

属性伝搬モデルを用いたマイクロブログのフォロー先推薦法

康 大樹[†] 島田 諭^{††} 関 洋平^{††} 佐藤 哲司^{††}

[†] 筑波大学情報学群知識情報・図書館学類 〒305-8550 茨城県つくば市春日 1-2

^{††} 筑波大学図書館情報メディア研究科 〒305-8550 茨城県つくば市春日 1-2

E-mail: [†]s0913157@u.tsukuba.ac.jp, ^{††}{sat,yohei,satoh}@slis.tsukuba.ac.jp

あらまし Twitter に代表されるマイクロブログでは、ユーザが他のユーザの投稿を簡単に閲覧するために、気に入ったユーザをフォローする機能がある。しかし、多くのユーザをフォローすると画面表示が煩雑になるという問題があり、フォロー先として適するユーザの発見や絞り込みを支援する機能が重要となる。本論文では、ユーザの興味や所属、居住地といった潜在的な属性の、ユーザ間でのフォロー関係を通じた伝搬に着目し、属性伝搬を考慮してフォロー先、ユーザを推薦する手法を提案する。本手法では、一定数の伝搬元ユーザを共有するユーザを、共通の属性を持つ確率が高いユーザとみなす。実際のユーザのフォロー関係を対象とする評価実験を行ったので報告する。

キーワード ユーザ推薦, Twitter, ソーシャルグラフ

A Method for Contact Recommendation in Micro-blogs using Transmission Model of Shared Attributes of Followers

Hiroki KOU[†], Satoshi SHIMADA^{††}, Yohei SEKI^{††}, and Tetsuji SATOH^{††}

[†] College of Knowledge and Library Sciences, School of Informatics University of Tsukuba
1-2, Kasuga, Tsukuba, Ibaraki, 305-8550 Japan

^{††} Graduate School of Library, Information and Media Studies, University of Tsukuba
1-2, Kasuga, Tsukuba, Ibaraki, 305-8550 Japan

E-mail: [†]s0913157@u.tsukuba.ac.jp, ^{††}{sat,yohei,satoh}@slis.tsukuba.ac.jp

Abstract In micro-blogs, e.g., Twitter, the function to follow other users is provided, and so contributions of other users followed by the user is displayed on one screen. However, there is a problem that the screen becomes confused when there are a lot of users followed, the function that supports discovery and narrowing suitable users to follow or contact becomes important. In this paper, we propose a method for contact recommendation that focuses transmission of shared latent attributes in social graph of users. The user who has constant number of shared followers is considered to be a user who indicates high probability of shared attributes, and who is suitable to contact in the method. We report on the results of the experiment using real social graph in Twitter.

Key words Contact Recommendation, Twitter, Social Graph

1. はじめに

マイクロブログとは短い文章を投稿し、時間軸に沿って表示するブログの一種である。近年ではマイクロブログの中でも代表的な Twitter^(注1)のユーザが急激に増加している。Twitter では投稿される短い文章をつぶやき、つぶやき一覧を投稿された時間順に表示するインターフェースをタイムラインと呼ぶ。また、フォローと呼ばれる機能を使えば自分のタイムラインに他ユーザの投稿を表示することができ、複数のユーザ間で簡単に

情報交換を行うことができる。一般に、友人や知人など身近で親しいユーザをフォローする場合と、興味のある情報を発信するユーザをフォローする場合とがある。ユーザは、自分のタイムラインを閲覧するだけで、所望の投稿をリアルタイムでチェックすることができる。しかし、フォローするユーザが増えるほど、多くの投稿が表示されてタイムラインが煩雑になるという問題がある。

本研究では、フォロー先のユーザが別のユーザをフォローするネットワーク構造によって、ユーザの属性が伝搬する属性伝搬モデルを構築し、新たなフォロー先候補を推薦するフォロー先推薦法を提案する。提案手法で扱う属性とは、ユーザの持つ

(注1): <http://twitter.com>

興味、所属を意味する。例えば、サッカー、野球、コンピュータなどに興味があるユーザは、サッカー、野球、コンピュータの属性を持つ。また、自分と同じ所属（会社、大学、高校など）のユーザは、仲間内で相互にコミュニケーションを行うためにフォローをすることから、共通の所属属性を持つ。ユーザは自分と同じ属性を持っている他ユーザをフォローする傾向があるとする。個々のユーザはいくつもの属性を持ち、属性毎に他のユーザをフォローしているとする属性伝搬モデルに基づいて、フォロー先ユーザの候補を絞り込めると考えられる。また、フォローしているユーザのフォロー先には同じ属性を持つ新たなユーザが存在すると考えられる。これを本研究では属性の伝搬と定義した。本研究ではこの属性の伝搬を利用したフォロー先ユーザの推薦を目指す。

