

リンクを含むつぶやきに着目した Twitter の分析

吉田 光男[†] 乾 孝司[†] 山本 幹雄[†]

[†] 筑波大学大学院システム情報工学研究科 〒 305-8573 茨城県つくば市天王台 1-1-1
E-mail: †m.yoshida@mibel.cs.tsukuba.ac.jp, ††{inui,myama}@cs.tsukuba.ac.jp

あらまし Web 2.0 は Web 上における情報の双方向性を強化し、新たな Web サービスを生み出した。その Web サービスの一種として Twitter をはじめとするマイクロブロッギング・サービスが挙げられる。マイクロブロッギング・サービスはオンライン・ソーシャル・ネットワーキング・サービスの一種であり、個人の情報プラットフォームとして機能する。本論文では、代表的なマイクロブロッギング・サービスである Twitter のリンクを含むつぶやきに着目し、その特徴を調査した。その結果、人間による投稿とボットによる投稿では投稿文字列長に顕著な差があること、ユニークユーザ数の多いニュースサイトが Twitter ではあまり注目されていないことなどが明らかになった。

キーワード マイクロブログ, コンテンツ解析, Web とインターネット, Web サイエンス

Analysis of Tweets including URLs on Twitter

Mitsuo YOSHIDA[†], Takashi INUI[†], and Mikio YAMAMOTO[†]

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

E-mail: †m.yoshida@mibel.cs.tsukuba.ac.jp, ††{inui,myama}@cs.tsukuba.ac.jp

Abstract Web 2.0 strengthened both tropism of the information to be able to put on Web and brought about new Web service. The microblogging service including Twitter is nominated for a kind of the Web 2.0 Web service. The microblogging service functions as a sort of personal information platform of the online social network service. In this paper, We pay our attention to tweets including URLs on Twitter, which is representative of the microblogging service, and investigated the characteristic. As a result, we show that news sites with much number of the unique users does not attract attention on Twitter, and the number of retweets is not even remotely related to the number of URLs posted, and so on.

Key words Microblogging, Analysis of Content, Web and Internet, Web Science

1. はじめに

インターネットが普及した今日、様々なサービス事業者が数多くの Web サービスをユーザに提供している。2005 年に提唱された“ Web 2.0 ”は、Web 上における情報の双方向性を強化するものであり [1]、近年では、オンライン・ソーシャル・ネットワーキング・サービス (Online Social Network Service, 以下「SNS」と呼ぶ) と呼ばれる社会的なネットワークをインターネット上で構築するサービスが提供されるようになってきた。SNS といえば、mixi^(注1) や GREE^(注2) のように招待制サービスとみなされることもあるが、招待機能は社会的なネットワークを構築しやすくするための機能の一種にすぎず、招待機能の

有無は SNS であるか否かの指標とならない。SNS には、そのサービスが情報プラットフォームとして機能するという特徴がある [2]。SNS の一形態として、マイクロブロッギング・サービス (Microblogging Service, 以下「マイクロブログ」と呼ぶ) が挙げられる。マイクロブログとは、チャットとブログの間のようなサービスであり、そのコンテンツは急増している。マイクロブログのコンテンツは、単なる生活記録や備忘録、独り言など、いわゆる「個人的」なコンテンツが多いといわれている [3]。本論文では、代表的なマイクロブログである Twitter^(注3) を取り上げ、分析対象とする。

Twitter (ツイッター) とは 2006 年 7 月に開始されたマイクロブログである。1 回のおつぶやき (公式には「ツイート」「tweet」と呼ぶが本論文では「つぶやき」と呼ぶ) が 140 文字に制限

(注1): <http://mixi.jp/>

(注2): <http://gree.jp/>

(注3): <http://twitter.com/>



図 1 ユーザ名 ceekz のタイムライン画面 . ceekz がフォローしているユーザのつぶやきが表示されている .

