

コミュニケーションに着目した Twitter フォロワーユーザ推薦

English title

北村 太一^{*1}
Taichi Kitamura小川祐樹^{*2}
Yuki Ogawa諏訪博彦^{*1}
Hirohiko Suwa太田敏澄^{*1}
Toshizumi Ohta^{*1} 電気通信大学

University of Elector-Communications

^{*2} 産業技術総合研究所

National Institute of Advanced Industrial Science and Technology

"Twitter", one of social media, is a service that users post daily "tweets" and the number of its users are spread increasing. Hereby, various services and researches were in the public eye. Researches about "follow", one of Twitter's function that a user can read other users' "tweet"(text), is no exception. However, a lot of these researches use method which recommend followees by existing follow relation or affinity of tweets, and they do not call into account communications between users. Whereat, we approached a method based on "involvement", density of communication between existing followee users, and test a verification of the effect on our method better than existing method in our research.

1. はじめに

本論文では、Twitter フォロワーユーザ推薦手法として、関与ベースに基づく推薦手法を提案し、その精度を評価することを目的とする。Twitter は、2006 年に Obvious 社によって開始されたマイクロブログサービスである。Nielsen の調査によると、2011 年 11 月現在の日本国内における Twitter ユーザ数は 1320 万人と日本国民の約 1 割が利用している計算になる。Twitter では、他者を「フォロー」(follow)することにより、情報収集やコミュニケーションが出来る。「フォロー」とは他者のツイートを購読する機能であり、A が B をフォローしている場合は「B は A のフォロワー(following または followee)」「A は B のフォロワー(follower)」と呼ぶ。Twitter ユーザは、仲間内でのコミュニケーションや著名人へのアクセス、情報発信・収集等、様々な目的に応じて Twitter を利用している。[富士通総研 10]

鳥海らは、Twitter ユーザの利用特性について①「広く浅いコミュニケーションツールとして利用」、②「特定ユーザ間のみでのコミュニケーションツールとして利用」、③「コミュニケーションツールとして利用せず」の 3 パターンに分類し、パターン①が全体の 40%以上を占めていることを報告している[鳥海 11]。友人・知人間のクローズドなコミュニケーション(パターン②)も含めると、コミュニケーションツールとして Twitter を利用するユーザは多く、そこで生成されるコミュニケーションネットワークには重要な意味があると考えられる。

Twitter におけるフォロワーユーザ推薦は、Twitter 公式「おすすめユーザー」などのサービスとして存在するほか、各種研究もなされている。これらの研究では、コンテンツベースフィルタリング及び協調フィルタリングを用いて推薦を行っている。一方「おすすめユーザー」は、「フォロー情報等に基づく」フォロワー推薦を行なっている。しかしこれらはユーザ間関与(リプライ等のコミュニケーション)を考慮に入れておらず、コミュニケーションツールとして Twitter を利用している場合、被推薦ユーザにとってフォロー価値の低いユーザ(著名人等のアカウント)が推薦される場合がある。

そのため、本研究ではリプライ等のユーザ間関与に基づく 2-hop ユーザ推薦モデルを提案し、1-hop 目と 2-hop 目の双方にユーザ間関与を用いた手法と、どちらか一方にユーザ間関与を

用いた手法を提案し、それぞれの手法と既存手法を比較し、その有効性を検証する。

本論文において「関与」とは、フォロワーユーザへのコミュニケーションと定義する。即ち、リプライ(reply)など「@ユーザ名」を含むツイートである。但し、公式リツイート(retweet; RT)は文章をフォローに広めるのみで RT 元とのコミュニケーションが存在しないので、「関与」から除外する。一方、非公式 RT はユーザ自身のコメントを RT 元の文章に付記するものであり、RT 元ユーザとのコミュニケーションがあるものとみなして「関与」に含めるものとする。

2. 関連研究

本章では、関連研究について述べる

2.1 Twitter のフォロワー関係に関する研究

Golder らは、twitter におけるユーザ同士の結びつき方について、2-hop の非対称ソーシャルグラフ 32=9 通り(図 1)に対してどのような条件でフォローが為される傾向にあるかを述べている[Golder10]。

