

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

下山 雄大^{1,a)} 小林 亜樹¹

概要：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刻に投稿することが有効である。このため、特定のユーザについてツイート行動を観測し、ツイート閲覧率を推定する手法を提案してきた。しかし、多数のユーザについての閲覧されやすい時刻を統合して推定する手法は未開発であったため、本稿ではその手法について提案する。本手法では、各ユーザの時間帯毎に観測した行動を基に、それぞれの時間帯でのツイート閲覧確率を求め、これらの総和としてツイートを閲覧するユーザ数の期待値を推定する。提案手法の効果を実験によって検証する。

キーワード：Twitter, ツイートタイミング, 閲覧確率

1. はじめに

Twitter^{*1}を代表とするマイクロブログサービスの普及に伴い、誰もが簡単にWeb上で情報を発信できるようになった。情報の内容も多岐に渡り、ニュースや広告のように他人に読んでもらうことを目的としたものもあれば、単なる自身の行動や感情などの雑記のようなものもあり、日々膨大な記事が投稿されている。

Twitterでは、企業の製品広告や、イベントの告知などの広告配信的に利用する人を対象に、プロモツイートという公式の広告ツイートサービスが提供されている。プロモツイートは、有料である代わりにユーザの目に留まるような工夫がされている。具体的には、ツイートの投稿時刻に関わらず、タイムラインと呼ばれるツイートが時系列順に表示されるリストの最新の場所に表示されるようになっていく。また、プロモツイートの内容に興味関心のありそうなユーザにのみ、タイムラインに表示されるようになっていく。

しかし、プロモツイートをしないユーザも広告配信的にツイートをを行うことがある。その際問題になるのがツイートを見逃されることである。プロモツイートではない通常のツイートは、タイムラインにおいて時系列順に表示されるため、古いツイートはある程度遡って閲覧するしかなく、他のツイートに埋もれ見逃されてしまう可能性がある。

このため筆者らは、特定のユーザの行動を観測し、時間帯毎にツイート閲覧確率を推定することで、目に留まるツイート投稿タイミングを推薦する手法を提案してきた [1]。しかし、多数のユーザについての閲覧確率を統合して推定する手法は未開発であった。

そこで本研究は、特定の1ユーザではなく、出来るだけ多くのユーザにツイートを閲覧してもらうことを目的とする。具体的には、各フォロワーのツイート閲覧確率を時間帯毎に推定することで、最も多くのフォロワーの目に留まるツイート投稿タイミングを推薦する手法を提案する。

多数のツイートに埋もれがちな情報から必要な情報を抽出するという観点からは、いくつかの研究が報告されている。竹村ら [2] は、ツイートの価値が時間に依存することに注目し、ツイートを「今見るべき」、「あとでも構わない」、「もう見なくてよい」の3つに分類して表示することで、価値のある記事の見逃しを防止している。古賀ら [3] は、ツイートに含まれる日付表現から未来のイベント情報を抽出し、ユーザに対して再通知するシステムを作成している。山村ら [4] は、フレンド毎にフィルターを設けて、フレンドのツイートの中でもユーザの選好に合うツイートのみをタイムラインに表示する手法を提案している。田中ら [5] は、不快なユーザのツイートを一時的に見づらくすることで、Twitter独自のつながりを保持したままタイムライン閲覧時の不快感を軽減している。

これらの手法はツイートに含まれる情報から多数のツイートを分類し、ツイートを見やすくする、見逃しを防止する手法であり、ツイートを閲覧する側のユーザが利用するものである。これに対し本研究は、ツイートを投稿する

¹ 工学院大学
Kogakuin University

^{a)} c509052@ns.kogakuin.ac.jp

^{*1} <https://twitter.com/>

側のユーザがツイートを見て欲しい時に使用するものであり、これらの手法とは利用するユーザと目的が異なる。

松尾ら [6] は、ユーザの興味は Twitter 上の話題に関連すると仮定し、ツイート内容から興味を抽出し、それにあった広告ツイートを投稿することで閲覧数の最大化を試みるシステムを提案した。

広告ツイートを投稿するユーザが利用する点、より多くのユーザに閲覧されることを目的とする点は本研究と同じであるが、松尾らの手法が広告ツイートの内容を推薦するのにに対し、本研究はツイートを投稿するタイミングを推薦する点異なる。

また、情報通知タイミングの観点では、林ら [7] が、ユーザの携帯電話の利用状況と使用率のアンケートから、ユーザが情報を見やすい通知タイミングの考察を行っている。林らは、「電車やバスを待つとき」のような実際のユーザの行動をもとにタイミングを考察しているのに対し、本研究では Twitter 上でのユーザの行動からタイミングを考察している。

