

スマートフォンにおけるコンテンツ閲覧と共有行動の分析

Analyzing Contents Viewing and Sharing User Activity On Smart-phone Application.

関 喜史 *1*2

Yoshifumi Seki

松尾 豊 *2

Yutaka Matsuo

*1株式会社 Gunosy

Gunosy Inc

*2東京大学

The University of Tokyo

When a user submits a web content to social media, he views some web contents and decides to submit it. Most studies of social media focused what a user submit, but didn't focus what he view before he submit to social media. In this paper, we analyze difference between viewing and sharing using user activity log in mobile news application. In result, we reveal that tendency of sharing different by topic of content. Therefor, we conclude that content a user viewed and content a user shared do not always go together.

1. 背景

ソーシャルメディアの普及に伴い、ウェブ上のコンテンツにとってどれだけソーシャルメディアに投稿されるかが重要な指標になってきている。ソーシャルメディアに対してリンクや画像などの他のユーザやメディアが生み出したコンテンツを投稿する行動は通常の投稿とは区別され「シェア」と呼ばれており、Facebookの機能がその語源となっている。Facebookには「いいね」と「シェア」の2つの機能が存在するが、「いいね」がその投稿に対しての共感を表すことに対して、「シェア」は「この情報を自分の友人にも知らせたい」という時に使用するものと言われており[Globis]、この「シェア」が現在ではソーシャルメディアに対して友人に対して広めたいコンテンツを投稿することを指す言葉となっている。近年注目を集めているバイラルメディアと呼ばれる新興メディアはシェアされやすいコンテンツを提供することにより、大手のポータルサイトなどに頼ることなくソーシャルメディア上から大量のページビュー(PV)を獲得することに成功しており、近年注目を集めている。

学術分野においてもシェアされたコンテンツは様々な研究の題材になっている。シェア件数の時間的推移からその出来事の注目度の推移の仕方を分析しようという試みや[Yang 11]、コンテンツのシェア数自体を推定しようという試み[Bandari 12]、またユーザのシェアの傾向からユーザのプロファイリングを行い、情報推薦に活用しようというものなどその方向性は様々である[Abel 11]。

しかしそれらの研究はほぼすべてにおいてユーザが何をシェアしたかということのみに着目しており、そのシェアに至るまでのユーザの行動については触れられていない。しかし大半のユーザはシェアするコンテンツ以外にも多くのコンテンツに触れており、その中からシェアするコンテンツを選別している。その選別は単にコンテンツが優れていることや、そのコンテンツに関心を持ったかという単純なものではないと我々は考えている。例えばソーシャルメディアにはユーザ個人が他人からどう見られたいかを考慮するという側面があることを踏まえれば、そのコンテンツをシェアするかどうかは「そのコンテンツに関心があるという事実」をシェアしたいかという判断が含ま

れる。そのためそのコンテンツに関心があることを知られなければ関心を持っていたとしてもシェアすることは無い。このようにユーザのシェアに至るまでの背景を考慮すると単純にシェアされた数がより多いほうがコンテンツとして人気があるということや、シェアされた数の時系列的な変化が注目度の変化を表すということは必ずしも言うことはできず、もちろんシェアしているコンテンツがユーザの興味関心を表しているとも言うことはできないと我々は考える。

本研究ではこのような背景から、スマートフォンニュースアプリにおける行動履歴を用いてシェアとその背後にあるユーザの行動について分析を行った。その結果としてニュースのシェアのされやすさはコンテンツによって異なり、ニュースのトピックの違いによってその傾向が異なることを明らかにした。またニュースアプリでは配信されたニュースを一覧できる画面から記事を読まなくてもニュースをシェアすることができる。このようなシェア行動が通常の読んだ上でシェアする行動とどのように異なるのかについても分析し、コンテンツの特性がどのように影響を及ぼすのかを示した。その上でユーザがクリックするコンテンツとシェアするコンテンツが必ずしも一致しないことを示し、コンテンツのトピックに影響を受けていることを明らかにした。

本稿の構成は以下の通りである。まず2.章で本研究で用いたデータについて紹介する。3.章で閲覧とシェアについてデータの概要を述べ、3.1章でコンテンツのトピックごとに閲覧とシェアの特性について分析し、4.章でユーザの閲覧とシェアの異なり方についてまとめ、5.章で本稿のまとめを行う。

