

同一ツイートをリツイートしたユーザ群の情報を用いた ユーザ推薦アルゴリズムの提案

黒柳 智士¹ 鈴木 浩¹ 服部 哲¹ 速水 治夫¹

概要: 近年、ブログの一種であるミニブログあるいはマイクロブログが注目されている。特に Twitter[1] と呼ばれるサービスが注目されている。Twitter とは 140 文字以内で自由につぶやくまたはつぶやき（以下、ツイート）をすることができる短文投稿型の Web サービスである。単につぶやくだけでなく、他のユーザをフォローすることにより、フォローしたユーザのつぶかきがタイムライン上に表示され、それによってリアルタイムなコミュニケーションを行うことができる。そのため、他のユーザをフォローすることは有益なことであり、あるユーザに対し別のユーザを推薦するアルゴリズムの有用性検討、またユーザ推薦研究が盛んであり、ユーザを推薦することは有用である。

本研究では、ユーザの行った公式・非公式リツイート（以下、リツイート）を潜在的な興味であると考えた。しかし潜在的な興味が共通であるとはいえ、実際に趣味が合うかどうかは不明であるため、そのユーザの話題傾向をツイート上から分析し、潜在的かつ共通の話題の二つの指標を持ってあるユーザを推薦するアルゴリズムの提案を行った。

Proposed recommendation algorithm using the information of the user group the user who retweeted the same tweet

SATOSHI KUROYANAGI¹ HIROSHI SUZUKI¹ AKIRA HATTORI¹ HARUO HAYAMI¹

1. はじめに

近年、ブログの一種であるミニブログあるいはマイクロブログが注目されている。これらの利用者は短いテキストを容易に投稿可能であり、手軽に利用ができるなど利点があるなど日々ユーザ数が増加している。特に 2006 年にアメリカでサービスが始まった Twitter[1] が最も爆発的な成長を見せており、こうしたサービスを支援する研究というものも有用ではないかと考えられる。そこで本研究ではユーザ間のつながりであるフォロー関係を支援を目指す。

Twitter とは 2006 年に始まった短文投稿型の Web サービスである。140 文字以内でリアルタイム且つ手軽に自由に呟くことができるため、その時々気分などにより手軽に素早くツイートすることが可能である。また、他のユーザをフォローすることでフォローしたユーザのツイートが

自分のタイムライン上に表示され、それによって会話のようなコミュニケーションをリアルタイムに行うことができる。そのコミュニケーションにより有益な情報を得ることができたり、趣味の発信、何気ない会話からの議論の発展などが広がるなど、単なるひとりごとのツールではなく、コミュニケーションツールとしても重宝されており、近年流行している [2]。

ユーザは他のユーザを自由にフォローすることができるが、実際どのようなユーザをフォローすればよいか迷うことがある。Twitter 公式サイトや各種サービス、各種研究においてユーザの検索、推薦するアルゴリズムはいくつか提案されているが、どのようなユーザを推薦するか、目的の段階で様々なアルゴリズムが存在するが、それらには問題点疑問点があり、詳細は 2 章 3 章にて述べる。

そこで本研究では、ユーザが行ったリツイートに着目した、ユーザがあるツイートをリツイートすることは、リツイートされた元ツイートに潜在的な興味を持っているのではないかと着目した、しかしリツイートだけでは潜在的な

¹ 神奈川工科大学大学院 情報工学専攻
Course of Information and Computer Sciences Graduate
School of Kanagawa Institute of Technology

興味だけでなく、単純に面白半分などで広めたツイートも含まれてしまい、ユーザの興味を探るには不十分である。そのため、同じツイートをリツイートしたユーザの話題傾向を分析することで話題が合うのではないかと予測し、リツイートによる潜在的な興味と話題傾向を用いたアルゴリズムを提案した。

2. 公式サイト・関連サービス

2.1 Twitter 公式サイトにおける問題点

Twitter 公式サイトにおいて、「フォロー情報等にも基づいたあなたへのおすすめユーザー」というものがある。そこでは、図 1 のように被推薦ユーザのフォローユーザがフォローしている第三者のユーザが多くの場合において推薦される。フォロワーのフォロワーなら比較的に親しいと思われるが、例えば被推薦ユーザとそのフォローユーザは同じ学校の人間であるが、フォローユーザと推薦されたユーザは近所の知り合いという場合も考えられ、コミュニティが全く異なる可能性も考えられる。

