

# オンラインメディア記事の読者の行動分析

## Analysis of readers in Online News Service

園田 亜斗夢<sup>\*1</sup> 関 喜史<sup>\*2</sup> 鳥海 不二夫<sup>\*1</sup>  
Atom Sonoda Yoshifumi Seki Fujio Toriumi

<sup>\*1</sup>東京大学 The University of Tokyo  
<sup>\*2</sup>株式会社 Gunosy Gunosy Inc.

Information is transmitted through websites, and the immediate reaction to various information is required. Hence, the efforts for readers to select information themselves have increased, which leads to the further improvement of recommendation services that can reduce such burdens. On the other hand, it is pointed out that filter bubbles that only provide biased information to users are generated due to the redundant recommendation. In this research, we analyzed behavioral changes and showed that behavior changes during the period as well as the existing research, and also consider the relationship between the diversity decrease and the period and the number of views.

### 1. はじめに

インターネットの登場により新聞社等のメディアの発信するニュースの媒体が新聞から web サイトに拡大し [1], 記事の速報性の向上や記事数の増加が進んでいる。これに伴い、ユーザが記事を選択する際に必要な時間と労力が増大している。このような状況下で、メディアを通じた情報接触において人々は何らかの基準で閲覧する情報を取捨選択せざるを得なくなっていると推測される。情報接触の過程で、取捨選択の際には2つの問題があると考えられる。1つは「見たいものだけを見る」といったユーザ主体の選択的接触による問題であり、もう1つはシステム主体で行われる検索システム、レコメンデーション、ランキング等の広義の推薦システムの利用による問題である。広義の推薦システムは、記事を選択する際に必要な時間と労力といった負担を減らし、満足度を向上させるため、多くの分野で導入されている。特に、技術の発達によってユーザの選好に応じて提示情報をパーソナライズするタイプの推薦システムの導入が進んでいる。一方で、過度の推薦によってユーザに偏った情報のみを提供するフィルターバブルが発生しているとの指摘もある [2]。フィルターバブルの問題点として、得られる情報が偏ることによって、社会の分極・分断の進行が挙げられる。つまり、過度の推薦は、極端な考えを生み出し、合意形成が困難となる状態を作りかねない。正確な情報が広く社会的に共有されることは民主的な意思決定の必要条件である。しかし、「見たいものだけを見る」ことが容易な情報環境下では、異なる意見を持つ他者に対する寛容性の低下やマイノリティに対する偏見の増大、フェイクニュースの無批判な受容などによってこの必要条件を満たすことが難しくなる [3]。

本研究では既存のニュースサイト上でのユーザの行動変容について分析を行い、筆者のこれまでの研究 [4], [5] と比較する。これにより、既存の結果と比較し、多様性の減少はメディアが異なっても起きているのかを確認する。そのために、本研究では、既存研究と同様にユーザの行動を閲覧した記事の多様性によって評価し、現実のデータで時間の経過に伴う多様性の指標が変化していることを確認し、多様性の変化と期間・閲覧数の関係についても考察する。また、多様性が増加するユーザ

の存在を確認し、そのようなユーザの閲覧数の変化の条件についても確認する。

### 2. 分析手法

#### 2.1 分析に用いるデータ

本研究では、2018年11月1日から12月31日までの2ヶ月間の株式会社 Gunosy が提供するニュース配信スマートフォンアプリケーションのデータを用いる。

この中で、11月1日～11月7日までの登録ユーザについて分析した。これは、既存ユーザではメディアに適合しており、短期間では変化しづらいことから、新規ユーザを扱うことで、期間の経過の影響を対象の2ヶ月間のデータで明確にするためである。また、11月15日～11月21日と12月24日～12月30日の期間それぞれで、閲覧数が一定以上のユーザを抽出した。これは分析する期間において十分な閲覧数があるユーザを対象とするためである。