2. 用いたデータについて

本章では本研究で利用したデータについて紹介する。本研究では株式会社 Gunosy の提供しているスマートフォンアプリケーションであるグノシーの行動データを用いる。グノシーは2015年3月時点でダウンロード数800万を記録している日本で最もよく知られたニュースアプリの1つである。今回は2013年のある期間にシェアを活発に行っていたユーザ12,768人の行動を分析対象とし、

- ニュースを閲覧した
- ニュースをシェアした (Facebook または Twitter にニュースのリンクを投稿した)

連絡先: 関 喜史, 株式会社 Gunosy, 東京都港区六本木 6-10-1 六本木ヒルズ森タワー 37F, 03-6455-4560, yoshifumi.seki@gunosy.com

上記の2つの行動について取り扱う。シェアに関してはユーザが対象アプリケーション上でシェアを試みたことを示すものであり、それが実際に投稿されたかや、その後削除されたかなどの対象アプリケーション外での行動については記録されていない。行動データは(ユーザ識別子, ニュース識別子, 行動種別, 行動時間)からなるレコードで構成されている。ユーザ識別子は匿名化されており, 上記の情報と公開されている情報を組み合わせたとしてもユーザ個人を特定することはできない状態になっている。

記事についてはそれぞれの記事識別子に対してタイトルとリード文が与えられている。本論文では分析の中で記事単位について言及する際にはタイトルを直接使用せずおおまかな内容の記述のみに留める。

3. 閲覧とシェアの概要

本章では行動データを用いて, 閲覧・シェアそれぞれについて簡単な集計を行い概要を述べる。行動データのうち86.2%が閲覧であり, 残りの13.8%がシェアという構成になっている。このシェアのうち30.6%は閲覧していないニュースに対するシェアであり, 閲覧をせずにシェアを行ったことのあるユーザ数は9,055人である。つまり今回の対象ユーザの内70.9%のユーザがニュースを読まずにシェアをしたことがあるということである。

次に実際にどのようなニュースが閲覧・シェアされていたのかをみる。

表 1: 閲覧数上位 5 件のニュース

1	企業経営者の不適切な発言についての記事
2	誤って子供がスマートフォンサイトで車を購入してしまった父親についての記事
3	iOS7 発表前の憶測をまとめた記事
4	男性から印象の悪い女性の服装についてのランキング
5	Web サービスやアプリについてまとめて紹介した記事

表 2: シェア数上位 5 件のニュース

1	ある地方に IT 企業の サテライトオフィスが集まっているという記事
2	報連相に関する記事
3	スマホレジについての記事
4	日本企業の課題についての記事
5	ある高校のサッカー部の PV を紹介した記事

表 1 に閲覧数の高いニュース上位 5 件, 表 2 にシェア数の高いニュース上位 5 件を示す。閲覧数の上位 5 件のニュースはシェア数の上位 5 件にはいっておらず, 同様にシェア数上位 5 件のニュースも閲覧数の上位 5 件にはいっていない。これを上位 10 件まで広げてもシェア数 2 位の報連相の記事が閲覧数 8 位に入っているのみである。仮にすべての記事が同じ確率でシェアされるのであれば閲覧数が多い記事はシェアが多く, またその逆も成立するはずである。このことから閲覧とシェアについては異なる傾向があることが予想される。

表 3 に閲覧せずにシェアした数上位 5 件を示す。上位 2 件はシェア数の 1 位, 3 位と一致している。3 位 5 位は iPhone に関する記事が占めており, 6 位以降も Apple や iPhone に関するニュースが多い。

表 3: 閲覧せずにシェアした数上位 5 件のニュース

1	ある地方に IT 企業の サテライトオフィスが集まっているという記事
2	スマホレジについての記事
3	iPhone アプリの無料キャンペーンについての記事
4	iPhone アプリのデザインをまとめた記事
5	iPhone の文字入力を シェイクで訂正する方法についての記事

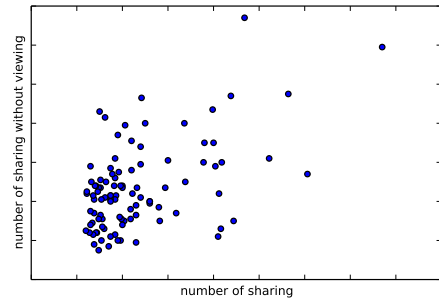


図 1: シェア数と閲覧せずにシェアした数の関係

図 1 に示したのはシェア数上位 100 件のニュースをシェア数と閲覧せずにシェアした数でプロットしたものである。同じシェア数であっても閲覧せずにシェアされていた数にはばらつきが多く, ニュースによって傾向が異なることがわかる。