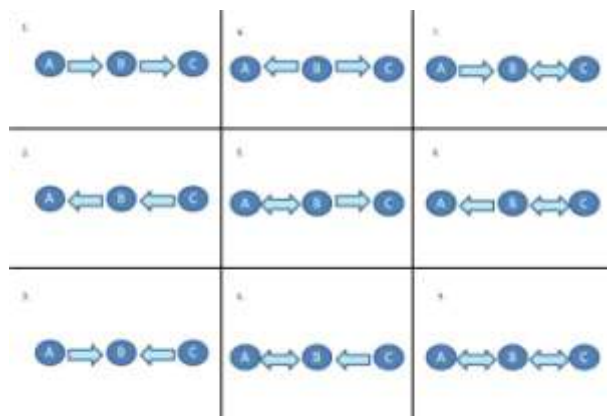


図 1 : 2-hop の非対称ソーシャルグラフ (Golder et al. より)

被験者 A に対して、様々な結びつきである C を推薦し、その結果を比較している。評価実験の結果、(A⇔B→C),(A→B⇔C),(A⇔B⇔C)の 3 パターンが有効という結果となり、推移性

(→:フォロー関係)及び相互性(⇔:相互フォロー関係)の双方が有効であることを示している。また、フォロー(←)と相互フォロー(⇔)の両関係を持った($A \leftrightarrow B \leftarrow C$)($A \leftarrow B \leftrightarrow C$)では逆効果であることが示されている。なお、フォローを推薦するとき用いたものは共通のフォロー・フォロー(図 1 の B)の数のみでツイートの中身を用いていない。

2.2 Twitter のフォロー推薦手法に関する研究

Hannon らは、Twitter のフォロー推薦に関する研究として、コンテンツベースフィルタリング(Content-Based Filtering; CB)と、協調フィルタリング(Collaborative Filtering; CF)及び、その両者を組み合わせたハイブリッド推薦手法を用いた Twitter フォローユーザの推薦手法の研究を行なっている[Hannon10]。

CB は、コンテンツ同士の内容・説明の類似性をベースとして推薦を行う手法で、情報推薦で一般的な推薦手法の一つである。被推薦ユーザ(A)及びそのフォロー・フォロー(B)のツイートの類似度、および B とそのフォロー(C)のツイートを計測することにより、ツイート内容の近い 2-hop ユーザ(C)を推薦する手法である。

一方 CF は、ユーザのプロファイルや行動履歴をベースに推薦する手法で、こちらも情報推薦において一般的な推薦手法の一つである。被推薦ユーザと類似した嗜好を持つコンテンツが推薦されるので、類似した嗜好を持つユーザが多ければ多いほど推薦精度が上がる。Hannon らは、フォロー推薦に CF を適用する上で、被推薦ユーザ(A)のフォロー・フォロー(B)によって最も多くフォローされている 2-hop ユーザ(C)を推薦する手法を採用している。これにより CB では追跡困難な嗜好・属性変化への追跡が行えるといったメリットがある。

さらに、Hannon らは、CB と CF の両者を組み合わせたハイブリッド推薦手法を提案し評価している。その結果、被験者実験においてハイブリッド推薦手法が最も良い結果であったことを報告している。しかし、被推薦ユーザ(A)のフォロー・フォロー(B)への関与および B のフォロー(C)に対する関与については触れられておらず、特に CF を用いた場合は各フォローユーザが均等に扱われるという問題点がある。

2.3 Twitter 公式推薦システム「おすすめユーザー」

Twitter には、フォローユーザを推薦する「おすすめユーザー」という推薦システムがあり、「おすすめユーザー」は大きく分けて 2 種類に大別される。

第一に、「芸能・文化」「音楽」「スポーツ」等ジャンル別に著名人・有名企業のアカントを列挙し、Twitter を始めたばかりのユーザに推薦する仕組みである。これらのアカウントは Twitter 運営元である Twitter Inc.の追加した著名人・有名企業アカウントである。この推薦手法は、Twitter を著名人のプライベート及び最新のニュース等に関する情報収集や著名人とのリプライ等によるコミュニケーションを欲するユーザに対しては有用であると考えられる。一方で、友人・知人あるいは「オタク」など趣味・嗜好の類似した Twitter ユーザとのコミュニケーションに利用するようなユーザに対しては有用性は低いものであると考えられる。