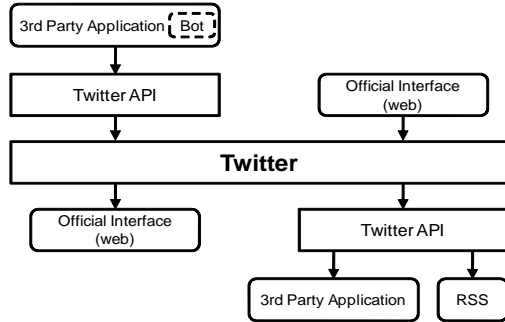


図 2 Twitter に投稿されるつぶやきの流れ (入出力インタフェース) を示す . 入出力ルートがそれぞれ複数存在する .

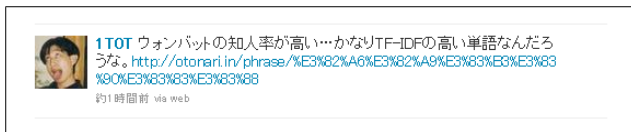


図 3 リンクを含むつぶやきの例 . 後半に URL を記述し , 前半にその URL に対するコメントやリンク先のタイトルを記述する傾向がある .

されていることが大きな特徴であり , たった今起こっていることを気軽に投稿できるように設計されている . 140 文字の制限は携帯電話のショート・メッセージ・サービス (SMS) をモデルとしている^(注4) . また , mixi などに見られる「友達」機能に相当する following/followers という機能が設けられている . 図 1 はユーザ名 ceekz のタイムライン画面である^(注5) . このページでは , ceekz がフォローしているユーザ (following) の最新のつぶやきを一覧することができる . 図 1 からわかるとおり , following の最新のつぶやきを一覧する画面の上部に , 自身のつぶやきを投稿するフォームがあるのも Twitter の大きな特徴である .

Twitter に投稿されるつぶやきの流れを表現したのが図 2 である . Official Interface とは図 1 のような , Twitter が公式に提供している投稿フォーム及び閲覧ページである . Twitter API とは , Twitter の情報の入出力を外部アプリケーション向

表 1 日本語リンクデータセット (jl データ)

期間	2009 年 5 月 8 日 2:10 ~ 12 月 2 日 17:40
データ概要	リンク (URL) を含む日本語のつぶやき
収集ページ	Twitter Search ^(注8)
収集間隔	10 秒毎
つぶやき数	19,699,831
ユーザ数	273,789

表 2 Public Timeline データセット (pt データ)

期間	2009 年 12 月 1 日 0:44 ~ 2010 年 2 月 12 日 23:59
データ概要	Public Timeline に出現するつぶやき
収集ページ	Twitter Timeline ^(注9)
収集間隔	2 分毎 (20 件)
つぶやき数	983,542
ユーザ数	683,441

けに提供するインタフェースであり , 第三者による 3rd Party Application 開発を容易にする . Official Interface を経由して投稿されるつぶやきは , 全てのつぶやきの 40%程度であり , 半数以上のつぶやきが Twitter API を介して行われる^(注6) . また , Twitter では情報の入出力インタフェースが多彩であるほかに , サービスの利用形態も多彩である . 公式には「いまだうして? (What 's happening?)」に答えるものであるが , それ以外にも「日記帳」「ソーシャル・ブックマーク (SBM)」「チャット」など様々な形で利用されている .

本論文では , Twitter を「ソーシャル・ブックマーク (SBM)」として利用しているユーザがいることを想定し , 図 3 のようなリンクを含むつぶやきに着目した Twitter の分析結果を報告する . なお , 本研究の目標は , 実際のソーシャル・ブックマーク^(注7)との比較を行い , その類似点などを明らかにすることであるが , 本論文ではそこまで立ち入らず , Twitter に投稿された URL の傾向などを報告する . 次節以降 , URL の投稿はプログラムによる自動投稿である比率が高いこと , 人間による投稿とプログラムによる自動投稿では投稿文字列長に顕著な差があること , ユニークユーザ数の多いニュースサイトが Twitter ではあまり注目されていないことなどを明らかにしていく .

2. 使用するデータセット

本論文では , 表 1 と表 2 に示した 2 種類のデータセットを利用している . 日本語リンクデータセット (以下「jl データ」と呼ぶ) は , リンク (URL) を含む日本語のつぶやきの集合である . また , Public Timeline データセット (以下「pt データ」と呼ぶ) は , Public Timeline に出現したつぶやきの集合である . jl データ , pt データともに , つぶやきを行ったユーザの情報も取得している . なお , 日本語のつぶやきは「ひらがな」または「カタカナ」のいずれかを 1 文字以上含むつぶやきと定義した .

