

情報発信の対象範囲に基づく Twitter ユーザの分類

竹村 光[†] 田島 敬史^{†‡}

^{†, ‡} 京都大学大学院情報学研究科 〒606-8501 京都府京都市左京区吉田本町

E-mail: [†]takemura@dl.kuis.kyoto-u.ac.jp, ^{‡‡}tajima@i.kyoto-u.ac.jp

あらまし Twitter は、従来の多くのソーシャルメディアの性質を兼ね合わせており、その利用目的は多岐に渡っている。そのため、ニュースなどの情報発信のために Twitter を利用するユーザもいれば、友人とのコミュニケーションのために Twitter を利用するユーザも存在し、ユーザによって情報発信の対象範囲が様々である。そこで本研究では、Twitter ユーザを、広く一般のユーザが興味を示す情報を発信するのか、それとも一部のユーザのみが興味を示す情報を発信するのかという観点から分類する手法を提案する。提案手法では、ユーザのフォロワーに着目し、フォロワー全体が何か共通の傾向を持っているかどうかに基づいて分類を行う。そして、提案手法を実装することで実験を行い、その有効性を評価する。

キーワード マイクロブログ、ツイート、対象局所性

1. はじめに

ブログや SNS というソーシャルメディアの普及に伴い、誰もが簡単に Web 上で情報を発信できるようになった。特に近年では、マイクロブログと呼ばれる、SNS の性質を併せ持ったブログサービスが、爆発的な成長を遂げている。

マイクロブログの中でも、特に Twitter^(注1) と呼ばれるサービスの成長が著しい。Twitter は、2012 年 12 月時点でアクティブユーザ数が 2 億人を超えており [1]、同年 6 月現在、1 日に 4 億以上の記事が投稿されていると言われている [2]。Twitter では、ツイートと呼ばれる、上限 140 字のテキストを記事として投稿する。この制限により、従来のブログサービスや SNS に比べて、ユーザが手軽かつ簡易に記事を投稿することを可能としている。また、Twitter には、フォローという、各ユーザが自分の読みたいユーザの記事をホームページに表示させる機能があり、図 1 のように、そのユーザがフォローをした全ユーザの記事が時系列順にソートされてホームページに表示される。これを、タイムラインと呼んでいる。また、あるユーザをフォローしているユーザのことを、そのユーザのフォロワーと呼ぶ。さらに、リプライと呼ばれる、ある記事への返信という形で記事を投稿する機能も有しており、これにより、ユーザ間のコミュニケーションも円滑に行うことができるようになっている。

Twitter は、従来の多くのソーシャルメディアの性質を兼ね合わせており、その利用目的が多岐に渡っている。社会のニュースのように、広い範囲に向けて情報発信を行う記事もあれば、単なる自身の状況や雑記のような記事もあったりと、多様なものが混在している。このような性質のため、Twitter は現在、新たな情報発信メディアとして大きな注目を集めている。

上記のように、Twitter は多様な目的で利用されるため、ユーザによって情報発信の対象範囲が様々である。ニュースなどの情報を発信するユーザや、名言を発信するユーザのように、広



図 1 Twitter ホームページのスクリーンショット

く一般のユーザが興味を示す情報を発信するユーザもいれば、一部のユーザに対して情報を発信するユーザや、友人とコミュニケーションを行うユーザのように、特定のユーザのみが興味を示す情報を発信するユーザも数多く存在する。

本研究では、このようにユーザによって情報発信の対象範囲が様々である点に着目し、Twitter ユーザを、広く一般のユーザが興味を示す情報を発信するのか、それとも一部のユーザのみが興味を示す情報を発信するのかという観点から分類する手法を提案する。提案手法では、ユーザのフォロワーに着目し、フォロワー全体が何か共通の傾向を持っているかどうかに基づいて分類を行う。もしあるユーザのフォロワー同士が何か共通の傾向を持っていれば、そのユーザは、ある特定のユーザが興味を示す記事を投稿していると考えられる。逆に、フォロワー同士の傾向がはらばらであれば、そのユーザは多様なユーザからフォローされている可能性が高く、広く一般のユーザが興味を示す記事を投稿していると考えられる。

