

リストを用いた Twitter ユーザのタグ付け手法

山口 祐人[†] 天笠 俊之^{†,††} 北川 博之^{†,††}

[†] 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

^{††} 筑波大学計算科学研究センター 〒305-8577 つくば市天王台 1-1-1

E-mail: †yuto ymgc@kde.cs.tsukuba.ac.jp, ††{amagasa,kitagawa}@cs.tsukuba.ac.jp

あらまし Twitter 上ではユーザが情報源となり、自らの興味や嗜好に基づいて情報を発信している。Twitter の普及に伴い、膨大な情報の中から有用な情報を発見したいという要求が高まりつつある。有用な情報を発見するには、各ユーザがどのようなトピックに関する情報を発信する傾向にあるかを知ることが重要であると考えられる。そこで本研究では、Twitter の機能であるリストを用いてユーザにタグを付けることによって、そのユーザが発信する情報のトピックを推定する手法を提案する。リストとは誰でも自由に作成し、共有できる“ユーザのリスト”である。リストに含まれるユーザ群は互いに関連を持ち、リストの名前がその関連を表す傾向にある。本稿では、リストの名前を用いてユーザのタグ付けを行う手法を提案し、評価実験によって提案手法の有効性を示す。

キーワード Twitter, トピック分析, タグ付け

A User Tagging Method using Twitter Lists

Yuto YAMAGUCHI[†], Toshiyuki AMAGASA^{†,††}, and Hiroyuki KITAGAWA^{†,††}

[†] Graduate School of Systems and Information Engineering, University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan

^{††} Center for Computational Sciences, University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577, Japan

E-mail: †yuto ymgc@kde.cs.tsukuba.ac.jp, ††{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、ブログと SNS の性質を併せ持つサービスが普及している。マイクロブログの主な特徴として、投稿できるメッセージの長さに制限があることが挙げられる。その制限によってユーザがメッセージを投稿する敷居が下がり、有用なものからそうでないものまで、様々な情報が日々リアルタイムに発信されている。また、マイクロブログは SNS のようにユーザ同士がコミュニケーションを取ることができるという特徴も持っている。

マイクロブログの中でも、Twitter [1] が特に普及している。Twitter 上では、膨大な数のユーザがそれぞれ多様な情報をリアルタイムに発信している。そのため、Twitter は新しい情報源として多くの注目を集めており、Twitter 上で発信される情報の中から有用な情報を得たいという要求が高まりつつある。しかし、発信される情報は膨大であり、トピックは多岐にわたっているため、得たい情報を発見するのは非常に困難である。

Twitter 上では、ユーザが情報源となり、自らの興味や嗜好

に基づいた情報発信を行っている。言い換えると、情報源によって、発信する情報のトピックは異なる傾向にある。そのため、ユーザがどのようなトピックに関する情報発信を行う傾向にあるかが分かれば、有用な情報を発見できる可能性が高い。しかし、ユーザが発信するメッセージは非常に短く、多くのユーザは複数のトピックに関する情報発信を行う。また、ユーザのプロフィールは使われていないことが多く、使われていたとしてもそこから得られるテキストは非常に短い。従って、メッセージやプロフィールの内容からユーザの情報発信の傾向をとらえるのは難しい。

そこで本研究では、Twitter の機能であるリストに着目し、ユーザの情報発信の傾向に関してタグ付けを行う。リストとは誰でも自由に作成、共有できる“ユーザのリスト”である。リストを作成するユーザは、何らかの関連を持つユーザ群を一つのリストにまとめ、リストに付けた名前でその関連を表すことが多い。図 1 はリストの例である。図のリストは@moon.the.cat というユーザが作成した weather という名前のリストであり、図の左下に表示されているユーザ群を含んでいる。従って、こ