(注6) : 詳細は 4.1 節で述べる .

(注7) : はてなブックマーク (<http://b.hatena.ne.jp/>) などが挙げられる .

(注8) : <http://search.twitter.com/>

(注9) : <http://twitter.com/public.timeline>

(注4) : Why 140 characters?

<http://business.twitter.com/twitter101/> (cited 2010-02-15)

(注5) : <http://twitter.com/ceekz> (cited 2010-02-15)

jl データは、Twitter を「ソーシャル・ブックマーク (SBM)」として利用されていることを想定して収集を行った。投稿されるつぶやきの数が非常に多いため、Twitter に投稿された該当するつぶやきが全て収集できているとは限らない。pt データは、Twitter のつぶやき全体をランダムサンプリングすることを目的に行った。jl データは日本人ユーザによるものが大半であると想定できることにに対し、pt データは世界各国のユーザによるものであることに注意する必要がある。

pt データを用いて、URL を含むつぶやきの割合と、日本語のつぶやきの割合を調べたところ、それぞれ 20.0% と 11.7% であった。このことから、jl データは Twitter のつぶやき全体のうち 2% 程度の収集であると考えられる。

jl データの各つぶやきには必ず URL が含まれている。19,699,831 件のつぶやきから抽出された URL は、20,073,051 件であり、重複を除外すると 13,792,809 種類であった。さらに、アクセス先が存在しない URL を除外すると 12,764,603 種類となった。

本研究では、プログラムによる自動投稿であると推測されるつぶやきを区別して集計を行う場合がある。図 2 に示した Twitter API を経由して投稿するアプリケーションは、図 4 に示す TweetDeck^(注10) のような Official Interface の代替となるアプリケーションと、twitterfeed^(注11) のようなプログラムによる自動投稿を行うアプリケーションとに大分できる。本論文では、後者のようなアプリケーションをボットと呼ぶ。ボットの判定は非常に難しい課題であるが、投稿元ソースをもとにボットの判定を行った。図 5 は Official Interface における投稿元ソースの表示例である。本論文では、投稿元ソースが「API」^(注12) 「twitterfeed」 「FriendFeed」 「Perl Net::Twitter」 「Twitter4J」 「twbot.rb」 「Twitter4R」 「Tweetbots」 「PEAR Services.Twitter」 「RSS2Twitter」 のいずれかである場合をボットによる投稿であると定義した。これらのソースは、jl データ及び pt データの出現頻度上位 100 ソースの中から筆者らによる私見により選定した。

3. 一般的な Twitter の分析

3.1 登録ユーザ数の増加

2008 年に行われた Twitter の分析 [2] では、約 10 万ユーザのデータを用いていたが、pt データには約 70 万ユーザのデータが存在する。このことから、最近の Twitter のユーザ登録数が急増していると考えられる。図 6 はユーザ登録数を年月別に示したグラフである。このグラフからわかるとおり、2009 年に入ってからユーザが急増しているが、2009 年後半になると登録数が減少しはじめる。

3.2 つぶやき数の増加

ユーザ数の増加に伴い、つぶやきの投稿数も増加していると考えられる。図 7 は、1 日あたりの投稿数がどのように伸びて



図 4 Official Interface の代替となる 3rd Party Application の例 (TweetDeck)。

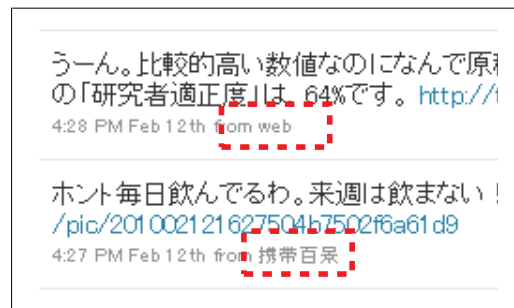


図 5 Official Interface における投稿元ソースの表示例 (破線部分)。この例では、上の投稿の投稿ソースは「web」であり、下の「携帯百景」である。