2. 関連研究

Twitter の投稿から有用な情報を発見する研究は多数存在する。Huberman ら [1] は、フォロー関係だけでなく、返信機能を用いてお互いに投稿しあうユーザのネットワークを抽出している。この研究では表向きには見えていないユーザ間のネットワークに注目している。吉本ら [2] は、フォローされている数でユーザの人気を計り、つぶやきの引用回数と引用先のユーザのフォローされている数で投稿の重要度を示す指標を定義している。田中ら [3] は、フォロー関係の意味を考慮することと、お気に入り付けられたつぶやきの割合から有用なユーザを判別できるとしている。ユーザは利用目的、有名度、好み、情報収集意識、情報発信意識の属性を持つと仮定し、情報発信意識の高いユーザの探索を試みている。

Java ら [4] は、Twitter 上の大規模なソーシャルグラフを分析し、Twitter のユーザは以下の 3 タイプの傾向に大別されるとしている。

Information Source 定期的に投稿し、他の多くのユーザからフォローされるユーザ

Friends 大部分のユーザ

Information Seeker 自らはほとんど投稿せず、他の多くのユーザをフォローするユーザ本研究では、上述の Friends に該当するユーザ、およびその周辺のコミュニティを対象とするフォロー先推薦の実現を目指す。

Wu ら [5] は、ソーシャルグラフ上の情報の流れと、ユーザの興味との関連を調べ、グラフ上での距離が長くなるにつれユーザ間での属性の類似度が低下することを報告している。Ziegler ら [6] は、ソーシャルグラフ上での信頼の伝搬をモデル化し、ローカルなグループ内における信頼性の指標を提案している。Scott ら [7] は、Twitter 上でのユーザ推薦の要件として、興味やフォロー元ユーザの共有を考慮することが重要であると指摘している。

これらの知見から、属するコミュニティが異なる場合には属性が伝搬しにくく、伝搬元ユーザを共有する同一のコミュニティ内では距離が長くなっても属性が伝搬しやすいことが推測される。そこで本研究では、ユーザが潜在的に持つ属性が複数であることを仮定し、属性が伝搬するには複数の伝搬元ユー

ザを共有することを条件とする。

従来研究では、有用な情報を多く発信する重要なユーザの発見を目指す研究が多い。これに対し、本研究では必ずしも重要なユーザに限らず、あるユーザにとって、自分と近い属性を持つと推定されるユーザを推薦することを目指す。具体的には、フォロー関係にあるユーザ間での属性伝搬を仮定し、一定数以上の伝搬元ユーザを共有するユーザを推薦する。

現在、Twitter でのフォロー先探索手法を大きく 2 つに分けると以下の二つがある。

ユーザ検索

ユーザ検索ではユーザ名、メールアドレスを図 1 に示ような検索窓に入力することで、ユーザを探すことができる。検索を行うためにユーザ名かメールアドレスをあらかじめ知っておく必要があるが、この検索の特徴は手間と時間がかからないという特徴がある。

フォロー追跡

Twitter ではユーザがどのユーザをフォローしているかを知ることができる。そこで、あるユーザがフォローしている中から自分が新たにフォローするユーザを探すことができる。この手法では自分が今まで知らなかった新たなユーザを発見することができる。また、自分がフォローしているユーザのフォローをたどる場合、より自分の興味がある分野の情報をつぶやくユーザや、自分と関わりのある所属に属しているユーザを発見できる確率が高くなる。ただし、ユーザ検索と違い、目的のユーザを発見できるとは限らない。

アカウントを見つけて、フォローしましょう!

図 1 Twitter のユーザ検索窓

3. 属性伝搬モデルを用いたフォロー先推薦手法

3.1 ユーザ間で共有される潜在的な属性

本研究で定義した属性について説明する。属性はユーザが持つ興味・所属のことを指す。ここで言う興味とはユーザが関心を持ち、積極的に情報を得たい、または発信したいと考えている分野を指す。

経験的に、ユーザは何らかの共通点を持つユーザを漠然とフォローすることが多く、フォローの理由を厳密に示すことがユーザ自身にも困難な場合が少なくない。このため本論文では、フォローされているという状態が、何がしかの属性を共有しているとフォロー元のユーザが判断した結果であるとみなす一方、属性の種類や内容は特定しない。例えば、サッカーに興味を持つユーザには、サッカーに関する情報を収集したいと考

えるユーザ、および、サッカーに関する情報を発信したいと考えるユーザの両方がいるが、本論文ではこの二者を区別せずに扱う。また、あるユーザ間でサッカーや野球といった興味が共通していることと、所属や居住地などが共通していることを等価とみなす。

Twitter ユーザは各々1つまたは複数の属性を持っており、そのユーザがフォローしているユーザも少なくとも一つは同じ属性を持っていると定義した。この関係性を図にしたものを図2に示す。図は対象ユーザをP、フォロー先のユーザをA、B、ユーザの持つ属性はX大学、Y高校、サッカーとし、フォロー関係を矢印で表した。そしてフォロー先のユーザA、Bは少なくとも1つはPと同じ属性を持つ。ユーザAはPと共通の属性としてX大学と持つ。