このようによく閲覧されるニュースとよくシェアされるニュースが一致しないだけでなく, シェアされるニュースと閲覧せずにシェアされるニュースの傾向も異なることがわかる。

3.1 トピックごとの傾向

前章ではデータの閲覧・シェア行動について簡単な集計を行うことで, ユーザの閲覧・シェアの傾向がその対象となるコンテンツによって異なることを示した。本章ではニュースを記事をトピック分析によって分類することで, コンテンツの特徴によって閲覧・シェアの行動がどのように異なるのかをみていく。

トピック分析には Latent Semantic Indexing (LSI) を用いた [Manning 08]。ニュース記事はタイトルとリード文を Mecab によって形態素に分割して一般語やメディア名などをストップワード処理により除去した。その後各形態素を TF-IDF によってスコア付けし, LSI による次元圧縮を行い, 最もスコアの高いトピックを各ニュースのトピックとする。トピック数は 300 とした。トピック数は最も記事数が多いトピックが全体の記事数の 5% を下回るように設定した。

図 2 はトピックごとの閲覧数とシェア数をプロットしたものである。これにより閲覧数の大小によってシェアされる割合が変化することはなく, 閲覧数とシェア数の関係は同じようなトレンドにあることがわかる。しかしこれはトピックごとにシェアされる割合が同じということではなく, 閲覧数に対してシェアの数は非常に小さいためトピックごとにシェアの割合はトピックごとに異なっている。

シェア率 (シェア数/閲覧数) について高いトピックを表 4 に, 低いトピックを表 5 に示す。このときクリック数が 1 万件を下回るトピックは集計から除いている。シェア率の最も高いトピックと最も低いトピックではシェア率に 2 倍の差があった。

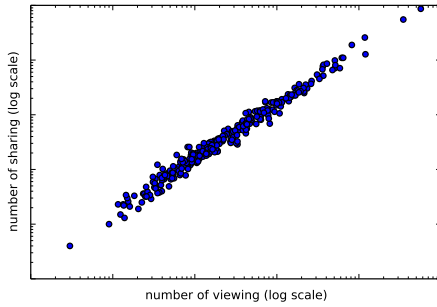


図 2: トピックごとの閲覧数とシェア数の関係

表 4: シェア率が高いトピック

トピックの概要	シェア率
科学や自然に関するトピック	15.9%
音楽に関するトピック	14.7%
健康に関するトピック	14.5%
プロダクトデザインに関するトピック	14.3%
Web デザインに関するトピック	14.1 %

このようにコンテンツによってシェアのされやすさが異なるのがわかる。シェア率の高いトピックは周りの人と話題にしやすいものも多く、低いトピックは読んでいることを周囲に知られたくないようなものであるといえるかもしれない。

同様にシェアのうち閲覧をしなくてシェアした割合についても高いトピックを表 6, 低いトピックを表 7 に示す。このときシェア数が 1000 件を下回るトピックは集計から除いている。最も低いトピックと最も高いトピックでは閲覧なしのシェアの割合が 2.5 倍にも登る。閲覧なしにシェアされた率が高いトピックは多くのコンテンツはあまりシェアされないが、まれにシェアされやすいコンテンツが出現しそれを閲覧せずにシェアすることを好むユーザが数多くシェアすることで、そのトピックのシェアのうち多くが閲覧なしのシェアになる。例えば野球に関するトピックでは WBC の際の選手の発言に関するニュースであったり、アニメに関するトピックでは日本アニメの海外での評価に関するニュースなどが多くシェアされていた。

このようにトピックごとに閲覧・シェアの傾向が異なることが明らかになった。

4. ユーザの閲覧とシェアに対する傾向の違い

本章ではユーザが最も閲覧したトピックと最もシェアしたトピックを比較し、ユーザの閲覧とシェアがどのように異なるかを分析する。ユーザが最も閲覧したトピックとシェアしたトピックが異なるユーザは 4,112 人で全体の 32.2% になる。

ここでどのトピックが閲覧とシェアが異なるのかを調べる。最も閲覧したトピックが t_{view} 、最もシェアしたトピックが t_{share} だったユーザ数を $N(t_{view}, t_{share})$ と表すこととする。ここでその組み合わせが最も閲覧したトピックと最もシェアしたトピックが等しいユーザに比べてどの程度多いのかを以下の指標で評価する。

$$R(t_{view}, t_{share}) = \frac{N(t_{view}, t_{share})}{N(t_{view}, t_{view})}$$