図 1 リスト weather の例

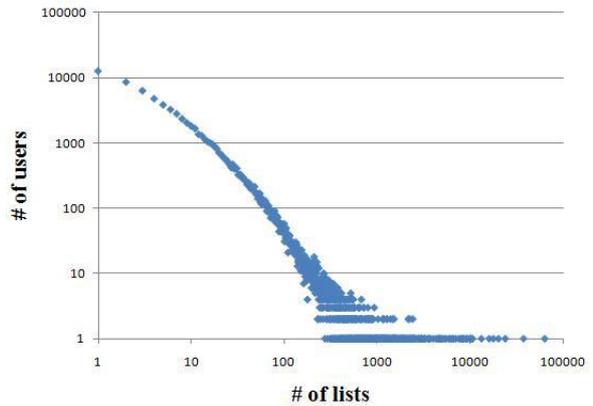


図 2 リストに含まれるユーザの頻度分布

これらのユーザ群は weather というトピックに関連すると考えられ、このリストを作成したユーザに weather というタグを付けられたと見なすことができる。

リストを作成したユーザは、そのリストに含まれるユーザが発信した情報を閲覧することができる。一般に、リスト作成者はあるトピックに関連した有用な情報を発信するユーザをリストに含める傾向にある。つまり、ユーザに付けられたタグは、ユーザがそのトピックに関して有用だという投票であると考えられることができる。

本稿では、リストの名前をタグとみなし、ユーザがどのようなリストに何回含まれるかによってタグ付けを行う手法を提案する。提案手法では、ユーザにタグが付けられた回数だけでなく、ユーザとタグの関連度も考慮する。多くのユーザに付けられているタグはそれぞれのユーザとの関連は小さいが、限られたユーザに付けられているタグはそれぞれのユーザとの関連は大きいと考え、関連度を算出する。

本稿の以降の構成は次のとおりである。まず、2. 節では Twitter についての説明をし、Twitter のリスト機能についての調査結果を示す。3. 節では提案手法の詳細について述べ、4. 節では提案手法の有効性を示すために行った評価実験について述べる。5. 節では本研究に関連する研究について概観し、6. 節で本稿の結論を述べる。

2. Twitter

Twitter [1] 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。2006 年のサービス開始から、2009 年末までに約 7500 万人ものユーザを獲得している [11]。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“購読”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^(注1)に即時に表示される。一般に、ユーザは自分が知りたい情報を発信するユーザや、著名人、知人など、ある一定の価値を見出したユーザをフォローする傾向にある。つまり、多くのユーザに有用であるとみなされたユーザは多くのフォロワーを集める。

2.1 リスト

Twitter ユーザは誰でも自由に“ユーザのリスト”を作成し、他のユーザと共有することができる。これは Twitter が提供している機能であり、非公開にしているリスト以外は誰でも閲覧することができる。1 ユーザ当たり最大 20 リストを作成することができ、1 リストには最大 500 ユーザまで含めることができる。

予備調査によると、103,127 ユーザ^(注2)のうち、リストを一つ以上作成したユーザは 24,887 ユーザであり、作成された全リスト数は 89,618 リストであった。また、一つ以上のリストに含まれているユーザは 78,098 ユーザであった。図 2 はユーザがどれだけのリストに含まれているかを示した図である。横軸はユーザがリストに含まれた数であり、縦軸はそのユーザ数である。この分布はフォロワー数の分布と非常によく似ている [12]。

リストを作成したユーザは、そのリストに含まれるユーザのツイートを閲覧することができる。すなわち、それらのユーザをフォローしていることとほぼ等価である。そのため、多くのフォロワーを集めるユーザは多くのリストに含まれる傾向にある。図 3 はリストに含まれる数とフォローされる数の相関を示した図である。横軸はフォローされる数を示し、縦軸はリストに含まれる数を示している。各点は個々のユーザを表す。図によると、フォロワー数とリストに含まれる数には非常に強い相