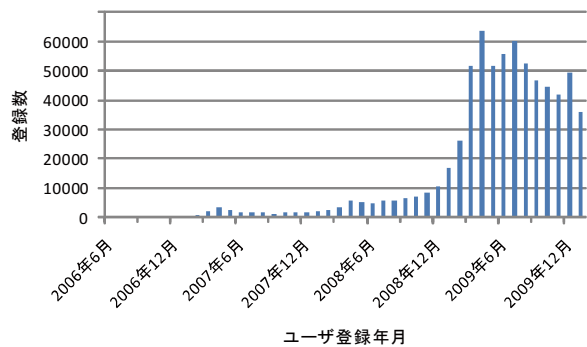


図 6 pt データをもとにした Twitter の年月別ユーザ登録数グラフ。2009 年に入ってからユーザ登録数が急増し、その後減少傾向にある。

いるのかを示したグラフである。グラフより、1 日あたりの投稿数が半年で 6 倍以上に伸びており、マイクロブログのコンテンツが急増していることを裏付ける。また、ボットによる投稿が急増している現象は確認できず、人間が作成したコンテンツが急増しているといえる。11 月前半に投稿数が一端減少しているのは、重複するつぶやきのブロックが開始されたことが影響していると考えられる^(注13)。

(注10): <http://www.tweetdeck.com/>

(注11): <http://twitterfeed.com/>

(注12): Twitter API 経由での投稿(図 2)の初期値。3rd Party Application が投稿元ソースを指定しない場合は「API」となる。

(注13): <http://techcrunch.com/2009/10/14/>

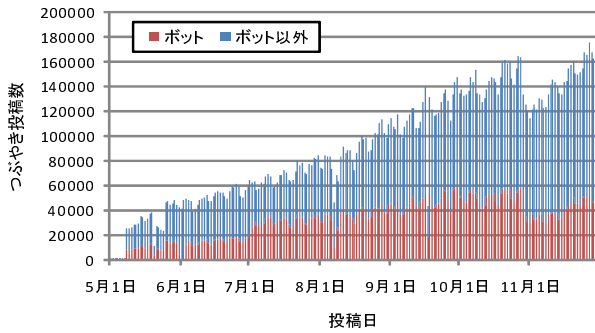


図 7 jl データをもとにした 2009 年の日別投稿数のグラフ。1 日あたりの投稿数が半年で 6 倍以上に伸びている。また、ボットによる投稿が全体の投稿数に大きな影響を与えていない。

表 3 投稿元ソースの上位 10 位 (%)

jl データ		pt データ	
web	22.02	web	40.89
API *	12.70	UberTwitter	7.59
twitterfeed *	12.64	TweetDeck	6.89
Tween	4.61	API *	4.93
movatwitter	4.24	Echofon	4.40
FriendFeed *	4.11	twitterfeed *	3.57
Echofon	3.52	mobile web	3.29
Perl Net::Twitter *	3.26	txt	3.08
Tumblr	2.77	Tweetie	2.01
TweetDeck	2.59	Twiterrific	1.45

投稿元ソースがボットの場合は、末尾に「*」を付与した。

4. ボットに着目した Twitter の分析

4.1 投稿元ソースの分布

2 節で述べたとおり、各つぶやきには投稿元ソースの情報が付随している。表 3 は、つぶやきがどの経路で投稿されたかを示した表である。jl データと pt データでは、投稿元ソースが web (Official Interface) である比率が大きく異なっている。通常 (pt データ)、4 割程度が web から投稿されるに対し、URL を含むつぶやき (jl データ) になると 2 割程度しか web から投稿されない。ボットによる投稿比率を調査すると、jl データでは 35.02% であるに対し、pt データでは 9.01% であった。URL を含むつぶやきから流行情報などを獲得する場合は、ボットの除外が重要であると考えられる。

4.2 つぶやき文字列長の分布

ユーザが URL を投稿する際、リンク先の情報をユーザに示すために、リンク先のタイトルやリンク先に対するコメントを併記する。jl データをもとに作成した、ボット以外の URL をのぞいたつぶやき文字列長グラフを図 8 に、ボットのみをグラフを図 9 に示す。グラフより、ボット以外の場合、10 文字 ~ 20 文字と比較的短いつぶやきが多く、「一言」のつぶやきが多い傾向が示唆される。