表 5: シェア率が低いトピック

トピックの概要	シェア率
恋愛に関するトピック	7.9%
ブログに関するトピック	8.8%
SNS の炎上に関するトピック	9.2%
仕事術に関するトピック	9.3%
マネージメントに関するトピック	9.4 %

表 6: 閲覧なしにシェアされた率が高いトピック

トピックの概要	閲覧なしシェア率
野球に関するトピック	52.3%
アニメに関するトピック	41.2%
ソーシャルゲームに関するトピック	38.0%
医療に関するトピック	37.8%
働き方に関するトピック	37.2 %

最も閲覧したトピックと最もシェアしたトピックが一致するユーザが少ないと大きく出やすくなるため、100 人以上いるトピックにおいて評価を行った。

表 8 に R が高い 5 つの組み合わせを示す。いずれの組み合わせも最も閲覧したトピックと最もシェアしたトピックの関連性は低いと考えられる。このようにユーザのコンテンツの閲覧とシェアの間には差異があるケースがある程度存在することが明らかとなった。

5. まとめ

本研究ではニュースアプリの行動履歴データを閲覧とシェアに着目して分析した。これまでコンテンツにシェアに関してはその結果しか分析対象として扱われていなかったが、本研究ではその背景にあるユーザの閲覧行動を分析することで、シェアの背景にあるユーザの行動原理を明らかにすることを試み、その結果として以下の様な知見が得られた。

- 閲覧していないコンテンツのシェアがある程度存在し、多くのユーザが経験していること
- コンテンツの内容によってシェアのされ方に違いがあること
- ユーザが閲覧するコンテンツとシェアするコンテンツが必ず一致しないこと

ソーシャルメディアへの投稿を利用した研究は数多くあるが、本研究で明らかになったようにユーザがシェアを行う際に行っている背景の行動によっては、誤った結果を招いてしまう場合も考えられる。例えばコンテンツの評価をどれだけシェアされたかで行う場合や、シェアしたコンテンツからユーザのプロファイリングを行う場合にはそのコンテンツのシェアされやすさを考慮する必要があると考えられる。

今後はコンテンツのトピック以外にも表現の仕方などの視点からシェアのされやすさを分析することで、そのコンテンツのシェアのされやすさを定量化することを目指す。またシェアの結果からその背景にあるユーザの消費行動を予測するという点についても検討していきたい。

本研究で用いたデータは特定のアプリケーション内でのものであるため、行動の割合や傾向などは一般化できる事実ではな

表 7: 閲覧なしにシェアされた率が低いトピック

トピックの概要	閲覧なしシェア率
仕事術に関するトピック	20.4%
おもしろ画像に関するトピック	21.7%
労働問題に関するトピック	21.9%
ベンチャーに関するトピック	23.5%
教育に関するトピック	24.0 %

表 8: 閲覧したトピックとシェアしたトピックとの異なり方

t_{view}	t_{share}	$R(t_{view}, t_{share})$
恋愛コラムに関するトピック	ライフハックに関するトピック	0.856
写真に関するトピック	ライフハックに関するトピック	0.291
恋愛コラムに関するトピック	健康に関するトピック	0.273
恋愛コラムに関するトピック	iPhone アプリに関するトピック	0.223
アニメに関するトピック	ライフハックに関するトピック	0.189

く、コンテンツが提供されるサービスの特性やそのユーザ層によって異なる。重要なのはコンテンツのシェアの背景にはシェアに至るまでのユーザの複雑な行動があり、コンテンツの特性やユーザの特性によってそれらは変わってくるということである。本稿がソーシャルメディアへの投稿の背景を考察する試みの一助になれば幸いである。

参考文献

- [Abel 11] Abel, F., Gao, Q., Houben, G.-J., and Tao, K.: Analyzing user modeling on twitter for personalized news recommendations, in *IMAP'11* (2011)
- [Bandari 12] Bandari, R., Asur, S., and Huberman, B. A.: The Pulse of News in Social Media: Forecasting Popularity, in *ICWSM'12* (2012)
- [Globis] Globis, : 「いいね」と「シェア」の使い方: <https://www.facebook.com/notes/globis/166256396757065>
- [Manning 08] Manning, C. D., Raghavan, P., and Schütze, H.: *Introduction to Information Retrieval*, Cambridge University Press, New York, NY, USA (2008)
- [Yang 11] Yang, J. and Leskovec, J.: Patterns of Temporal Variation in Online Media, in *WSDM'11* (2011)