2. 提案手法

本研究では、各フォロワーの時間帯毎のツイート閲覧確率を用いて、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する。

本章ではまず、村上らの提案したツイート閲覧確率の推定手法について述べる。ツイート閲覧確率とは、どのタイミングで、どの程度ツイートを閲覧しているかを数値として表したものである。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は次に示す 6 段階である。

- (1) Twitter ユーザ u のモデル化を行う。
- (2) ユーザ u のアクション密度 f 、タイムライン速度 s 、閲覧ツイート数 w を推定する。
- (3) ユーザ u のツイート閲覧確率 r を算出する。
- (4) 1~3 を各ユーザ、各時間帯ごとに行う。
- (5) 閲覧ユーザ数の期待値 E を算出する。
- (6) E から適切なツイート投稿タイミングを推薦する。

2.1 Twitter

Twitter は、ツイートと呼ばれる上限 140 字の短文を投稿できるサービスである。一般的に長文の記事を投稿する従来のブログサービスと比べて、手軽に投稿できるという特徴がある。

2.1.1 フォロワー

Twitter は他のユーザをフォローすることで、そのユーザのツイートを自分のタイムラインに表示させることができる。図 1 のように、ユーザ u がフォローしている他のユーザのことを u のフレンドといい、 u のフレンド集合を

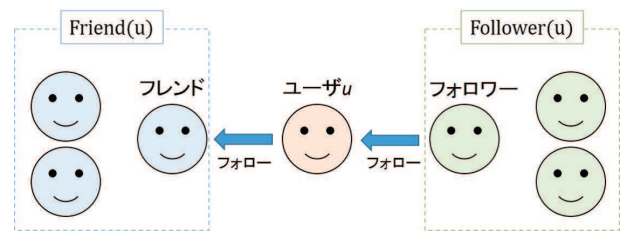


図 1 フォロー関係

$\text{Friend}(u)$ と表す。逆に、 u のことをフォローしている他のユーザのことを u のフォロワーといい、ユーザ u のフォロワー集合を $\text{Follower}(u)$ と表す。

2.1.2 ツイート

ツイートとは、上限 140 字の短文の記事のことである。ツイートにはそれぞれ固有の id が割り当てられており、 id を割り当てられたツイートを t_{id} と表す。ツイートの種類は以下の 3 つに分けることができる。

- 通常ツイート
- リツイート
- リプライ

通常ツイートは投稿欄に書いた文章をそのまま投稿する通常のツイートである。リツイートは他のユーザが投稿したツイートをもう一度自分が投稿し直すツイートのことである。リプライはあるツイートに対して返信を行うツイートのことである。

2.1.3 タイムライン

タイムラインとは、ツイートを時系列順に並べて表示したリストのことである。最新のツイートが一番上に表示され、古いツイートは下に流れていく。タイムラインの種類は主に以下の 3 つに分けることができる。

- パブリックタイムライン
- ユーザタイムライン
- ホームタイムライン

ツイートの項で記した 3 種のツイートの集合をそれぞれ通常ツイート T 、リツイート R 、リプライ M とする。パブリックタイムラインは非公開設定を除く全ユーザの T 、 R で構成される。ユーザタイムラインは自身の T 、 R 、 M で構成される。ホームタイムラインは自身の T 、 R 、 M と、フレンドの T 、 R で構成される。基本的にフレンドの M は含まれないが、自身に対する M 、他のフレンドに対する M のみ含まれる。

ユーザ u を例に説明する。 u のユーザタイムライン $UT(u)$ は式 (1) で表せる。ここで $\text{user}(t_{id})$ は t_{id} を投稿したユーザを返す関数である。

$$UT(u) = \{t_{id} \mid \text{user}(t_{id}) = u\} \quad (1)$$

また、 u のホームタイムライン $HT(u)$ は式 (2) で表せる。ここで $\text{isReply}(t_{id})$ は t_{id} がリプライのときに true を、リプライ以外のときに false を返す関数である。また、re-

$\text{ply_to_user}(t_{id})$ は t_{id} の返信先のツイートを投稿したユーザを返す関数である。

$$HT(u) = UT(u) + \{t_{id} \mid \text{user}(t_{id}) \in \text{Friend}(u), t_{id} \notin X\} \quad (2)$$

$$X = \{t_{id} \mid \text{isReply}(t_{id}) = \text{true}, \text{reply_to_user}(t_{id}) \notin \{u \cup \text{Friend}(u)\}\}$$

