

ツイートの閲覧者数事前推定法

下山 雄大† 小林 亜樹†

†工学院大学

1 はじめに

多くのユーザにツイートを届けるためには、ツイートの閲覧されやすい時間帯に投稿することが有効である。そのためには、自身のツイートの閲覧者数を事前に推定することが重要である。このため、観測したフォロワーの行動を基に、自身のツイートを閲覧するフォロワー数の期待値を推定する手法を提案してきた [1]。

しかし、Twitter にはリツイートと呼ばれる機能があり、自身の投稿したツイートがフォロワーにリツイートされ、フォロワー外のユーザに閲覧されることもある。リツイートによるツイートの拡散性は高く [2]、リツイートによる閲覧者数の増加（以後、RT 効果と呼ぶ）を無視することは出来ない。そこで本研究では、RT 効果を考慮したツイートの閲覧者数の期待値を推定する手法を提案する。

2 提案手法

本研究では、各フォロワーのツイート閲覧確率とリツイート率を用いて、ツイートの閲覧者数の期待値を推定する手法を提案する。

本章では 2.2 節までで、既存のツイート閲覧確率の推定手法と、それを用いた閲覧フォロワー数の期待値の推定手法について述べる。ここで、ユーザの行動は 24 時間周期で周期性があると仮定し、過去の行動履歴から、1 時間毎のツイート閲覧確率を推定する。

提案手法の大まかな処理の流れを図 1 に示す。

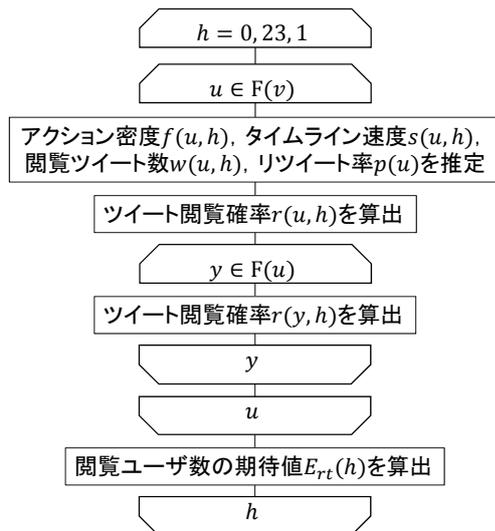


図 1 フローチャート

2.1 ツイート閲覧確率

ツイート閲覧確率は 1 時間あたりにホームタイムライン上を流れるツイートの内、何割のツイートを閲覧したかを表したもので、ある時間帯 h におけるユーザ u のツイート閲覧確率を $r(u, h)$ と表す。Twitter ユーザは閲覧とツイートの 2 つのアクションをし、各アクション毎に一定数のツイートを閲覧するものとモデル化する。1 時間あたりのアクション数をアクション密度 $f(u, h)$ 、アクション 1 回あたりの閲覧数を $w(u, h)$ とすると、1 時間あたりに閲覧するツイート数 $c(u, h)$ は式 (1) で表せる。

$$c(u, h) = w(u, h)f(u, h) \quad (1)$$

ここで 1 時間あたりにホームタイムライン上を流れるツイート数をタイムライン速度 $s(u, h)$ とすると、ツイート閲覧確率 $r(u, h)$ は式 (2) で表せる。

$$r(u, h) = \begin{cases} \frac{c(u, h)}{s(u, h)}, & \text{if } s(u, h) > c(u, h) \\ 1, & \text{if } s(u, h) \leq c(u, h) \end{cases} \quad (2)$$

$c(u, h)$ が $s(u, h)$ より大きいときは、ユーザ u はホームタイムライン上を流れるツイートを全て閲覧していることになるため、 $r(u, h)$ は上限値として 1 をとる。

2.2 閲覧フォロワー数の期待値

閲覧フォロワー数の期待値 $E_f(h)$ は、時間帯 h にツイートを投稿した際に閲覧するフォロワー数を表している。ツイートを投稿するユーザを v 、 v の各フォロワーを u とすると、 $E_f(h)$ は式 (3) で表せる。

$$E_f(h) = \sum_{u \in F(v)} r(u, h) \quad (3)$$

2.3 RT 効果を考慮した閲覧者数の期待値

以上の手法を用いて、RT 効果を考慮した閲覧者数の期待値 $E_{rt}(h)$ を推定する。図 2 にユーザのフォロー関係を示す。ここで u が閲覧したツイートをリツイートした場合、 u のフォロワーである y のタイムラインにも v のツイートが流れる。 u が閲覧したツイートをリツイートする確率を $p(u)$ とすると、閲覧者数の期待値は式 (4) となる。

$$E_{rt}(h) = \sum_{u \in F(v)} \left\{ r(u, h) \left(1 + p(u) \sum_{y \in F(u)} r(y, h) \right) \right\} \quad (4)$$

Prior estimation method of tweet viewers
†Yudai Shimoyama †Aki Kobayashi
†Kogakuin University

