

# マイクロブログにおけるユーザ専門性を考慮した Web ページ推薦

山口 祐人<sup>†</sup> 天笠 俊之<sup>††,†††</sup> 北川 博之<sup>††</sup>

<sup>†</sup> 筑波大学大学院システム情報工学研究科 〒305-8573 つくば市天王台 1-1-1

<sup>††</sup> 筑波大学システム情報系 〒305-8573 つくば市天王台 1-1-1

<sup>†††</sup> 宇宙航空研究開発機構 宇宙科学研究所 宇宙科学情報解析研究系

E-mail: <sup>†</sup>yuto\_ymgc@kde.cs.tsukuba.ac.jp, <sup>††</sup>{amagasa,kitagawa}@cs.tsukuba.ac.jp

あらまし マイクロブログ上では、興味深い Web ページなどを他のユーザと共有するために、メッセージに外部の Web ページの URL を含めることが多い。それらの Web ページは、生成されて間もないことが多いことが知られており、マイクロブログからは鮮度の高い Web ページを得られる可能性が高い。一方、マイクロブログユーザは自らの興味や嗜好に基づいた情報発信を行う。そのため、それぞれのユーザがどのような分野に詳しいかという専門性は、ユーザの重要な特性の一つである。本研究では、ある Web ページがどのような専門性を持つマイクロブログユーザからどれだけ共有されたかという指標を用いて、Web ページのトピックや重要度を推定し、マイクロブログユーザに対して Web ページを推薦する手法を提案する。

キーワード マイクロブログ, Twitter, 情報推薦, ユーザモデリング

Yuto YAMAGUCHI<sup>†</sup>, Toshiyuki AMAGASA<sup>††,†††</sup>, and Hiroyuki KITAGAWA<sup>††</sup>

<sup>†</sup> Graduate School of Systems and Information Engineering, University of Tsukuba  
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan

<sup>††</sup> Faculty of Engineering, Information and Systems University of Tsukuba  
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan

<sup>†††</sup> Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency

E-mail: <sup>†</sup>yuto\_ymgc@kde.cs.tsukuba.ac.jp, <sup>††</sup>{amagasa,kitagawa}@cs.tsukuba.ac.jp

## 1. はじめに

近年、マイクロブログと呼ばれるソーシャルメディアが急激に普及している。マイクロブログは、ユーザ同士が短いメッセージを投稿、共有するサービスである。投稿できるメッセージの長さが制限されているという性質から、ユーザは気軽にメッセージを投稿する。そのため、有益な情報からそうでないものまで、多種多様な情報がリアルタイムに発信されている。したがって、マイクロブログは有益な情報をリアルタイムに獲得できる新しい情報源として注目を集めている。

マイクロブログの中でも、Twitter<sup>(注1)</sup>と呼ばれるサービスが特に普及している。Twitter は約 1 億ものアクティブユーザを獲得しており<sup>(注2)</sup>、世界最大のマイクロブログとなっている。また、データを公開しており、だれでも自由に利用することがで

きることから、本研究では Twitter を対象とする。

Twitter におけるメッセージはツイートと呼ばれ、ツイートを投稿することをポストという。Twitter ユーザはツイートの外部の Web ページの URL を含め、他のユーザと共有することが多い。これは、自らが投稿したブログなどの外部ページの URL をツイートに含めることで他のユーザに知らせたり、閲覧して面白い、参考になると感じた Web ページを他のユーザと共有したりすることによる。予備調査によると、ポストされたツイートのうち約 14% は URL を含んでいることが明らかになった。共有された Web ページは、生成されて間もないページであることが多いと報告されている [5]。本稿では、生成されて間もない Web ページを鮮度の高いページと呼ぶ。したがって、Twitter からは鮮度の高い Web ページを得ることができると考えられる。しかし、Twitter 上で共有される Web ページは膨大であり、そのトピックも多岐にわたっているため、その中から自らが得たいものを得るのは困難である。

そこで本研究では、膨大な Web ページの中から重要なもののみをユーザの興味に合わせて推薦する手法を提案する。まず、

(注1): <http://twitter.com>

(注2): <http://techcrunch.com/2011/10/17/twitter-is-at-250-million-tweets-per-day/>