2.2 ユーザのモデル化

ユーザの Twitter 上での行動をモデル化する。ユーザの行動はタイムラインの閲覧とツイートの投稿の2つに分類することができ、これらをユーザのアクションと呼ぶ。1時間あたりのアクションの回数をアクション密度 f といい、ある時間帯 h におけるユーザ u のアクション密度は $f(u, h)$ と表す。

アクション密度 $f(u, h)$ は、閲覧のアクション $f_v(u, h)$ と、閲覧+ツイートのアクション $f_t(u, h)$ の2つに分けることができ、 $f(u, h)$ は式 (3) で表される。

$$f(u, h) = f_v(u, h) + f_t(u, h) \quad (3)$$

ここで時間帯 h は1時間間隔とし、0~23の整数をとるものとする。また時間帯内においては一定間隔でアクションをすると仮定する。

ユーザは、各アクション時にその時点のホームタイムラインの最新ツイートから一定件数を遡って読むものとする。この遡って閲覧するツイート件数を閲覧ツイート数 $w(u, h)$ と呼ぶ。

$w(u, h)$ は使用するデバイスやアプリケーションに依存するためユーザによって大きく異なることが予想されるが、村上ら [1] の手法では簡単のため $w(u, h)$ は20件と固定していた。本手法では、閲覧したツイートに対して返信を行うリプライの特徴を利用して、閲覧ツイート数 $w(u, h)$ を推定する。詳細は3.4節で述べる。

したがって、本手法で扱うユーザ u は $f_v(u, h)$, $f_t(u, h)$, $w(u, h)$ の3つのパラメータを持つとモデル化する。

2.3 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は $r(u, h)$ と表す。

本手法のモデル化では、ユーザ u は1時間に $f(u, h)$ 回のアクションをし、アクション1回につき $w(u, h)$ 件のツイートを閲覧するとしている。したがって1時間あたりに閲覧するツイート数 $c(u, h)$ は、 $f(u, h)$ と $w(u, h)$ の積となり、式 (4) で求めることができる。ただしこの1時間で閲覧するツイート数 $c(u, h)$ は、同一のツイートを複数回閲覧した際に、複数件のツイートを閲覧したと数えている

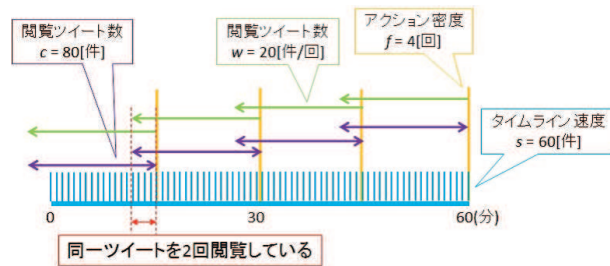


図2 1時間で閲覧するツイート数 $c(u, h)$

(図2参照)。

$$c(u, h) = w(u, h)f(u, h) \quad (4)$$

また、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 といい、ある時間帯 h におけるユーザ u のタイムライン速度は $s(u, h)$ と表す。

$c(u, h)$ と $s(u, h)$ から閲覧確率 $r(u, h)$ は式 (5) で表せる。 $c(u, h)$ 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。

$$r(u, h) = \begin{cases} \frac{c(u, h)}{s(u, h)} & (s(u, h) > c(u, h)) \\ 1 & (\text{otherwise}) \end{cases} \quad (5)$$

2.4 閲覧ユーザ数の期待値

閲覧ユーザ数の期待値 $E(h)$ は、時間帯 h にツイートを投稿した際に閲覧してくれるユーザ数を表している。ツイートを投稿するユーザを v とすると、 $E(h)$ は式 (6) より求められる。ここで u は v の各フォロワーを表している。

$$E(h) = \sum_{u \in \text{Follower}(v)} r(u, h) \quad (6)$$

この閲覧ユーザ数の期待値 $E(h)$ の最も高い時間帯 h を式 (7) より求め、適切なツイート投稿タイミング δ として推薦する。

$$\delta = \arg \max_h E(h) \quad (7)$$

3. パラメータ推定

本手法で用いるアクション密度 $f(u, h)$ 、ホームタイムライン速度 $s(u, h)$ 、閲覧ツイート数 $w(u, h)$ は、モデル化やAPIの制限により直接観測することはできず、推定を行う必要がある。そこで本章では各パラメータの推定方法について述べる。