閲覧記事の評価には、既存のカテゴリを用いた。既存研究では記事のクラスタリングを行い、そのクラスタによって評価していたが、本研究で扱うデータはカテゴリ数が50と分析に十分な細分化がなされていたため、既存のカテゴリを採用した。

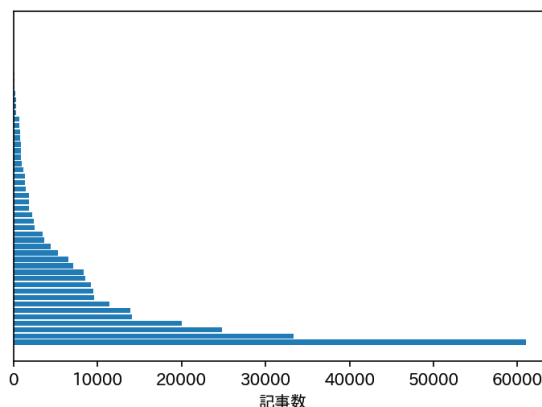


図 1: カテゴリごとの記事数の分布

ここで、図 1 に横軸に記事数を取り、記事数が多い順に並べたものを示す。ここから、カテゴリごとの記事数は異なり、一部のカテゴリに多くの記事が集中していることが分かる。

## 2.2 閲覧行動の多様性の評価

本研究では閲覧行動の多様性の変化を行動変容として、ある一定の期間経過後、読んだ記事の種類が変化したかどうかを分析する。つまり、読んでいる記事のカテゴリの多様性によって、ユーザの行動を評価する。カテゴリの集中度は、情報量エントロピーを用いて次のように求められる。

$$H(u) = - \sum_i p_i \cdot \log p_i$$

ただし、 $p_i$  は各ユーザにおけるカテゴリ  $c_i$  の存在確率で、閲覧数  $N$  のとき、 $p_i = c_i/N$  と表されるものである。カテゴリエントロピーの高いユーザは、様々なカテゴリに所属する記事を読んでいるユーザであり、カテゴリエントロピーの低いユーザは、特定のカテゴリに所属する記事を集中的に読んでいるユーザである。一般に、フィルターバブルとは推薦システムによりユーザがその人の観点に合わない情報から隔離され、ユーザ自身の興味があると判断された範囲に集約されていくことである。したがって、読んでいる記事のカテゴリの集中度でその影響を測ることができると期待される。例えば、もしフィルターバブルによって観点に合う情報のみと接するようになれば、閲覧記事は特定のクラスタに集中するため、多様性が失われ、カテゴリエントロピーは低下すると考えられる。

そこで、一定期間ごとに閲覧行動の多様性を評価し、多様性に変化が見られるか確認する。その際、期間を複数比較することで、多様性が安定するまでの期間を確認する。

## 3. 期間の経過によるカテゴリエントロピーの変化

### 3.1 分析結果

期間ごとのカテゴリエントロピーは表 1 のようになった。11 月 15-21 日と 12 月 13-19 日を比較すると、約 1ヶ月の間に有意水準 1% で有意にカテゴリエントロピーが減少していた。よって、この分析から、株式会社 Gunosy のサービスにおいても、期間の経過によって閲覧するカテゴリが集中していっていると考えられる。また、カテゴリエントロピーの減少幅は 11 月 15-21 日と 11 月 22-28 日の間が最も大きく、期間の序盤ではより閲覧するカテゴリの幅が狭まりやすいと考えられる。一方、有意な差ではないが、11 月 29-12 月 5 日と 12 月 6-12 日の間では多様性が増加し、その後、また減少に転じている。このことから、安定しつつも、カテゴリエントロピーの減少は続いていることがわかる。

また、この期間の記事全体のエントロピーは 1.26 であることから、閲覧されている記事のエントロピーはおおよそ半減しているが、これは表示記事や閲覧数の制約の影響があると推測される。

