

Twitterのプロフィール変更時における ユーザの行動と関係のあるキーワードの抽出

島 仁誠 吉田 光男^a 梅村 恭司^b

豊橋技術科学大学

j163336@edu.tut.ac.jp a) *yoshida@cs.tut.ac.jp* b) *umemura@tut.jp*

概要

ソーシャルメディアにおいて、ユーザのことを知るためにアカウント名、自己紹介文、ロケーション名などを見る。ユーザの生活環境が変化すると、これらのプロフィールも変更される可能性がある。本稿では、Twitterでのプロフィールの変更に着目し、プロフィールの変更時期とその理由について分析した。その結果、日本語ではプロフィールの変更時期のピークが4月1日にあるという結果が得られたが、英語では1年をとおして変更時期のピークは観測されなかった。英語ではプロフィールを変更するときに、誕生日に関するキーワード (BIRTHDAY, birthday, BDAY) がアカウント名に挿入される傾向があり、日本語ではプロフィールを変更するときに、Twitterでの活動を控えるときは「低浮上」、他のユーザに何かをアピールしたいときは「固定ツイ」のキーワードをアカウント名に挿入される傾向があることが分かった。

キーワード Twitter, プロフィール, キーワード抽出

1 はじめに

Twitter¹は近年よく利用されているソーシャルメディアの1つである。フォロー・フォロワーの大規模なネットワークやリアルタイムな短文の投稿 (ツイート) など Twitter の特性を利用した研究が数多く行われている。プロフィールはユーザのことを知るために利用され、そのユーザはいつでも自由に変更することができる。Twitter は世界各地で普及しており、プロフィールを変更する時期は言語圏によって異なる可能性がある。また、プロフィールを変更するときは特定のイベントや生活環境の変化と関係のあるキーワードを挿入している可能性がある。

多くの研究者は Twitter のプロフィールを用いた分析を試みている。プロフィールについての分析として、Semertzidis ら [1] は自己紹介文について詳細に分析し、Julia ら [2] はユーザの属性による自己紹介文の書き方の違いについて調査している。さらに Mislove ら [3] は、プロフィールに基づき、米国におけるユーザの人口分布について調査している。プロフィールを活用する研究として、プロフィールを用いたユーザの属性ごとの分類の研究 [4, 5, 6] が挙げられる。また、石野 [7] と鈴木ら [8] は、プロフィールを用いて大学生や公式アカウントなどの特定アカウントの自動検出をしている。さらに、鈴木ら [9] は、自己紹介文を用いて観光地の類型化手法を提案している。これらの先行研究では、ユーザのプロフィールは変更しないと仮定している。しかしながら、現実ではユーザはしばしばプロフィールを変更している。

本稿では、Twitter におけるユーザのプロフィールの変更に着目して、Twitter 上でのプロフィールの変更時期とその理由について分析する。本稿では、次の2つの疑問に答える。

RQ1 Twitter ユーザのプロフィールはいつ変更されるのか。

RQ2 プロフィールが変更されたときにどのようなキーワードが挿入されるのか。

分析の結果、プロフィールを変更する時期について、日本語と英語とで違いがあることが分かった。また、プロフィールが変更されたときに挿入されるキーワードとして、興味深いキーワードを観測できた。

2 データ

本研究では、2015年から2016年の間にTwitterから収集した日本語と英語のリツイートデータを使用した。データの詳細を表1に示す。これらのデータはTwitter Search API²を用いて収集した。本分析では、1年をとおしてユーザがどの時期にプロフィールを変更するのかを分析する。年の途中で登録したユーザを除外するため、2015年に1度でもリツイートしたユーザに限定した上で、2016年のデータでプロフィールの変更を分析した。この分析に使用したユーザは、日本語では8,083,170ユーザ、英語では19,485,950ユーザであった。

収集済みのリツイート数を図2.1の実線に示す。収集済みのデータには、2月4日、4月27日、9月6日など、

Copyright is held by the author(s).