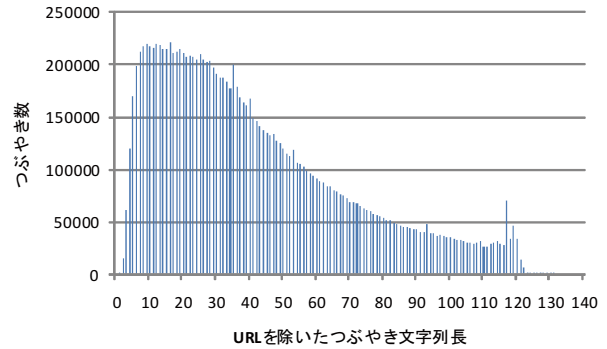


図 8 ボット以外の URL を除いたつぶやき文字列長のグラフ。12 付近が山のピークとなり、平均文字列長は 41.51 である。

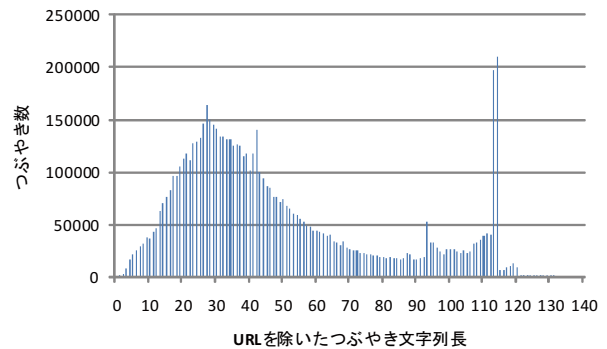


図 9 ボットのみ URL を除いたつぶやき文字列長のグラフ。27 付近が山のピークとなり、平均文字列長は 48.89 である。

図 9 の最頻値となっている 113 文字 ~ 114 文字のつぶやきは、「テキストの途中.. http://www.example.com/」のように「..」で文が終了するなど、文が途中で打ち切られている傾向があった。ボットはテキストの自動要約を行うことができず、URL を含め、丁度 140 文字になるように超過部分を単純に削って投稿した結果である可能性が高い。

テキストが長ければ長いほど、そのつぶやきを見たユーザに多くの情報を与える。また、Twitter のつぶやきにはそれぞれパーマリンク (Permalink) が設けられており、テキストが長いほど外部の検索エンジン (Google など) で検索されやすくなる。これらから、ボットはインターネット利用者に対して多くの情報を与えようとしていると考える。

5. リンクに着目した Twitter の分析

5.1 投稿された URL の FQDN 分布

Twitter に投稿された URL の傾向を知るために、jl データを使い FQDN (Fully Qualified Domain Name, 完全修飾ドメイン名) の分布を調べた。FQDN とは、完全な形で記述されたドメイン名であり、例えば「http://www.example.com/」の FQDN は「www.example.com.」である。なお、本論文では、ルートノードを示す最後のドットは省略する。

Twitter では、1 回のおつぶやきが 140 文字に制限されているため、投稿される URL は「URL 短縮化サービス」を利用したも

のが大半になると推測できる。URL 短縮化 (URL shortening) とは、長い文字列の URL を短い文字列の URL に変換することである。変換後の短い URL にアクセスすると、元の長い URL にアクセスすることができる。2010 年 2 月現在、Twitter では、ある文字列長以上の URL が投稿された場合、自動的に bit.ly^(注14) を用いて URL 短縮化が行われる^(注15)。

Twitter に投稿された URL から抽出した FQDN は 241,261 種類であり、短縮化された URL を復元した、最終的なアクセス先 URL から抽出した FQDN は 374,548 種類であった。

最終的なアクセス先 URL から抽出した FQDN の上位 30 件を表 4 に示す。ポット以外を集計した上位 5 件は、いずれも画像投稿サービス及び動画投稿サービスであり、ニュースなどよりも娯楽的なサービスが好まれているといえる。