図 2 に 11 月 15-21 日と 12 月 13-19 日のカテゴリエントロピーの分布を示す。これより、カテゴリエントロピーの平均だけでなく、全体としてエントロピーが減少し、閲覧するカテゴリの幅が狭まっていることがわかる。また、図 3 に 11 月 22-28 日と 12 月 13-19 日のカテゴリエントロピーの分布を示す。図 2 と比較すると、分布の差がわかりづらくなっており、安定しはじめていることがわかる。

一方、期間 1 と期間 2 を比較すると、カテゴリエントロピーが増加しているユーザも 32% 存在し、全体の分布としては減

表 1: 期間ごとのカテゴリエントロピーの平均

期間	平均カテゴリエントロピー
11 月 15-21 日	0.727
11 月 22-28 日	0.676
11 月 29-12 月 5 日	0.673
12 月 6-12 日	0.682
12 月 13-19 日	0.643

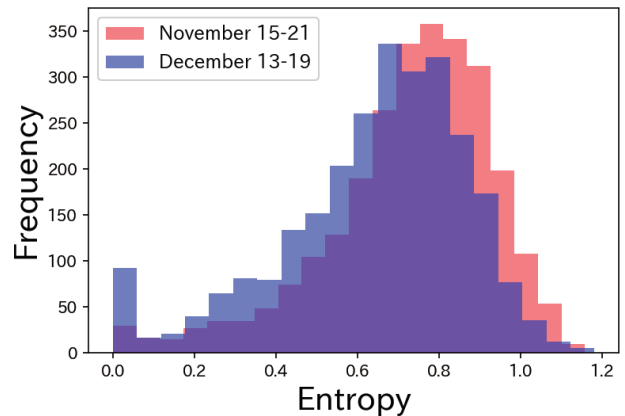


図 2: カテゴリエントロピーの分布 (11 月 15-21 日と 12 月 13-19 日)

少しているが、増加しているユーザも一定程度いることがわかった。

### 3.2 考察

既存研究 [4] では日本経済新聞社の提供する日経電子版において多様性が減少することを確認したが、以上の結果より株式会社 Gunosy のサービスでも多様性の指標として採用したカテゴリエントロピーが減少していることがわかった。また、複数の期間を比較することで、カテゴリエントロピーは序盤では大きく減少することがわかった。一方、期間の後半でも増加と減少が起きていることから、真に安定し出すのは、より期間が経過した後であり、より長期間での分析が必要であると考えられる。さらに、エントロピーが増加しているユーザも 32% 存在したことから、期間を経ることで、全体としてはエントロピーは減少しているが、増加するユーザも十分に存在し、ユーザごとの違いが大きくなってくると考えられる。

## 4. カテゴリエントロピーと閲覧数の関係

### 4.1 相関分析

本章では、閲覧記事数が多様性の変化に与える影響を明らかにするために、カテゴリエントロピーと閲覧数の関係を分析した。そのために、11 月 15-21 日 (期間 1) と 12 月 13-19 日 (期間 2) のカテゴリエントロピーと閲覧数を比較した。この期間中に閲覧数は 14% 減少していた。まず、エントロピーと閲覧数の相関分析を行った。表 2 にカテゴリエントロピーと閲覧数の相関係数を示す。この表から、期間 1 のエントロピーと期間 2 のエントロピーには相関があることが分かる。つまり、一定の期間の経過後もカテゴリエントロピーが高いユーザは高いままのことが多いと考えられる。また、各期間のエント

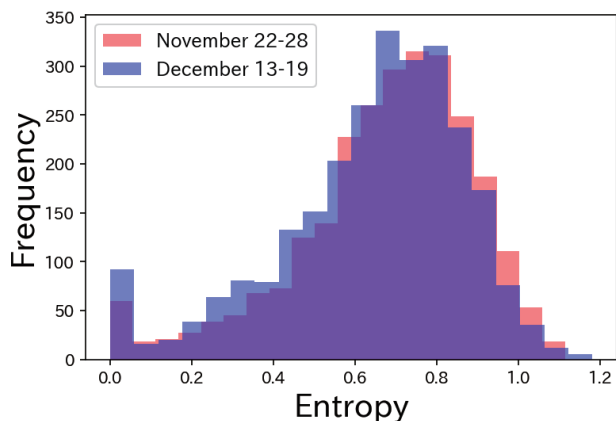


図 3: カテゴリエントロピーの分布 (11月 22-28 日と 12月 13-19 日)