3.1 TwitterAPI

TwitterはAPIを利用してタイムラインの取得やツイートの投稿が可能である。しかし特定のアカウントに関する操作や情報の取得には、そのアカウントのユーザの許可が必要となる。図3のように、第三者が取得できる情報に

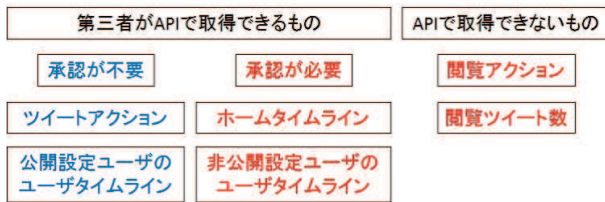


図 3 API で取得できるもの

は制限がある。ホームタイムライン、非公開設定ユーザのユーザタイムラインの取得には、それぞれのアカウントのユーザの許可が必要となる。一方で公開設定ユーザのユーザタイムラインは許可なく取得することができる。また、Twitter を閲覧しただけのような情報は API を用いて取得することはできない。

3.2 アクション密度の推定

アクション密度 $f(u, h)$ は 1 時間当たりのアクションの回数を表したものであり、本手法のモデル化におけるアクションとはタイムラインの閲覧とツイートの投稿の 2 つである。したがって $f(u, h)$ 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がわかれば計算できる。

しかし図 3 の通り、API で取得可能な情報はツイートアクションのみであり、タイムラインを閲覧したという情報は取得することができない。そのためアクション密度は推定する必要がある。

本手法のモデル化では、ツイートを投稿するアクションの際にもタイムラインを閲覧したとしている。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 $f_t(u, h)$ を推定アクション密度 $\hat{f}(u, h)$ として推定する。

ツイートアクション数 $f_t(u, h)$ は、ユーザ u のユーザタイムライン $UT(u)$ を取得し、投稿時刻が時間帯 h のツイートのみを取り出すことで得られる。式 (8) に示す。ここで $\text{hour}(t_{id})$ は t_{id} を投稿した時間帯を返す関数である。

$$\begin{aligned} \hat{f}(u, h) &= f_t(u, h) = UT(u, h) \\ UT(u, h) &\equiv \{t_{id} \mid t_{id} \in UT(u), \text{hour}(t_{id}) = h\} \end{aligned} \quad (8)$$

3.3 タイムライン速度の推定

タイムライン速度 s は 1 時間当たりにホームタイムライン上を流れるツイート数を表したものである。ある時間帯 h におけるユーザ u のタイムライン速度 $s(u, h)$ は、ユーザ u の普段閲覧しているホームタイムラインを取得し、投稿時刻が時間帯 h のツイートのみを取り出すことで得られる。

しかし図 3 の通り、ユーザ u のホームタイムラインを取得するには u の承認が必要であり、第 3 者である本システム利用者は取得することが出来ない。そのためユーザ u のホームタイムラインを推定する必要がある。

2.1 節で述べた通り、ホームタイムラインはユーザタイ

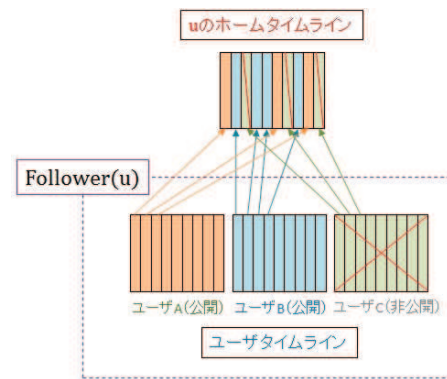


図 4 推定ホームタイムライン

ムラインの集合である。したがって図 4 のように、 u と $\text{Friend}(u)$ 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。式 (9) に示す。

$$\begin{aligned} \hat{s}(u, h) &= UT(u, h) + \{t_{id} \mid \text{user}(t_{id}) \in \text{Friend}(u), \\ &\quad t_{id} \notin X, \quad (9) \\ &\quad \text{hour}(t_{id}) = h\} \end{aligned}$$

$$X = \{t_{id} \mid \text{isReply}(t_{id}) = \text{true},$$

$$\text{reply_to_user}(t_{id}) \notin \{u \cup \text{Friend}(u)\}\}$$

また図 3 の通り、ツイートを非公開に設定しているユーザのユーザタイムラインは取得することができないため、本来のホームタイムラインとは誤差が生じる (図 4 のユーザ C 参照)。

3.4 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。しかし図 3 の通り、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できないため、他の方法で推定する必要がある。