訪れるユーザ数が多いウェブサイトほど Twitter に URL が投稿される可能性が高いと考え、ウェブサイトの媒体情報が入手でき、表 4 にも出てくる毎日新聞 (mainichi.jp) と朝日新聞の比較を行った。自社発表による月間ユニークユーザ数は、毎日新聞が 1,723 万^(注16)、朝日新聞 (www.asahi.com) が 900 万^(注17)であり、また、ネット視聴率白書においても同様の差があることが報告されている [4]。しかし、表 4 からわかるとおり、毎日新聞よりも朝日新聞の URL の方が Twitter に投稿されやすい傾向がある。このことから、Twitter 利用者層に大きな偏りがある可能性が示唆される。

URL 転送先の調査は 2009 年 12 月 12 日に行った。このため、2 節で述べたとおり 7.45% の URL がアクセス先が存在しなくなっている。アクセス先が存在しなくなっている比率がポットによる投稿と、それ以外の投稿ではどの程度差があるかを調べたところ、ポットによる投稿では 9.28% がアクセス不能であるに対し、ポット以外の投稿では 5.42% であった。ポットによる投稿の方が、リンク切れを起こす可能性が高いと考えられる。

5.2 URL 投稿と Retweet の関係

Twitter コミュニティには、Retweet と呼ばれる他人のつぶやきを再配信する文化がある。2010 年 2 月現在では、公式機能として取り入れられているが、本研究のためのデータ収集時には公式機能として採用されていなかった。Retweet は「再配信の際のコメント RT @UserName 他人のつぶやき」を基本的な書式としている^(注18)。「再配信の際のコメント」は省略される場合があるほか、字数制限の兼ね合いで「他人のつぶやき」が抜粋になる場合もある。また、Retweet を Quote Tweet と呼ぶ運動も一部で行われている (省略表記は QT)^(注19)。本論文では、Quote Tweet も Retweet の一種であるとみなす。

表 5 Retweet されやすいユーザの上位 5 件

ユーザ名	Retweet 回数	つぶやき数	フォロワー数
mainichijpedit	23,322	3,551	225,564
47news	22,578	29,035	4,874
hatebu	17,856	5,885	15,146
watch_akiba	12,340	5,303	10,290
asahi	11,484	2,885	238,046

Retweet による URL 投稿の割合を調べたところ、8.89% が Retweet によるものであった。ポット以外の投稿に絞れば 12.95%、ポットのみの投稿に絞れば 1.44% が Retweet によるものであり、相対的にポットから Retweet される割合は低いことがわかるほか、ユーザが URL を拡散しようとする動きが認められる。

Retweet されやすいユーザの上位 5 件を表 5 に示す。上位から順に、毎日新聞、共同通信 (非公式)、はてなブックマーク (非公式)、AKIBA PC Hotline!、朝日新聞が並ぶように、ニュース等を一方的に配信するユーザが上位に入っている。ただし、表 4 とあわせて見ればわかるとおり、Retweet されやすい媒体が必ずしも URL が多数投稿されるとは限らない。Retweet 回数 1 位の毎日新聞はつぶやき数 (ポット除外) で 29 位 (mainichi.jp)、5 位の朝日新聞はつぶやき数で 18 位 (www.asahi.com) になるなど、Retweet 回数は関連 URL の投稿にあまり影響を与えていないと考えられる。また、つぶやき数が多ければ多いほど、フォロワー数が多ければ多いほど、つぶやきがユーザに認知されやすい。しかし、表 5 からわかるとおり、認知されやすさが Retweet に直接繋がっておらず、投稿されたつぶやきの内容の影響が大きいと考えられる。

6. 関連研究

マイクロブログに関する研究は、近年数多く行われるようになってきた。本節では、マイクロブログ (特に Twitter) を対象とした研究について紹介する。

Krishnamurthy ら [2] は、2008 年 1 月時点での Twitter のユーザ特性について報告を行った。Twitter を利用するユーザには、情報を提供するものが主の「broadcasters」、知人とのコミュニケーションを中心行う「acquaintances」、閲覧を中心とした「miscreants or evangelists」の 3 グループがあると報告している。pt データを用いた追試により、同様の 3 グループを確認した。また、彼らは投稿元ソースの調査も行っており、約 60% が web (Official Interface) から投稿されていると報告している。pt データを用いた追試では、4.1 節に示したとおり、web からの投稿は 40% 程度まで減少している。Twitter の普及に伴い、3rd Party Application のバリエーションが増えたものと考えられる。