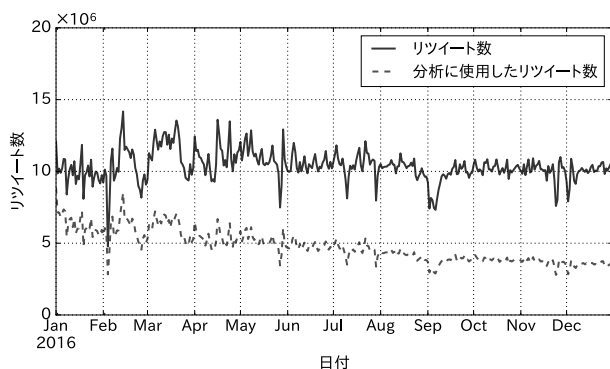
The article has been published without reviewing.

¹<https://twitter.com/>

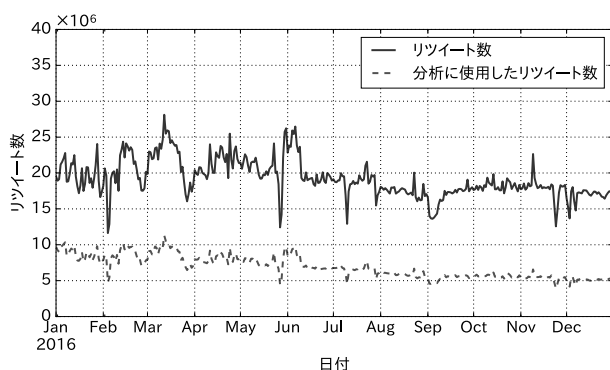
²<https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets>

表 1 収集済みのリツイート数とユーザ数

言語	年	リツイート数	ユーザ数
日本語	2015	9,441,746,992	24,166,859
日本語	2016	7,605,744,928	28,432,609
英語	2015	16,170,390,938	93,618,760
英語	2016	14,058,551,334	93,903,809



(a) 日本語



(b) 英語

図 2.1 リツイート数の推移

データの欠損が数日ある。これは、Twitter Search API と収集サーバの不具合によるものである。図 2.1 の破線は分析に利用したユーザのリツイート数である。ユーザの利用が徐々に減少していることが分かる。上記の状況が分析結果に影響を与える可能性がある。

分析するプロフィールとしてアカウント名、自己紹介文、ロケーション名に着目する。Twitter では、アカウント名には実名もしくはニックネームを書くユーザが多い。また現在の状態が付与されることもある（テスト期間、留学中など）。自己紹介文には仕事や趣味などの簡単な自己紹介、ロケーション名には居住地が書かれていることが多い。なお、リツイートデータのタイムゾーンは日本標準時 (JST) に変換しているため、英語の場合、時差により現地時間とのずれが生じる。

3 分析手法と結果

3.1 プロフィールを変更する時期

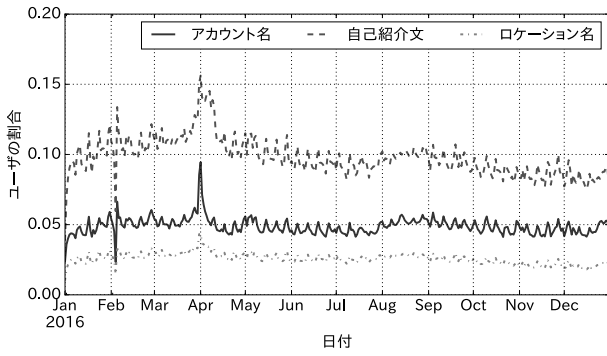
ユーザがプロフィールを変更する時期について調査する。日本では大学入学や就職のために 4 月から新たな環境で生活を始める人が多い。もしユーザの生活環境の変化がプロフィールに影響を与えるならば、4 月にプロフィールの変更が多く観測されると考えられる。