本手法ではリプライの特徴を用いて推定する手法を導入する。リプライとは、あるツイート (以下、元ツイート) に対して関連づけられ、元ツイートの投稿者を宛先とするメンションツイートである。Twitter は、リプライをツイートする機能を提供しており、単なるメンションツイートとは区別される。リプライは関連づけられる元ツイートが必要であるため、通常はこの元ツイートを閲覧した後に、そのツイートへの返信といったインタフェースをなされて投稿されると考えられる。したがって、このリプライの投稿者は、元ツイートを閲覧したと推定できる。

元ツイートをどのような形で閲覧するかには、複数の方法が存在する。例えば、元ツイートユーザのユーザタイムラインを遡って閲覧するなどである。しかし、多くの Twitter クライアントでは、ホームタイムライン表示を基本としているといえるため、もっとも手間の少ないリプライユーザが自身のホームタイムラインを遡って閲覧したも

ユーザ u のホームタイムライン

ユーザ	ツイート内容
bbb	\(^o^)/
u	@ccおはよー
aaaa	バクバクモグモグ
bbb	資料作成終わらない!
cc	起きました
aaaa	朝ごはんを食べよう
u	どろりっちなう
aaaa	今日もいい天気だ

図 5 閲覧ツイート数の推定

のと仮定する。

すると、このリプライの投稿時点において、リプライユーザは少なくともリプライ元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ホームタイムライン上で、最新ツイートから元ツイートまで遡るツイート件数を、その時点での閲覧件数として推定する手法を提案する。

具体例を示す。図 5 はユーザ u のホームタイムラインを表している。ここで上から 2 つ目のユーザ u のリプライに注目する。このリプライは、 u のホームタイムライン上で 3 件前のユーザ cc のツイートに対しての返信である。したがって、 u はこのリプライを投稿する際に、少なくとも cc のツイートまでは閲覧したと推定できる。

この閲覧したツイート数を最低閲覧ツイート数 $\theta(u, t_{id})$ といい、式 (10) より求める。ここで $\text{reply_to}(t_{id})$ は t_{id} のリプライ元ツイートを表す関数である。また、 $\text{line}(u, t_{id})$ は、 t_{id} が u のホームタイムライン上で最新ツイートから数えて何件目に表示されているかを表す関数である。したがって、この場合の u の最低閲覧ツイート数 $\theta(u, t_{id})$ は 3 件だと推定できる。

$$\theta(u, t_{id}) = \text{line}(u, \text{reply_to}(t_{id})) - \text{line}(u, t_{id}) \quad (10)$$

上記手法による最低閲覧ツイート数 $\theta(u, t_{id})$ 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 (11) のように、過去 n 件のリプライからそれぞれ $\theta(u, t_{id})$ を求め、それらの最大値を推定閲覧ツイート数 $\hat{w}(u, h)$ として利用するのが妥当だと考えられる。 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。

$$\hat{w}(u, h) = \max_{t_{id} \in M(u)} \theta(u, t_{id}) \quad (11)$$

$$M(u) = \{t_{id} \mid \text{user}(t_{id}) = u, \text{isReply}(t_{id}) = \text{true}\}$$

しかし前述の通り、元ツイートを閲覧するには複数の方法が存在する。ユーザタイムラインやリスト、キーワード検索などのタイムラインを利用した場合、最新ツイートか

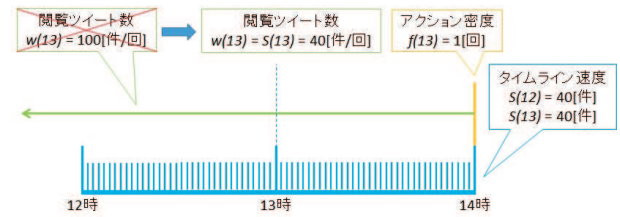


図 6 閲覧ツイート数の上限

ら時系列順に表示されていても、数時間から数日前のツイートがタイムラインの上位に出現する可能性がある。これらを利用した場合、 u のホームタイムライン上では 1000 件以上前に表示されたツイートに対しても容易にリプライを投稿することが出来る。

したがって、最低閲覧ツイート数 $\theta(u, t_{id})$ が極端に大きい値となった場合は、ホームタイムライン以外のタイムラインを利用してリプライを投稿したと判断し、推定閲覧ツイート数 $\hat{w}(u, h)$ を計算する際には一種の外れ値として除外する必要がある。

本研究では、1 時間幅での時間帯毎に閲覧確率を推定し、多数の利用者による期待値を算出する方針としている。これは、Twitter のようなリアルタイムに近いメディアの特性を活かし、情報を新鮮なうちに多くの人目に触れるような状況を現出することを目的としているためである。そこで、本研究における閲覧は、ツイート後おおむね 1 時間の範囲内における閲覧を指すこととする。

