

ニュース記事の配信形態が読者の閲覧行動に与える影響の分析

Analysis of the Impact of Distribution Formats for News Articles on Readers' Browsing Behavior

鈴木香帆^{1*} 中条雅貴¹ 鳥海不二夫¹ 並木亮² 石原祥太郎²
Kaho Suzuki¹ Masaki Chujo¹ Fujio Toriumi¹ Ryo Namiki² Shotaro Ishihara²

¹ 東京大学大学院工学系研究科システム創成学専攻

¹ Graduate School of Engineering, The University of Tokyo

² 日本経済新聞社

² Nikkei, Inc.

Abstract: Abstract (English) We examine how presentation format and update schedule condition engagement and exposure diversity by contrasting two formats within the digital products of Nikkei, a Japanese business newspaper. The bundle format is a digital replica of the print edition released twice daily in the morning and evening, offering a finite package of articles, while the news website is a continuously updated stream. Using a year-long dataset of on-platform reading logs, we embed article titles and evaluate topical spread with Average Pairwise Distance (APD) and concentration with the Generalist?Specialist score (GS-score). For engagement, readers who predominantly use the bundle sample a larger number of articles, devote less time to each article, and concentrate their usage in the morning, whereas website reading is distributed more evenly across the day. For diversity, bundle-oriented readers exhibit higher APD and lower GS-scores than website-oriented readers, indicating broader topical dispersion and weaker concentration around a single interest centroid; these differences are statistically reliable under standard tests. Taken together, the twice-daily bundle functions as structural scaffolding that lowers search and switching costs and systematically increases incidental exposure to heterogeneous topics, suggesting that maintaining print-like bundles alongside continuously updated sites can broaden information diets and support media pluralism in digital news.

1 はじめに

近年まで、人々は主として紙媒体に印刷された新聞を通じてニュースを消費していた。これらは人間の編集者によって編集・構成されており、生産スケジュールや誌面構成の制約のため、発行時刻は固定されていた。したがって、読者は人間による推奨記事に接していたといえる。しかし今日では、記事はオンラインでアクセス可能となり、日本を含む世界中の主要メディア企業がデジタルニュースコンテンツを配信している [1]。

インターネットの広範な普及に伴い、人々は従来のマスメディアだけでなく、オンラインニュースサイトやソーシャルメディアプラットフォームを通じて情報を摂取するようになってきている [2]。人々のオンライン

活動が活発化するにつれ、**エコーチェンバー**および**フィルターバブル**が重要な社会問題として浮上してきた [3] [4]。エコーチェンバーとは、同質性（ホモフィリー）と選択的曝露によって駆動される情報接触の分断を指し、フィルターバブルとはアルゴリズムによるパーソナライゼーションの結果、利用者の接触範囲が狭められる現象である。いずれも異なる立場の意見への接触を減少させる [4] [7]。プラットフォームの推薦アルゴリズムは既存の見解を強化し、結果として意見の極化を深める可能性がある [8]。理論的には、このようなアクセス容易性の向上は、より多様なトピックや視点への広範な曝露を可能にするはずである。しかし実際には、デジタル環境——とりわけアルゴリズム的パーソナライゼーションによって形成された環境——は、利用者を自らの既存の関心や信念と整合的な内容へと導く傾向がある。その結果、近年の研究では、アルゴリズム推奨によってエコーチェンバー

*連絡先：（東京大学大学院工学系研究科システム創成学専攻）
（東京都文京区本郷 7-3-1）
E-mail: kaho@torilab.net

が形成され、異質な視点への偶発的曝露が減少していることが報告されている [5] [9] ? [12]。

デジタルコンテンツの提示方法が利用者の選択行動に与える影響については、多くの研究が行われてきた [13] ? [15]。しかし、提示構造が利用者の実際のトピック多様性にどのような影響を及ぼすのかについては、十分な検討がなされていない。特に、推薦アルゴリズムが多様性に与える影響に焦点を当てた研究は多い一方で、記事の**構成**や**提示形式**そのものが、利用者の曝露するトピックの多様性にどのような影響を与えるのかについては、ほとんど注目されてこなかった [2] [16]。本研究はこのギャップを埋めることを目的とし、異なる編集提示形式が閲覧行動およびトピック多様性への曝露にどのような影響を与えるかを検討する。