本研究の応用例としては、記事を検索する際に、広く一般の人々が興味を示す情報を発信するユーザの記事を優先して表示するシステムが考えられる。Twitter の記事を検索し、ある特定の情報を得ようとしても、その情報が、一部のユーザにしか

(注1) : <http://twitter.com/>

関連しないような多くの記事に埋もれてしまうといった事態も頻繁に生じる。そこで、このシステムを用いることで、情報を広く一般の人々が興味を示す情報を発信しているユーザの記事が優先して表示され、ユーザの記事検索の効率向上につながると思われる。また、情報発信の対象範囲を考慮したユーザ推薦にも応用できると考えている。

本研究では、Twitter から実際に収集したデータを用いて、提案手法の評価実験を行った。

2. 関連研究

近年の爆発的なマイクロブログの普及により、マイクロブログに関する研究は盛んに行われるようになった。記事を様々な観点から分類する研究 [3] や、記事をその有用性によりランキングする研究 [7]、マイクロブログのリアルタイム性に着目した研究 [17] など、その内容は様々である。

マイクロブログのユーザの特性に着目した研究も盛んに行われている。Java ら [11] は、Twitter のネットワーク構造を分析し、ユーザが主に、情報発信・受信や、他のユーザとのコミュニケーションのためにマイクロブログを利用していることを明らかにした。Twitter ユーザを分類する研究としては、Twitter ユーザを人間なのか、bot^(注2)なのか、あるいはその両方を兼ねているのかにより分類する研究 [6] や、Twitter ユーザのツイートやネットワーク構造に着目し、政治的思想や民族性によりユーザを分類する研究 [16] など、その分類方法は様々である。本研究と近い分類方法としては、Yan ら [12] による、Twitter ユーザを、広報などに関する記事を投稿するユーザ (オープンアカウント) なのか、生活ログなどに関する記事を投稿するユーザ (クローズドアカウント) なのかにより分類する研究が挙げられるが、オープンアカウントだからといって広く一般のユーザが興味を示すわけではなく、本分類方法とは異なるといえる。Cha ら [5] は、Twitter において、ソーシャルグラフにおける入次数・リツイート・リプライの 3 つの尺度からユーザの影響度を推測する手法を提案した。本研究における情報発信の対象範囲は、その範囲が広がるほど、影響力が大きくなる可能性は高いと考えられるが、対象範囲が広いからといって、一概に影響力が大きいとは言うことはできない。

マイクロブログにおいて、ある一部のユーザのみに関連する記事を発見するのに有用であるような研究も行われている。Sakaki ら [9] は、ユーザの投稿から地震や台風などのイベントをリアルタイムに取得し、さらにその発生地・中心地を推測する手法を提案している。Ikawa ら [10] は、マイクロブログ上で発信されるメッセージの内容を基に、そのメッセージが発信された場所を推定する手法を提案している。これらの研究は、一部の地域のユーザに関連のある記事を発見するのに有用であると考えられる。また、Sriram ら [4] や Nishida ら [15] による、ツイートをトピックにより分類する研究は、各トピックにおいてそれぞれ専門性の高い記事を発見しているといえる。

このように、様々な観点において、ある一部のユーザのみに

関連のある記事は存在し、それらが発見するのに有用である研究は行われている。しかし、これらの観点は複数存在するため、それぞれの観点においてそのような記事を発見するというのは、対象範囲を推定するための適切な手法とはいえない。そこで本研究では、対象範囲を一元的に判定することを試みた。

Ehrlich ら [8] は、ある職場において、職場内でのみ利用することのできるマイクロブログと、Twitter のような一般に利用することのできるマイクロブログでの投稿内容の違いに着目することにより、それらの使われ方の違いを分析した。その結果、前者では、技術的なサポートを求める、会話を行うといった使い方が多く見られたのに対し、後者では、一般的な情報の共有といった使い方が多く見られた。近年では、Twitter 内でも、特定のコミュニティのユーザと関わりを持つためだけに利用するといったユーザも多く見られ、1 つのマイクロブログで上記の双方の使い方がされているといえる。本研究は、Twitter 内において、このように複数の使われ方がされていることに着目したものであるといえる。