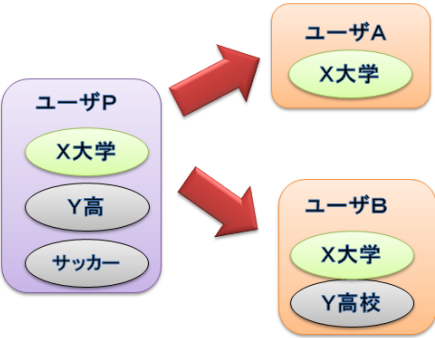


図2 フォロー先と共通する属性

3.2 属性伝搬モデル

本研究における属性伝搬について説明する。上記に述べたようにユーザは属性の一つまたは複数持ち、そのユーザがフォローしているユーザは少なくとも一つは同じ属性を持っていると考えられる。これを本研究では属性の伝搬と呼ぶ。伝搬した属性はフォロー先のユーザも持っているため、さらにそのフォロー先をたどることで、新たな同属性を持ったユーザを取得できる可能性がある。本研究では属性の伝搬を繰り返すことで、特定の属性を持ったユーザの集合を検出できると考えた。しかし、ユーザが持っている、または伝搬している属性を特定するには実際にそのユーザのプロフィール、つぶやきを確認しなければならないため、膨大な時間と手間を費やす必要がある。そこで本研究では条件を満たしたユーザを属性が伝搬したものと定義した。これを図にしたものを図3に示す。図では対象ユーザであるユーザPがフォローしているユーザA、Bが共通のユーザSをフォローしている。この時、ユーザA、Bは共通のX大学属性を持っていることから共通のフォロー先であるユーザDもX大学属性を持っていると考えられる。

対象ユーザが直接フォローしているユーザは少なくとも一つの共通する属性を持つ。そのなかからX大学属性を持つユーザに着目する。この時、X大学属性を持つユーザが2人以上いた場合、そのユーザ2人がフォローしているユーザは同じ属性が

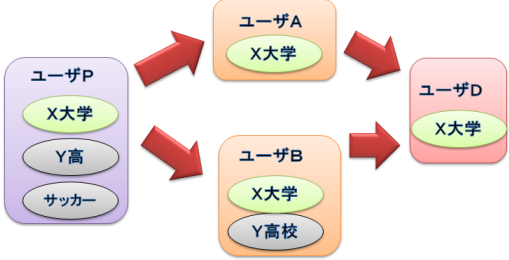


図3 属性の伝搬モデル

伝搬している可能性が高い。このユーザを本研究ではこの状態を属性伝搬したものとする。このように属性伝搬を繰り返すことで、図4のような共通の属性を持ったユーザ集合が取得できる。この集合をコミュニティと呼ぶ。コミュニティは対象ユーザごとに、そのユーザが持つ属性の数だけ取得できるとする。

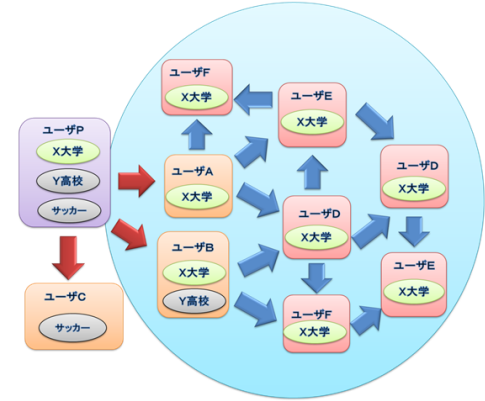


図4 推測される属性伝搬の広がり

また、先ほどの例では二人以上のフォローにより属性が伝搬したが、図5のように伝搬元の人数を2人から3人にすることでより正確な属性伝搬が行えると考えた。また、同じように3人より4人、4人より5人と伝搬元の人数を増やすことでより属性伝搬は正確になっていくと考える。属性が正確に伝搬した確率を伝搬確率と呼ぶ。伝搬確率は以下の式により算出する。