特筆すべきは、これらの動態が**政治的領域**におけるイデオロギー的同質性を助長する懸念を生じさせる点である。娯楽やライフスタイル分野では、オンライン形式が多様性を支える可能性がある一方で、政治的内容は自己選択的な閲覧によって極化に陥りやすい [16]。紙媒体における形式的制約が生み出す均一な曝露とは異なり、デジタルな提示形式は、情報摂取の断片化およびメディア多元性の低下を引き起こすおそれがある [13] [17] ? [19]。それにもかかわらず、既存研究の大半はアルゴリズムによるキュレーションの影響、あるいは紙媒体とデジタル媒体の二項的比較に焦点を当てており [7] [20]、**人間が設計した編集デザイン**が曝露の多様性にどのような影響を及ぼすかについては、ほとんど明らかにされていない。

本研究では、2種類のデジタルニュース提示形式を比較する。第一に、**バンドル形式 (bundle format)** は、紙の新聞をデジタルで再現したものであり、同一の記事を決まった時刻（通常は朝と夕方）に配信する。第二に、**ニュースサイト形式 (news website)** は、紙面レイアウトの制約を受けず、記事を継続的に更新して逐次的に提示する。これらの形式が利用者行動および曝露にどのような差異をもたらすかは、これまで十分に解明されてこなかった。実際、通勤者は朝にバンドル形式を参照する傾向がある一方で、継続的に更新されるニュースサイトは、1日を通して関心主導的な閲覧を促すと考えられる。

本研究は、編集提示形式が利用者のエンゲージメントおよびトピック多様性への曝露をどのように形成するかを明らかにするものである。分析には、日本経済新聞社 (Nikkei Inc.) が運営する「日本経済新聞 電子版 (Nikkei Online Edition)」の大規模な閲覧データを用いた [21]。バンドル形式とニュースサイト形式のユーザ群を比較し、曝露多様性を **平均ペアワイズ距離 (Average Pairwise Distance; APD)** [22] および **Generalist?Specialist スコア (GS-score)** [23]

によって測定した。エンゲージメントについては、**読了記事数**および**閲覧時間**を指標とした。その結果、バンドル形式の利用者はニュースサイト利用者と比べて APD が高く、GS スコアが低い傾向がみられた。これは、伝統的な紙面型のバンドル形式が、より多様なトピックへの曝露をもたらすことを示唆している。

総じて、本研究の結果は、朝夕の定時配信を特徴とするバンドル形式が、デジタルニュースにおけるトピック曝露の拡大に有効な設計上の手段となりうることを示している。継続的に更新されるニュースサイトと併存する形で紙面型のバンドル版を維持することは、**検索および切替コストを低減し、異質なトピックとの偶発的な出会いを増加させる構造的手がかりを提供する**。これらの示唆は、編集的構造を保持しつつ流通形態を現代化する**ハイブリッド型プロダクトデザイン**の方向性を指し示すものである。

2 仮説の設定

本研究では、分析を導く仮説を明示する。ニュースの提示形式は、利用者のエンゲージメントおよびトピック多様性への曝露の双方に影響を与えられられる。

バンドル形式 (bundle format) では、編集者が有限で非パーソナライズされた記事群を構成し、一定の時刻に配信する。この形式では複数のトピックが並列して提示されるため、異なるジャンル間の検索および切替コストを低減し、読者の即時的関心外のトピックへの偶発的曝露を促進する可能性がある。

一方、継続更新型のニュースサイト形式 (continuously updated website) では、固定された日次境界が存在せず、記事は絶えず更新される。利用者は多様な時刻にアクセスし、しばしば同一トピック領域内で新しい記事へのリンクを辿るため、関心が限定されたトピック群に集中しやすくなる。