また、上記の推薦以外に共通のフォロワーが居ないのにも関わらず、推薦されるユーザというものが存在し、どのようなアルゴリズムでユーザを推薦しているか不明な部分が存在している。一部では、別のユーザを閲覧するとその閲覧されたユーザ側に「おすすめユーザー」として提示されることや、ユーザの IP アドレス情報を元に近くにいるユーザを「おすすめユーザー」として提示している [3] など、Twitter 社においてもユーザ推薦への試行錯誤が行われていることが分かる。

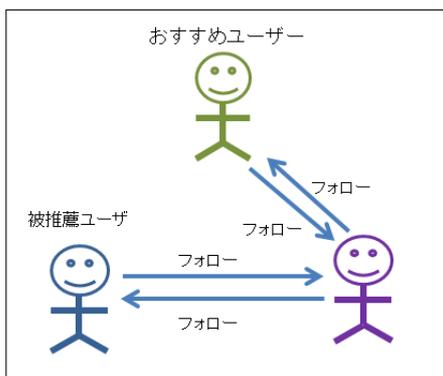


図 1 公式サイトにおけるおすすめユーザー一例

Fig. 1 Simplified diagram algorithm

2.2 Twitter 公式アプリにおけるおすすめユーザー

iPhone や Android に向けて作られた Twitter 専用アプリというものがあるが、それらが携帯内に登録されているメールアドレスを勝手に検索をかけて「おすすめユーザー」として提示している事が現在言われている [4].

2.3 その他ユーザ検索サービス

ユーザを推薦するものではないが、きあつたー、ツイナビ、ツイプロフィールなどのユーザ検索するサービスが存在する。これらは Twitter の基本的な情報以外に、キーワードなどの情報を付加することが出来るものである。フォローしたいユーザを検索するとき、キーワードとなる情報を自分で入力し見つけるというものである。

3. 関連研究

3.1 Twitter におけるリツイート経路を重ね合わせによるユーザ発見支援

太田らの研究では [5], リツイートの伝搬経路を可視化することで、同一の興味を持つユーザの発見支援をした。オーバーラップグラフにおいて、2 回以上登場しているユーザは、被推薦ユーザの知らないところで同じツイートをリツイートしていることからこれらが潜在的なユーザとなり、同じ興味を持っているということを直感的に発見できると述べている。

実験結果から潜在的なユーザというものが同じ興味を持っている可能性が高いことが判断されているが、ユーザのプロフィールなどでそこから確認しなければならないなど、経路を可視化しているだけでは情報が少ない事が報告されている。

3.2 Twitter User Recommender ~Twitter のお気に入り機能を用いたおすすめユーザー推薦システム~

渡部らの研究では [6], 従来の推薦システムはフォロー関係を利用したものが多く、それらは 2 値表現でしかなく、ユーザの評価度合を表現できないと述べている。そこで、Twitter のお気に入り機能について着目した。お気に入りをを用いることでフォロー関係で見えないユーザの評価とし、評価の高いユーザを推薦するものである。

推薦ユーザ提示の流れとして、ユーザが推薦対象とした分野で評価の高いユーザを入力することで、そのユーザのお気に入り関係が取得され、そのユーザ群を推薦対象としてチェックボックスで絞り込むものである。これにより、フォロー関係ではわからない局所的なコミュニティ内で評価の高いユーザあるいは、推薦対象としたい分野に特化したユーザ群を得られていることがあげられる。

3.3 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

北村らの研究では [7], 既存の手法では、ユーザ間の関与を考慮されておらず、趣味・嗜好の異なるユーザが推薦されてしまう可能性があるとして述べ、ユーザ間に存在する「フォロー」、「リプライ」の 2 つを指標とし、それらを組み合わせることで 2-hop 先のユーザを推薦し、どれが一番推薦の度合が高いか評価を行った。結果として、被推薦ユーザからリブ

ライ・フォローの関係が最も推薦の度合いが高いと確認することができた。

いわば、友人の友人という関係を推薦する方式であり、気が合うかどうかや興味の対象が一緒かどうか考慮されおらず、またユーザは比較的探しやすいことがあげられる。

4. 本研究における着目点

本研究では、リツイート情報と話題傾向を用いた推薦アルゴリズムを提案する。それを実現する上での着眼点を述べる。

4.1 ツイート

ツイートとは、ユーザのつぶやきの事を指し、ユーザは140文字以内で自由につぶやくことができる。また文字だけではなく画像や動画も同時に投稿可能である。本研究ではユーザのツイートから話題傾向の分析を行う。

