

ソーシャルメディアにおける情報伝播ネットワークの可視化とアプリストアに与える影響についての統計的解析

細江 成洋^{1,a)} 菊池 祥太郎¹ 安田 孝美¹ 水野 政司²

概要: 昨今、ソーシャルメディアの特徴である、誰もが情報を発信でき、その情報が瞬時に広範囲に伝わるという特性から、口コミというものが期待されており、その重要性や効果に注目が集まっている。本研究では、ソーシャルメディアの中でも、マイクロブログサービス「Twitter」を取り上げ、Twitter上での口コミ効果が、どのくらい商品の売れ行きに関わっているのを調査するために、Apple社のモバイル端末iPhone・iPod touch向けアプリケーションのダウンロードサービスである「iTunes App Store」を対象とした。統計の結果、iTunes App Store ランキング急上昇前にはリツイートの増加が見られたが、ゲームなど特定のアプリにしか見られないことが判明した。さらに、ランキングが急上昇するには、リツイートの数よりも、リツイートの元となるつぶやきを発信した特定のユーザーが重要だということが明らかとなった。

キーワード: ソーシャルメディア、ツイッター、モバイルアプリ、統計、ソーシャルグラフ、マーケティング

Visualization spread of information network on Social media and statistical analysis about the influence to the application store

NARIHIRO HOSOE^{1,a)} SHOTARO KIKUCHI¹ TAKAMI YASUDA¹ MASASHI MIZUNO²

Abstract: Recently, which is a feature of social media, from the characteristic anyone can disseminate information, the information that transfer extensively instantly, a word of mouth is expected and attent to the effect and importance. In this study, among the social media, we focused on the micro-blogging service "Twitter". To investigate a word-of-mouth effect on Twitter, how much is related to the sales of goods, we targeted the download service application "iTunes App Store". As result of the study, the followings are found out. Increase in Retweet was observed in iTunes App Store rankings soaring before. However, it is only found in a particular application, such as the game. In addition, soar of ranking, a particular user who tweets that are based retweet than the number of retweet was proved to be important for ranking to be soaring.

Keywords: Social media, Twitter, Mobile application, Statistics, Social graph, Sales Strategy

1. はじめに

今日において、インターネットサービスの中でも「ソーシャルメディア」は急速に普及し、今や情報インフラとしての役割が注目されている。例えば、2011年3月11日に起こった東日本大震災では、緊急時の情報インフラとして

利用され注目を浴びたことが挙げられる。また、アラブの春と呼ばれる、2010年から2012年にかけてアラブ世界において発生した民主化運動でも、デモ活動に係るソーシャルメディアによる抗議の呼びかけに利用された [1]。

このようにソーシャルメディアは、誰もが情報を発信できる特徴と、情報が瞬時に広範囲に伝わる特性がある。加えて、情報の共有が容易な点も大きい。例を挙げると、マイクロブログサービス「Twitter」のリツイート、SNSサービス「Facebook」のいいね！ボタンの機能は、ユーザーが

¹ 名古屋大学大学院 情報科学研究科 社会システム情報学専攻

² クエリーアイ株式会社

^{a)} h-nari@nagoya-u.jp

情報の価値が高いと判断した時に、手軽に他者へその情報を伝達することができる。

さらに、ソーシャルメディアの発展に伴い、口コミというものが期待されている、マーケティングにおける口コミの効果は古くから関心を集めてきた主題であるが [2]、主にソーシャルメディアの登場により、その重要性や効果に再度注目が集まっている。事実、約 70% の人が、今までにネットクチコミを参考にして商品やサービスを購入している [3]。

本研究では、マイクロブログ「Twitter」での口コミ効果が、どのくらい商品の売れ行きに関わっているかを調べるために、売れ行きの指標として Apple 社のモバイル端末 iPhone・iPod touch・タブレット端末 iPad 向けアプリケーションのダウンロードサービスである「iTunes App Store」のダウンロード数ランキングの統計的分析を行うことで、両者との間にどのような相関が存在するかを明らかにする。さらに、その結果からソーシャルメディアマーケティングへの有効性についても検討したいと思う。