伝搬確率 = $\frac{\text{属性が伝搬したユーザ数}}{\text{検出したユーザ数}}$

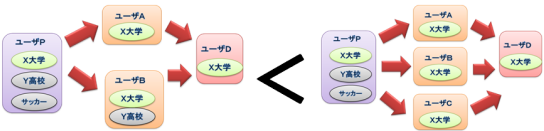


図5 伝搬元人数による属性伝搬への影響

3.3 フォロー先推薦アルゴリズム

本論文では、3.1 節で定義した属性伝搬モデルに基づき、ユーザを推薦する手法を提案する。本手法で用いるユーザ探索アルゴリズムを以下に詳述する。これを図に示したものを図 6 に示す。今回は TwitterAPI よりユーザ間のフォロー関係を取得した。このアルゴリズムにより取得したデータを用いて、仮説検証を行う。手順としては、まず入力としてユーザのユーザ名を入力する。その後、TwitterAPI から対象ユーザとフォロー関係にあるユーザ名を取得する。この時、取得したユーザを本研究では第 1 世代ユーザと呼ぶ。本研究ではフォロー関係にあるユーザは少なくとも一つは共通の属性を持っていると定義しているため、第 1 世代ユーザは対象ユーザと共通の属性を持っているものとする。次に第 1 世代ユーザとフォロー関係にあるユーザを取得する。ここで取得したユーザを第 2 世代ユーザと呼ぶ。第 2 世代ユーザは第 1 世代ユーザと少なくとも 1 つ属性が伝搬しているが、対象ユーザから属性が伝搬していないユーザも取得してしまう。そこで、属性が伝搬していると考えられるユーザに絞り込む。本研究では上記で定義したように同属性を持つユーザから 2 人以上とフォロー関係にあるユーザは属性が伝搬しているものと判断する。そこで、第 1 世代ユーザと 2 人以上フォロー関係にある第 2 世代ユーザを対象ユーザから属性が伝搬したものとする。この先、ここで厳選した第 2 世代ユーザは属性が伝搬していると考えているため、擬似的に第 1 世代ユーザと同じように扱う。次に対象ユーザから属性が伝搬しているとした第 2 世代ユーザとフォロー関係にあるユーザを取得する。このユーザを第 3 世代ユーザと呼ぶ。その後、取得した第 3 ユーザを対象ユーザから属性が伝搬しているものに厳選する。ここで厳選したユーザも擬似的に第 1 世代ユーザとして扱う。このようにして属性の伝搬を繰り返すことで、対象ユーザと同属性を持ったユーザである擬似的な第 1 世代ユーザを推薦する。

4. 評価実験

本論文では、3.2 節で述べた属性伝搬モデルの有効性を確認するため評価実験を行う。以下、4.1 節で実験の概要を説明し、4.2 節で 実験結果を詳述する。

4.1 評価実験の概要

本研究では以下の実験を行う。

- (1) 伝搬元の設定人数による属性伝搬への影響
- (2) 属性伝搬モデルの推薦システムとしての有効性
- (3) 世代ごとの検出ユーザ数

前章で考案したアルゴリズムを用いて実際のデータを取得し、定義した属性の伝搬が行われているか評価を行った。データを取得した際に作成したシステムの概要図を図 7 に示す。

まず対象ユーザのユーザ名をシステムに入力する。システムはそれを Twitter-API に引き渡し、対象ユーザのフォローしているユーザを取得する。この時取得したユーザは前章で定義した第 1 世代ユーザである。次に第 1 世代ユーザを入力としてそのフォロー先である第 2 世代ユーザを取得する。そして第 2 世代ユーザを入力として第 3 世代ユーザを取得する。この時、

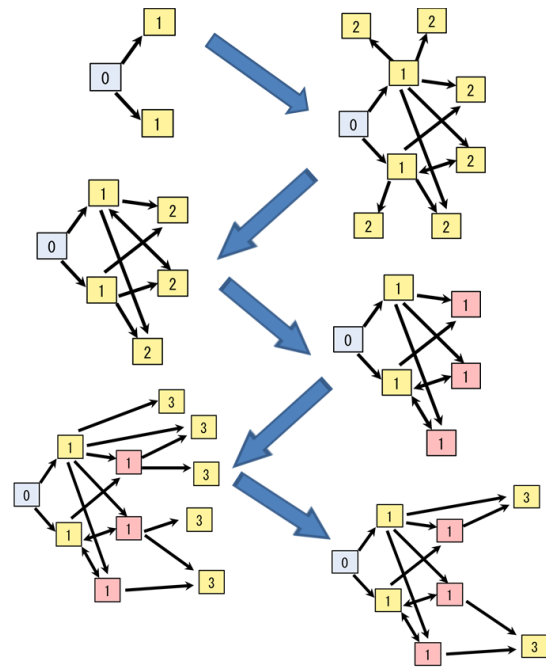


図 6 フォロ 先推薦アルゴリズム

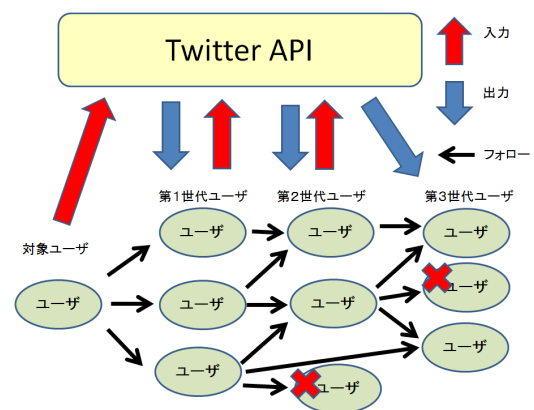


図 7 フォロ 先推薦システムの概要図

対象ユーザから見て属性の伝搬が行われていないであろう第 2 世代ユーザは入力としてしない。これまでに取得した第 2 世代ユーザ、第 3 世代ユーザの中から属性が伝搬していると考えられるユーザを推薦する。ここで推薦されたユーザが実際に属性の伝搬が行われているか評価する。本来ならこの先も属性の伝搬が繰り返され、さらなるユーザが推薦できることが予想できるが、本研究ではこの時点で属性の伝搬を評価するのに十分な数なユーザが取得できると考えたのでここまでを対象とした。評価手法としてはネットワークの可視化を用いた。本研究ではネットワーク可視化ソフトである Cytoscape^(注2)を使用した。