4.2 リツイート

リツイートとは、RTとも呼ばれ、他ユーザのツイートをそのままの形で自分がツイートを行うことができ、これによって情報の拡散を行うことができる。尚、リツイートには公式、非公式リツイートの2種類があるが、本研究では公式リツイートを対象とする。以下図2として公式リツイートの一例を提示する。本研究ではユーザの潜在的な興味がリツイートに隠されているのではないかと考察する。



図2 公式リツイートの一例
Fig. 2 Official Retweet Sample

4.3 ツイートの話題傾向

著者の過去の研究からツイート内にあるキーワード情報単体では類似度が非常に下がってしまう。そこで、はてなキーワードのカテゴリ [8] 分類を利用する。単語レベルからカテゴリ分類を利用することで精度は落ちるが、大まかな傾向を用いる。以下図3として本研究で用いる18のカテゴリを提示する。

一般	動植物	マンガ
読書	テレビ	アイドル
音楽	アニメ	社会
映画	食	地理
ウェブ	スポーツ	アート
コンピュータ	ゲーム	サイエンス

図3 はてなカテゴリ一覧
Fig. 3 Hatena Catecole List

5. 提案アルゴリズム概要

提案アルゴリズムの説明として以下図4の説明を行う。

図4では、実線・一点破線・破線をそれぞれあるツイートのリツイート経路として各ユーザに伝達している。はじめに被推薦ユーザのツイートを取得し、その中からリツイートを取得する。その後、被推薦ユーザがリツイートしているツイートを同様にリツイートしているユーザをAPIにて取得し、ここで取得したユーザを推薦候補ユーザとし、共通のリツイートを行っているユーザを降順に並べ、上位ユーザのツイートを同様に取得する。

そして、その取得したツイートをはてなキーワードカテゴリを組み込んだ形態素解析器にてユーザの話題傾向を分析し、被推薦ユーザに対しても同様に形態素解析をかける。個々の例では取得したユーザのツイート群から、料理・アーティストの2つが話題傾向として高いことがわかるため、このユーザはこれらの話題に興味を持っていることが考えられる。そして、両者のカテゴリにおける類似度を計算し、似ているユーザの推薦を行う。

6. 提案システム構成

6.1 ツイート取得部

はじめに、被推薦ユーザは本システムに OAuth 認証をすることでログインする。その後、被推薦ユーザのツイートを一定数取得する。また、その中からリツイートも別途抽出する。

6.2 推薦候補ユーザツイート取得部

リツイート取得部で取得したリツイートの元ツイートを取得し、そのツイートを同様にリツイートしているユーザ(推薦候補ユーザ)を取得する。

6.3 ツイート話題解析部

推薦候補ユーザのツイート取得部で取得したツイートに対して形態素解析器にかけ、話題傾向を分析を行う。話題傾向は第4.3項で示した通り、形態素解析器の単語辞書にはてなキーワード及び、はてなキーワードカテゴリを組み合わせたものを利用し、カテゴリデータを話題傾向して取得する。また、被推薦ユーザとのカテゴリにおける類似度を計算部も設ける。