第二に、Twitter ユーザのフォロー情報などに基づいて他ユーザをフォロー推薦する仕組みである。Twitter Inc.では推薦アルゴリズムを明らかにしていないが、フォローやフォローの多くがフォローを行なっているユーザが優先的に推薦されるようになっている。そのためユーザは、フォローの多い有名名人を中心に推薦されることとなり、コミュニケーションツールに Twitter を利用するユーザに対しては、有効な推薦がなされていない。なお、ここで「有名名人」はフォローが 10000 人を超えるユーザ、オピニオンリ

ーダなど特定分野・属性に関する情報発信力の強いユーザとし、著名人・企業には限らない。

2.4 Twitter のリプライ機能を用いた研究

田沼らは、特定分野に関する情報をツイートするユーザを、Twitter 機能の一つである“リプライ”に着目して発見する手法を提案している[田沼 11]。この研究はユーザ間コミュニケーションを基にした研究で、特定分野に関するツイートのリプライが多ければ多いほど発見精度が上がることを報告している。アカウント発見に関して、「分野に関するワードを含む割合の高いアカウント(core アカント)を発見する手法」と「core アカントのリプライに着目し、同様の分野に関するワードがリプライ内で含まれる割合を用いる手法」の 2 手法を統合することにより、core アカントであるユーザ自身のツイートの形態素解析結果では発見が困難な「特定分野に濃いユーザアカウント」を発見できることを報告している。

2.5 関連研究まとめ

2.1~2.3 節の研究・システムはフォローユーザ間のコミュニケーション(=ユーザ間関与)を推薦に用いておらず、被推薦ユーザにとって関係性の薄いユーザが推薦されることがある。一方、2.4 節の研究ではフォロー推薦ではないが、リプライというユーザ間関与を用いてユーザを発見する手法を提案している。本論文では田沼ら[7]の考え方を一部踏襲し、ユーザ間関与に基づくフォローユーザ推薦手法を提案する。

3. 関与に基づくフォローユーザ推薦

2 章で述べた通り、Hannon らや Golder らの手法では、フォローユーザに対する被推薦ユーザの関与については考慮されておらず、フォロー及びフォローとの繋がりなどで趣味・嗜好の異なるユーザが推薦される可能性がある。そこで本研究では、ユーザのフォローへの関与を考慮したユーザ推薦を行う。ユーザ A がフォローユーザ B に関与を示して入れば A と B の関係が密であることが予想される。また、B とそのフォロー C との間にも同様のことが考えられる。この場合において、A にとって未知な 2-hop ユーザ C を推薦する。

また、ここでは A から B 及び B から C へのフォロー・関与についてのみ注目し、逆方向(C から B 及び B から A)のフォロー・関与については考慮しないものとする。これは、他のユーザを「フォロー・関与している」ことはそのユーザに対して関心や興味を持っていることを意味するのに対し、他人に「フォロー・関与されている」ことは逆に関心や興味を示されていることである。即ち被推薦ユーザの関心や興味ではない。よって、A-B 間及び B-C 間ではそれぞれ A, B が B, C に「フォロー・関与されている」だけの関係を排し、「フォロー・関与している」という関係のみを利用することとする。これにより推移性(transitivity)を担保することができる。

提案手法の概要を図 2~4 に示す。図 2(提案手法 1)および図 3(提案手法 2)は、A-B 間と B-C 間の一方にユーザ間関与、他方にフォロー関係を利用したモデルである。提案手法 1 では 2-hop 目の B-C 間、提案手法 2 では 1-hop 目の A-B 間に関与を利用している。図の矢印は、フォローを示す。例えば“A→B”は、A が B をフォローしていることを意味する。一方図 4 は、A-B 間と B-C 間の双方にユーザ間関与を利用したモデル(提案手法 3)である。