また、マイクロブログ記事の検索は、リアルタイムに情報が得られるという点で、Web 検索とは異なる点が多く、研究対象にされることが多い [18] [13] [14]。本研究も、記事を検索する際に、広く一般のユーザが興味を示す情報を優先して表示するシステムの作成を目標としており、マイクロブログ記事の検索に着目した研究であるといえる。

3. 情報発信の対象範囲

本章では、情報発信の対象範囲を表す指標として、Twitter ユーザの対象局所性を定義する。また、情報発信の対象範囲が局所的となる要因についても述べる。

3.1 対象局所性の定義

本論文では、ある Twitter ユーザの対象局所性の定義として、そのユーザの情報発信の対象となりうるユーザの数が少ないほど大きくなる指標を用いる。形式的には、ユーザ u の対象局所性 $TargetLocality(u)$ を、以下のように定義する。

$$TargetLocality(u) = \log \frac{N}{|S_u|}$$

ここで、 N は Twitter のユーザ数、 S_u は u の情報発信の対象となりうるユーザ集合を表す。

例えば、就職活動に関する情報を発信するユーザは、主に就職活動生を対象として記事を投稿するため、全 Twitter ユーザの中で見ると、その情報に興味を示すユーザは一部に偏っている。よって、このユーザは、対象局所性が高いと考えられる。

逆に、社会のニュースに関する情報を発信するユーザは、広く一般に有用な記事を投稿するため、その情報に興味を示すユーザは非常に多い。よって、このユーザは、対象局所性が低いと考えられる。

本論文では、Twitter ユーザの対象局所性がある閾値 δ より大きければ、そのユーザをターゲット型ユーザと呼び、そうでないユーザを非ターゲット型ユーザと呼ぶ。形式的には、以下の条件により分類を行う。

(注2) : Twitter に自動的に投稿するアカウント及びそのプログラムのこと

$$\begin{cases} \text{if } \text{TargetLocality}(u) > \delta & u \text{ はターゲット型ユーザ} \\ \text{otherwise} & u \text{ は非ターゲット型ユーザ} \end{cases}$$

3.2 対象範囲が生じる要因

本節では、情報発信の対象範囲が局所的となる要因について述べる。分析の結果、これには大きく分けて3つの要因があることが分かった。以下では、この3つの要因に分けて考察する。

(1) 個人属性

性別・年齢・居住地などの個人の特徴により、情報発信の対象範囲が局所的になる場合が多く見られる。若い女性を対象にファッションの情報を発信するユーザの記事は、その対象範囲が局所的であるといえる。

(2) コミュニティ

大学、サークルといった、そのユーザが属しているコミュニティにより、情報発信の対象範囲が局所的になる場合も多い。このようなコミュニティのユーザと関わりを持つために、Twitterを使用しているようなユーザは、自分の投稿した記事を広く一般に伝えることを目的としておらず、そのコミュニティのユーザと会話をしたり、そのユーザとの間で情報共有を行ったりすることを主な目的とする。例えば、大学内の友人同士で形成されているコミュニティであれば、「今日の試験できた？」のような、そのコミュニティのみを対象とするような記事が頻繁に投稿される。

(3) 趣味・興味

様々なトピックの専門性が、情報発信の対象範囲を局所的にする場合も多い。例えば、あるプログラミング言語に関して有用な情報を発信するユーザは、情報を広く一般に伝えることを目的とするのではなく、その言語に興味がある人を対象に記事を投稿していると考えられる。このように、興味を持ってもらえるであろう一部のユーザを対象として情報を発信するユーザは、数多く存在する。

