

スマートフォンアプリとWikipediaにおける ニュースに関する閲覧行動の予備調査

Preliminary Investigation for Analysis of Browsing News Pages
on Smartphone App and Wikipedia

吉田 光男 ^{*1}
Mitsuo Yoshida

関 喜史 ^{*2}
Yoshifumi Seki

^{*1}豊橋技術科学大学
Toyohashi University of Technology

^{*2}株式会社 Gunosy
Gunosy Inc.

We conducted a preliminary investigation for the relationship between the number of browsing pages by the news application (impressions and clicks) and page views of Wikipedia (PC accesses and mobile accesses). As a result, we confirmed the tendency that the number of mobile accesses is larger than the number of PC accesses for the Wikipedia entry title included in the news article titles distributed in the news application. The correlation coefficients of these numerical values were not high, on the click numbers, the coefficients of mobile accesses were larger. When utilizing Wikipedia page view data for news applications, it is necessary to separate PC access frequency and mobile access frequency. Furthermore, it suggests that mobile access frequency is more useful.

1. はじめに

スマートフォンの普及に伴い、ニュースの閲覧行動も大きく変化している。2016年3月に発表された意識調査によれば、84%がニュース閲覧にスマートフォンを利用しており、テレビの61%を大きく上回っている[LINe 18]。また、ユーザがコンテンツを投稿するソーシャルメディアも普及し、自身の生活に関わる情報のほかに、ニュースなどの情報を「シェア」する行動も一般的になった。ユーザがあるニュースを気に入ったとき、そのニュースをシェアするだろうか。あるいは、あるニュースをより深く知るために、そのニュースに関するキーワードでウェブ検索エンジンを利用するだろうか。我々は、ニュースの閲覧と、ニュースを閲覧する以外の行動との関係について興味を持っている。

ニュースの閲覧、ソーシャルメディアの投稿、ウェブ検索エンジンの利用、それぞれの関係の分析は、ウェブリソースを横断した分析と言える。吉田と荒瀬[吉田 16]は、各種のウェブリソースにおいて、トレンドキーワードがどのように出現したのかを調査しており、ニュースの記事数とソーシャルメディアでの投稿数、あるいは検索頻度とは、緩やかな相関があることを示している。また、Wikipediaの閲覧数(アクセス回数)とソーシャルメディアでの投稿数および検索頻度との間には、高い相関があることも示している。一方、ニュースに関する報告は、記事数のみが取り上げられており、読者の行動となる「閲覧数」については触れられていない。

ニュースの閲覧とそれ以外の行動との関係を明らかにすることを目指し、本稿では、グノシーにおけるニュースの閲覧数とWikipediaの閲覧数との関係に関する予備調査を報告する。

2. データ・調査方法

本稿では、スマートフォンアプリにおけるニュースの閲覧数とWikipediaの閲覧数との関係について調査する。スマートフォンアプリにおける記事閲覧数のデータとして、情報キュレーションサービスであるグノシー^{*1}の閲覧データを利用

し、Wikipediaによる記事閲覧数のデータとしてWikistats pageview files^{*2}の閲覧データを利用する。利用するデータの期間は2017年6月1日から2017年11月30日の6ヶ月間である。

グノシーのデータからは、記事タイトル、表示回数、クリック数、配信タイプを用いる。表示回数はスマホアプリ上で記事のタイトル(およびサマリー画像)が表示された回数であり、クリック数はそのタイトルをクリックするなどして記事本文を表示した回数である。表示回数は、アプリ上でスクロールするなど、ユーザに対して実際に記事タイトルが表示した回数が記録されている。配信タイプはスマホアプリ上で記事を配信したか、プッシュ通知によって配信したかの二値の情報である。地震などの速報ニュースについては、1日に数件、プッシュ通知によって配信されており、この場合、スマホアプリ上で記事タイトルを表示しないため、表示回数にはカウントされない。なお、プッシュ通知によって配信し、ユーザが記事を読んだ場合には、クリック数にカウントされる。これらの仕様により、正確なクリック率の算出が困難であることから、本稿ではクリック率を指標として用いていない。

Wikipedia(Wikistats pageview files)のデータからは、見出し語、PCアクセス回数、モバイルアクセス回数を用いる。見出し語とアクセス回数との対応の取り方については、[吉田 15]を参照されたい。本稿では、対応をとる際、言語情報(元データの第一カラム)が“ja”である場合はPCアクセス回数、“ja.m”である場合はモバイルアクセス回数とした。

グノシーおよびWikipediaのデータは、それぞれ、その性質が異なるため、何らかの方法で照合が必要である。今回、グノシーの記事タイトルに含まれるWikipediaの見出し語に着目し、その見出し語のアクセス回数の合計をニュースに関連するアクセス回数とした。例えば、記事タイトルが「人工知能が地球を救う」であり、Wikipediaの見出し語として「人工知能」「地球」が存在し、それぞれのアクセス回数が10と5であるならば、この記事に関するWikipediaのアクセス回数は15であると見なす。

グノシーおよびWikipediaのデータは、いずれも1時間単位で存在するが、本稿では、記事ごとに日別に再集計して利用する。この際、記事ごとに、グノシーにおけるクリック数が10

連絡先: 吉田光男, 豊橋技術科学大学, 愛知県豊橋市天伯町雲雀ヶ丘 1-1, yoshida@cs.tut.ac.jp

^{*1} <https://gunosy.com/>

^{*2} <https://dumps.wikimedia.org/other/pagecounts-ez/>

回以上の時間のみを集計対象とし、その時間帯の Wikipedia のアクセス回数のみを利用する。例えば、6 月 1 日 13 時に 5 クリック、14 時に 50 クリック、15 時に 20 クリックされた記事の場合、10 クリックを下回る 13 時のクリック数を算入せず、6 月 1 日のクリック数は 70 であると見なす。また、Wikipedia におけるアクセス回数の再集計においては、14 時と 15 時のデータのみを用いる。Wikipedia におけるアクセス回数の再集計で、PC アクセス回数またはモバイルアクセス回数が 0 の記事は除外した。