図 3.1 に日本語と英語とのプロフィール変更時期の推移を示す。縦軸は日あたりにプロフィールを変更したユーザの割合であり、割合の分母はその日にリツイートしているユーザ数である。自己紹介文は日本語と英語との両方で最も変更されている。結果を見る限り、日本語のユーザは英語のユーザよりもプロフィールを頻繁に変更している。日本語では 4 月 1 日にプロフィール変更のピークが観測された。これは、新生活もしくはエイプリルフールの影響によるものだと考えられる。英語では日本語の場合と同じような変更のピークは観測されなかったが、11 月にアカウント名の変更の小さなピークが観測された。これはアメリカの大統領選挙の時期と一致している。

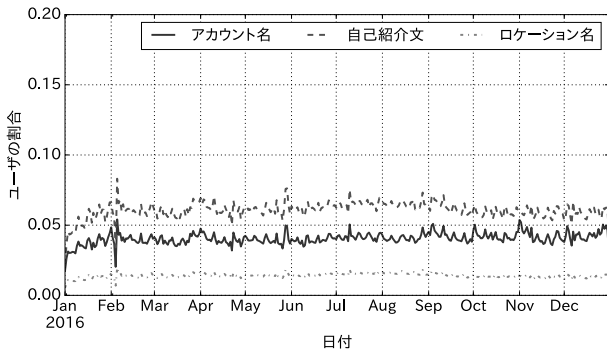
エイプリルフールのような一時的なイベントのために、ユーザはプロフィールを変更してからすぐに元に戻しているかもしれない。そこで、プロフィールを A → B → A のように元に戻している時期を調査した。その結果を図 3.2 に示す。縦軸はプロフィールを元に戻した日のリツイート数である。アカウント名は日本語と英語の両方で最もよく元に戻されている。日本語のアカウント名において、4 月 1 日にピークが観測された。これは 4 月 1 日だけ特別な名前が使用されることを表している。言い換えると、4 月 1 日のプロフィールの変更は、エイプリルフールの影響があったと考えられる。英語のアカウント名においては、11 月 1 日（日本標準時で集計しているため、アメリカでは 10 月 31 日に対応する）にピークが観測された。これはハロウィンの影響によるものだと考えられる。次のピークは 11 月 9 日（アメリカでは 11 月 8 日に対応する）に現れた。これはアメリカの大統領選の日と一致する。

3.2 プロフィール変更時に挿入されるキーワード

プロフィールの変更理由を明らかにするために、プロフィールが変更されたときに挿入されるキーワードを調査する。プロフィールが変更されたときに挿入されるキーワードを抽出するために、プロフィールの変更前後の文字列を用いて、文字列の差分をキーワードと抽出した。以降、アカウント名の変更について報告する。自己紹介文とロケーション名については一般的な語や地名が抽出され、特徴的な結果は得られなかった。