したがって、本研究では構造的提示形式が以下の2つの結果変数にどのような影響を与えるかを検討する。

1. 利用者のエンゲージメント
2. 曝露するトピックの多様性

H1: 構造的影響仮説 (Structural Influence Hypothesis)

ニュース提示の構造的形式——すなわち記事がバンドル形式で配信されるか、継続的に更新されるウェブサイトで配信されるか——は、利用者のエンゲージメントに影響を及ぼす。具体的には、バンドル形式においては、より多くの記事を短時間で読む傾向が見られると予測する。

H2:多様性優位仮説 (Diversity Advantage Hypothesis)

編集者が選定したバンドル形式に主として接触する利用者は、ニュースサイト経由で記事を読む利用者に比べて、より高いトピック多様性に曝露される。

3 データセット

3.1 データセット概要：記事分布とサンプル構築

本仮説を検証するため、日本経済新聞社 (Nikkei Inc.) が運営する日本経済新聞 電子版 (Nikkei Online Edition) の閲覧ログおよびメタデータを使用する。このデータセットは、2024 年 1 月 1 日から 12 月 26 日までの 361 日間にわたる利用者の閲覧行動を記録している。

日本経済新聞は、ビジネスおよび金融分野に焦点を当てる日本有数の報道機関であり、政治的には概ね中立的と見なされている。オンライン版には、自己選択的な閲覧が可能な継続更新型ニュースサイトと、紙媒体を再現した朝刊・夕刊のバンドル版（それぞれ午前 6 時前後および午後 5 時前後に更新）が存在する。

バンドル版の記事が更新されると、その構成は基本的に固定され、紙の新聞と同一の内容に対応する。そのため、紙面制約により記事が要約される場合が多く、編集上の優先順位が反映される形式となっている。このバンドル形式は、主に朝の起床後や通勤時間帯にニュースを読むビジネス層を対象として設計されている。

表 1: 会員ステータスの分類

ステータス	説明
1	匿名、未ログイン、または不明
2	非会員
3	無料会員として登録済み
4	有料会員

表 2: データセットの主要フィールド

フィールド名	説明
ig_usr_user_id	匿名化されたユーザ ID
ig_usr_user_status	会員ステータス (表 1 参照)
er_dat_jp_timestamp	記事閲覧イベントのタイムスタンプ (日本時間)
spent_time	閲覧時間 (秒単位)
er_met_title	閲覧された記事のタイトル

サンプル構築に際しては、継続的かつ積極的に閲覧する利用者を対象とするため、分析期間中に少なくとも 300 本の記事を閲覧した有料会員 1,979 名を選定し

た。年間 300 記事という閾値は、高いエンゲージメントを示す上位層を抽出する目的で設定したものである。

図??および図??が示すように、利用者ごとの閲覧記事数分布はヘビーテール型を示しており、少数の利用者が極めて多数の記事を閲覧する一方、大多数の利用者は比較的少数の記事しか読まない。同様に、記事ごとの閲覧数分布も偏っており、一部の記事が多く読者を集める一方で、多くの記事は限定的な関心しか得ていない。

したがって、本研究では年間 300 本以上の記事を読む上位約 10 %の利用者に焦点を当て、安定した閲覧行動を分析対象とする。非会員はペイウォールにより閲覧制限を受けるが、有料会員は全記事に制限なくアクセスできるため、プラットフォーム上の自己選択的エンゲージメントをより正確に推定できる。

最終的なデータセットには、1,979 名の利用者による 4,442,251 件の閲覧イベントが含まれる。これらのイベントは 296,412 本のユニークな記事に対応し、それぞれ発行タイムスタンプ、記事 ID、記事タイトルやカテゴリなどのメタデータを持つ。閲覧イベントは匿名化されたユーザ ID と正確なアクセス時刻によって索引化され、個人情報を一切収集せずに、ユーザごとの閲覧系列を再構築することが可能である。また、600 秒 (10 分) を超える閲覧セッションは放置時間とみなし、分析から除外した。