実験に用いるデータ

本研究では Twitter ユーザを対象に属性の伝搬が行われているか評価する。しかし、ユーザの中には bot と呼ばれる機械

(注2): <http://cytoscapeweb.cytoscape.org>

的につぶやきを行うアカウントが存在する．例えば，天気予報やニュースなどを定期的に知らせる bot やキャラクターのセリフをランダムにつぶやくものがある．この bot の中にはフォロー返しと呼ばれる，自分をフォローしてきたユーザを無条件でフォローし返すものがある．このように bot が機械的にフォローした場合，非常に属性の伝搬が見えにくく，本研究の仮説を評価するには適さないと判断したので，bot を本研究の対象外とした．bot は一般的に膨大な量のフォローをされており，フォロー返しをしているものはその分膨大なフォローを行っている．そこで本研究ではユーザのフォロー数の傾向調査を行った．その結果を図 8 に示す．

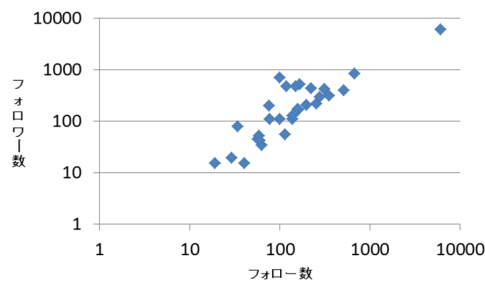


図 8 Twitter ユーザのフォロー数傾向調査

提案手法は，共通の属性を持つユーザ間でのユーザ推薦を目的としている．このため，共通の属性を持つことが明確でない，非常に多くのユーザ からフォローされているユーザは事前に除外する．図 8 は Twitter ユーザ 30 人のフォロー数とフォロー数（フォローされている数）を示したものである．図より Twitter ユーザの分布は 1,000 人以下に多い傾向があることがわかった．よって本研究では，予備調査に元づき，フォロワー数が 1,000 人以上のユーザを除外することとした．またユーザのフォロー関係を Twitter API を通じて取得するため，フォロー関係を非公開に設定しているユーザは含まない．

属性の伝搬元による影響

フォロー先推薦システムを用いて，定義した属性の伝搬の評価実験を行った．属性の伝搬元人数の影響を評価するために入力としてフォロー数 81 人（本研究での対象外ユーザ含む）のユーザ A を使用した．図 9 は対象ユーザと第 1 世代ユーザの関係をネットワーク図で表したものである．

本研究で可視化した図では，ノードがユーザ（黄色いものは対象ユーザ），そのラベルがユーザ名，エッジがユーザ間のフォロー関係を示している．また，図 10 は対象ユーザに第 1 世代ユーザを自分の持つ属性別に分けてもらったものである．

結果として 3 属性に分類することができた．ここで分類できたコミュニティをそれぞれ A，B，C と呼ぶことにする．

次に検出コミュニティの中からコミュニティ B に注目して伝搬元人数が検出ユーザにどのような影響を及ぼすか評価した．図 11 は検出したコミュニティ B のネットワークを可視化したものである．