3. 調査結果

まず、ニュースに関する Wikipedia の閲覧数について、PC アクセスが多いのかモバイルアクセスが多いのかを調査した。その結果を、表 1 に示す。総記事数は、別の日に同じ記事が出現した場合は二重にカウントされるものの、本稿における調査対象となる記事数と同等である。大半に記事について、モバイルアクセスの方が多く、スマートフォンによるニュースの閲覧と、スマートフォンによる検索や Wikipedia の閲覧が強く関係していることがうかがえる。

次に、ニュースに関する Wikipedia の閲覧数について、表示回数やクリック数との相関があるかを、ピアソン積率相関係数を用いて調査した。その結果を、表 2 に示す。全体的に値は低く、相関しているとは言いがたいことが分かった。アプリ配信について見ると、表示回数では PC アクセスの方が値が高くなるものの、クリック数ではモバイルアクセスの方が値が高くなる傾向がある。また、クリック数について、アプリ配信とプッシュ配信を比べると、プッシュ配信の方が値が高くなる傾向がある。これらは、記事一覧で表示されるという受動的な行動と、記事をクリックするという能動的な行動という風に見ることもでき、また、Wikipedia の閲覧も能動的な行動であることから、スマートフォン上での能動的な関係同士は、そうでない関係よりも、なんらかの相関が存在することが示唆される。

前節で述べたように、グノシーのデータと Wikipedia のデータとは、記事タイトルに見出し語が含まれるかどうかで、関連付けている。この照合処理において、見出し語が極端に短い場合、適切な関連付けができていない可能性があることから、使用する見出し語の長さを 2 文字以上、3 文字以上と制限をかけて調査した。この調査結果を、表 3、表 4 に示す。文字数の制限をかけるほど、値が下がる結果となった。これは、誤った照合よりも、照合されないことの方が影響が大きいことが考えられる。

4. おわりに

本稿では、スマートフォンアプリによる閲覧数（表示回数およびクリック数）と Wikipedia における閲覧数（PC アクセス回数およびモバイルアクセス回数）との関係について、予備調査を行った。その結果、スマートフォンアプリで配信されたニュースの記事タイトルに含まれる Wikipedia 見出し語について、PC アクセス回数よりもモバイルアクセス回数の方が多く傾向が確認できた。また、それらの数値の相関を調べたところ、相関は高くないものの、クリック数については、モバイル回数の方が多くことが確認できた。スマートフォンアプリに対して Wikipedia の閲覧数データを活用する際には、PC アクセス回数とモバイルアクセス回数とを分ける必要があり、モバイルアクセス回数の方が有用であることを示唆している。

今回の調査では、ニュースのカテゴリを考慮していない。例えば、政治に関するニュースとエンターテインメントに関する

表 1: Wikipedia における PC アクセス回数よりもモバイルアクセス回数の多かった記事数

	総記事数	モバイルが多い記事数
アプリ配信	917,740	813,542 (88.6%)
プッシュ配信	4,726	4,469 (94.6%)

表 2: スマートフォンアプリの閲覧数と Wikipedia のアクセス回数との相関（積率相関係数・見出し語の制限なし）

		PC	モバイル	PC/モバイル
アプリ	表示	0.166	0.162	0.169
	クリック	0.132	0.149	0.149
プッシュ	クリック	0.163	0.196	0.190

表 3: スマートフォンアプリの閲覧数と Wikipedia のアクセス回数との相関（積率相関係数・見出し語 2 文字以上）

		PC	モバイル	PC/モバイル
アプリ	表示	0.129	0.124	0.130
	クリック	0.108	0.121	0.121
プッシュ	クリック	0.157	0.189	0.184

表 4: スマートフォンアプリの閲覧数と Wikipedia のアクセス回数との相関（積率相関係数・見出し語 3 文字以上）

		PC	モバイル	PC/モバイル
アプリ	表示	0.108	0.109	0.112
	クリック	0.098	0.113	0.112
プッシュ	クリック	0.154	0.187	0.181

ニュースとでは、閲覧の性質が異なる [関 17]。このような性質の異なりは、Wikipedia へのアクセス（あるいはウェブ検索エンジンの利用）に影響を与えている可能性があり、今後、カテゴリを考慮した調査を進める。また、日別に相関係数を算出すると、0.428（11 月 30 日・表示・モバイル）を記録するなど、比較的、高い値が出ることもあった。今回の調査では、すべての日を一括して相関係数を算出しており、日別の全体傾向（グノシーの利用者の増加など）が考慮されていない。この点についても、データを細分化するなどして、より詳細な調査を進める。

参考文献

- [LINE 18] LINE 株式会社：世代間のニュースサービス利用に関する意識調査を実施 (2018) <https://linecorp.com/ja/pr/news/ja/2016/1267>
- [関 17] 関 喜史, 潮 旭, 米田 武, 松尾 豊：クリックベイトなニュース記事の特定に向けたユーザ行動分析, 信学技報, Vol. 117, No. 207, pp. 65–70 (2017)
- [吉田 15] 吉田 光男, 荒瀬 由紀, 角田 孝昭, 山本 幹雄：検索頻度推定のための Wikipedia ページビューデータの分析, 第 29 回人工知能学会全国大会 (2015)
- [吉田 16] 吉田 光男, 荒瀬 由紀：トレンドキーワードに関するウェブリソースの横断的分析, 情報処理学会論文誌. データベース, Vol. 9, No. 1, pp. 20–30 (2016)