4. 提案手法

本章では、前章で述べた対象局所性の定義に基づき、情報発信の対象範囲の観点から Twitter ユーザを分類する手法を提案する。

4.1 対象局所性の判定手法

本研究では、あるユーザのフォロワーは、そのユーザの情報発信の対象範囲となりうるユーザをサンプリングしたものであると仮定する。そして、分類の対象となるユーザのフォロワーに着目する。社会のニュースのような、広く一般のユーザが興味を示す情報を発信するようなユーザは、多様なユーザからフォローされていると考えられる。一方、一部のユーザのみが興味を示す情報を発信するユーザは、そのフォロワーに何か共通の傾向が見られると考えられる。本研究では、この共通の傾向のことを、一貫性と呼ぶ。例えば、就職活動に関する情報を発信するユーザは、主に就職活動生にフォローされていると考えられるし、友人とのコミュニケーションのために Twitter を利用しているようなユーザは、そのユーザが属しているコミュ

ニティ内のユーザに主にフォローされていると考えられる。このように、ユーザのフォロワーに一貫性が見られるのか、それとも一貫性が見られず、ばらばらであるのかによって分類を行う。一貫性が高いほど、フォロワーの偏りが大きく、そのユーザの対象局所性が大きいことを意味する。

また、ここで考える一貫性とは、あるユーザのフォロワー全員があらゆる面で類似しているかを考えるのではなく、全員、あるいはその多くが一つでも共通点を持っていれば、一貫性が高いと考える。例えば、ある特定の地域の地震情報を発信するユーザの場合、そのフォロワー同士は、「住んでいる地域」という1点でのみ一貫している可能性が高く、年齢、性別、趣味や属するコミュニティなどは全てばらばらであると考えられるためである。したがって、フォロワー同士の特徴を表す複数の指標の中で、最も一貫性の高いものを取得する。

4.2 フォロワーの一貫性

本節では、分類の対象となるフォロワー間の一貫性を求める際に本研究で用いたフォロワーの属性を示す。

(1) プロフィール・位置情報

居住地などの同じ個人属性を持つユーザ同士や、同じコミュニティに属しているユーザ同士、同じ趣味を持つユーザ同士などは、プロフィールに同じ語が出現する可能性が高いので、そのような語を抽出する。ここで取得する語は、プロフィールや位置情報の内容の特徴付けると考えられる、名詞に限定するものとする。ここでは、以下のように、対象ユーザ u におけるフォロワー間のプロフィール・位置情報に関する一貫性 $\alpha(u)$ を求めている。

$$\alpha(u) = \frac{1}{n} \max_{t \in T} |\{S_{ui} | t \in S_{ui}, i = \{1, 2, \dots, n\}\}|$$

ここで、 T は、Twitter ユーザのプロフィール・位置情報に出現する語の集合、 S_{ui} は、 u のフォロワー f_i のプロフィール・位置情報に含まれる語の集合、 n は u のフォロワー数を表す。

(2) フォローの傾向

一貫性の高いユーザ同士は、フォローにも共通の傾向が見られる場合が多い。例えば、現実世界の同じコミュニティに属しているユーザ同士は、ソーシャルグラフ上で密に結びついていることが多く、そのコミュニティ内の別のユーザを共通してフォローしている可能性が高い。また、就職活動に関して有用な情報を発信するユーザのフォロワー同士は、就職活動において有用な、別のユーザを共通してフォローしていることが多いと考えられる。したがって、フォロワー同士のフォローの傾向に着目する。ここでは、以下のように、対象ユーザ u におけるフォロワー間のフォローの傾向一貫性 $\beta(u)$ を求めている。

$$\beta(u) = \frac{1}{n} \max_{a \in A} |\{U_{ui} | a \in U_{ui}, i = \{1, 2, \dots, n\}\}|$$