ある Web ページがどのような専門性を持つユーザに共有されたかによって、その Web ページの重要度を推定する。ユーザの専門性とは、ユーザが主にどのようなトピックに関しての情報を発信するかという属性である。例えば、スポーツに関する情報をよく発信する多くのユーザがある Web ページを共有していれば、その Web ページは重要であり、かつスポーツに関するものであると考えられる。次に、ユーザの興味を推定し、その興味に見合った Web ページを推薦する。ユーザの興味は、そのユーザがフォローするユーザ群の専門性によって推定する。例えば、スポーツに関して専門性の高いユーザを多くフォローするユーザは、スポーツに興味を持っていると考えられる。

本稿の構成は以下のとおりである。まず、2. 節でマイクロブログについて説明し、3. 節で Web ページの共有に関する予備調査について述べる。4. 節で関連研究を概観し、5. 節で Web ページの推薦手法を提案する。そして、6. 節で提案手法の有効性を評価する実験について述べ、7. 節で本稿をまとめる。

## 2. マイクロブログ

マイクロブログとは、ソーシャルメディアの一つで、短いメッセージを投稿、共有するサービスである。マイクロブログユーザはメッセージを投稿することにより情報を発信し、メッセージを閲覧することにより情報を収集する。また、他のユーザとコミュニケーションを取ることにも可能である。投稿できるメッセージの長さが短いという性質から、マイクロブログからは膨大な情報が日々リアルタイムに発信されている。そのため、マイクロブログは様々な情報をリアルタイムに獲得できる新しい情報源として注目を集めている。

### 2.1 Twitter

Twitter は、マイクロブログサービスの中でも最も注目を集め、爆発的に普及しているサービスである。Twitter が有するユーザは、数が膨大であるだけでなく、その性質も多種多様である。世界各国からのユーザが存在するのはもちろん、近年では企業や政治家などの著名人、主要なニュースサイトまでもが Twitter の利用を開始し、情報を発信している。このように、Twitter からは様々な情報が発信されているため、多岐に渡るトピックに関する情報を得ることができる。

Twitter では、投稿できるメッセージの文字数が 140 文字以内に制限されている。メッセージをツイートと呼び、ツイートを投稿することをポストと呼ぶ。また、Twitter にはフォローという機能がある。フォローとは他のユーザを登録し、そのユーザのツイートを継続的に受信、閲覧する機能である。あるユーザをフォローしているユーザは、そのユーザのフォロワーであるという。ユーザがポストしたツイートは、そのユーザの全てのフォロワーのタイムライン<sup>(注3)</sup>に即時に表示される。一般に、ユーザは自分が得たい情報を発信するユーザをフォローする。本研究では、ユーザ間のフォロー関係によって構築されるグラ

(注3): タイムラインとは、ユーザ自らのツイートと、そのユーザがフォローする全てのユーザのツイートを時系列順に表示させたものであり、各々のユーザが一つずつ持つ。

フをソーシャルグラフと呼ぶ。

Twitter ユーザはユーザリスト（以下、リストと呼ぶ）を作成し、他のユーザと共有することができる。これは、Twitter が提供している一種のユーザグループ機能である。ユーザはリストを作成することにより、リストに含まれるユーザ（以下、リストメンバと呼ぶ）のツイートをまとめて閲覧することができる。また、リストを作成したユーザだけでなく、他のユーザもそのリストをフォローすることにより、リストメンバのツイートの閲覧が可能である。

リスト作成者は、ある共通の属性を持つユーザ群を一つのリストにまとめ、リストの名前でその属性を表すことが多い。例えば、スポーツに関する情報をよく発信するユーザを一つのリストにまとめ、そのリストに “sports” という名前を付けることがある。こうすることにより、リスト作成者はスポーツに関するツイートをまとめて閲覧できるようになる<sup>(注4)</sup>。よって、リスト名はユーザの専門性を表していると考えられる。先行研究 [14] では、ユーザが含まれるリストの名前を用いてユーザの専門性を推定する手法を提案した。例えば、“politics” や “government” などのリストに多く含まれているユーザは政治に関して専門性が高いと考えられる。