2. 関連研究

ソーシャルメディアの中でも、マイクロブログ「Twitter」を対象とした研究について紹介する。

吉田ら [4] は、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。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、また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を報告している。

Saez-Trumper ら [5] は、PageRank アルゴリズムに時間軸を導入した新たなアルゴリズムを提案している。Twitter ネットワーク内で、ある特定のトピックにおける情報拡散のピーク前にアラートを鳴らし、トレンドセッターと呼ばれる、トレンドや流行を打ち出す人を探し出す。また、有名人などのフォロワーが多いユーザーは、あるトピックに対して既に話題になっている場合、その話題になったものに対してほとんど関与していないことを報告している。

Kooti ら [6] は、Twitter の社会的慣習の出現について報告している。リツイート機能が正式に実装されるまでは、リツイートの表記記号が主に 7 つ (via, HT, Retweet, Retweeting, RT, R/T, recycle icon) の候補が挙げられ、Twitter ネットワーク内で複数のコミュニティに分かれ、様々な議論があったことを明らかにしている。

荒牧ら [7] は、Twitter のつぶやき内容からインフルエンザの流行を予測した。つぶやいた人物が、インフルエンザにかかっているかどうかを判別するために、ニュートラルネットワークの一種であるサポートベクターマシン (SVM) を利用している。その結果、Google が提供する「Google Flu Trends」というインフルエンザ予測サービスと同じくらいの精度が出た。

表 1 Twitter のデータセット

収集方法	各言語のパブリックタイムラインに表示されるつぶやき [9]
収集間隔	1 時間毎
対象国	日本
期間	2010 / 12 / 1 00 : 00 ~ 2012 / 9 / 30 23 : 00
集計対象	特定の URL を含むつぶやき

表 2 iTunes App Store のデータセット

収集方法	iTunes Search API [10]
収集間隔	1 時間毎
期間	2010 / 12 / 1 00 : 00 ~ 2012 / 9 / 30 23 : 00
集計対象	各ランキング・カテゴリの 1 位 ~ 300 位のアプリケーション
調査対象	モバイルアプリ有料総合・無料総合・トップセールスの 3 種類のカテゴリ

Bollen et al. ら [8] は、NY ダウ平均株価の上下予測を行うために、Twitter のつぶやき本文から、ネガティブとポジティブを判断する OpinionFinder と、6 種類の感情因子 (Calm, Alert, Sure, Vital, Kind, Happy) を分解する Google-Profile of Mood States (GPOMS) を利用して解析した。その結果、ダウ平均 (DJIA) との相関を取ることにより、87.6% の精度で日ごとの変動を予想できることを報告した。

3. データセット

本研究では、2 種類のデータセットを利用している、これらのデータセットは表 1 と表 2 に示した。

3.1 Twitter のデータセット

表 1 の集計対象、特定の URL を含むつぶやきは、以下の条件を満たすものを集計している。

- つぶやき本文中に、あるアプリケーションの固有 ID を含む iTunes App Store へのリンク URL が存在する。
- つぶやき本文中に、あるアプリケーションの固有 ID を含む iTunes App Store へのリンクを含むブログ記事の URL が存在する。

これに該当する、つぶやきの投稿時間、投稿者の Twitter アカウント名、つぶやき本文、つぶやき本文中に含まれるアプリケーションの固有 ID をデータベースに集計し、調査対象とした。

3.2 iTunes App Store のデータセット

表2の調査対象、各ランキング・カテゴリとは、以下のことを指している。

- ジャンル別に23種類 + 総合ランキングの計24種類に対して、料金体系別に3種類
- さらに、対応デバイスとしてiPhone / iPod touch 向けのアプリケーションと、タブレット端末 iPad 専用のアプリケーションの、2種類が分けられる

つまり、 $24 \times 3 \times 2 = 144$ 種類ものランキング・カテゴリが存在する。

本研究では、iPhone / iPod touch 向けのアプリケーションを対象としており、また、その中でも表2のように、有料総合・無料総合・トップセールスを調査対象としている。