ここで、 A は Twitter ユーザの集合、 U_{ui} は、 u のフォロワー f_i がフォローしているユーザの集合を表す。

5. 実験と考察

本章では、提案手法を実装することで評価実験を行い、その

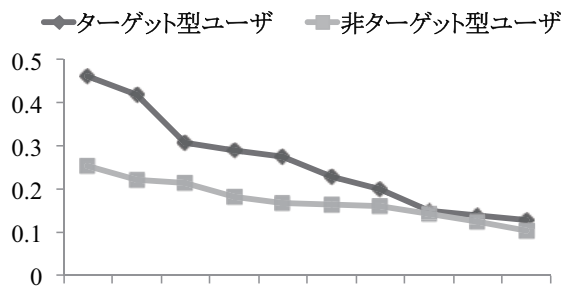


図2 フォロワー内でのプロフィール・位置情報における最頻出語の割合

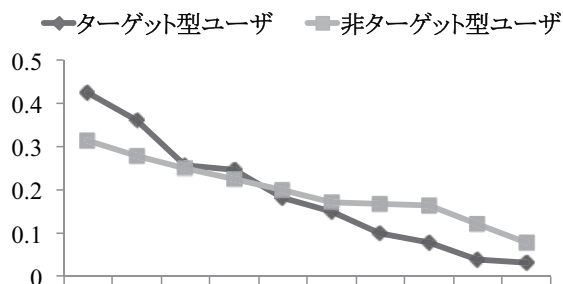


図3 フォロワー内で最も多くフォローされているユーザの割合

結果を基に考察を行う。また、実験結果における詳細を示し、その分析を行う。

5.1 実験の概要

本実験では、Twitter API を用いて、Twitter から実験に必要なデータを取得した。以下、あるユーザがフォローしているユーザのことを、そのユーザのフォロイーと呼ぶ。まず、50 人の日本人ユーザをランダムに選択し、これらのユーザが最近フォローした、情報発信を行うユーザを最大 2 件取得した。ただし、50 人のユーザを取得する際、アクティブユーザを取得するために、フォロワー数、フォロイー数、投稿数のいずれかが 0 のユーザは除外した。これにより収集されたユーザの中から、ターゲット型ユーザ、非ターゲット型ユーザをそれぞれ 10 件ずつ、計 20 件を取得し、本実験の正解セットとした。そして、それら 20 件のフォロワー、さらにフォロワーのフォロイーまでを取得し、収集された 13,598,213 件のユーザを、本実験のデータセットとして用いている。

正解セットのユーザ 20 件それぞれのフォロワーについて、前章で述べた 2 つの属性である、プロフィール・位置情報における最頻出語の割合と、そのユーザとは別に最も多くフォローされているユーザの割合を求めた。本実験では、MeCab^(注3)を用いて形態素解析を行っている。

また、情報発信のために Twitter を利用するユーザと、コミュニケーションのために Twitter を利用するユーザの分類に関する研究は既に行われており [11]、これらはその手法により分類できると仮定して、本実験では、情報発信のために Twitter を利用するユーザのみを対象とした。

5.2 実験結果と考察

実験結果を図 2、図 3 に示す。図 2 は、ユーザのフォロワー

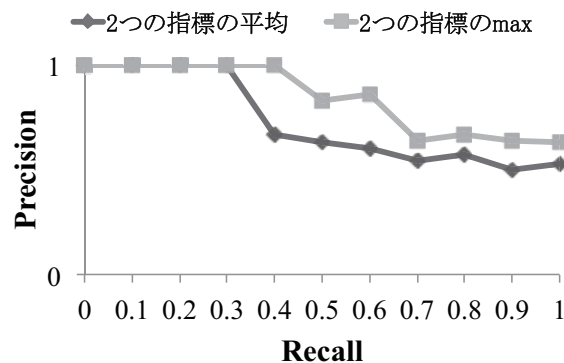


図4 2つの手法における適合率・再現率曲線

内でのプロフィール・位置情報における最頻出語の割合、図 3 は、フォロワー内で最も多くフォローされているユーザの割合を表しており、それぞれスコアの低いものからソートしている。