このコミュニティ内のユーザを元に本研究で提案した属性伝搬アルゴリズムにより，第 3 世代ユーザまでを取得した．伝搬

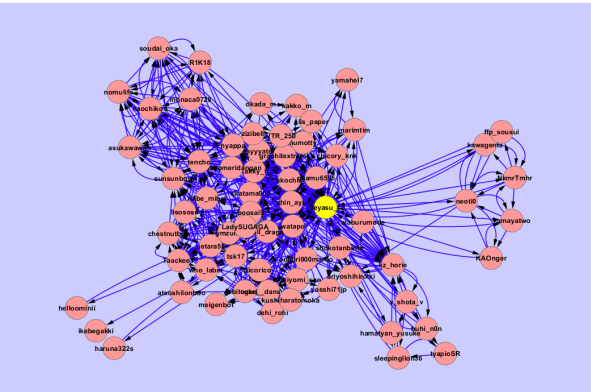


図 9 対象ユーザと第 1 世代ユーザのフォロー関係

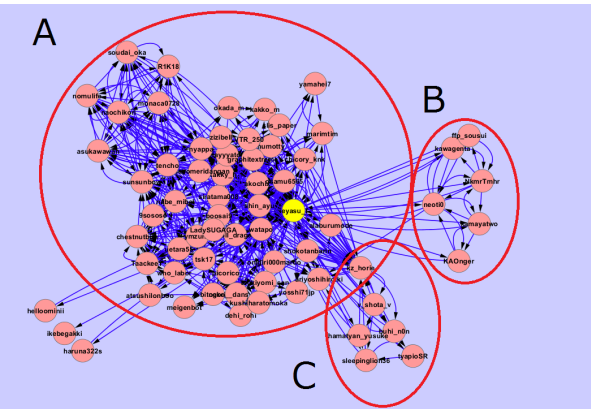


図 10 ネットワーク可視化によって検出されたコミュニティ

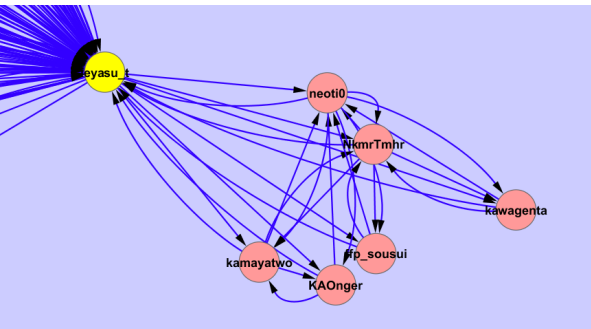


図 11 コミュニティ B の可視化結果

元人数を 2 人，3 人，4 人でユーザ取得を行った結果を表 1 に示す．

表 1 伝搬元の設定人数による影響			
伝搬元のユーザ数	2	3	4
検出したユーザ数	29	7	4
属性伝搬したユーザ数	5	5	4
伝搬確率	17.2%	71.4%	100.0%

伝搬元を 2 人以上を条件に取得したものを図 12 に示す．伝搬元の人数を 2 人にした検出結果では，新たに 29 人のユーザが検出された．このユーザの中で実際にどれだけのユーザに属性が伝搬されているかを評価するために，対象ユーザに属性が伝搬しているユーザを検出ユーザのタイムラインを見て評価し

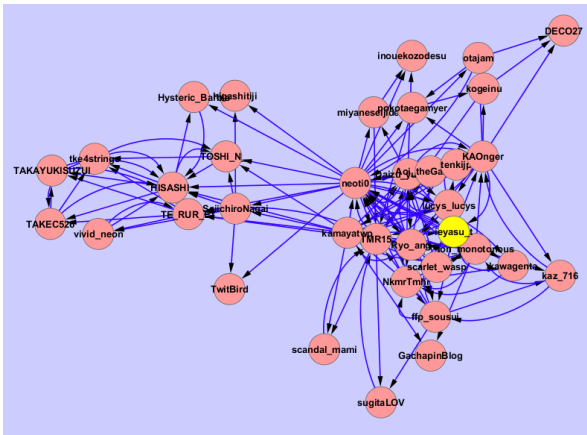


図 12 伝搬元ユーザ 2 人以上における属性伝搬の可視化結果

てもらった。その結果、5 人のユーザに属性が伝搬していることがわかった。

伝搬元ユーザが 2 人以上の時は 29 人中の 5 人に属性が伝搬していることから伝搬確率が 17.2%となる。

伝搬確率が低い原因の 1 つは、有名人アカウントの存在である。有名人アカウントとはメディア露出が多いユーザを意味しており、特徴としてフォローされている数が極端に多い。これは多くのユーザが有名人アカウントを持っているユーザをメディアを通して知っており、フォローするためである。このため、有名人アカウントを持つユーザはフォローしている数より、フォローされている数が極端に多い傾向がある。有名人アカウントは多くのユーザがフォローしているため、コミュニティ内のユーザ数が多い場合、その中の複数ユーザが同一の有名人アカウントをフォローしている場合があり、同コミュニティ内のユーザとして判断されてしまうことがある。実際、不適切ユーザの多くは有名人アカウントであった。

伝搬元を 3 人以上を条件に取得したものを図 13 に、伝搬元を 4 人以上を条件に取得したものを図 14 に示す。

伝搬元のユーザが 3 人以上では検出ユーザは 7 人、その中で属性が伝搬したものは 5 人、よって伝搬確率は 71.4%となった。伝搬元のユーザが 4 人以上では検出ユーザは 4 人、その中で属性が伝搬したものは 4 人、よって伝搬確率は 100.0%となった。

以上の結果から伝搬元人数を多くすることで、より正確な伝搬をする傾向があることがわかった。伝搬元人数が 4 人以上の場合では伝搬確率が 100%となり、正確な伝搬もしていることがわかった。また、伝搬元ユーザを 5 人以上としたところ、本実験に用いたデータではユーザを検出できなかった。

属性の伝搬評価

属性の伝搬が他のユーザでも行われるか実験した。今回は 3 種類のコミュニティを用いて実験を行った。次にその結果を示す。実験に使用したコミュニティをそれぞれ C1、C2、C3 とする。

図 2 はコミュニティ C1、C2、C3 の伝搬確率が最高になった時の値を示している。このことから、伝搬元人数の調整を行えば、7 割近く、またはそれ以上の確率で伝搬を行うことがで

