

リンク構造解析による Twitter ユーザのランキング手法

山口祐人[†] 高橋翼[‡] 天笠俊之[‡] 北川博之[‡]

[†] 筑波大学第三学群情報学類

[‡] 筑波大学大学院システム情報工学研究科

1 はじめに

Web 上で人々が情報を発信する機会が増加し、様々な人々の意見や主張を閲覧することが可能になった。一方で、我々は多様な情報の取捨選択を強いられている。近年、Twitter[1] というマイクロブログサービスが爆発的に普及し、注目を集めている。非常に多くの Twitter ユーザ (以下ユーザ) が Twitter 上で情報を発信しているため、様々なユーザの意見や主張を閲覧することができる。そのため、有用な情報を発信するユーザを発見できれば、価値ある情報を入手する機会が増える。しかし、ユーザの数は膨大であり、有用なユーザの発見は困難である。本稿では、ユーザ同士の Twitter 上での様々な関係に着目し、リンク構造解析を用いてユーザの有用性を評価、ランキングすることで、有用なユーザの発見の支援を行う。

2 Twitter

Twitter は近年注目を集めているマイクロブログサービスの一つである。マイクロブログサービスとは、ブログの一種であり、非常に短いメッセージを投稿することができるサービスである。また、ユーザ間のリアルタイムなコミュニケーションにも利用される。

Twitter では、140 文字以内のメッセージを投稿することができる。このメッセージを tweet と呼び、tweet を投稿することを post と呼ぶ。Twitter には follow という機能がある。follow とは他のユーザを登録し、そのユーザの tweet を自分のアカウントのホーム画面に表示させる機能である。ユーザは follow により他のユーザが発する情報を得ることができる。一般に、ユーザはある一定の価値を見出したユーザを follow する傾向にある。また、ユーザが形成した文化として Retweet (以下 RT) がある。RT とは、自分の tweet に他の tweet を引用することである。その際、引用元 tweet を post

したユーザの ID も付加される。他の tweet を RT することで、自分を follow しているユーザにも伝えることができる。一般に、ユーザはある一定の価値を見出した tweet を RT する傾向にある。RT はユーザが形成した文化であるが、広く行われるようになったため、Twitter が正式に RT 機能を提供し始めている。

3 関連研究

本研究と密接な関係にある研究として、Weng らによる研究が挙げられる。Weng ら [2] はユーザ間の follow の関係に着目し、ユーザ間の類似度やリンク構造を考慮し、ユーザを評価、ランキングする手法 TwitterRank を提案している。また、TwitterRank と他の手法との比較実験を行っている。本研究は、リンク構造を考慮したランキングを行う点は同じだが、follow に加えて RT に着目している点と、ユーザ間の類似度を考慮しない点で Weng らの研究とは異なる。

4 提案手法

Twitter 上でのユーザ間の様々な関係を、有用性を評価する指標として捉えた、ユーザの評価手法 TURank (Twitter User Rank) を提案する。TURank ではリンク構造を考慮し、単に被 follow 数、被 RT 数だけでなく、follow 元、RT 元のユーザの有用性を加味した評価を行う。PageRank をベースに、エッジの異種性を考慮する ObjectRank[3] の概念を用いたリンク構造解析を、ユーザ間のリンク構造に適用する。

RT は有用な tweet の推薦を意図する引用と、他のユーザとの会話を意図する引用とに区別される。本研究では、前者を有用性を測る指標として捉え、後者をノイズとみなす。ユーザが RT を行うとき、その引用元に自分の tweet が含まれる場合がある。この RT を self-RT と呼ぶ。会話を意図した RT の多くが self-RT であるため、本研究では self-RT を除去する。

4.1 グラフの構築

リンク構造解析の対象とするグラフを構築する。図 1 に、グラフの構造や、エッジの重みを表す Schema graph を示す。Schema graph は、ユーザ、tweet を要素

Twitter User Ranking by Link Structure Analysis

Yuto YAMAGUCHI[†] (yuto_ymgc@kde.cs.tsukuba.ac.jp),

Tsubasa TAKAHASHI[‡] (tsubasa@kde.cs.tsukuba.ac.jp),

Toshiyuki AMAGASA[‡] (amagasa@cs.tsukuba.ac.jp) and

Hiroyuki KITAGAWA[‡] (kitagawa@cs.tsukuba.ac.jp)

[†] College of Information Sciences, University of Tsukuba

[‡] Graduate School of Systems and Information Engineering, University of Tsukuba