4. 定義

本章では、統計的処理を行うにあたって、用いた定義を述べる。

4.1 欠損値

3.2にて集計したデータセットには、App Storeのシステムダウンなどにより、一部欠損値がある。こうした不連続になっている欠損時間については、対象となる欠損時間からみて一つ前の時間を検出し、その値を欠損時間の補間値とした(最近傍補間)。また、この補間処理は全てのアプリケーションIDに対して行った。

4.2 急上昇検出

各々のランキング・カテゴリにおける「急上昇」は、各々のアプリケーションの1時間前のランキングに対して、ランキングが10以上上昇した時間を採用し、またその中でも、ランキングの変動幅が大きいものを急上昇と定義した。

さらに、その急上昇時点から±24時間(前後1日)のランキング時間も収集し、急上昇前と後の計49時間分を統計データとして扱う。

4.3 リツイート

本研究では、Twitterのつぶやきの中でもリツイート(以下、RTと呼ぶ)に絞っている。リツイートは4.2と同じく、急上昇時点から±24時間(前後1日)のRTを収集し、急上昇前と後の計49時間分を統計データとして扱う。

5. 因果と相関

統計的処理を行う上で、以下3点の相関関係について述べる。

5.1 iTunes App Store ランキングと実ダウンロード数との因果関係

iTunes App Storeのランキングアルゴリズムが式(1)を用いられていると仮定する。この数式[11]は、Apple社の公式な発表によるものではないが、実際のモバイルアプリケーションの販売実績に基づくもの[12]であり、信頼できる数式だと十分考えられる。近い時間における実ダウンロード数が増えればランキングスコアが上昇し、結果としてランキングも上昇する、と考えられる。これにより、iTunes App Store ランキングと実ダウンロード数との間には、因果関係があると思われる。

$$S_i = \sum_{j=0}^{23} n_{i-j} * 8 + \sum_{j=24}^{47} n_{i-j} * 5 + \sum_{j=48}^{71} n_{i-j} * 5 + \sum_{j=72}^{95} n_{i-j} * 2 \quad (1)$$

5.2 Twitter の RT 数とアプリケーション参照数との因果関係

あるアプリケーションIDを含むURLが書かれたつぶやきのRT数が増加することで、そのURLが参照される確率が高まるため、両者との間には因果関係がある。

5.3 Twitter の RT 数と iTunes App Store ランキングとの相関関係

本研究では「ダウンロード数が増えれば、iTunes App Storeのランキングが増加する」という因果と、「アプリケーションへのURLを本文中に含むTwitterのつぶやきがRTされることで、そのアプリケーションへの参照機会が増える」という因果の両者について、その間にある相関関係を明らかにすることを目的としている。

6. 統計結果と考察

本章では、統計処理の結果と、その結果に基づいた考察について述べる。

検出できたランキングの急上昇点は以下の通りである。

有料アプリ・・・3968 (アプリ総数: 5874)

無料アプリ・・・4931 (アプリ総数: 4931)

トップセールス・・・3427 (アプリ総数: 3427)

6.1 時間帯別の RT 数

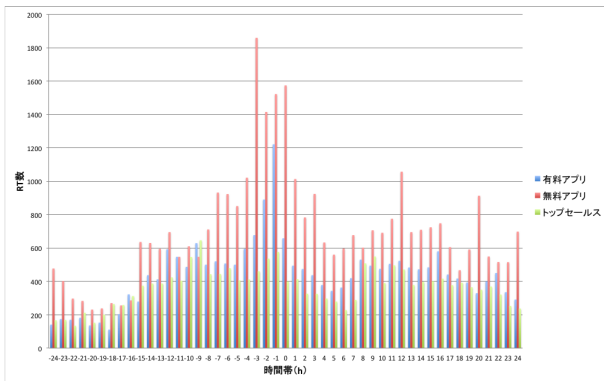


図 1 App Store ランキング急上昇時点から前後 24 時間以内に RT が起きた時間分布

図 1 は、App Store ランキング急上昇時点から前後 24 時間（合計 49 時間）を時間毎に RT 数を収集したものである。グラフより、全てのランキングアプリにおいて、17 時間前から急上昇時点に近づくに伴い、RT 数が増加して行くのを見て取れる。細かく見ると、有料アプリと無料アプリに関しては、3 時間前に極端な RT 数の増加が見られるのに対し、トップセールスはそれほど増加していない。また、7 時間後辺りから緩やかな上昇を見せているのは、ランキングが急上昇したことにより、話題になったことで RT 数が増加したのではないかと考えられる。