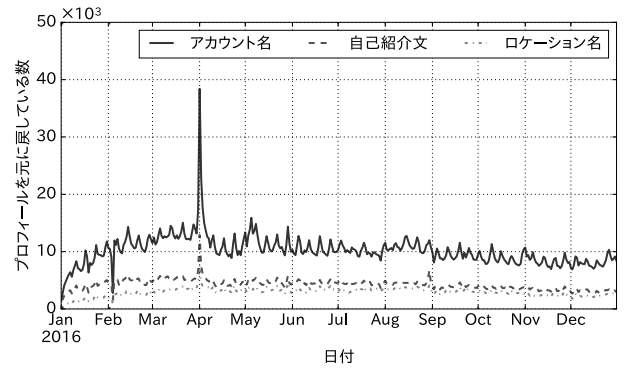


(a) 日本語

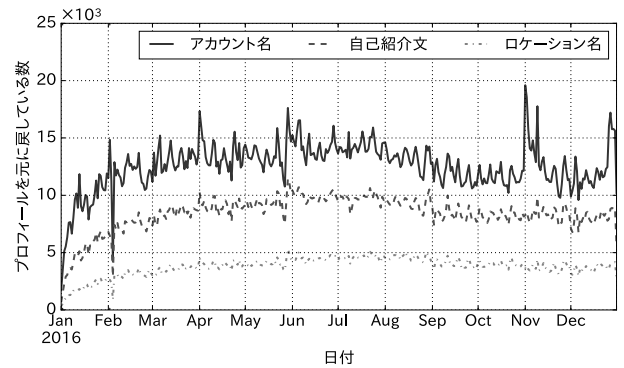


(b) 英語

図 3.1 プロフィールの変更時期の推移



(a) 日本語



(b) 英語

図 3.2 プロフィールを元に戻している時期の推移

3.2.1 日本語

日本語におけるキーワード抽出は次の手順で行う。まず、プロフィールの変更前後の文字列組に対してPythonの“difflib.SequenceMatcher”を使い、共通部分の文字列を抽出する。次に、プロフィールの変更後の文字列を共通部分の文字列で分割する。分割されたそれぞれの文字列を挿入キーワードの候補とし、文字数が少ない場合には一般的な語や記号が多くなるため、候補のうち4文字以上の文字列をキーワードとして抽出した。たとえば、プロフィールの変更前の文字列が「しま@大学生」、プロフィールの変更後の文字列が「しま@試験期間」の場合、挿入キーワードは「試験期間」である。

日本語のアカウント名が変更されたときに挿入されるキーワードの上位20件を表2に示す。注目すべきキーワードについては下線を引いた。表のキーワードは出現数(リツイート数)でソートし、ユニークなユーザ数も示してある。多くのユーザは「低浮上」、「固定ツイ」を挿入していた。この言葉は学生などの若いユーザによってしばしば使われている。学生は試験期間などの忙しい時期にはTwitterを利用しなくなる。「低浮上」は、忙しくてTwitterを利用できない状況を友人に示すために使われる。また、固定ツイート(ツイートを自分のタイムラインの一番上に常に表示する機能)に近況や宣伝などを記述し、アカウント名に「固定ツイートを見て」の

ような文字列を追加することで、アカウント名だけでは記述できない内容を友人にアピールしている。このような場合に「固定ツイ」がよく挿入されている。

月ごとに挿入されるキーワードを調査したところ、上位5件のキーワードは概ね固定されているものの、「ポケモンGO」、「の名は。」、「ハロウィン」、「クリスマス」などの一時的なイベントに関するキーワードは上位20位以内に出現した。

3.2.2 英語

英語におけるキーワード抽出は次の手順で行なう。まず、プロフィールの変更前後の文字列をそれぞれスペースで分割する。挿入キーワードの候補は、プロフィールの変更前には出現せず、プロフィールの変更後だけに出現している文字列とする。候補のうちアルファベットを含んでいる4文字以上の文字列を挿入キーワードとした。たとえば、プロフィールの変更前の文字列が「shima」、プロフィールの変更後の文字列が「shima birthday」の場合、挿入キーワードは「birthday」である。

英語のアカウント名が変更されたときに挿入されるキーワードの上位20件を表3に示す(表の見方は前節を参照)。多くのユーザは誕生日に関するキーワード(BIRTHDAY, birthday, BDAY)を挿入していた。これは日本語では観測されなかった傾向である。

表2 アカウント名が変更されたときに挿入されるキーワードの上位 20 件 (日本語)

挿入キーワード	リツイート数	ユーザ数
@低浮上	24,659	20,903
ありがとう	12,088	10,820
お疲れ様でした	9,745	8,150
固定ツイ	8,340	6,990
は低浮上	6,511	5,734
当選祈願	5,175	4,653
(低浮上)	5,159	4,710
テスト期間	5,070	4,573
お疲れ様でした!	4,951	4,182
@固定ツイ	3,250	2,767
[반동걸]	3,193	2,675
(アローラのすがた)	2,762	2,671
固定ツイ見てね	2,640	2,415
楽しかった	2,488	2,342
おじさん	2,342	2,059
お疲れ様	2,332	2,116
【低浮上】	2,311	2,045
しんどい	2,311	2,173
† 冨墓 †	2,201	1,955
低浮上気味	2,125	1,986