したがって本手法では、図 6 のように、時間帯 h におけるユーザ u の推定閲覧ツイート数 $\hat{w}(u, h)$ は、 u のホームタイムラインの 1 時間分、つまりタイムライン速度 $s(u, h)$ を上限とする。最低閲覧ツイート数 $\theta(u, t_{id})$ が $s(u, h)$ より大きい場合は、外れ値として計算から除外する。以上より推定閲覧ツイート数 $\hat{w}(u, h)$ は式 (12) となる。計算上、時間帯を跨ぐ部分において誤差が生じるが、推定値は観測値の最大値を使用することで十分吸収できると考える。

$$\hat{w}(u, h) = \max_{t_{id} \in M(u)} \theta(u, t_{id}) \quad (12)$$

$$\tilde{M}(u) = \{t_{id} \mid t_{id} \in M(u), \theta(u, t_{id}) \leq s(u, h)\}$$

3.5 推定元データの期間

各パラメータを推定するためには過去数週間分の情報から計算することが望ましい。この期間に関しては、生活リズムに大きな変化を及ぼす可能性のある数ヶ月単位ではなく、あまり変化の起こらないであろう 2 週間程度が妥当だと考えられる。しかし、期間に関する最適な数値の決定は本稿では扱わない。

4. 実験

2.4 節で述べた閲覧ユーザ数の期待値 $E(h)$ は、時間帯 h に投稿したツイートを閲覧するユーザ数を推定している。

表 1 協力者の基本データ

ユーザ	フレンド数	フォロワー数	実験期間中の 1日の平均ツイート数
A	147	211	25.6
B	70	5	0.3
C	773	1235	21.5
D	62	50	28.0
E	331	281	23.2
F	31	26	20.8

表 2 ツイート時刻と件数

時間帯	ツイート件数
0	9
3	8
6	9
9	9
12	9
15	9
18	9
21	9

本章では、これが実際に成り立っているかを実験を通じて評価する。

4.1 評価方法

評価するには、実際に実験用アカウントにて実験ツイートを投稿し、実験用アカウントのフォロワーのうち、実験ツイートを閲覧したユーザ数を観測すればよい。しかし 3.1 節で述べた通りユーザの閲覧アクションは取得することができず、実験ツイートを閲覧したユーザ数を直接観測することはできない。

そこで本研究では、フォロワーが実験ツイートを閲覧した際には、その実験ツイートに対してリプライを投稿してもらうこととする。このリプライの有無によってフォロワーが実験ツイートを閲覧したかどうかを観測することができる。したがって、実験ツイートに対してリプライを投稿したユーザ数が、実験ツイートを閲覧したユーザ数を表すこととなる。

この実験によって求めたユーザ数と、提案手法によって推定したユーザ数の期待値 $E(h)$ を比べ、 $E(h)$ は実際のユーザのアクションとどのような関係があるかを評価する。

4.2 実験方法

実験用 Twitter アカウントを用意し、実験協力者にフォローしてもらった。協力者は普段から Twitter を利用している大学生 6 人である。表 1 に協力者のアカウントの基本情報を示す。

実験用アカウントにて適当な時間帯に実験ツイートを投稿した。協力者が実験ツイートを閲覧した際には、その実験ツイートに対してリプライを投稿してもらった。実験は 20 日間行った。表 2 に実験ツイートを投稿した時間帯と件数を示す。実験ツイートを投稿する時刻の分と秒については毎回ランダムとした。

収集したデータより各パラメータの推定を行った。推定元データは実験期間である 20 日間分とした。閲覧ツイート数 $\hat{w}(u, h)$ の推定において、実験期間中のリプライ数が 20 件未満のユーザについては十分な推定が出来ないため、 $\hat{w}(u, h)$ は 20 件と仮定した。

式 (13) より実際にツイートを閲覧したユーザ数 $D(h)$ を求め、本手法によって計算された期待値 $E(h)$ と比べ、本

表 3 時間、ユーザ別推定アクション密度 $\hat{f}(u, h)$

時間帯	A	B	C	D	E	F
0	4.4	0.0	0.1	1.6	2.2	1.1
3	0.2	0.0	0.6	0.6	0.9	0.6
6	0.1	0.0	0.5	0.1	0.2	0.7
9	0.6	0.0	1.7	0.6	1.6	0.3
12	1.9	0.0	1.2	2.6	0.7	1.0
15	1.1	0.0	1.8	1.8	1.8	1.5
18	1.1	0.1	1.7	0.5	0.4	1.4
21	2.6	0.0	0.5	1.7	0.8	1.1

