキングにおいても、フォロワー数とリスト数の 2 つの特微量について、ランキングの順位との相関が認められることが確認できる。また、これら 2 つの特微量と順位との順位相関係数はいずれも負の値をとっており、負の相関であることが認められる。

5. 考察

表 3, 表 4 より、提案手法により作成されたランキング

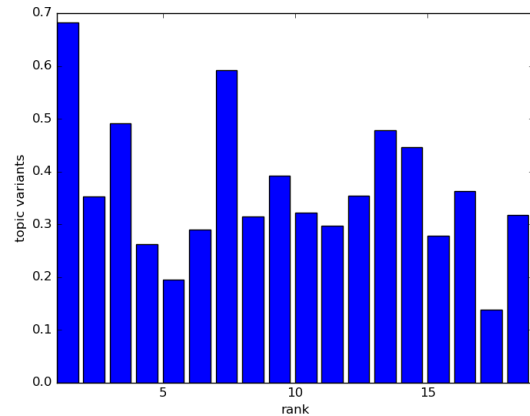


図 2 ランク VS トピックの分散 (機械学習)

の順位とフォロワー数とリスト数の間には、負の相関が認められたことが確認できた。またツイート数やフォロワー数、アカウント作成日については順位との相関が認められなかったことが確認できた。これらは、より早期にフォローされるフォロワーほど、フォロワー数が多く、より晚期にフォローされるフォロワーほど、フォロワー数が少ないという仮説や、より早期にフォローされるフォロワーほど、アカウント作成日が早く、より晚期にフォローされるフォロワーほど、アカウント作成日が新しいというような傾向は、強くは見られないという仮説を支持する、もしくは仮説と矛盾しない結果であり、仮説の正しさが部分的には支持されたと考えられる。

一方で、トピックの分散についてはランキングの順位と相関が認められず、順位相関係数においても両方の話題において負の値が示された。この結果は、より早期にフォローされるフォロワーほど、幅広いトピックについての情報を発信し、より晚期にフォローされるフォロワーほど、その話題に特化した情報を発信するという仮説とは矛盾する。これは、仮説の誤りが示唆されたと考えられるほか、LDA の構築が適切でなかった可能性を示唆しているとも考えられる。

ここで、図 2, 図 3 に各話題におけるランキングの順位と、その順位ユーザのツイートのトピックの分散を示す。図中において横軸はランキング中におけるユーザの順位を、縦軸はそのユーザのツイートのトピックの分散の値を示している。図 2, 図 3 より、それぞれの話題において、トピックの分散の値は、ランキングの最上位では高く、上位では低く、中位では高く、下位では低く、最下位また少し高くなるというおおまかな傾向を確認することができる。このことより、ランキングの順位、つまり被フォロー順序と、そのユーザの発信する情報のトピックの偏りの間には非線形な関係があることが示唆されたと考えられ、今後適切なモデルを構築することが課題として浮き彫りになったといえる。

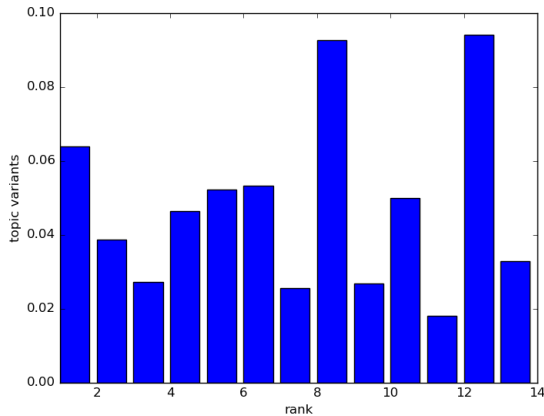


図 3 ランク VS トピックの分散 (ポータブルオーディオ)

6. おわりに

本稿では、ある話題においてユーザがフォローされる順序には、ユーザがその話題において果たしている役割が反映されているという仮説を立てた上で、フォロー順序を用いてある話題における Twitter ユーザの役割を推定する手法を提案した。実データを用いた実験の結果、提案法により作成される被フォロー順序のランキングの順位と、ユーザのフォロワー数や、そのユーザが登録されているリスト数の間には相関があることが確認された。

今後の課題としては、提案手法により作成されるランキングがユーザの役割を推定するのに妥当であるかのより進んだ検証と、フォロワーがフォローされる順序とそのフォロワー果たしている役割についてのより適切なモデルの構築が考えられる。

謝辞

本研究は、JSPS 科研費 16H02904 の助成を受けたものです。ここに記して謝意を示します。

参考文献

- [1] Cha, M., Haddadi, H., Benevenuto, F. and Gummadi, P. K.: Measuring User Influence in Twitter: The Million Follower Fallacy., *ICWSM*, Vol. 10, No. 10-17, p. 30 (2010).
- [2] Erlandsson, F., Bródka, P., Borg, A. and Johnson, H.: Finding Influential Users in Social Media Using Association Rule Learning, *Entropy*, Vol. 18, No. 5, p. 164 (2016).
- [3] Ralph Allan Bradley, M. E. T.: Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons, *Biometrika*, Vol. 39, No. 3/4, pp. 324-345 (online), available from <http://www.jstor.org/stable/2334029> (1952).
- [4] Weng, J., Lim, E.-P., Jiang, J. and He, Q.: Twitter-rank: finding topic-sensitive influential twitterers, *Proceedings of the third ACM international conference on Web search and data mining*, ACM, pp. 261-270 (2010).

- [5] 久米雄介, 打矢隆弘, 内匠逸: 興味領域を考慮した Twitter ユーザ推薦手法の提案と評価, 研究報告知能システム (ICS), Vol. 2015, No. 1, pp. 1-8 (オンライン), 入手先 (<http://ci.nii.ac.jp/naid/110009882486/>) (2015).
- [6] 黒柳智士, 山田泰宏, 鈴木浩, 服部哲, 速水治夫: 著名人情報に基づいた Twitter フォロワーユーザ推薦システム, 研究報告コンシューマ・デバイス&システム (CDS), Vol. 2013, No. 25, pp. 1-4 (オンライン), 入手先 (<http://ci.nii.ac.jp/naid/170000073754/>) (2013).
- [7] 大村涼, 赤石美奈, 佐藤健: 語彙連鎖構造を用いた twitter ユーザー推薦手法の提案, 第 75 回全国大会講演論文集, Vol. 2013, No. 1, pp. 609-610 (オンライン), 入手先 (<http://ci.nii.ac.jp/naid/170000091263/>) (2013).
- [8] 竹内啓: スポーツの数理科学: もっと楽しむための数字の読み方, 共立出版, 東京, Japan (1988).
- [9] 坪田啓司, 小林亜樹: Twitter にて会話しやすいユーザを推薦する手法の評価, 研究報告データベースシステム (DBS), Vol. 2013, No. 15, pp. 1-8 (オンライン), 入手先 (<http://ci.nii.ac.jp/naid/110009632334/>) (2013).
- [10] 北村太一, 小川祐樹, 諏訪博彦, 太田敏澄: ユーザ間関与に基づく Twitter フォロワーユーザ推薦, 日本社会情報学会全国大会研究発表論文集, Vol. 26, pp. 229-232 (オンライン), DOI: 10.14836/jasi.26.0.229.0 (2011).