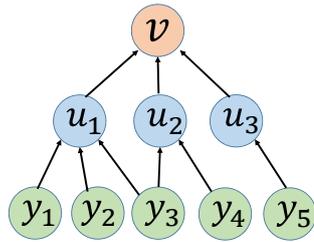


図2 フォロー関係

3 パラメータ推定

3.1 アクション密度の推定

アクション密度 $f(u, h)$ は、1時間当たりの閲覧アクションとツイートアクションの回数である。しかし、TwitterAPI ではツイートアクションの回数は取得出来ても、閲覧アクションの回数は取得することが出来ない。したがって本手法では、1時間当たりのツイートアクション数を推定アクション密度 $\hat{f}(u, h)$ として推定する。

3.2 タイムライン速度の推定

タイムライン速度 $s(u, h)$ を求めるには、ユーザ u のホームタイムラインを取得する必要がある。しかし第三者は本人の承認なしに他ユーザのホームタイムラインを取得することは出来ないため、推定する必要がある。 u のホームタイムラインは、 u のツイートと u のフレンドのツイートで構成されている。したがって各ユーザのユーザタイムラインをそれぞれ取得し、それらを合わせることで u の推定ホームタイムラインが得られる。時間帯 h における推定タイムライン速度を $\hat{s}(u, h)$ と表す。

3.3 閲覧ツイート数の推定

閲覧ツイート数 $w(u, h)$ はアクション1回当たりに閲覧するツイート数を表したものである。しかしツイートを何件閲覧したかという情報はAPIでは取得することが出来ないため、閲覧したツイートに対しての返信の形をとるリプライを利用して推定する。リプライ先のツイートの閲覧は、もっとも手間の少ない自身のホームタイムラインを遡って閲覧したものと仮定する。

すると、リプライの投稿時点において、ユーザは少なくともリプライ先ツイートまで自身のホームタイムラインを遡って閲覧したと推定できる。そこで、リプライ投稿時点のツイートを遡る件数を最低閲覧ツイート数 θ とする。

θ は少なくとも閲覧したツイート数であり、本来はより多くのツイートを閲覧したと予想される。そのため、式(5)のように、過去 n 件中 i 件目のリプライから求めた θ_i の最大値を推定閲覧ツイート数 $\hat{w}(u, h)$ とする。 n については推定の精度を上げるため、20件以上が妥当だと考えられるが、最適な数値の決定は本稿では扱わない。

$$\hat{w}(u, h) = \max \theta_i, 1 \leq i \leq n \quad (5)$$

4 実験

実験用 Twitter アカウントを用意し、協力者である大学生6人にフォローしてもらった。協力者の Twitter 上の行動履歴を20日間観測し、観測したデータより提案手法にて各パラメータの推定を行った。同時に、実験用アカウントにて適当な時間帯にツイートを投稿し、返信機能を用いてフォロワーのうち何人のユーザが閲覧するかを観測した。

図3に結果を示す。閲覧フォロワー数の期待値と実測値は、全般として近い値を出していることがわかる。しかし、6時においては実測値が期待値を大きく上回っている。協力者の中でもこの傾向が顕著なユーザにインタビューした結果、6時は閲覧のみのアクションを多く行っている時間帯であることがわかった。本手法におけるツイート閲覧確率は、ツイートアクションのみから計算しているため、閲覧確率が実際の行動より低い値となってしまった。

また、RTによる閲覧者数の増加は、各フォロワーのフォロワー数やリツイート率が影響するため、時間帯によって増加率は異なっている。本実験条件においては、0時がリツイートによる拡散性が最も高い時間帯であった。RT効果を考慮した閲覧者数については、フォロワー外のユーザ、すなわち実験協力者以外の閲覧行動を観測する方法がないため、評価を行うことが出来なかった。

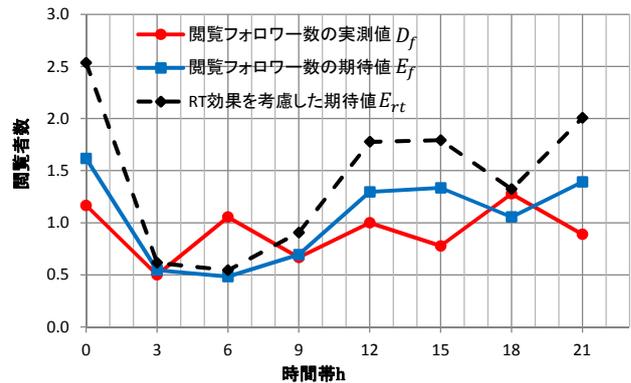


図3 時間帯毎の閲覧者数

5 おわりに

各フォロワーのツイート閲覧確率とリツイート率を用いて、ツイートの閲覧者数の期待値を推定する手法を提案した。RT効果を考慮した閲覧者数の評価や、モデル化とパラメータの推定による誤差の改善が今後の課題である。

参考文献

- 下山 雄大, 小林 亜樹, “フォロワーの目に留まるツイート投稿タイミング推薦手法”, 情報処理学会研究報告データベースシステム (DBS), 2013-DBS-158(17), 1-8, 2013.
- 高橋 美和, 久木 章江, “即時性・拡散性に着目した Twitter 利用者による東日本大震災時の情報伝達に関する研究: 市民の防災力向上に向けて その46”, 日本建築学会学術講演梗概集, 2012(都市計画), 801-802, 2012.
- 村上 雄亮, 小林 亜樹, “目に留まるツイート投稿タイミング推薦手法”, DEIM Forum 2013, P1-4, 2013.