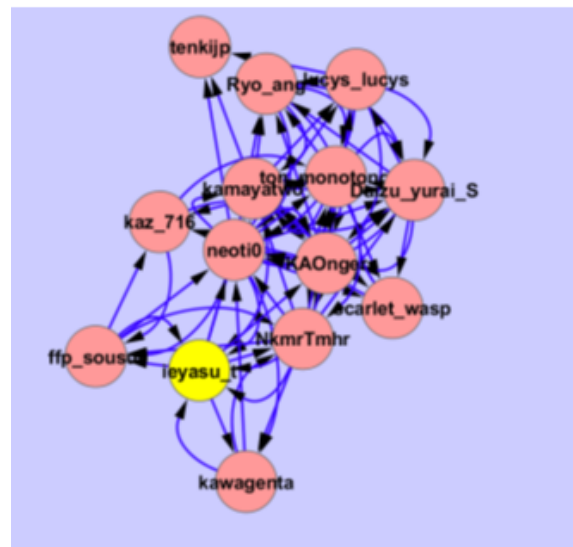


図 13 伝搬元ユーザ 3 人以上における属性伝搬の可視化結果

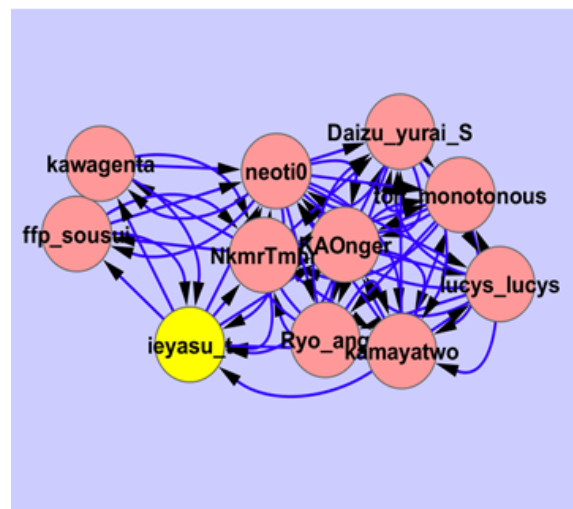


図 14 伝搬元ユーザ 4 人以上における属性伝搬の可視化結果

表 2 コミュニティ別の伝搬確率

コミュニティ	C 1	C 2	C 3
伝搬元のユーザ数	4	3	5
検出したユーザ数	6	4	14
属性伝搬したユーザ数	4	6	11
伝搬確率	66.7%	66.7%	78.6%

ることが分かった。

世代別のユーザ

世代別ユーザがどのように検出されるか傾向を調査するために、ユーザ A のコミュニティ全てをフォロー先推薦アルゴリズムを用いてユーザ検出を行った。表 3 はその結果を示したものである。

この結果より、膨大なユーザが検出できることがわかる。また、第 2 ユーザよりも第 3 ユーザの方がユーザが多くなっていることがわかる。これは伝搬元のコミュニティ内のユーザが第

表 3 ユーザ A に対する属性伝搬モデルを用いた世代別の検出ユーザ数

伝搬元人数	第 1 世代ユーザ	第 2 世代ユーザ	第 3 世代ユーザ
2 人以上	65	625	1,374
3 人以上	65	313	1,049
4 人以上	65	188	884

2 世代ユーザの場合は第 1 世代ユーザ数，であるのに対し，第 3 世代ユーザの場合は，第 2 世代ユーザになっていることが考えられる．表より，第 2 ユーザ数は第 1 ユーザ数をいずれも上回っているためこのような結果になったと推察できる．

5. 考 察

本研究では伝搬元ユーザの人数による検出ユーザへの影響を評価した．実験の結果から，属性伝搬元の人数を多くすることでより正確な伝搬ができることが分かった．しかし，属性伝搬元の人数が多くなるほど検出されるユーザは少なくなる傾向があることもわかった．これは伝搬元のユーザが多いほど，条件にあてはまるユーザが少なくなるため，検出人数が減ってしまうからだと考えられる．伝搬元の人数を増やすと最終的にどのコミュニティも，検出ユーザが 0 人になる．推薦できるユーザが見つからなかった原因も条件に当てはまるユーザがいなかったためであると言える．また，属性が伝搬していないユーザを検出してしまった原因の一つに有名人アカウントを多く検出してしまったことが考えられる．有名人アカウントは幅広いユーザからフォローされるため，属性伝搬モデルの評価実験にはあまり適していないユーザと言える．解決策としては，有名人アカウントに対して多くのユーザは片方向のフォローになるので，コミュニティを検出する際に，双方向フォローをしているユーザにしなければ有名人アカウントの検出を避けることができると考えられる．または伝搬元の人数を多く設定することも有効であると考えられる．しかし，有名人アカウントは，多くのユーザからフォローされているという点を見れば，推薦ユーザとしては十分利用できると思われる．また，有名人アカウントはその特性から現在の流行に大きく影響される傾向があるのではないかと予想される．属性の伝搬の評価では 7 割近く，またはそれ以上の確率で伝搬を行っているという結果になった．これは推薦システムとして有効と言える値たと言える．しかし実験結果からはそれぞれのコミュニティで伝搬確率が最大となる伝搬元の設定人数が異なった．このことから，属性伝搬元の人数設定はそのコミュニティにあった最適なものにする必要があると言える．具体的に属性伝搬元の人数を設定する際に考慮すべき点は以下のものが考えられる．