(注1): 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、各々のユーザに対して一つずつ存在する。

(注2): 全て日本語圏のユーザ

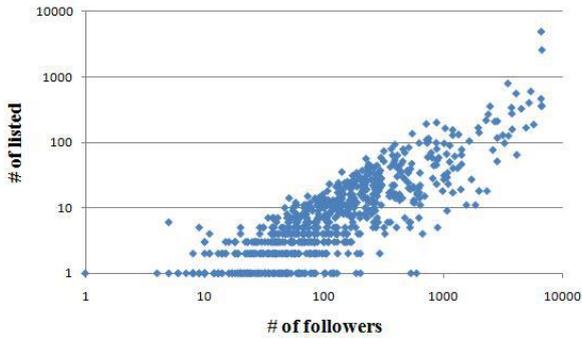


図 3 リストに含まれる数対フォロワー数

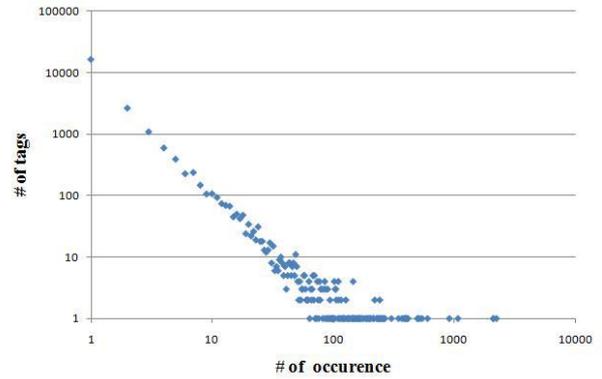


図 4 タグの頻度分布

関があることが分かる．このことから、影響力の強い、有用な情報を発信するユーザは多くのリストに含まれる傾向にあることが分かる．

通常、あるユーザが他のユーザをフォローした時、どのようなトピックの情報を得るためにフォローしたかは分からない．しかし、リストはその名前を見ることで、リスト作成者がどのようなトピックの情報を得るために他のユーザをリストに含めたかを知ることができる．リストに含めることをタグ付けと考え、ユーザにどのようなタグが何回付けられたかによって、そのユーザがどのようなトピックに関して有用であるという評価を得ているかが分かる．

3. 提案手法

本節では、リストを用いたユーザのタグ付け手法を提案する．まずリスト名を単語ごとに分割し、それをタグとして抽出する．次に抽出したタグをそのリストに含まれるユーザへ付加する．最後にユーザとタグとの関連度を算出し、ユーザへのタグ付けを行う．各ステップの詳細を以下の小節で示す．

3.1 リスト名からのタグの抽出

本研究ではリストの名前をタグとみなし、それをリストに含まれるユーザに与える．リストの名前は最大 25 文字であり、半角英数字またはハイフンのみを使用できる^(注3)．多くのリストの名前は複数の単語をハイフンでつないだ形式になっている．例えば、actor-jp という日本の俳優のリストが存在する．そこで、リストの名前から単語を抽出するために、リストの名前をハイフンで区切り、それぞれの単語をタグとした．あるリストの名前からどのようなタグが抽出されたかを表すリストタグ行列 S を定義する．全てのリスト集合を L 、リスト名から抽出した全てのタグ集合を G とすると、リストタグ行列 S は $|L| \times |G|$ 行列となり、次の式で与えられる要素 s_{ij} を含む．

$$s_{ij} = \begin{cases} 1 & \text{if } g_j \in \text{tags}(l_i) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

ここで、 g はタグ、 l はリストを表し、 $\text{tags}(l)$ はリスト l の名前をハイフンで区切って抽出したタグの集合を表す．リスト名に同じタグが二度含まれても S の要素は 1 となる．

表 1 出現頻度の高い上位 10 タグ

| rank | tag | rank | tag |
|------|---------|------|--------|
| 1 | list | 6 | friend |
| 2 | bot | 7 | shop |
| 3 | news | 8 | info |
| 4 | music | 9 | famous |
| 5 | friends | 10 | media |