6.2 Twitter におけるユーザーの出現回数

有料アプリ	出現回数	無料アプリ	出現回数	トップセールス	出現回数
touch_lab	7654	touch_lab	8466	touch_lab	5852
appbank	1748	presentnowjp	3262	appbank	1200
famitsu	851	appbank	2318	famitsu	497
socialadpress	568	hatebu	1243	hatebu	458
hatebu	506	NEWS_0	419	iphonejoshi	395
twinabi	379	torino175_live	413	sasakitoshinao	296
wataru152	333	touhokujishin	395	kirinnao	231
torino175_live	299	NHK_PR	383	torino175_live	207
asip2k25	288	APPZ_JP	343	asip2k25	183
idanbo	260	iphonejoshi	343	APPZ_JP	179

図 2 App Store ランキング急上昇時点から前後 24 時間における出現回数が多い Twitter のユーザー（トップ 10 位）

図 2 は、App Store ランキング急上昇時点から前後 24 時間（合計 49 時間）の間に出現した Twitter のユーザーを集計したものである。グラフより、一番出現回数が多かったのは touch_lab であることが分かった。また、上位に出現したユーザー（appbank や famitsu など）は、主にモバイルアプリに関するニュースなどの情報を発信するユーザーである。個人ユーザーもいるが、IT・メディアの関係者であることが判明した。

6.3 RT 数とランキング差による回帰分析

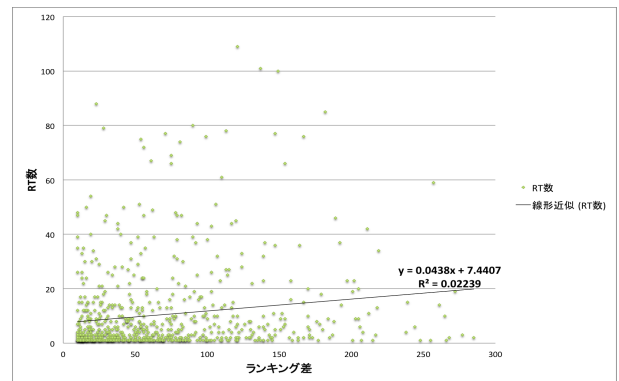


図 3 RT 数とランキング差による回帰分析（トップセールス）

ここでは、「RT 数が増えれば増える程、急上昇したランキングと急上昇する前のランキングとの差も広がるのではないか」という仮説を検証するために、RT 数とランキング差による回帰分析を行った。その結果、全てのランキング・カテゴリにおいて、決定係数 R^2 が 0 に限りなく近く、非常に当てはまりの悪い結果となり、効果的な予測ができないことが分かった。

さらに、RT 数は 24 時間前～急上昇時点の間の RT をカウントしているが、RT が一切行われぬモバイルアプリも多いことが分かり、想定していなかった結果が浮き彫りになった。図 3 のトップセールスでは、21.45% (735/3427) しか RT が抽出できず、また有料アプリ、無料アプリにおいても、20%強しか見られなかった。