Courtenay ら [5] は、つぶやきの言及行為に着目した分析を行っている。他のユーザに対する言及は、つぶやき中に「@UserName」を含むかどうかで確認することができ、約 30% のつぶやきに「@」が含まれていると報告した。pt データを用いた追試では、47.2% のつぶやきに「@」が含まれており、言及行為

(注14): <http://bit.ly/>

(注15): 変換基準値は非公開であり、また、たびたび変更されている。

(注16): http://mainichi.jp/info/etc/pdf/mainichijp_baitai0904.pdf (cited 2010-02-15)

(注17): <http://www.asahi.com/advertising/guide.html> (cited 2010-02-15)

(注18): ユーザによる文化であり、厳密な形式は定義されていない。例えば「再配信の際のコメント RT 他人のつぶやき (@UserName)」とする流儀もある。

(注19): http://www.greenspace.info/mt/2009/10/20/_rt.qt.html (cited 2010-02-15)

表 4 最終的なアクセス先 URL から抽出した FQDN の上位 30 件 (%)

つぶやき数 (ボット以外)		つぶやき数 (ボットのみ)	
twitpic.com	11.90	www.nicovideo.jp	5.80
f.hatena.ne.jp	5.59	twitter.com	5.65
www.nicovideo.jp	3.97	ameblo.jp	2.43
movapic.com	3.82	friendfeed.com	2.27
www.youtube.com	3.43	d.hatena.ne.jp	2.04
blog.livedoor.jp	1.74	book.akahoshitakuya.com	1.85
d.hatena.ne.jp	1.68	detail.chiebukuro.yahoo.co.jp	1.66
live.nicovideo.jp	1.67	www.amazon.co.jp	1.56
twitter.com	1.16	blog.livedoor.jp	1.32
www.amazon.co.jp	0.99	tsushima.2ch.net	1.27
ameblo.jp	0.88	news.google.co.jp	1.18
brightkite.com	0.87	sankei.jp.msn.com	1.13
www.pixiv.net	0.80	yutori7.2ch.net	0.85
www.ustream.tv	0.75	sinkan.net	0.84
www.flickr.com	0.72	dailynews.yahoo.co.jp	0.83
www.itmedia.co.jp	0.69	www.plurk.com	0.79
bcphotoshare.com	0.51	japan.cnet.com	0.75
www.asahi.com	0.50	fieldnotes.sytes.net	0.75
sankei.jp.msn.com	0.49	www.youtube.com	0.74
ja.wikipedia.org	0.48	weather.livedoor.com	0.73
gigazine.net	0.41	dic.nicovideo.jp	0.70
www.47news.jp	0.39	mixi.jp	0.69
maps.google.com	0.37	jp.reuters.com	0.67
japan.cnet.com	0.37	www.47news.jp	0.63
twitter.g.hatena.ne.jp	0.36	live.nicovideo.jp	0.62
natalie.mu	0.28	www.itmedia.co.jp	0.58
alfalfa.livedoor.biz	0.28	www.asahi.com	0.57
www.gizmodo.jp	0.27	news.searchina.ne.jp	0.49
mainichi.jp	0.26	mediamarker.net	0.48
tsushima.2ch.net	0.25	oshiete1.goo.ne.jp	0.48

が増加してきていると考えられる。

Bernard ら [6] は、マイクロブログではどの程度ブランドに言及するかの調査を行った。彼らは、ブランドに対する言及の 20% に感情情報が付いており、50% 以上がポジティブ、33% がネガティブな内容であったと報告している。

Maxim ら [7] は、マイクロブログに適した検索システムの提案を行っている。マイクロブログではコンテンツの単位が小さいため、検索されたキーワードが盛り上がった時間帯のつぶやきを、まとめて表示するシステムを提案した。

Owen ら [8] は、フィード (RSS) のフィルタリングにマイクロブログを利用するシステムを提案した。フィードの購読には、しばしば有益な情報を見逃してしまうという問題が存在する。彼らは、マイクロブログにおける following の最新のつぶやきをもとにフィードのフィルタリングを行うことで、その問題を解消するシステムを提案した。