この関与を測る指標として、“関与度合”を導入する。これは、あるユーザに対してどれくらいの関与を示したか数値化するもので、単位ツイートあたりの関与(公式 RT を除く@tweet)数で定

義する。A から見た B への関与度合は $w(ab)$ と表記し、100tweet あたり 7 つの @tweet であれば、関与度合は $w(ab)=0.07$ である。提案手法 1 では A からのフォローが認められる B からの関与度合の総和の多寡をユーザ推薦に用いる。一方提案手法 2 では、A から B への関与が認められる(即ち $w(ab)$ が 0 でない) B からのフォロー数の多少をユーザ推薦に用いる。

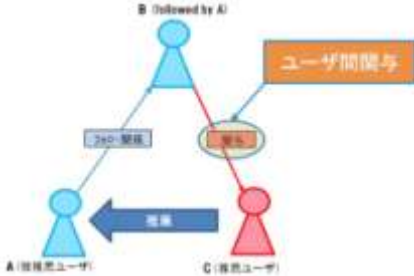


図 2 提案手法 1 (B-C 間に関与利用)

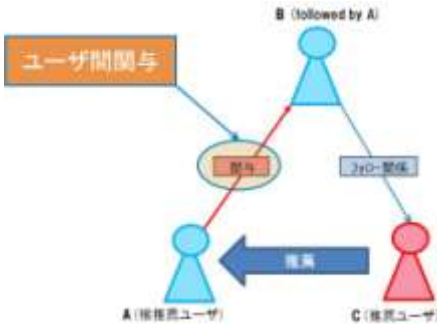


図 3 提案手法 2 (A-B 間に関与利用)

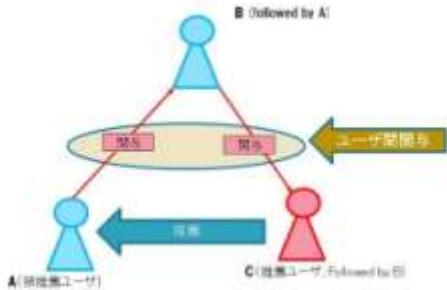


図 4 提案手法 3 (関与のみを用いた推薦手法)

A-B 間及び B-C 間の双方に関与を用いている提案手法 3(図 4)では、A から見て C は、複数のユーザ B を介して繋がっている場合も考えられるので A-B 間の関与度合と B-C 間の関与度合の積を、C と繋がっている各々の B に対して合算させた値で表す多少で比較を行う。(図 5)

推薦に用いる指標を"関与スコア" (Involvement Score) と呼び、提案手法 1 での関与スコアは $IS(1)$ 、提案手法 2 では $IS(2)$ 、提案手法 3 では $IS(3)$ と表す。それぞれの定義は以下の通りである

$$IS(1) = \sum w(bc) \quad (and f(a,b)=1) \quad \dots(1)$$

$$IS(2) = \sum f(b,c) \quad (and w(ab) > 0) \quad \dots(2)$$

$$IS(3) = \sum w(ab) \cdot w(bc) \quad \dots(3)$$

図 5 の関与スコアは、提案手法 3 における C_1, C_2, \dots, C_i の関与スコア $IS(3)$ を表している。また B から見た C への関与度合を $w(bc)$ 、A→B 間及び B→C 間のフォロー関係を $f(a,b), f(b,c)$ と

定義し、A→B 間にフォロー関係があれば $f(a,b)=1$ 、なければ $f(a,b)=0$ とする。

表 1 に提案手法と既存手法を整理する。提案手法 1 は A-B 間についてはフォロー関係を用い、B-C 間の関与を考慮して推薦する手法(フォロー・関与型)である。提案手法 2 は、A-B 間は関与を考慮、B-C 間はフォロー数で推薦する手法(関与・フォロー型)である。提案手法 3 は、A-B 間・B-C 間共に関与を考慮する手法(関与・関与型)である。なお、すべての手法において、B は A のフォロー、C は B のフォローに限定する。既存手法は、関与を考慮せず A-B 間・B-C 間共にフォロー数を考慮する手法(フォロー・フォロー型)である。