## 3. 予備調査

Twitter における Web ページの共有について調査を行った。まず、Twitter Streaming API の sample メソッドを用いて 2011 年 12 月 13 日 20 時から 14 日 20 時までの 24 時間にポストされたツイートを取得したところ、URL を含むツイートの割合は約 14 % であった。Twitter 上で一日にポストされるツイートの数が約 2.5 億である<sup>(注5)</sup>ことを考えると、マイクロブログ上では日々膨大な量の Web ページが共有されていることが分かる。

また、Web ページの共有頻度の調査も行った。図 1 に共有頻度分布を示す。縦軸、横軸ともに対数スケールである。共有頻度分布はべき分布にしたがっており、限られた数の Web ページのみ頻繁に共有されていることが分かる。[5] によると、ユーザは一部の有益な Web ページを頻繁に共有していると報告されている。

## 4. 関連研究

近年、Twitter に関する研究が多く行われている。ユーザ間のフォロー関係によって構築されるソーシャルグラフに関する研究 [7] [8] や、発信された情報が Twitter 上をどのように伝搬するかを調査する研究 [2] [9] [16]、Twitter ユーザをランキングする研究 [13] [10] [15] など、その内容は様々である。

マイクロブログを用いて、ユーザに対してニュース記事などの Web ページを推薦する研究が多く行われている。Abel ら [1] は、ユーザが過去に共有したニュース記事の内容を用いてユー

(注4): リストに含まれるユーザが必ずしもスポーツに関するツイートをポストするとは限らない。

(注5): <http://techcrunch.com/2011/10/17/twitter-is-at-250-million-tweets-per-day/>

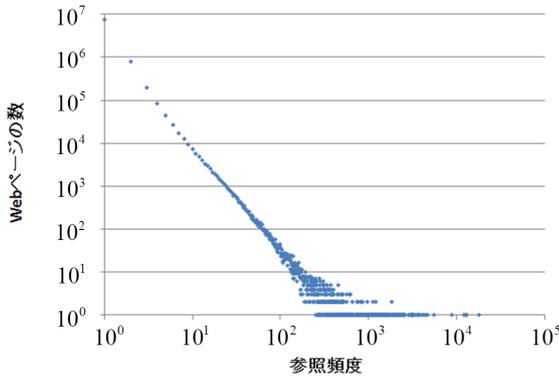


図1 Web ページの共有頻度分布

ザをモデリングし、また別のニュース記事を推薦する手法を提案した。また、ユーザの興味や関心が、時間的要因によって変化すると報告している。例えば、週末とそれ以外の曜日では、ユーザが持つ興味や関心は異なると報告している。

Chen ら [4] は、Twitter ユーザに対して Web ページを推薦する手法を複数提案し、それらを評価実験で比較している。実験によると、ソーシャルグラフ上において自分に近いユーザが参照した Web ページを推薦対象とし、ユーザの興味や関心をそのユーザがポストしたツイートの内容から推定し、Web ページの重要度をそのページを共有したユーザ数によって推定した手法が最も有効であることが示されている。

Phelan ら [11] は、Twitter ユーザに対してニュース記事を推薦する手法を提案した。ユーザのプロファイリングはユーザがポストしたツイートに含まれる単語の TF-IDF を用いて行い、ニュース記事のプロファイリングは記事に含まれる単語の TF-IDF を用いて行っている。

以上の研究は、本研究とは以下の点で異なる。一つは、本研究では Web ページを共有したユーザの専門性を考慮している点である。これにより、共有された Web ページがどのようなトピックに関するものであるかを推定することができる。もう一つは、本研究では Web ページやツイートの内容を用いない点である。Web ページのトピックやユーザの興味は、リストの名前のみを用いて推定している。

また、Web ページの推薦ではないが、Hannon ら [6] は Twitter ユーザに対して、そのユーザが興味を持つトピックの情報をよく発信するユーザを推薦する手法を提案した。評価実験によると、ユーザ間のフォロー関係を用いた協調フィルタリング [12] 的な手法が、ツイートの内容を用いる内容ベースの手法より有効であることが示されている。また、Dong ら [5] は、Twitter 上で共有された Web ページを用いて、鮮度の高い Web ページのランキングを行っている。Twitter を情報源として用いることで、従来の方法で Web ページをクロールするよりも効率よく鮮度の高い Web ページを収集できると報告している。