6.4 RT 数が多いモバイルアプリのジャンル

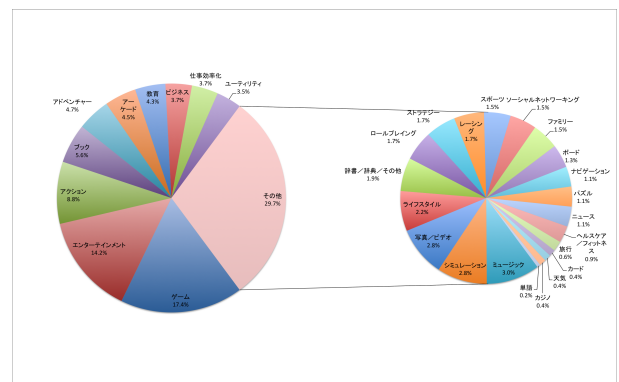


図 4 RT 数が 10 以上のモバイルアプリのジャンル（有料アプリ）

6.3 では、RT が一切行われぬモバイルアプリも多いことが分かった。しかしそれに対して、「RT が多いモバイルアプリはどのようなジャンルか」という仮説が出てきた。そこでその仮説を実証すべく、RT 数が 10 以上のモバイルアプリのジャンルを調査した。

一つのアプリには、以下の例の通り、複数のジャンルが

付与されている場合がある。今回は付与されている複数のジャンルも調査対象とした。

例: "genres":["ゲーム", "パズル", "アクション"] [13]

図4が有料アプリの調査結果であるが、ゲームとエンターテインメントのジャンルが多いことが分かる。また、細かく見ると、アクション、アドベンチャー、アーケードなど、ゲーム関連のジャンルが多いことが判明した。以上のことは、無料アプリ、トップセールスにおいても同様のことが言える。このことから、RT数が多い傾向のあるモバイルアプリは、ゲーム関連であることが明らかとなった。

6.5 中心性

6.2では、急上昇時点から前後2時間において、出現頻度が高いユーザーを調査したが、誰が重要人物なのかの特定ができないため、ネットワーク分析の中でも中心性の概念を使用することにした。使用した中心性指標は以下の通りである。

- 出次数中心性: 出次数が高ければ、情報の送り手として中心的な役割を果たす。
- 近接中心性: 近接性が高ければ、ネットワーク内の情報を素早く獲得する機会に富む。
- 媒介中心性: ネットワーク上における、複数のコミュニティを結ぶ架け橋として機能する。

これらの各中心性の高いユーザーを発見すれば、ネットワーク上において、新たな情報を獲得するのに有利なポジションに位置しているため、重要人物が浮かび上がる。また、Twitterのユーザー間のつながりは、有向グラフで示されるため、入次数中心性と出次数中心性が存在するが、ここでは「誰が多く発信しているか」を調査することを目的としたため、出次数中心性を採択した。

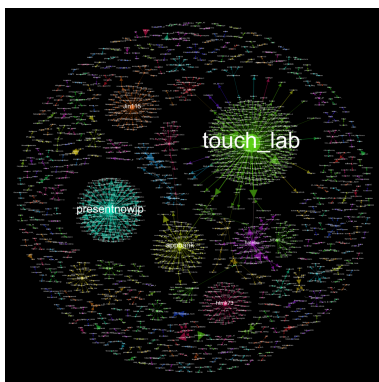


図5 急上昇時点における無料アプリのソーシャルグラフ (出次数中心性)

今回、各中心性指標を測定するにあたって、Gephi^{*1}というソフトウェアを利用した。Gephiは、あらゆる種類のネットワークを可視化するソフトウェアである。このソフトウェアを使い、24時間前から急上昇時点において1時間毎の各指標値のトップは誰なのかを調べた結果、全てのランキングカテゴリにおいて、ほぼ同様の傾向が見られた。

まず、出次数が一番高いユーザーは24時間前から急上昇時点まで、touch_labのユーザーがほぼ占めていることがわかった(図5)。次に、近接性が一番高いユーザーは、毎時違うユーザーが出現したため、入れ替わりが激しいことが分かった。最後に、媒介性が一番高いユーザーは、24時間前～大体14時間においてはユーザーの入れ替わりが激しいことが分かったが、それ以降～急上昇時点までtouch_labが独占していることが判明した。

これらから、touch_labというユーザーが情報伝播のネットワークの中で、重要人物であることが浮き彫りとなった。また、このユーザーは、情報を獲得するユーザーではなく、主にアプリを中心とした情報を発信するユーザーであることから、近接性が低いのではないかと推測される。