元々のコミュニティに存在するユーザ数

属性の規模

今回の実験から属性伝搬元の人数が多いほど高い確率で属性が伝搬する傾向があることがわかった．よって，元々のコミュニティ内のユーザ数が多いほど伝搬元の人数を多く設定できる．属性の規模とは，Twitter に存在する共通の属性を持ったユーザの集合を指す．今回の実験から属性ごとに大きさがあるのではないかと予想される．例を挙げると，X 大学という属性と，

Y サークルという属性があるとする．Y サークルは X 大学のサークルだとすると，必然的に Y サークルの属性を持つユーザは X 大学の属性も持っていることになる．つまり，属性の規模は X 大学 > Y サークルとなる．この時，Y サークルというコミュニティのユーザは X 大学の属性も持つため，Y サークルという属性で属性伝搬を行った場合，Y サークルの属性を持たない X 大学属性を持つユーザを検出してしまう場合が想定される．この問題も属性伝搬元の人数を多く設定することで解決できると考えられる．

世代別のユーザ検出調査では，検出ユーザは世代が増すごとに検出ユーザが多くなるという結果が出た．この結果から，属性の伝搬を繰り返すことで十分な検出ユーザ数が得られることが分かった．しかし，今回の実験では検出ユーザが推薦するには多すぎるので，ユーザに重みづけを行い，重みのあるユーザから優先して推薦を行うことでスムーズな推薦を目指す．

6. おわりに

現在，Twitter は新たなコミュニケーションツールとして注目されている．その理由の一つとして，情報の取捨選択を簡便にするフォローが挙げられる．しかし，フォロー先のユーザを発見する手法が乏しく，フォロー機能を生かせていないのが現状である．本研究では，ユーザは同じ属性を持つユーザをフォローすると定義し，属性伝搬モデルを構築することで，従来にない，新しいフォロー先推薦手法を目指した．具体的には，提案した属性伝搬モデルの有効性を検証を行った．行った検証は以下の 3 つである．

- 属性伝搬モデルの伝搬元の人数が属性伝搬に及ぼす影響
- 属性伝搬モデルの推薦システムとしての有効性
- 世代ごとの検出ユーザ数

伝搬元の人数が属性伝搬に及ぼす影響の評価実験では，属性伝搬元の人数を多く設定することで，属性伝搬がより正確に行われる傾向があることが分かった．しかし，検出ユーザ数は属性伝搬基の設定人数と反比例することが分かった．伝搬元の人数を多くするといずれは条件にあてはまるユーザがいなくなり，ユーザを検出できなくなることが分かった．このように属性ごとに適切な伝播基人数を設定する必要があることが分かった．次に同じ実験を 3 コミュニティに対して行ったところ，7 割近い確率で属性が伝搬することが分かった．

今後の課題として，今回は一方的なフォローのみを考慮して属性の伝搬を評価したが，相手からもフォローされる双方向フォローが属性伝搬に与える影響の評価が考えられる．また，提案手法ではユーザの属性を具体的には特定せずにフォロー先を推薦するが，ユーザの投稿内容も加味して，属性の類似度を考慮して推薦するユーザを絞り込むことも有用する，ユーザの重みづけが考えられる．そこで，優先して推薦するユーザを決めるために重みづけが必要になってくる．重みづけの指標としては，他ユーザとのフォロー関係の数が適切であると考えられる．また，ユーザのつぶやきに注目し，自然言語処理を用いて投稿の内容を新たな属性検出の指標とすることが考えられる．また，属性の規模を考慮して，属性伝搬元の人数を設定する必

要があることがわかった．この問題を解決するために，さらに調査を進め，属性の大きさを判定するシステム開発を目指す．

謝辞

本研究の一部は科研費（21500091）の助成を受けたものである．ここに記して謝意を示す．

文 献

- [1] Bernardo A. Huberman, Daniel M Romero, and Fang Wu. Social networks that matter: Twitter under the microscope. *First Monday*, Vol. 14, No. 1 - 5, 2009.
- [2] 吉本和紀, 鈴木優, 吉川正俊. マイクロブログにおける他者への影響を考慮した投稿者の重要度推定手法. *DEIM Forum 2010*, No. C3-4, 2010.
- [3] 田中淳史, 田島敬史. Twitter のツイートに関する分類手法の提案. *DEIM Forum 2010*, No. A5-4, 2010.
- [4] Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng. Why we twitter: understanding microblogging usage and communities. *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, 2007.
- [5] Fang Wu, Bernardo A. Hubermanm, Lada A. Adamic, and Joshua R. Tyle. Information flow in social groups. *Statistical and Theoretical Physics*, Vol. 337, No. 1-2, 2004.
- [6] Cai-Nicolas Ziegler and Georg Lausen. Propagation models for trust and distrust in social networks. *Information Systems Frontiers*, Vol. 7, No. 4/5, 2005.
- [7] Scott A. Golder, Sarita Yardi, Alice Marwick, and danah boyd. A structural approach to contact recommendations in online social networks. *SIGIR*, 2009.