## 5. 提案手法

Web ページの重要度は、その Web ページを共有したユーザ数によって推定できる。しかし、たとえ多くのユーザが共有し

ていたとしても、その Web ページを共有したユーザが自分と全く異なるトピックに関心を持つ場合は、自分がその Web ページに関心を持つ可能性は低い。そのため、Web ページがどのようなトピックにおいて重要であるかを考慮する必要がある。

この問題に対して、本研究ではある Web ページを共有したユーザの専門性に着目する。あるトピックに関して専門性の高いユーザが共有する Web ページは、そのトピックにおいて重要であるという仮定に基づき、Web ページのトピックごとの重要度を推定する。例えば、スポーツに関して専門性の高い多くのユーザが共有する Web ページは、スポーツに関して重要である可能性が高い。これにより、トピックごとに重要な Web ページを得ることができるため、ユーザの興味に応じた Web ページ推薦が可能になる。

提案手法では、あるユーザがどのような専門性を持つユーザをフォローしているかによってユーザの興味を推定する。また、Web ページがどのような専門性を持つユーザに共有されたかによって、Web ページのトピックごとの重要度を算出する。そして、ユーザの興味と Web ページのトピックの類似度によって推薦するか否かを決定する。

提案手法の手順を以下に示す。

(1) ユーザの興味の算出

(2) ユーザが興味を持つ Web ページを共有する可能性の高い候補ユーザ集合の作成

(3) ユーザに対して Web ページを推薦するか否かを決定する推薦スコアの算出

### 5.1 ユーザの興味の算出

ユーザ  $u$  がフォローしているユーザが属すリストを用いて  $u$  の興味を推定する。例えば、リスト名に “politics” や “government” などの単語が含まれるリストに属すユーザを多くフォローするユーザは、政治に興味を持つという推定を行う。

具体的な算出方法を示す。ユーザを  $u$  とし、リスト名に含まれる単語を  $t$  とする。まず、ユーザが属すリストを用いて、ユーザの専門性を表す Expertise score  $e_{ut}$  を次の式で算出する。

$$e_{ut} = \frac{\sum_{l \in \text{lists}(u)} \delta_{lt}}{\sum_u \sum_{l \in \text{lists}(u)} \delta_{lt}} \quad (1)$$

ここで、 $\text{lists}(u)$  は  $u$  が属すリストの集合を表し、 $\delta_{lt}$  はリスト  $l$  の名前が単語  $t$  を含むとき 1 を取り、それ以外の時 0 を取る。Expertise score  $e_{ut}$  が大きいほど、 $u$  が  $t$  に関して専門性が高いことを意味する。

次に、Expertise score を用いて、ユーザの興味を表す Preference score  $p_{ut}$  を以下の式で算出する。

$$p_{ut} = \frac{idf(t) \sum_{v \in \text{follow}(u)} e_{vt}}{\sum_t idf(t) \sum_{v \in \text{follow}(u)} e_{vt}} \quad (2)$$

ここで、 $\text{follow}(u)$  は  $u$  がフォローしているユーザの集合であり、 $idf(t)$  は単語  $t$  の大局的重みである。 $idf(t)$  は、リスト集合  $L$  と単語  $t$  を含むリストの数  $N_t$  を用いて次のように算出される。

$$idf(t) = \log \frac{|L|}{N_t} \quad (3)$$

$p_{ut}$  が大きいほど,  $u$  が  $t$  に強い興味を持つ<sup>(注6)</sup>.

## 5.2 候補ユーザ集合の作成

あるユーザが興味を持つ Web ページを共有する可能性の高いユーザを抽出し, そのユーザから成る集合を候補ユーザ集合と呼ぶ. 提案手法では, 候補ユーザ集合に属すユーザが共有した Web ページのみを推薦対象とする.