リストの名前から抽出したタグの頻度分布を図 4 に示す．図に示すように、タグの頻度分布は冪分布に従っており、ユーザが生成するタグは一部に集中する傾向にある．

表 1 に出現頻度の大きい上位 10 タグを示す．リスト名から抽出されたタグはほとんどが名詞であった．タグには次の 4 種類が存在すると考えられる．

- (1) ユーザの情報発信のトピックを表すもの (music, sports)
- (2) ユーザそのものの性質を表すもの (famous, politician)
- (3) リストを作成したユーザにとってのみ意味をなすもの (friend, conversation)
- (4) タグとして意味をなさないもの (list, and)

一つ目、二つ目は本研究において抽出を目指す種類のタグである．三つ目は本研究においてはノイズとなる種類のタグで、リストを作成したユーザの友人や、頻繁に会話をするユーザなどへ付けられる．四つ目も本研究においてノイズとなるもので、リスト名をハイフンで区切ったために生じるタグである．例えば、politicians-list や iphone-and-android といったリストから生じる．一つ目、二つ目の種類のタグは特定のユーザに多くつけられる傾向にあるが、三つ目、四つ目のタグはいろいろなユーザにまんべんなく付けられる傾向にある．

3.2 ユーザへのタグ付け

リスト名から抽出したタグを用いてユーザへのタグ付けを行う．まず、あるユーザがどのリストに含まれるかを表すユーザリスト行列 C を定義する． C は全てのユーザ集合を $|U|$ とすると、 $|U| \times |L|$ 行列として定義される． C の要素 c_{ij} は次の式で与えられる．

$$c_{ij} = \begin{cases} 1 & \text{if } u_i \in \text{lfollow}(l_j) \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

(注3): 全角文字も使用できるが、使用した時の挙動は不安定である．

ここで、 u はユーザを表し、 $l \in follow(l)$ はリスト l が含むユーザの集合を表す。

リストタグ行列 S とユーザリスト行列 C を用いて、ユーザタグ行列 A を作成する (式 3)。

$$A = CS \quad (3)$$

A の各要素 a_{ij} には、ユーザ u_i にタグ g_j が付けられた回数が格納される。

3.3 関連スコアの算出

ユーザに付けられた回数が多いタグをそのユーザのタグとして抽出するだけでは、3.1 節で議論した、リストを作成したユーザにとってのみ意味をなすもの、タグとしての意味をなさないものが抽出されてしまう。例えば、list というタグは頻度が最も大きく (表 1)、多くのユーザに付けられる。しかし、大部分のユーザに付けられるタグは、それぞれのユーザとあまり関連しているとは言えない。そこで提案手法では、それらのタグの重みを下げるためにタグとユーザの自己相互情報量を用いる。ユーザ u とタグ g の自己相互情報量 $PMI(u, g)$ は次の式で与えられる。

$$PMI(u, g) = \frac{n(u, g)N}{n(u)n(g)} \quad (4)$$

$$N = \sum_i \sum_j a_{ij} \quad (5)$$

$$n(u_i) = \sum_j a_{ij} \quad (6)$$

$$n(g_j) = \sum_i a_{ij} \quad (7)$$

$$n(u_i, g_j) = a_{ij} \quad (8)$$

ここで、 N はユーザタグ行列の全ての要素の和、 $n(u)$ は全てのタグがユーザ u に付けられた回数の和、 $n(g)$ は全てのユーザにタグ g が付けられた回数の和、 $n(u, g)$ はユーザ u にタグ g が付けられた回数を表す。 $PMI(u, g)$ は、ユーザ u とタグ g のそれぞれの出現回数に対する u と g の共起回数、つまり u に g が付けられた回数の割合が大きいほど大きな値を示す。