表 2: カテゴリエントロピーと閲覧数の関係

比較対象	相関係数
期間 1 のエントロピーと期間 2 のエントロピー	0.609
期間 1 のエントロピーと期間 1 の閲覧数	0.273
期間 2 のエントロピーと期間 2 の閲覧数	0.345
エントロピーの変化と閲覧数の変化	0.299

ロピーと閲覧数の間にも弱い相関があることが見て取れる。さらに、エントロピーの変化と閲覧数の変化間にも弱い相関があることが分かり、読んでいる記事数の変化が多様性の変化に影響していると考えられる。つまり、新規登録後日数がたつにつれ、全体として閲覧数が減少し、カテゴリエントロピーもその影響を受けていると考えられる。

#### 4.2 ユーザグループごとの分析

ここでは、ユーザをクリック数のしきい値で 3 分割し、その 3 分割したユーザ群の期間中での遷移を基に、エントロピーの変化を分析した。具体的には、閲覧数が平均より少ないユーザをグループ A、閲覧数が平均に近いユーザをグループ B、閲覧数が平均の 2 倍以上多いユーザをグループ C として分析した。期間 1 と期間 2 を比較し、これらのグループ間の移動について分析したところ、グループ A はユーザ数が 72% 増加し、グループ B は 15% 減少、グループ C は 22% 減少していた。ここから、利用初期は多くの記事を読覧するが、期間の経過に伴い安定してくることがわかる。

表 3 に、グループの移動ごとのユーザ数の割合と期間ごとのカテゴリエントロピーを示す。グループを移動しない場合 (A → A, B → B, C → C) のうち、A → A, B → B に関しては有意水準 1% で有意にカテゴリエントロピーが減少していることが分かるが、C → C では有意に減少はしていなかった。つまり、グループを移動しない場合で閲覧数が少ない場合はカテゴリエントロピーは減少するが、閲覧数が十分にある場合はカテゴリエントロピーは減少しにくいと考えられる。このことから、多様性の減少を抑制するには、閲覧数の確保が重要であると考えられる。

次に、グループを閲覧数が増加する方向に移動する場合のうち、分析に十分なユーザ数を含む A → B, B → C について分

表 3: カテゴリエントロピーとグループの遷移の関係

比較対象	ユーザ割合 [%]	期間 1 のエントロピー	期間 2 のエントロピー
A → A	11.0	0.625	0.509
A → B	7.2	0.624	0.677
A → C	0.0	0.528	0.449
B → A	20.2	0.724	0.525
B → B	48.2	0.752	0.699
B → C	3.5	0.767	0.794
C → A	0.3	0.858	0.529
C → B	5.5	0.808	0.719
C → C	4.0	0.855	0.804

析する。このうち、A → B に関しては有意水準 1% で有意にカテゴリエントロピーが増加していることが分かるが、B → C では有意に増加はしていなかった。しかしながら、閲覧数が増加する方向に移動する場合、カテゴリエントロピーも増加する傾向にあり、減少はしないことが分かる。また、平均カテゴリエントロピーと比べても高い水準であると言える。このことから、閲覧数が増えることは、カテゴリエントロピーの減少を抑制するために良い影響があると考えられる。

さらに、グループを閲覧数が減少する方向に移動する場合のうち、分析に十分なユーザ数を含む B → A, C → B について分析する。B → A, C → B どちらについても、有意水準 1% で有意にカテゴリエントロピーが減少していることが分かる。このことから、閲覧数が減ることは、カテゴリエントロピーの減少を抑制するために悪い影響があると考えられる。