フォロワー内でのプロフィール・位置情報における最頻出語の割合に関しては、ターゲット型ユーザの方が、非ターゲット型ユーザよりも、平均的にフォロワーの一貫性を表すスコアが高くなった。このことから、ターゲット型ユーザのフォロワーのプロフィール・位置情報には同じ語が出現することが多く、分類における有力な手がかりであることが分かった。最も多くフォローされているユーザの割合に関しては、非ターゲット型ユーザに比べ、ターゲット型ユーザのスコアのばらつきが大きくなった。これは、どのような要因で一貫性が高いのかによって、フォローの傾向が大きく異なることが原因であると考えられる。つまり、同じコミュニティに属しているユーザ同士や、同じ趣味を持つユーザ同士などは、同じユーザを共通してフォローしていることが多いと考えられるが、例えば同じ居住地などでつながっているユーザ同士などは、コミュニティや趣味など、居住地以外の指標が全てばらばらであるため、同じユーザを共通してフォローしているとは限らないと考えられる。

次に、フォロワー同士の特徴を表す複数 (本実験では 2 つ) の指標の中で、最も一貫性のスコアが高いものを取得する手法が有効であることを確認するために、各指標のスコアの平均を取ったものをベースラインとして、比較を行った。これら 2 つの手法それぞれで求めたスコアを高いものからソートし、ターゲット型ユーザを正解として、適合率・再現率曲線を求めた。すると、図 4 にように、複数の指標の中で最も一貫性のスコアが高いものを取得する手法の方が、性能がよい結果となった。また、これら 2 つの手法それぞれの 11 点平均適合率を求めると、複数の指標のスコアの平均を取る手法が 0.73、最大のものを取る手法が 0.84 と、最大のものを取る手法の方が 10 ポイント以上高くなった。これらの結果より、フォロワー同士の特徴を表す複数の指標の中で、最も一貫性のスコアが高いものを取得する手法の有効性が示された。

5.3 実験結果の詳細と分析

本節では、実験結果から、正解データとして用いたユーザの分析を行った結果を示す。表 1 に、正解データ内の一部のユーザにおける詳細を示す。分析の結果、ターゲット型ユーザに関しては、フォロワー内でのプロフィール・位置情報における最頻

(注3) : <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

表 1 実験結果における一部のユーザの詳細

ユーザ型	ユーザ名	最頻出語	最多フォロイー
ターゲット型	@minnanomachi	写真	@camerapeople
	@ShILCjp	英語	@funfunreibun
	@CANELE_JP	大阪	@t.ishin
	@alife_sapporo	札幌	@alifesapporo
	@USJ-BGM	USJ	@USJ-Official
非ターゲット型	@wonderrocketweb	好き	@pamyurin
	@his_call_center	東京	@masason
	@NTV_OnDemand	大好き	@Myojo_henshu
	@1242_PR	好き	@ariyoshihiroiki
	@guragura2011	東京	@kirinjisinken

出語と、最も多くフォローされているユーザ（最多フォロイー）は、そのユーザと関係性のあるケースが多く見られた。例えば、写真に関するイベント告知のアカウント’@minnanomachi’は、フォロワー内での最頻出語が「写真」であったり、最多フォロイーも同じく写真に関するアカウントであったりと、その関連性が見られた。一方、非ターゲット型ユーザに関しては、多くのケースで、フォロワー内での最頻出語や最多フォロイーが、そのユーザと関連性が見られなかった。具体的には、フォロワー内での最頻出語が「好き」「東京」といった一般的に出現頻度の高い語であったり、最多フォロイーが有名人であったりするようなケースが多く見られた。このことから、一般に出現頻度の高い語や、広く一般に知られているユーザのスコアを下げることで、ターゲット型ユーザと非ターゲット型ユーザをさらに明確に分けられると考えられるため、今後改善を行う予定である。

6. ま と め

本論文では、広く一般のユーザが興味を示す情報を発信するのか、それとも一部のユーザのみが興味を示す情報を発信するのかという、情報発信の対象範囲の観点から、Twitter ユーザを分類する手法を提案した。提案手法では、ユーザのフォロワーに着目し、フォロワー全体が何か共通の傾向を持っているかどうかに基づいて分類を行った。そして、Twitter のデータを用いて評価実験を行い、提案手法が、ターゲット型ユーザと非ターゲット型ユーザの分類に有効であることを示した。