ユーザとタグの関連スコアを、付けられた回数と、自己相互情報量を用いて次のように定義する。

$$score(u, g) = \log n(u, g) \cdot PMI(u, g) \quad (9)$$

タグが付けられた回数は自己相互情報量に対して非常に大きな値になるため、対数を取った。関連スコアが高い上位 k 件のタグをユーザのタグとする。

4. 評価実験

本節では、提案手法の有効性を示すために行った評価実験について述べる。本実験では以下の三つの手法によるタグ付けの精度を比較する。

$$(1) \quad score(u, g) = \log n(u, g) \cdot PMI(u, g) \text{ (proposed)}$$

$$(2) \quad score(u, g) = n(u, g) \cdot PMI(u, g) \text{ (pmi)}$$

$$(3) \quad score(u, g) = n(u, g) \text{ (baseline)}$$

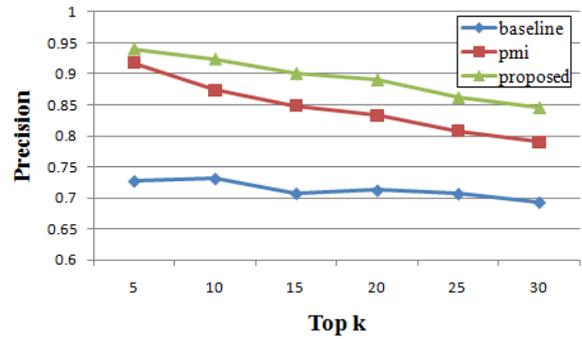


図 5 Top k 件のタグの α 適合率

proposed は提案手法、pmi はタグが付けられた回数に自己相互情報量を掛け合わせ、スコアとする手法、baseline はタグが付けられた回数をスコアとする手法である。

本実験の目的は、自己相互情報量の取り入れの有効性の検証と、回数の対数を取ることを有効性の検証である。4.1 節では本実験で用いたデータについて、4.2 節では実験の方法について説明する。4.3 節では結果に対する考察を行う。

4.1 データ

本実験では、日本語圏ユーザが作成した 694,005 リストを用いた。全リストに含まれたユーザの異なり数は 2,005,635 ユーザであり、全リストから抽出されたタグの異なり数は 110,925 タグである。ユーザが発信した直近 200 件のツイートのうち、ひらがなを含むツイートが 10 件以上あるユーザを日本語圏のユーザとした。対象としたリストは日本語圏ユーザが作成したリストであるが、リストが含むユーザには日本語圏以外のユーザも含まれる。

リストの取得には TwitterAPI [2] の user/lists メソッドを用い、リストに含まれるユーザの取得には user/list_id/members メソッドを用いた。

4.2 実験方法

まず上記のデータを用いて、比較する三つの手法によるタグ付けを行った。次に被験者 8 名によって、ユーザとユーザに付けられたタグとの関連度の評価を行った。各被験者が知っている可能性の高いユーザを評価対象とするため、リストに含まれる数が多い上位 20 ユーザを評価対象とした。また、被験者が評価できるタグの数には限界があるため、ユーザとの関連スコアが大きい上位 30 タグを評価対象とした。被験者はユーザに付けられた各タグについて、それがユーザと関連するか否か、関連するならばそのタグがユーザの性質を表すのに有用であるか否かの二点についての評価を行った。

本実験では、タグがユーザと関連すると判断されれば α 適合、関連し、かつ有用であると判断されれば β 適合と定義する。図 5 では、評価対象の各ユーザに付けられた上位 k 件タグの α 適合率の平均、図 6 では、 β 適合率の平均を示す。次節で結果に対する考察を行う。