まず, ユーザ  $u$  に対して Preference score  $p_{ut}$  が大きい上位  $k_1$  件のタグを抽出し, 興味タグ集合  $C_u^T$  とする. 興味タグ集合  $C_u^T$  に属すタグ  $t$  はユーザ  $u$  の興味を表している. そして,  $C_u^T$  に属すそれぞれのタグ  $t$  に対して Expertise score  $e_{vt}$  の大きい上位  $k_2$  件のユーザを抽出し, それらの和集合をユーザ  $u$  の候補ユーザ集合  $C_u^U$  とする.

ユーザ  $u$  に対して Preference score  $p_{ut}$  が大きい上位  $k_1$  件のタグのみを対象とすることで, ユーザの興味を反映したタグのみを対象とし, それぞれのタグに対して Expertise score  $e_{vt}$  が大きい上位  $k_2$  のユーザのみを対象とすることで, ユーザ  $u$  が興味を持つタグに対して強い影響力を持つユーザを候補ユーザ集合とすることができる.  $k_1$  及び  $k_2$  の値は実験的に決定する.

## 5.3 推薦スコアの算出

推薦対象ユーザ  $u$  に対して Web ページ  $w$  を推薦するか否かを決定する推薦スコア  $s_{uw}$  を次の式で算出する.

$$s_{uw} = \sum_{v \in C_u^U} \sum_{t \in C_u^T} p_{ut} \cdot e_{vt} \cdot \theta_{vw} \quad (4)$$

$$\theta_{vw} = \begin{cases} 1 & \text{if } v \text{ posts } w \text{ in time window } TW \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

ここで,  $\theta_{vw}$  はユーザ  $v$  が Web ページ  $w$  をある特定の時間窓  $TW$  の間に共有したか否かを表し, 共有していれば 1, していなければ 0 となる. 式によると, ユーザ  $u$  が強い興味を持つタグについて, 影響力の強い多くのユーザが参照した Web ページ  $w$  は推薦スコアが大きくなる. 算出した推薦スコアの大きい上位  $N$  件の Web ページを推薦する.

## 6. 評価実験

本節では, 提案手法の有効性を示すために実施した評価実験について述べる. 6.1 節で本実験に用いたデータセットについて説明し, 6.2 節で実験方法について述べる. そして 6.3 節で結果について考察する.

### 6.1 データセット

本節ではデータセットの構築について説明する. データの取得には Twitter API<sup>(注7)</sup>を用いる.

#### 6.1.1 ユーザ及びリストデータの取得

Preference score や Expertise score を算出するために, ユーザ及びリストデータの収集を行った. まず, データ収集の起点となるシードユーザを何人が取得する. そして, シードユーザを起点とし, ソーシャルグラフ上のフォロー関係をたどって

表 1 データセットの詳細

	size
$ U_L $	501,777
$ L $	1,179,129
$ T_e $	56,819
$ U_M $	3,873,979
$ T_w $	12,766,847
$ W $	8,778,855
$ U_P $	4,508,166
$ U_M \cap U_P $	703,890
$ W_R $	1,826,702

ユーザの幅優先探索を行い, ユーザを取得する. 取得したユーザをユーザ集合  $U_L$  とする. また,  $U_L$  に属するユーザが作成するリストを取得し, リスト集合  $L$  とする.

#### 6.1.2 単語及びリストメンバの抽出

$L$  に属するリストのリスト名から単語を抽出し, 単語集合  $T_e$  とする.  $T_e$  から長さ 1 の単語を除去する. また,  $L$  に属するリストからリストメンバを抽出し, リストメンバ集合  $U_M$  とする.

#### 6.1.3 URL を含むツイートの取得

推薦対象とする Web ページを取得するために, Twitter Streaming API を用いて URL を含むツイートを取得した. ツイートの取得期間は 2011 年 12 月 15 日 14 時から 18 日 14 時までの三日間である. 取得したツイートをツイート集合  $T_w$  とし, それらのツイートに含まれる URL の Web ページを Web ページ集合  $W$  とする. また,  $T_w$  に含まれるツイートをポストしたユーザをユーザ集合  $U_P$  とする.

#### 6.1.4 データセットの詳細

取得したデータセットの詳細を表 1 に示す.  $U_P$  に属するユーザのうち,  $U_L$  に含まれるユーザ, すなわち, Expertise score が計算可能なユーザは 703,890 である. また,  $W$  のうち,  $U_L$  に属すユーザに共有されたものは 1,826,702 であるが, 本実験ではこれを推薦対象の Web ページ集合  $W_R$  とする.