今後の展望について述べる。まず、本研究の評価実験は、決して十分なデータ数とは言えないので、さらにデータ数を増やすことを検討している。また、5.3 節で示したような、提案手法の改善についても検討を行う。さらに、提案手法を用いて、記事を検索する際に、広く一般のユーザが興味を示す情報を発信しているユーザの記事を優先して表示するシステムの実装も行う予定である。

文 献

- [1] Matt Brian. Twitter now has 200 million monthly active users, up 60 million in 9 months. <http://thenextweb.com/twitter/2012/12/18/twitter-now-has-200-million-monthly-active-users-up-60-million-in-9-months/>, 12 2012.
- [2] Matt McGee. With 400 million tweets per day, twitter spending “inordinate resources” on improving content discovery. <http://marketingland.com/twitter-400-million-tweets-daily-improving-content-discovery-13581>, 6 2012.
- [3] D. Irani, S. Webb, C. Pu, and K. Li. Study of trend-stuffing on twitter through text classification. In *Collaboration, Electronic messaging, Anti-Abuse and Spam Conference (CEAS)*, 2010.
- [4] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas. Short text classification in twitter to improve information filtering. In *Proceeding of the 33rd international ACM SIGIR conference on research and development in information retrieval*, pp. 841–842. ACM, 2010.
- [5] M. Cha, H. Haddadi, F. Benevenuto, and K.P. Gummadi. Measuring user influence in twitter: The million follower fallacy. In *4th international aaai conference on weblogs and social media (icwsm)*, Vol. 14, p. 8, 2010.
- [6] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia. Who is tweeting on twitter: human, bot, or cyborg? In *Proceedings of the 26th Annual Computer Security Applications Conference*, pp. 21–30. ACM, 2010.
- [7] Y. Duan, L. Jiang, T. Qin, M. Zhou, and H.Y. Shum. An empirical study on learning to rank of tweets. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pp. 295–303. Association for Computational Linguistics, 2010.
- [8] K. Ehrlich and N.S. Shami. Microblogging inside and outside the workplace. In *Proc. ICWSM*, Vol. 10, 2010.
- [9] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web*, pp. 851–860. ACM, 2010.
- [10] Y. Ikawa, M. Enoki, and M. Tatsubori. Location inference using microblog messages. In *Proceedings of the 21st international conference companion on World Wide Web*, pp. 687–690. ACM, 2012.
- [11] A. Java, X. Song, T. Finin, and B. Tseng. Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, pp. 56–65. ACM, 2007.
- [12] Y. Liang, M. Qiang, and Y. Masatoshi. Classifying twitter users for spatio-temporal entity retrieval. No. 15. 2012-DBS-156, 2012.
- [13] Kamran Massoudi, Manos Tsagkias, Maarten de Rijke, and Wouter Weerkamp. Incorporating query expansion and quality indicators in searching microblog posts. In *Advances in Information Retrieval*, pp. 362–367. Springer, 2011.
- [14] R. Nagmoti, A. Teredesai, M. De Cock, et al. Ranking approaches for microblog search. IEEE Computer Society, 2010.
- [15] K. Nishida, R. Banno, K. Fujimura, and T. Hoshide. Tweet classification by data compression. In *Proceedings of the 2011 international workshop on DETecting and Exploiting Cultural diversiTy on the social web*, pp. 29–34. ACM, 2011.
- [16] M. Pennacchiotti and A.M. Popescu. A machine learning approach to twitter user classification. In *Fifth International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2011.
- [17] H. Takemura and K. Tajima. Tweet classification based on their lifetime duration. In *Proceedings of the 21st International Conference on Information and Knowledge Management (CIKM)*, pp. 2367–2370. ACM, 2012.
- [18] J. Teevan, D. Ramage, and M.R. Morris. # twittersearch: a comparison of microblog search and web search. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pp. 35–44. ACM, 2011.