表 4 時間、ユーザ別推定タイムライン速度 $\hat{s}(u, h)$

時間帯	A	B	C	D	E	F
0	66	22	1298	76	404	48
3	55	4	312	56	152	39
6	40	5	177	49	88	38
9	47	11	455	59	159	40
12	66	31	746	69	256	44
15	56	28	656	67	230	43
18	65	28	810	72	306	44
21	67	17	1154	75	367	45

手法の有効性について評価した。

$$D(h) = \frac{\beta(h)}{\alpha(h)} \quad (13)$$

$\alpha(h)$: 時間帯 h に投稿した実験ツイート数

$\beta(h)$: 時間帯 h に投稿した実験ツイート
に対する総リプライ数

4.3 閲覧状況評価

実験のデータより推定した各パラメータを表 3、表 4、表 5、図 7、図 8、図 9 に示す。

ユーザ C とユーザ F に注目する。図 7 より、 C と F の推定アクション密度 $\hat{f}(C, h)$ 、 $\hat{f}(F, h)$ は近い値をとっていることがわかる。一方、 C の推定タイムライン速度 $\hat{s}(C, h)$ が 100 件から 1200 件と非常に高い値をとるのに対し、 F の推定タイムライン速度 $\hat{s}(F, h)$ はどの時間帯も 40 件前後となっている。その結果、 C の推定ツイート閲覧確率 $\hat{r}(C, h)$ は、全ての時間帯において 0.1 以下という低い値となった。対して、 F の推定ツイート閲覧確率 $\hat{r}(F, h)$ は、

表 5 時間, ユーザ別推定ツイート閲覧確率 $\hat{r}(u, h)$

時間帯	A	B	C	D	E	F
0	1.00	0.00	0.00	0.15	0.06	0.40
3	0.06	0.00	0.05	0.08	0.07	0.28
6	0.06	0.00	0.06	0.01	0.03	0.32
9	0.28	0.00	0.09	0.08	0.12	0.13
12	0.57	0.00	0.04	0.26	0.03	0.39
15	0.38	0.00	0.07	0.19	0.10	0.60
18	0.33	0.08	0.05	0.05	0.02	0.53
21	0.77	0.00	0.01	0.16	0.03	0.42

表 6 各時間帯における閲覧ユーザ数の期待値 $E(h)$ と実測値 $D(h)$

時間帯	期待値 $E(h)$	実測値 $D(h)$
0	1.62	1.17
3	0.54	0.50
6	0.48	1.06
9	0.69	0.67
12	1.30	1.00
15	1.33	0.78
18	1.05	1.28
21	1.39	0.89