6.6 ヒートマップ

この節と次の6.7節では、iTunes App Storeのみを分析対象としたため、5.3で述べたような相関関係は明らかにしていない。

	月	火	水	木	金	土	日
0	4.05%	4.03%	6.45%	4.60%	2.62%	3.77%	9.18%
1	9.12%	7.98%	7.65%	5.58%	9.65%	14.53%	8.40%
2	3.62%	8.23%	5.10%	2.65%	9.01%	3.77%	3.58%
3	5.07%	5.97%	3.30%	4.88%	2.62%	1.82%	2.80%
4	3.04%	2.42%	1.50%	5.72%	2.17%	3.35%	0.78%
5	0.14%	2.42%	0.75%	2.09%	1.94%	1.82%	2.02%
6	4.49%	0.97%	4.50%	2.23%	1.25%	2.51%	5.13%
7	0.87%	0.65%	0.75%	1.53%	5.93%	4.89%	2.02%
8	1.45%	0.32%	3.30%	0.70%	0.91%	2.23%	0.62%
9	4.92%	1.94%	1.95%	3.63%	2.39%	4.47%	1.40%
10	2.03%	2.98%	3.30%	2.65%	3.53%	5.17%	0.62%
11	7.38%	2.42%	3.00%	4.32%	1.94%	1.54%	2.33%
12	1.74%	0.97%	2.25%	2.51%	3.65%	3.21%	4.98%
13	1.86%	5.16%	2.70%	4.46%	1.71%	3.07%	4.98%
14	5.79%	4.68%	5.70%	4.74%	4.56%	2.79%	3.73%
15	3.18%	1.45%	4.05%	4.32%	2.28%	5.31%	0.78%
16	1.16%	3.39%	2.85%	2.37%	1.37%	2.93%	0.47%
17	0.29%	0.97%	1.35%	1.29%	1.14%	1.12%	2.02%
18	4.34%	3.71%	2.95%	4.32%	4.56%	4.75%	4.20%
19	8.25%	8.87%	8.23%	8.51%	6.73%	6.84%	10.73%
20	5.93%	6.13%	6.45%	5.16%	7.18%	4.75%	7.00%
21	7.24%	9.68%	8.85%	8.79%	9.69%	8.24%	10.42%
22	5.50%	9.35%	6.45%	7.25%	6.61%	3.21%	3.89%
23	8.54%	6.13%	7.05%	5.72%	6.50%	3.91%	7.93%
合計	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

図6 曜日・時間別ランキング急上昇時のヒートマップ表示 (無料アプリ)

この節では、ランキングの急上昇はいつ起こるのかを調査した。図6は、赤色の濃さがランキングの急上昇の多さを示しており、反対に、赤色の淡さがランキングの急上昇の少なさを示している。結果、全てのランキング・カテゴリにおいて18時～3時の間が急上昇しやすいポイントであることが分かり、特にゴールデンタイムと呼ばれる、19時～21時台が最も急上昇しやすい時間帯であることが判明した。

*1 "Gephi". <https://gephi.org/>

6.7 ランキング急上昇 1 時間前のランキング分布

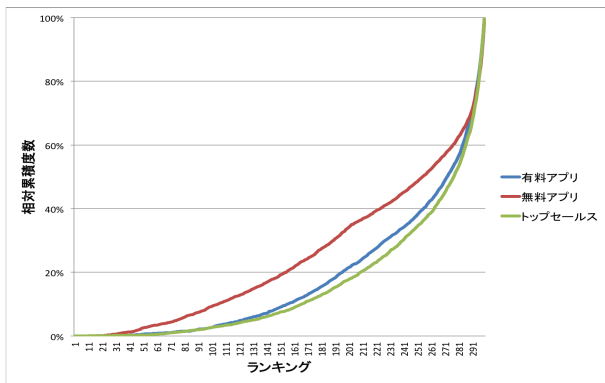


図 7 ランキングの急上昇が起きた 1 時間前のランキング分布 (相対累積分布)

この節では、「上位のモバイルアプリはランキングが上がりやすく、下位のモバイルアプリほど大きな変動を見せるのではないか」という仮説を実証するために、ランキング急上昇前のランキングの順位を調査した結果、急上昇しやすいランキング順位は 200 位台であることが判明した。また図 7 からは、以下のことがいえる。