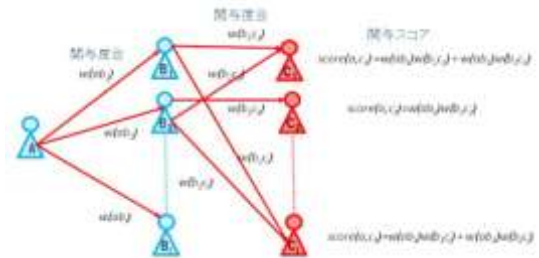


図 5 提案手法 3 における関与度合及び関与スコア

表 1 既存手法と提案手法

	概要	関与A→B	関与B→C
提案手法1 (フォロー・関与型)	$f(a,b)=1$ における $\sum w(bc)$ の大きいCを推薦(2hop目は関与)	-	○
提案手法2 (関与・フォロー型)	$\sum f(b,c)$ (但し $w(ab)>0$)の多いCを推薦(1hop目は関与)	○	-
提案手法3 (関与・関与型)	関与スコアでCを推薦(ユーザ間関与で推薦)	○	○
既存手法 (フォロー・フォロー型)	$f(a,b)=1$ かつ $\sum f(b,c)$ の多いCを推薦。(フォロー数で推薦)	-	-

(○:関与考慮 - :関与考慮せず)

4. 実験

4.1 実験目的

3 種類の提案手法がどのような特徴を持っているかを検証するために、1-hop 目(A-B 間)にフォロー数、2-hop 目(B-C 間)に関与を用いた推薦手法「フォロー・関与型」(図 2, 提案手法 1)、1-hop 目(A-B 間)に関与、2-hop 目(B-C 間)にフォロー数を用いた「関与・フォロー型」(図 3, 提案手法 2)、1-hop 目(A-B 間)、2-hop 目(B-C 間)共に関与を用いた「関与・関与型」(図 4, 提案手法 3)、1-hop 目(A-B 間)、2-hop 目(B-C 間)共にフォロー数を用いた「フォロー・フォロー型」である既存手法を用いて比較する。

4.2 データセット

被推薦ユーザ A 集合として、任意の 11 人(フォロー数:14 人~252 人 平均 109.45)を抽出した。ユーザ A 集合の年代は 20 代 9 名、30 代 2 名であり、性別は全て男性である。各ユーザ A について、フォロー情報(ユーザ B 集合及びユーザ C 集合)及びツイート(ユーザ A のツイート、ユーザ B 集合のツイート)の収集を行った。収集期間は 2011/12/19~2011/12/22 で、ツイートに関しては、それぞれ最新 100 ツイートを収集している。ユーザ A 集合、ユーザ B 集合の全フォローは延べ 571466 件である(ユーザ B 集合: 1204 ユーザ、ユーザ C 集合:570262 ユーザ)。但し、ユーザ C 集合には B 集合のフォローを全て収集したため、ユーザ A 集合及びユーザ B 集合と重複するユーザも存在する。

4.3 実験方法

表 1 で示した 4 種類の推薦手法を比較するための実験を行う。既存手法では、ユーザ B 集合からより多くフォローされている上位 10 人となるユーザ C 集合を推薦する。提案手法 1 では、B-C 間の関与度合の上位 10 人となるユーザ C 集合を推薦する。提案手法 2 では、A-B 間の関与が存在するユーザ B 集合によってフォローされているユーザ C 集合のうち、ユーザ集合 B からより多くフォローされている上位 10 人となるユーザ C 集合を推薦する。提案手法 3 では、関与スコアの上位 10 人となるユーザ C 集合を推薦する。評価手法は 10 分割交差検証法を用いた。