4.3 考察

図 5 によると、proposed が他の二手法を上回る α 適合率を示している。この結果の原因を以下のように考察する。baseline

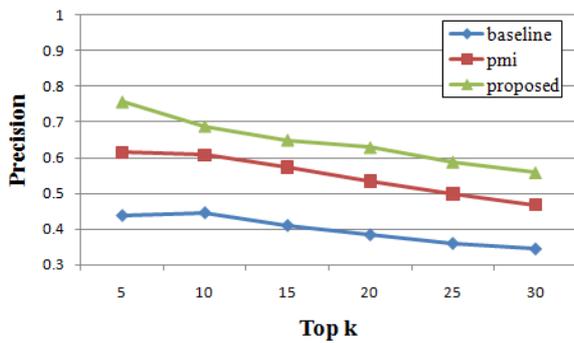


図 6 Top k 件のタグの β 適合率

の上位には list や people といったタグが抽出され、それらが被験者によってユーザとは関連しないとされたため、 α 適合率が小さくなっている。被験者によって関連がないとされたタグは、多くのユーザに対して付けられる傾向にあったため、各ユーザとの自己相互情報量が小さな値を示す傾向にあった。そのため、自己相互情報量を考慮する pmi や proposed ではこれらのタグの順位が下がったと考えられる。しかし、list などの頻出タグは各ユーザに付けられる回数が非常に大きな値を示すため、回数対数の対数を取らない pmi では比較的上位に抽出されてしまった。よって、pmi は proposed に比べてわずかに低い α 適合率を示していると考えられる。list や etc, people などのタグは、リスト名をハイフンで区切ったものをタグとしたために発生してしまったものであるが、自己相互情報量の考慮により、それらの関連スコアを下げることでできている。また、タグの出現頻度は冪分布に従うため、頻出タグは非常に大きな値を示してしまうが、対数を取ることでその影響を抑えることができている。

図 6 によると、 β 適合率においても proposed が他の手法を上回る値を示している。本実験において評価対象とした、多くのリストに含まれるユーザは、famous や celebrity といったタグを多くつけられていた。これらのタグは被験者によって各ユーザに関連はするが、ユーザの性質を表すのに有用ではないとされた。famous などの β 適合ではないタグの多くは、著名人などの公式アカウントであればほとんどのユーザに付けられているため、自己相互情報量は小さな値を示す。よって前述した list などのタグと全く同様の理由で、proposed が他の二手法を上回ったと考えられる。

以上により、提案手法は次の二点で有効であると考えられる。

- (1) 自己相互情報量の考慮により、リスト名を用いたがために生じてしまった、意味をなさない頻出タグの関連スコアを下げることでできている
- (2) タグの出現頻度は冪分布に従うため、頻出タグの出現回数は非常に大きな値を示してしまうが、回数対数の対数を取ることでその影響を抑えている。

しかし、リスト名を用いたタグ付けには二つの問題点が見られた。一つは、リスト名から抽出したタグには同義語やスペルミスなど、同じ意味を持つが違う表記をされるものが多く存在した点である。もう一つは、有名アカウントにはユーザ名がタグとして付けられるという点である。表 2 に、提案手法

表 2 ユーザ hatoyamayukio の上位 10 タグ

| rank | tag | rank | tag |
|------|---------------|------|-------------|
| 1 | hatoyamayukio | 6 | politicians |
| 2 | hatoyama | 7 | seijika |
| 3 | politician | 8 | hato |
| 4 | politics | 9 | statesman |
| 5 | seiji | 10 | government |

によって日本の政治家である鳩山由紀夫氏の公式アカウント hatoyamayukio に付けられた上位 10 件のタグを示す。上位 10 タグは政治という意味を持つタグと、ユーザ名を表すタグで占められている。これらのタグは、被験者によってユーザの性質を表す有用なタグであるとされたが、同じ意味を持つタグが頻出してもユーザの性質を十分に表すことはできない。よって、何らかの方法でこれらのタグを一つにまとめる必要があると考えられる。

5. 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 [6] [5] 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 [3] [15] [7], Twitter ユーザをランキングする研究 [12] [8] [14] など、その内容は様々である。

中でも特に本研究と関連する研究として、Wu らによる、Twitter ユーザにタグ付けを行う研究が挙げられる [13]。Wu らはユーザがポストしたツイートに対して様々な前処理を行った後、TFIDF 法や TextRank [10] を用いてキーワードを抽出し、それをユーザのタグとしている。また評価実験により、TextRank を用いた方が精度が高いことを示している。Wu らの研究ではユーザ自身が発信した情報に基づいたタグ付けを行っているため、他のユーザがリストの名前という形で与えた情報を基にタグ付けを行う本研究とは異なる。また、Huang ら [4] はハッシュタグに関する研究を行っている。ハッシュタグとは #hashtag の形式でツイートに付加されるタグであり、主に何らかの出来事に関するツイート群には同じハッシュタグが付けられる。Huang らは、ハッシュタグは他のソーシャルメディアにおいてコンテンツに付けられるタグとは異なり、数日の間使用されたのちに消えていくと報告している。Huang らの研究はハッシュタグに関するものであり、ユーザへのタグ付けを行う本研究とは異なる。