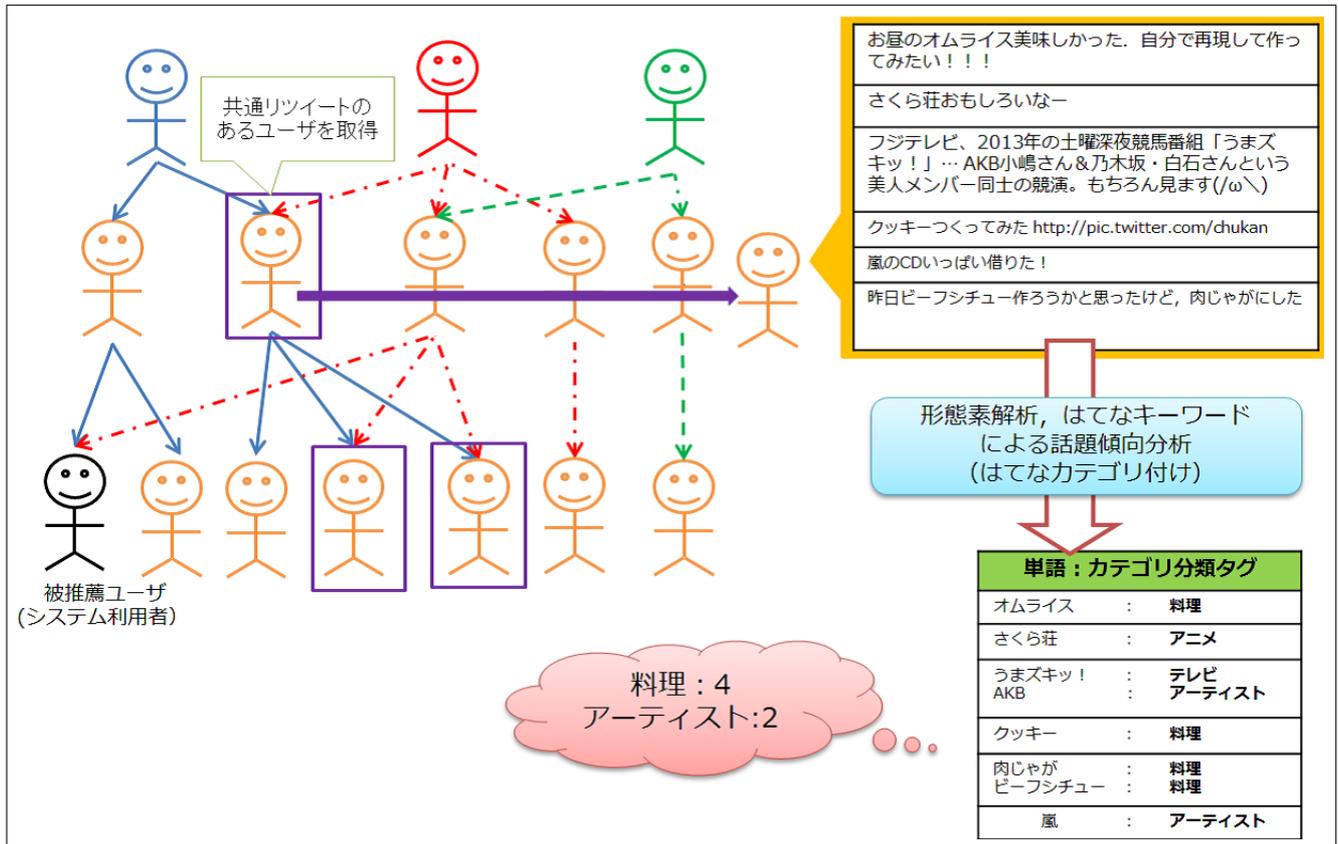


図 4 提案アルゴリズム概要図

Fig. 4 Proposal algorithm schematic diagram

6.4 推薦ユーザ表示部

推薦ユーザ表示部では、被推薦ユーザとの共通リツイート数と被推薦ユーザとの話題の類似度で推薦候補ユーザから推薦ユーザを決定し提示する。

7. 今後の展開

今後の展開としてアルゴリズムの有用性の評価を行う。そのためのシステムの構築を行い、完了しだい順次有用性確認のための実験を行う。

8. おわりに

本研究では、ユーザの推薦アルゴリズムとして、ユーザの行ったリツイートを潜在的な興味の可能性と考え、同様の興味をもつユーザを推薦対象とし、またそれにユーザの話題傾向という分類を用いたユーザの類似度による似ているユーザを推薦するアルゴリズムを提案した。

今後、アルゴリズムの有用性確認のための実験を行い、どのようなユーザに有用性があるか確認し、また推薦されたユーザの傾向についても調査を行う。

参考文献

[1] Twitter, 入手先 (https://www.twitter.com/) (2013.04.18).
 [2] Twitterとは 国内で“再流行”、一般化の兆しも, 入手先

(http://www.itmedia.co.jp/news/articles/0907/28/news011.html) (2013.05.05).

[3] Twitter”おすすめユーザー”とはどのような人物がでてくるか, 入手先 (http://matome.naver.jp/odai/2135979903999287601) (2013.04.25).
 [4] ツイッターが携帯のアドレス帳を抜き取っているという驚愕の事実, 入手先 (http://magazine.gow.asia/life/column_details.php?column_uid=00000371) (2013.04.25).
 [5] 太田侑介, 寺田実, 丸山一貴: Twitterにおけるリツイート経路の重ね合わせによるユーザ発見支援, 情報科学技術フォーラム FIT2011 p.73-76
 [6] 渡部将太, 宮森恒 Twitter User Recommender ~Twitterのお気に入り機能を用いたユーザ推薦システム~, DEIM Forum 2012 B3-4
 [7] 北村太一, 小川祐樹, 諏訪博彦, 太田敏澄: 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, JSAI2012 3E1-R-6-5
 [8] キーワードカテゴリーとは -はてなキーワード-, 入手先 (http://d.hatena.ne.jp/keyword/%A5%AD%A1%BC%A5%EF%A1%BC%A5%C9%A5%AB%A5%C6%A5%B4%A5%EA%A1%BC) (2013.04.28).