5. 実験結果

評価実験の結果を表 2 に示す。F 値、精度、再現率ともに提案手法 2 (関与・フォロー型) が最も高く、既存手法 (フォロー・フォロー型)、提案手法 1 (フォロー・関与型) と続き、提案手法 3 (関与・関与型) は、F 値、精度、再現率ともに最も悪い数値となっている。つまり、A-B 間は関与を重視し B-C 間はフォロー数を考慮して推薦する関与・フォロー型が他の 3 手法よりも良い結果となっている。一方で、提案手法 3 (関与・関与型) が、既存手法 (フォロー・フォロー型) より精度・再現率ともに低い結果となっている。

表 2 各手法における精度・再現率及び F 値

推薦パターン	F 値	精度	再現率
提案手法1 (フォロー・関与型)	0.057	0.053	0.063
提案手法2 (関与・フォロー型)	0.118	0.115	0.123
提案手法3 (関与・関与型)	0.035	0.032	0.039
既存手法 (フォロー・フォロー型)	0.087	0.084	0.091

6. 考察

評価実験によって、提案手法 2 (関与・フォロー型) で最も良い結果が確認されている。一方で提案手法 3 (関与・関与型) では最も悪い結果が確認されている。また、提案手法 3 ほどではないが、提案手法 1 (フォロー・関与型) も既存手法 (フォロー・フォロー型) より悪い結果となっている。

提案手法 2 で推薦されるユーザは、被推薦ユーザ A がコミュニケーションを取っているユーザ B 集合から、多くフォローされているユーザである。被推薦ユーザ A 関与度合いが強い人の多くがフォローしているユーザを推薦することができ、推薦精度の向上が図られたものと考えられる。

一方、提案手法 3 で推薦されるユーザは、被推薦ユーザ A がコミュニケーションを取っているユーザ B 集合のユーザが、多くコミュニケーションをとっているユーザである。このため、ユーザ A と仲の良いユーザ B の親友であるユーザ C が推薦されることもある。この時、ユーザ A とユーザ C が同一のコミュニティに所属していれば問題ないが、ユーザ A とユーザ B の関係が大学の研究室仲間、ユーザ B とユーザ C の関係が幼馴染などで

ある場合、ユーザ A とユーザ C の繋がりほとんどないこととなる。このため、推薦精度が低下したものと考えられる。

7. おわりに

本論文では、Twitter フォローユーザ推薦手法として、関与ベースに基づく推薦手法を提案し、その精度を評価することを目的とした。2hop 先のユーザを推薦する手法として、1hop 目、2hop 目共にユーザ間関与を用いる手法と、どちらか一方にユーザ間関与を用いる手法を提案している。既存手法を含め、4 つの手法について、評価実験を行ったところ、1hop 目に関与を利用し、2hop 目にフォロー関係を利用する提案手法 2 (関与・フォロー型) が最も良い結果となることを確認している。

今後は、より被験者数を増やし、被験者の特性を考慮したうえで各提案手法の精度の比較を行うことで、各被験者の特性に合った推薦手法の発見を行いたい。

参考文献

- [富士通総研 10] 富士通総研: Twitter(ツイッター)利用状況調査 ～若年層と 40 代で異なる利用目的～, <http://jp.fujitsu.com/group/fri/report/cyber/research/twitter/>, 2010 年 3 月発表.
- [鳥海 11] 鳥海不二夫, 神谷達幸, 石井健一郎: Twitter におけるつぶやきを用いたユーザの特性分析, 情報処理学会, 第 7 回 ネットワーク生態学シンポジウム, pp.1234-1237, 2011.
- [Golder 10] Scott A. Golder, S. Yardi: Structural Predictors of Tie Formation in Twitter: Transitivity and Mutuality, 2010 IEEE Second International Conference on Social Computing, socialcom, pp.88-95, 2010.
- [Hannon 10] John Hannon, Mike Bennett, Barry Smyth: Recommending Twitter Users to Follow Using Content and Collaborative Filtering Approaches, RecSys '10 Proceedings of the fourth ACM conference on Recommender systems, pp. 199-206, 2010.
- [田沼 11] 田沼勇輝, 鈴木政巳, 小林亜樹: Twitter における特定分野に「濃い」アカウントの発見手法, DEIM Forum 2011 A10-4, 2011.