他のソーシャルメディアにおけるタグに関する研究も盛んに行われている。Li ら [9] はソーシャルブックマークサービス del.icio.us におけるタグを用いてユーザの共通興味を発見する研究を行っている。del.icio.us において Web ページに付けられたタグは、Web ページから抽出したキーワードより語彙数が小さいにも関わらず、非常に表現力が高いことを報告している。

6. 結論

本稿では、Twitter ユーザがどのようなトピックに関する情報発信を行う傾向にあるかをとらえるために、リストを用いて

ユーザにタグ付けを行う手法を提案した。提案手法は、リストに含まれるユーザ群は互いに何らかの関連を持ち、リストの名前がその関連を表す傾向にあることに着目した。従来のタグ付け手法はユーザが発信したツイートを分析しているのに対して、提案手法は他のユーザが作成したリストを用いるため、より客観的な評価に基づくタグ付けが可能である。

評価実験によって、提案手法が他の手法よりも有効であることを示した。提案手法はタグの出現回数の対数をとることと、自己相互情報量の考慮により、リスト名から抽出したタグのうち、頻出するノイズのようなタグの関連スコアを下げることでできている。しかし、同じ意味を持つ異なる表記をされているタグを上位に多く抽出してしまうという問題点があった。今後の展望として、それらのタグのクラスタリングが挙げられる。同じ意味を持つタグは同じユーザに頻繁に付けられるため、共起頻度を用いてのクラスタリングが可能であると考えられる。

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。

文 献

- [1] *Twitter*, <http://twitter.com>.
- [2] *Twitter API*, <http://apiwiki.twitter.com/Twitter-API-Documentation>.
- [3] 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).
- [4] Huang, J., Thornton, K. M. and Efthimiadis, E. N.: Conversational tagging in twitter, *Proceedings of the 21st ACM conference on Hypertext and hypermedia*, HT '10, New York, NY, USA, ACM, pp. 173–178 (2010).
- [5] 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).
- [6] 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).
- [7] Kwak, H., Lee, C., Park, H. and Moon, S.: What is Twitter, a social network or a news media?, *World Wide Web Conference* (2010).
- [8] 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).
- [9] Li, X., Guo, L. and Zhao, Y. E.: Tag-based social interest discovery, *Proceeding of the 17th international conference on World Wide Web*, WWW '08, New York, NY, USA, ACM, pp. 675–684 (2008).
- [10] Mihalcea, R. and Tarau, P.: TextRank: Bringing Order into Texts, *Proceedings of EMNLP-04 and the 2004 Conference on Empirical Methods in Natural Language Processing* (2004).
- [11] Moore, R. J.: *New data on Twitter's users and engagement*, <http://themetricssystem.rjmetrics.com/2010/01/26/new-data-on-twitters-users-and-engagement/> (2010).
- [12] Weng, J., Lim, E., Jiang, J. and He, Q.: TwitterRank: Finding Topic-sensitive Influential Twitterers, *WSDM* (2010).
- [13] Wu, W., Zhang, B. and Ostendorf, M.: Automatic generation of personalized annotation tags for Twitter users, *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association*

- for Computational Linguistics*, HLT '10, Stroudsburg, PA, USA, Association for Computational Linguistics, pp. 689–692 (2010).
- [14] Yamaguchi, Y., Takahashi, T., Amagasa, T. and Kitagawa, H.: TURank: Twitter User Ranking Based on User-Tweet Graph Analysis, *WISE*, pp. 240–253 (2010).
- [15] Ye, S. and Wu, S. F.: Measuring Message Propagation and Social Influence on Twitter.com, *SocInfo 2010* (2010).