以上より、カテゴリエントロピーと閲覧記事数グループの遷移を分析することで、閲覧記事数が増加することはカテゴリエントロピーの増加に貢献することがわかった。また、閲覧数に変化がない場合は、カテゴリエントロピーは減少しやすいが、閲覧数が十分にある場合はカテゴリエントロピーの減少を抑制できることがわかった。

一方、期間 2 のカテゴリエントロピーの絶対値に注目すると、期間 1 の閲覧数が多いほど期間 2 の値が大きいたことが分かる。このことから、閲覧記事数の増減はカテゴリエントロピーの変化に影響を与えるが、全体としては、前節の相関分析で見たように期間 1 のエントロピーの影響が強いことが分かる。つまり、一定の期間の経過後もカテゴリエントロピーがもともと高いユーザは比較的高い状態を維持していると考えられる。しかしながら、閲覧数がこれ以上に減少していった場合に、一定期間前の影響と閲覧数減少による影響が逆転することもあると考えられ、より長期での分析が今後の課題である。また、閲覧数が一定以上の場合には他のカテゴリを読覧する確率が増えると推測され、エントロピーは増えやすいはずである。また、カテゴリごとに閲覧される期待値が異なる。そこで、閲覧数とカテゴリごとの期待値を表現する指標を導入し比較することで分析がより進むと考えられ、これは今後の課題である。

## 5. 結論

株式会社 Gunosy が提供するニュース配信スマートフォンアプリケーションの閲覧行動について、記事のカテゴリのエントロピーの平均によりユーザの行動の多様性を評価し、複数の期間の閲覧行動を比較し、分析した。これらの結果と既存研究 [4] の比較から、メディアが変わっても、多様性の指標であ

るエントロピーの減少が起きていることが確認できた。また、複数の期間を比較することで、カテゴリエントロピーは序盤の方が大きく減少すること、新規登録後1ヶ月以上経過してもカテゴリエントロピーの減少は続いていることがわかった。一方、増加するユーザも一定数存在し、期間を経ることで、ユーザごとの差異が大きくなっていくと考えられる。

また、記事の閲覧数とカテゴリエントロピーの関係を分析したところ、閲覧数とエントロピーの間には相関があり、また、閲覧数の変化とエントロピーの変化の間にも相関が存在することがわかった。

今後の課題として、より長期間の分析を行い、多様性の減少幅が小さくなり、安定期に入るまでの期間の解明が必要である。また、初期の多様性と閲覧数の変化が一定期間経過後のエントロピーに与える影響の変化に対する分析が今後の課題である。他にも、行動変容を導く要因の分析のために、閲覧記事だけではなく、表示されたが読まれなかった記事と表示されなかった記事についても分析を進めていく必要がある。

## 参考文献

- [1] 総務省. 平成 28 年情報通信メディアの利用時間と情報行動に関する調査報告書. [http://www.soumu.go.jp/main\\_content/000492877.pdf](http://www.soumu.go.jp/main_content/000492877.pdf).
- [2] Eli Pariser. *The filter bubble: What the Internet is hiding from you*. Penguin UK, 2011.
- [3] 小林哲郎ほか. ソーシャルメディアと分断化する社会的リアリティ (特集 twitter とソーシャルメディア). 人工知能学会誌, Vol. 27, No. 1, pp. 51-58, 2012.
- [4] Atom Sonoda, Fujio Toriumi, Hiroto Nakajima, and Miyabi Gouji. Analysis and modeling of behavioral changes in a news service. In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pp. 73-80. IEEE, 2018.
- [5] 園田亜斗夢, 鳥海不二夫, 中島寛人, 郷治雅. ネットワーク構造に基づく新聞記事の分類による読者の行動分析. 人工知能学会全国大会論文集 2018 年度人工知能学会全国大会 (第 32 回) 論文集, pp. 3O1OS1a03-3O1OS1a03. 一般社団法人人工知能学会, 2018.