## 6.2 実験方法

本実験では, 被験者 12 名による評価実験を行った. 被験者それぞれに対し, 三つの比較手法によって Web ページを推薦し, 被験者が推薦された Web ページに興味があるか否かを評価した.

### 6.2.1 比較手法

本実験では, 以下の三つの手法を比較する.

- Proposed: 提案手法 (本実験では  $k_1 = k_2 = 30$  と設定する)
- Follow: あるユーザ  $u$  と,  $u$  がフォローするユーザ  $v \in follow(u)$  は類似する興味を持つと報告されている [4]. よって,  $u$  は  $v$  がフォローするユーザ  $z \in follow(v)$  が発する情報にも興味があると考えられる. 本比較手法では, この性質に基づき,  $u$  がフォローするユーザがフォローするユーザ (以下, FF と呼ぶ) が頻繁に参照する Web ページを  $u$  に推薦する. まず,  $u$  の FF を  $FF = \{z \in U \mid \exists v \in follow(v), z \in follow(v)\}$  と定義する. また, FF の中でも, より多くフォローされているユーザの重みを大きくするために, FF に含ま

(注6): 正確には, 単語  $t$  に関するトピックに強い興味を持つ.

(注7): <https://dev.twitter.com>

れるユーザ  $z$  それぞれに対して重み  $trust(z) = \sum_{v \in C_u^U} \phi_{vz}$  を与える．ここで、 $\phi_{vz}$  は  $v$  が  $z$  をフォローしているとき 1 を取り、していないとき 0 を取る．そして、 $u$  に対する Web ページ  $w$  の推薦スコアを  $s_{uw} = \sum_{z \in FF} trust(z) \cdot \theta_{zw}$  とする．

- Popular: マイクロブログ上で頻繁に共有される Web ページは重要であるという性質に基づき、 $u$  に  $w$  を推薦する推薦スコアを  $s_{uw} = \sum_{v \in U} \theta_{vw}$  とする．この場合、 $u$  の候補ユーザ集合は全体のユーザ集合  $U$  である．

Proposed と Follow を比較することにより、リストに基づく手法とフォローに基づく手法のどちらがより正しくユーザの興味を表すことができるかを検証する．また、Popular と比較することにより、単純にマイクロブログ上での共有頻度のみを用いた場合よりどれだけ精度が向上するかを示す．

### 6.2.2 被験者

提案手法における Expertise score は、有名ユーザに対してはうまく算出できると考えられる．これは、有名ユーザは多くのリストに含まれる傾向にあり、Expertise score を算出する情報量が多いためである．したがって、有名ユーザを多くフォローする被験者に対しては、Preference score を上手く算出できる可能性が高い．そこで、本実験では、12 名の被験者を次のように 2 つのグループに分割する．まず、被験者がフォローするユーザのフォロワー数の合計で被験者をソートする．そして、上位 6 名の被験者をグループ 1 とし、下位 6 名の被験者をグループ 2 とする．この分割により、グループ 1 には有名ユーザを多くフォローする被験者が属し、グループ 2 には有名ユーザをあまりフォローしない被験者が属することとなる．なお、有名ユーザを抽出する方法は数多く提案されているが [13] [15] [3] [16]、本実験では、最も単純な指標であるフォロワー数を採用した．

### 6.2.3 評価方法

12 名の被験者の Twitter アカウントに対して、推薦スコアの大きい上位 20 ページ、合計で 60 ページを推薦する．それぞれの手法による推薦結果に重複がある場合は 60 ページより少なくなる．被験者は推薦されたそれぞれのページに対して、興味があるか否かの評価を行う．評価基準は次の 5 段階である．

- (1) 全く興味がない
- (2) 興味がない
- (3) どちらかというに興味がある
- (4) 興味がある
- (5) とても興味がある

被験者による評価結果は Precision@k で表す．被験者が 3 以上の評価をしたページは、その被験者に対して適合したと判定し、Precision@k を算出する．Precision@k はそれぞれの被験者に対して推薦した上位  $k$  件のページの適合率の平均である．図 2 に被験者全体による実験結果を、図 3 にグループ 1 の被験者による実験結果を、図 4 にグループ 2 の被験者による実験結果を示す．