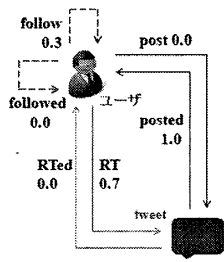


図 1: Schema graph

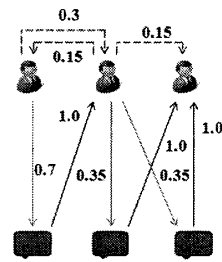


図 2: Data graph

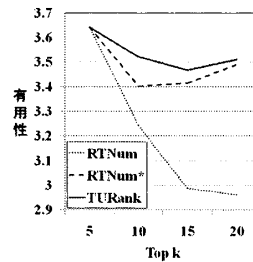


図 3: 有用性評価

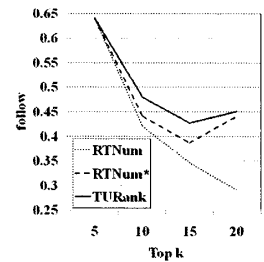


図 4: follow の割合

とするノード集合 V , post, posted, follow, followed, RT, RTed を要素とするエッジ集合を用いる. post はユーザからそのユーザが投稿した tweet へのエッジ, follow はユーザからそのユーザが follow しているユーザへのエッジ, RT はユーザからそのユーザが RT している tweet へのエッジである. また, posted, followed, RTed はそれぞれの逆を表す. 6 種類のエッジに評価値伝搬の重みを与えることで, エッジの異種性を考慮する. 図 2 は, Schema graph に基づく Data graph である. Data graph の各エッジには, エッジの種類ごとに Schema graph で各エッジに設定された重みを, 同種類の発エッジ数で正規化した重みを与える. ただし, 図 2 では重み 0 のエッジを省略している.

4.2 評価値の計算

PageRank におけるリンク構造解析と同様に, Data graph の各ノードの評価値を伝搬させ, ユーザ, tweet の評価値 r を算出する (式 1). 隣接行列 A には, 各エッジの重みが格納される. 式 1 を繰り返し計算し, 全てのユーザ, tweet の評価値を収束させることで, リンク構造を考慮した評価が可能になる. 算出した評価値を降順に並べ, ユーザをランキングする.

$$r = dAr + \frac{(1-d)}{|V|}e \quad (1)$$

5 評価実験

被 RT 数 (RTNum), self-RT を除いた被 RT 数 (RT-Num*), TURank の 3 つのランキングの有効性を比較する. 各手法でランキングを作成する際には 2009 年 12 月 27 日~29 日の三日間に post された tweet のうち, 一回以上 RT され, 日本語で記述された 248684tweet と, それらの tweet を post または RT した 68818 ユーザを対象とした.

評価実験では, 各手法の上位 20 ユーザを評価対象とする. 評価対象のユーザの直近 200tweet を 5 人の被験者に提示し, 以下の二項目について評価を依頼した. 一つは各ユーザの有用性の 5 段階評価 (1~5) であり, もう一つは各ユーザを follow したいか否かである.

図 3 は, 各手法の Top k 件のユーザの有用性評価の平均値をプロットした図である. また, 図 4 は各手法の Top k 件のユーザのうち, 被験者が follow したいと思うユーザの割合の平均値をプロットした図である.

提案手法では被 RT 数が比較的少ないユーザの順位の上昇がみられ, 被験者によってそれらのユーザが有用であることが示された. 順位が上昇したユーザは, 上位のユーザに follow, RT されたユーザであった. 本実験の結果により, 有用なユーザは他の有用なユーザを follow, RT する傾向にあると推測される. また, self-RT 除去の有効性が示された. これは, 有用でないユーザ同士の会話が除去されたためであると推測される.

6 まとめ

Twitter 上でのユーザ間の様々な関係にリンク構造解析を適用し, ユーザを評価する手法 TURank を提案した. 今後は RT の利用形態や, その頻度, 確率などの分析を行い, 提案手法の精度向上を目指す.

謝辞

本研究の一部は科学研究費補助金特定領域研究 (# 21013004) による.

参考文献

- [1] <http://twitter.com>
- [2] J. Weng, E. Lim, J. Jiang, Q. He. Twitter-Rank: Finding Topic-sensitive Influential Twitterers. WSDM 2010
- [3] A. Balmin, V. Hristidis, Y. Papakonstantinou. ObjectRank: Authority-Based Keyword Search in Databases. VLDB 2004
- [4] L. Page, S. Brin, R. Motwani, T. Winograd. The PageRank Citation Ranking: Bringing Order to the Web. Technical report, Stanford Digital Library Technologies Project, 1998