表3 アカウント名が変更されたときに挿入されるキーワードの上位 20 件 (英語)

挿入キーワード	リツイート数	ユーザ数
loves	63,865	23,086
HAPPY	33,267	20,276
happy	26,037	15,935
BDAY	21,677	13,866
bday	19,069	13,793
BIRTHDAY	16,351	13,742
TODAY	16,200	12,316
birthday	15,849	14,239
girl	14,885	12,877
days	14,465	10,199
pinned	13,221	7,251
misses	12,112	7,588
Happy	11,248	8,290
baby	10,796	7,962
love	10,289	8,457
spooky	10,148	8,493
7/27	9,390	5,030
PLEASE	9,246	5,413
TOMORROW	9,232	7,424
stan	9,146	6,263

月ごとに挿入されるキーワードを調査したところ、上位 5 件のキーワードは概ね固定されているものの、「spooky」、「Christmas」、「Merry」などの一時的なイベントに関するキーワードは上位 20 位以内に出現した。大統領選挙が行われた 2016 年 9 月から 11 月には、選挙に関するキーワードとして「Deplorable」、「VOTE」、「Trump」が上位に出現した。

4 おわりに

本稿では、Twitter ユーザのプロフィールの変更に着目し、変更の時期とその理由について分析した。日本語ではプロフィールの変更時期のピークが 4 月 1 日にあるという結果が得られたが、英語では 1 年をとおして変更時期のピークは観測されなかった。英語ではアカウント名を変更するときに、誕生日に関するキーワード (BIRTHDAY, birthday, BDAY) が挿入される傾向にある。日本語ではアカウント名を変更するときに、「低浮上」と「固定ツイ」が挿入される傾向にあり、日本語と英語とでは、現在の活動の状況を示すのか、あるいは、誕生日などの属性の状況を示すのかという違いが観測された。

参考文献

- [1] Semertzidis, K., Pitoura, E. and Tsaparas, P.: How People Describe Themselves on Twitter, DBSocial 2013, pp. 25–30, 2013.
- [2] Julia, V. and Kim, H.: Highly tweeted science articles: who tweets them? An analysis of Twitter user profile descriptions, Scientometrics, Vol. 112, No. 1, pp. 345–366, 2017.
- [3] Mislove, A., Lehmann, S. and Ahn, Y.: Understanding the Demographics of Twitter Users, ICWSM-11, pp. 554–557, 2011.
- [4] Jalal S. Alowibdi, Ugo A. Buy and Yu, P.: Empirical Evaluation of Profile Characteristics for Gender Classification on Twitter, ICMLA 2013, pp. 365–369, 2013.
- [5] Marco, V., Joao P. and Fernando, B.: Using Unstructured Profile Information for Gender Classification of Portuguese and English Twitter Users, SLATE 2015, pp. 57–64, 2015.
- [6] Uddin, M., Imran, M. and Sajjad, H.: Understanding Types of Users on Twitter, arXiv.org, arXiv:1406.1335, 2014.
- [7] 石野亜耶: 大学生の Twitter アカウントの自動検出, 広島経済大学研究論集, Vol. 38, No. 3, pp. 23–29, 2015.
- [8] 鈴木祥平, 倉田陽平, 石川博: Twitter のユーザプロフィールを用いた公式アカウントの検出とフォロー形成パターンの分析, DEIM 2017, G2-3, 2017.
- [9] 鈴木祥平, 池田拓生, 倉田陽平ほか: Twitter のユーザプロフィールを用いた観光地の類型化, DEIM 2015, A2-1, 2015.