### 6.3 考察

図 2, 3, 4 によると、Proposed, Follow 共に Popular を大きく上回る適合率を示していることがわかる．これは、マイクロブログ全体での参照頻度が大きい Web ページを推薦しても、

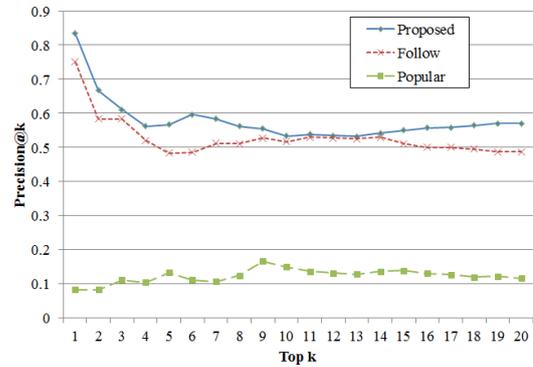


図 2 被験者全体の Precision@k

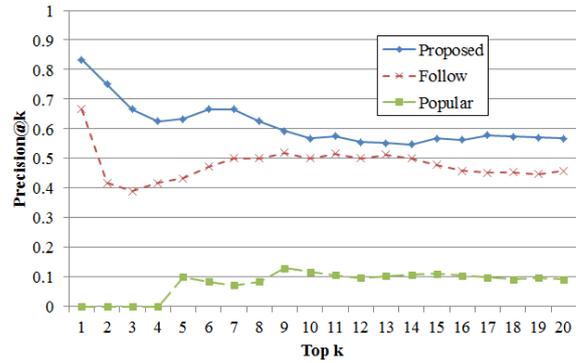


図 3 グループ 1 の被験者の Precision@k

個人の興味を考慮していないため、有効でないことを示している．また図 3 と図 4 を比較すると、グループ 1 においては、Proposed が Follow を上回り、グループ 2 においては Follow が Proposed を上回っている．以下では、Proposed と Follow の結果について考察する．

Follow では、被験者がフォローするユーザが被験者と同じような興味を持つ場合に良い結果となると考えられる．本実験の 12 名の被験者は Twitter においては一般的といえるユーザであり、実際の友人とコミュニケーションを取ることが多い．興味の似ている友人とフォローし合っているため、Follow が比較的良い結果を示したと考えられる．一方で、Proposed は仮説の通り、被験者が有名ユーザを多くフォローする場合には有効であった．これは、有名ユーザは多くのリストに含まれる傾向にあり、専門性が正しく推定されたためであると考えられる．例えば、スポーツに興味のある被験者が、スポーツに関する情報発信を行う有名ユーザを多くフォローしていれば、被験者の興味を上手く推定できる可能性は高い．

以上の考察より、主に Twitter をコミュニケーションを取るために用いているユーザは、興味の似ている友人を多くフォローする傾向があるために Follow の手法が有効に働き、主に Twitter を情報収集に用いているユーザに対しては、興味のある分野における有名ユーザを多くフォローする傾向があるために Proposed が有効に働くと考えられる．

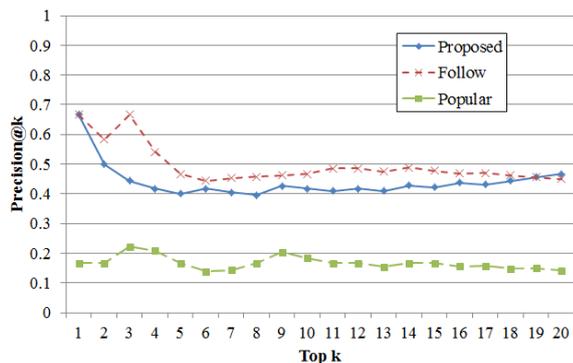


図 4 グループ 2 の被験者の Precision@k

## 7. 結 論

本稿では、マイクロブログ上で共有される Web ページをユーザに推薦する手法を提案した。提案手法はまずユーザがどのようなリストに含まれるかによってユーザの専門性を推定し、推定した専門性を用いてユーザの興味や Web ページの重要度を推定した。実験結果により、Twitter を主に情報収集に利用しており、有名ユーザを多くフォローするようなユーザに対しては提案手法が有効であることが示された。今後は、フォローに基づく Web ページの推薦手法とリストに基づく Web ページの推薦手法がそれぞれどのような場合に有効であるかをより詳細に実験考察し、それらを組み合わせた Web ページ推薦手法の提案をする予定である。