- ランキング変動の約 60%以上が 210 位より下のランキング (無料アプリ)
- ランキング変動の約 60%以上が 250 位より下のランキング (有料アプリ・トップセールス)

以上より、下位のモバイルアプリほど大きな変動が起こることが明らかとなった。

7. 考察と今後の課題

6. の結果より、ランキング急上昇前には RT の増加が見られたが、ゲームを例とした特定のアプリにしか見られないことが判明した。また、RT 数が増加すれば、ランキングもそれに伴い増加するのではなく、むしろ RT 本文に登場する touch_lab のようなユーザーが重要だということがわかり、単純な RT 数の増加によって、情報伝播が進み、多くの人ダウンロードしたという事実は確認できなかった。さらに、ランキングの急上昇は主にゴールデンタイム時に起きており、ランキング位置が 200 位以下でないことが判明した。これらの結果から、ソーシャルメディアマーケティングを行う上で、口コミの広がりとアプリケーションの購入頻度の多さとは、必ずしも関係がないため、どのユーザーにモバイルアプリの宣伝をしようかが重要なポイントとなる。また、アプリケーションの種類によっては、情報伝播が起こらないことも考えられるため、アプリケーションをソーシャルメディア上で宣伝する場合には、ある程度の慎重さが求められると考える。

今後の課題としては 3 点ある。1 点目は、図 3 では、各ランキング・カテゴリに外れ値が存在する。その外れ値を

扱うかが今後のポイントになってくるため、今後の課題としたい。2 点目は、今回は RT の元となったツイートはいつ行われたかを検証していない。これが明確になることで、元のツイートから何時間後に RT による情報伝播が起こったのかを推測することができると思われる。今後の課題としたい。3 点目として、分析した期間には東日本大震災という特別な事象が含まれている。こういった特異な出来事が起きた場合、人は情報に敏感になると予測できるため、その出来事の前後の期間を取り除く、あるいはこの出来事の前後の期間だけを分析対象として扱えば、今回とは違う結果になることが予想される。さらに、また今回の結果との比較検証を行うことで、人が情報に敏感になった時にどのような行動を示すかが浮き出てくるのではないかと考える。

参考文献

- [1] 総務省. "平成 24 年版 情報通信白書". <http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h24/html/nc1212c0.html>
- [2] ネット上の消費者情報探索とネット・クチコミのマーケティング利用, <http://www.yhmf.jp/pdf/activity/adstudies/vol.20-01-03.pdf>
- [3] 「ネットクチコミ」に関する意識調査 <http://www.softbankhc.co.jp/press/080219.html>
- [4] 吉田光男, 乾孝司, and 山本幹雄. "リンクを含むつぶやきに着目した Twitter の分析." (2010).
- [5] Saez-Trumper, Diego, et al. "Finding trendsetters in information networks." Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2012.
- [6] Kooti, Farshad, et al. "The Emergence of Conventions in Online Social Networks." ICWSM. 2012.
- [7] 荒牧英治, 増川佐知子, and 森田瑞樹. "Twitter catches the flu: 事実性判定を用いたインフルエンザ流行予測." 情報処理学会研究報告. SLP, 音声言語情報処理 2011.1 (2011): 1-8.
- [8] Bollen, Johan, Huina Mao, and Xiaojun Zeng. "Twitter mood predicts the stock market." Journal of Computational Science 2.1 (2011): 1-8.
- [9] "Twitter - public timeline". https://twitter.com/public_timeline
- [10] Apple, "iTunes search api". <http://www.apple.com/itunes/affiliates/resources/documentation/itunes-store-web-service-search-api.html>
- [11] "How to successfully market your iPhone Application? #appstudy". <http://www.slideshare.net/misteroo/how-to-market-your-app>
- [12] "fabernovel.com". <http://www.fabernovel.com/>
- [13] "Super Monkey Ball". <http://itunes.apple.com/lookup?id=281966695&country=JP>