Marc ら [9] は、地域、性別、趣味など属性別にマイクロブログのコンテンツを分析した。属性ごとにコンテンツの内容が大きく異なることを報告したほか、コンテンツ内容の可視化手法を提案した。

7. おわりに

本論文では、代表的なマイクロブログである Twitter の分析を行った。特に「ソーシャル・ブックマーク (SBM)」として利用しているユーザがいることを想定し、リンク (URL) を含むつぶやきを中心に分析した。

まず、一般的な Twitter の分析により、2 点明らかになった。1 点目は、2009 年前半から登録ユーザ数が急増しているが、後半になると陰りがみえることである。2 点目は、5 月から 11 月末までの 6 ヶ月間の間に、1 日あたりに投稿つぶやき数が 6 倍以上に伸びていることが明らかになり、マイクロブログのコンテンツの増加が裏付けられたことである。

次に、ボットに着目した分析により、2 点明らかになった。1 点目は、URL を含むつぶやきはボットによる自動投稿の割合が比較的高いことが明らかになり、流行情報等の分析にはボットの除外が必要になることである。2 点目は、ボット以外の投稿では 10 文字 ~ 20 文字の「一言」のつぶやきが比較的多いこと、ボットは自動要約ができないため、超過部分を単純に削って投稿されやすい傾向があることである。

さらに、URL を含むつぶやきの分析から 2 点の知見を得た。

1 点目は、ニュースなどよりも写真や動画などの娯楽的サービスの URL が投稿されやすい傾向があること、ユニークユーザ数の多いニュースサイトが Twitter ではあまり注目されていないことである。2 点目は、URL 投稿に占める Retweet の割合が 13%程度あり、ユーザが URL を拡散しようとする動きが認められることである。

今後、実際のソーシャル・ブックマークとの比較を行い、その類似点などを明らかにする。さらに、その結果を踏まえ、マイクロブログをはじめとする SNS が情報プラットフォームとして機能しているかを調査し、今後のインターネットのあり方を検討したいと考えている。

文 献

- [1] Tim O'Reilly. Web 2.0: Compact definition?
<http://radar.oreilly.com/archives/2005/10/web-20-compact-definition.html> (cited 2010-02-15), 2005.
- [2] Balachander Krishnamurthy, Phillipa Gill, and Martin Arlitt. A few chirps about twitter. In *Proceedings of the first workshop on Online social networks*, pp. 19–24, Seattle, WA, USA, 2008. ACM.
- [3] 岩木祐輔, ヤフトアダム, 田中克己. マイクロブログにおける有用な記事の発見支援. 第 1 回データ工学と情報マネジメントに関するフォーラム (DEIM2009), March 2009.
- [4] 衣袋宏美. ネット視聴率白書 2008-2009. 翔泳社, August 2008.
- [5] Courtenay Honeycutt and Susan C. Herring. Beyond microblogging: Conversation and collaboration via twitter. In *Hawaii International Conference on System Sciences*, pp. 1–10, Los Alamitos, CA, USA, 2009. IEEE Computer Society.
- [6] Bernard J. Jansen, Mimi Zhang, Kate Sobel, and Abdur Chowdury. Micro-blogging as online word of mouth branding. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, pp. 3859–3864, Boston, MA, USA, 2009. ACM.
- [7] Maxim Grinev, Maria Grineva, Alexander Boldakov, Leonid Novak, Andrey Syssoev, and Dmitry Lizorkin. Sifting micro-blogging stream for events of user interest. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pp. 837–837, Boston, MA, USA, 2009. ACM.
- [8] Owen Phelan, Kevin McCarthy, and Barry Smyth. Using twitter to recommend real-time topical news. In *Proceedings of the third ACM conference on Recommender systems*, pp. 385–388, New York, New York, USA, 2009. ACM.
- [9] Marc Cheong and Vincent Lee. Integrating web-based intelligence retrieval and decision-making from the twitter trends knowledge base. In *Proceeding of the 2nd ACM workshop on Social web search and mining*, pp. 1–8, Hong Kong, China, 2009. ACM.