謝辞 本研究の一部は科学研究費補助金・基盤研究 A (#2124005) による。

## 文 献

- [1] Abel, F., Gao, Q., Houben, G.-J. and Tao, K.: Analyzing user modeling on twitter for personalized news recommendations, *Proceedings of the 19th international conference on User modeling, adaption, and personalization, UMAP'11*, Berlin, Heidelberg, Springer-Verlag, pp. 1–12 (2011).
- [2] Boyd, D., Golder, S. and Lotan, G.: Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter, *Hawaii International Conference on System Sciences*, Vol. 0, pp. 1–10 (2010).
- [3] Cha, M., Haddadi, H., Benevenuto, F. and Gummadi, K. P.: Measuring User Influence in Twitter: The Million Follower Fallacy, *In Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM)* (2010).
- [4] Chen, J., Nairn, R., Nelson, L., Bernstein, M. and Chi, E.: Short and tweet: experiments on recommending content from information streams, *Proceedings of the 28th international conference on Human factors in computing systems, CHI '10*, New York, NY, USA, ACM, pp. 1185–1194 (2010).
- [5] Dong, A., Zhang, R., Kolari, P., Bai, J., Diaz, F., Chang, Y., Zheng, Z. and Zha, H.: Time is of the essence: improving recency ranking using Twitter data, *Proceedings of the*

- 19th international conference on World wide web, WWW '10*, New York, NY, USA, ACM, pp. 331–340 (2010).
- [6] Hannon, J., Bennett, M. and Smyth, B.: Recommending twitter users to follow using content and collaborative filtering approaches, *Proceedings of the fourth ACM conference on Recommender systems, RecSys '10*, New York, NY, USA, ACM, pp. 199–206 (2010).
- [7] Huberman, B. A., Romero, D. M. and Wu, F.: Social networks that matter: Twitter under the microscope, *First Monday, Volume 14. Number 1. January 5* (2009).
- [8] Java, A., Song, X., Finn, T. and Tseng, B.: Why We Twitter: Understanding microblogging usage and communities, *Joint 9th WEBKDD and 1st SNA-KDD Workshop, San Jose, CA* (2007).
- [9] Kwak, H., Lee, C., Park, H. and Moon, S.: What is Twitter, a social network or a news media?, *World Wide Web Conference* (2010).
- [10] Leavitt, A., Burchard, E., Fisher, D. and Gilbert, S.: The influentials: New approaches for analyzing influence on twitter, *a publication of the Web Ecology project* (2009).
- [11] Phelan, O., McCarthy, K. and Smyth, B.: Using twitter to recommend real-time topical news, *Proceedings of the third ACM conference on Recommender systems, RecSys '09*, New York, NY, USA, ACM, pp. 385–388 (2009).
- [12] Su, X. and Khoshgoftaar, T. M.: A survey of collaborative filtering techniques, *Adv. in Artif. Intell.*, Vol. 2009, pp. 4:2–4:2 (2009).
- [13] Weng, J., Lim, E., Jiang, J. and He, Q.: TwitterRank: Finding Topic-sensitive Influential Twitterers, *WSDM* (2010).
- [14] Yamaguchi, Y., Amagasa, T. and Kitagawa, H.: Tag-based User Topic Discovery using Twitter Lists, *ASONAM* (2011).
- [15] Yamaguchi, Y., Takahashi, T., Amagasa, T. and Kitagawa, H.: TURank: Twitter User Ranking Based on User-Tweet Graph Analysis, *WISE*, pp. 240–253 (2010).
- [16] Ye, S. and Wu, S. F.: Measuring Message Propagation and Social Influence on Twitter.com, *SocInfo 2010* (2010).