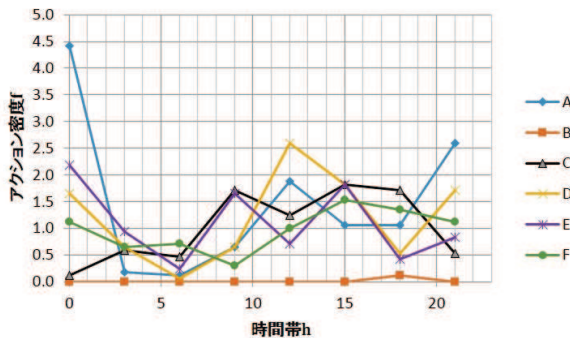


図 7 推定アクション密度 $\hat{f}(u, h)$

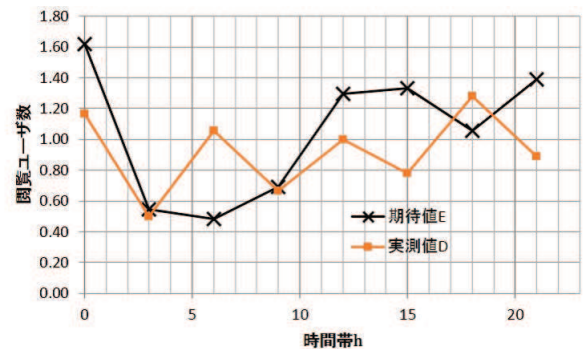


図 10 閲覧ユーザ数の期待値 $E(h)$ と実測値 $D(h)$ の関係

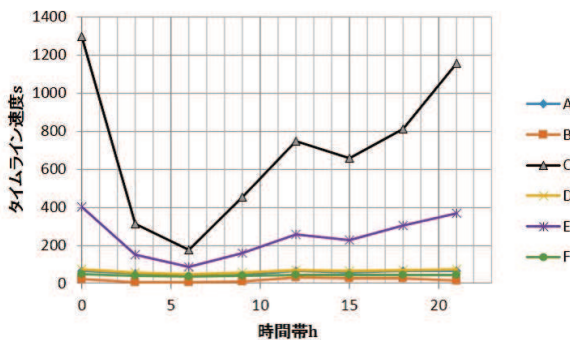


図 8 推定タイムライン速度 $\hat{s}(u, h)$

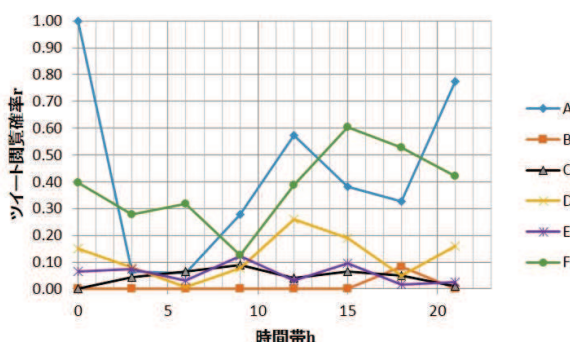


図 9 推定ツイート閲覧確率 $\hat{r}(u, h)$

0.13 から 0.60 をとり, 全ての時間帯において C より高い値となった.

このように, ツイート閲覧確率はタイムライン速度に影響されるため, アクション密度が高い, つまりツイート数

が多いからといって, ツイート閲覧確率も高い値をとることは限らない.

また, 実験のデータより求めた各時間帯における閲覧ユーザ数の期待値 $E(h)$ と実測値 $D(h)$ を表 6, 図 10 に示す.

図 10 において, 黒い線は提案手法により推定した期待値 $E(h)$ を, 橙の線は実験により求めた実測値 $D(h)$ を表している. スケールが合っており, 全般として近い値を出していることがわかる. 個々の点に注目すると, 3時と9時において青い点と赤い点が重なっており, 閲覧ユーザ数の期待値と, 実際にツイートを閲覧したユーザ数がほぼ等しかったことがわかる.

6時においては実測値が期待値を大きく上回っている. これは, 普段はあまりツイートを投稿していないが, 実験ツイートへはリプライを投稿してくれたとすると説明がつく. すなわち, ツイートを投稿せず, ツイートの閲覧のみを行っている時間帯であった場合である. 例えば, 通勤電車の中で携帯端末を利用してツイートの閲覧のみを行うことが日常であったが, 本実験では実験ツイートに対してリプライを投稿してくれたというような実験バイアスの可能性が示唆される. これは検証が必要である. この理由については不明であり, 今後の課題である.

5. おわりに

各フォロワーのツイート閲覧確率を用いて, 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

た。実験により、期待値と実測値が全般として近い値をとることが確認できたが、個々の値については大きく違う値をとる時間帯もあった。本実験では実験ツイート数が少ないため、リプライ1件あたりの期待値の変化量が大きいことも原因だと考えられる。より大規模な実験を行うことが今後の課題である。

謝辞

本研究の一部は JSPS 科研費 23240110 の助成を受けたものです。

参考文献

- [1] 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).
- [2] 竹村 光, 田島 敬史: 価値の時間依存性に基づくマイクロブログ記事の分類, DEIM Forum 2012 F2-6 (2012).
- [3] 古賀 健士郎, 庄司 功樹, 小林 亜樹: タイミングを考慮した将来イベント tweet の再通知, 情報処理学会 第 74 回全国大会講演論文集, 2012(1), 711-713 (2012).
- [4] 山村 悟, 佐藤 哲司: フォロー別フィルタによるツイートフィルタリングの提案, DEIM Forum 2013 B1-1 (2013).
- [5] 田中 優美, 伊藤 久洋: Twitter におけるタイムライン閲覧時の不快感軽減の試み, 情報処理学会 第 73 回全国大会講演論文集, 2011(1), 167-169 (2011).
- [6] 松尾 潤, 川村 秀憲, 鈴木 恵二: Twitter のコメント分析による広告の推薦, 情報処理学会 研究報告知能システム (ICS), 2011-ICS-162(4), 1-6 (2011).
- [7] 林 智天, 川原 圭博, 田村 大, 森川 博之, 青山 友紀: コンテキスト適応型コンテンツ配信における情報通知タイミングに関する検討, 電子情報通信学会総合大会講演論文集, 2005 年 通信 (1), 727 (2005).