3.2 利用者グループの分類

分析のため、各利用者の形式別利用傾向を数量化する。具体的には、観測期間中に閲覧した記事のうち、バンドル形式の記事が占める割合 (bundled share) を算出する。

図??は利用者ごとのバンドル利用率の分布を示す (横軸: バンドルシェア、縦軸: 利用者数)。分布は 0 付近に強く偏っており、多くの読者がバンドル版をほとんど利用していないことを示す。同時に、値が 1 に近い利用者群も存在し、バンドル版を主に利用する読者が一定数いることが確認できる。中間層の利用者は両形式を混合的に利用している。

明確な比較を行うため、分布の上位 25%と下位 25%の利用者に限定する。上位四分位群は主にバンドル版を利用する読者、下位四分位群は主にウェブサイトを利用する読者である。このように極端な四分位群に焦点を当てることで、形式間の行動差を最大化し、解釈を単純化する。中央値付近の利用者は両形式を混在して利用する傾向があり、行動の差異を曖昧にするため、分析対象から除外した。

4 多様性指標と意味埋め込み手法

本研究では、既存研究 [?, ?] に従い、記事読者のトピック多様性を評価するために二つの補完的指標——平均ペアワイズ距離 (Average Pairwise Distance: APD) およびジェネラリスト?スペシャリストスコア (Generalist?Specialist score: GS-score) ——を用いる。両指標は共通の基盤として、記事タイトルを高次元の意味空間に埋め込み、その類似度をコサイン類似度により算出するものである。GS-score はユーザ固有の重心ベクトルとの平均類似度を算出するため計算効率が高い。一方、APD はすべてのペア間距離を平均するため計算負荷は高いが、読者が接触した記事群の語義的な分散性をより直接的に捉える。分析の信頼性と効率性の両立を図るため、本稿では両指標を併用して報告する。

4.1 記事タイトルの意味埋め込み

意味的多様性を定量化するため、全記事タイトルを OpenAI の text-embedding-3-small モデルを用いて高次元ベクトル表現に変換した。同モデルは多言語対応の意味埋め込みモデルであり、日本語テキストを直接入力可能である。そのため、追加の前処理やトークン化を行わずに、記事タイトルをそのまま入力して埋め込みベクトルを生成した。各タイトルに対応する出力は 1,536 次元ベクトルであり、記事の語義的特徴を連続空間内に表現する。このベクトル群を用いて、記事間および各ユーザの閲覧履歴内でのコサイン類似度を計算した。

この類似度分布をもとに、語義的多様性を操作的に定義した。すなわち、タイトルレベルの埋め込みを利用することで、ユーザが記事を選択する際に実際に参照可能な情報集合に基づいた分析を行うことが可能になる。

4.2 多様性の測定指標

利用者の閲覧行動における多様性を測定するため、次の二つの代表的指標を導入する。

- 平均ペアワイズ距離 (Average Pairwise Distance; APD) [?]
- ジェネラリスト?スペシャリストスコア (Generalist?Specialist score; GS-score) [?]

APD は、利用者 u が読んだ記事集合内の全記事ペア間の語義的距離を平均することで、閲覧セットの分散性を定量化する。値が大きいほど、より広範なトピックに接触していることを示す。一方、GS-score は読者

の重心ベクトル (centroid) に対する平均類似度を測るものであり、読者の関心がどの程度特定トピックに集中しているかを反映する。値が高いほど「専門化」、低いほど「一般化」傾向を示す。

利用者 u が n 本の記事を閲覧したとする。記事タイトルの埋め込みベクトル集合を

$$X^{(u)} = \{x_1, x_2, \dots, x_n\}$$

とおく。ここで、各 x_i は OpenAI の埋め込みモデルにより得られた記事タイトルの意味ベクトルである。二つのベクトル x_i, x_j の類似度は、コサイン類似度によって定義される。

$$\cos(x_i, x_j) = \frac{x_i \cdot x_j}{\|x_i\| \|x_j\|} \quad (1)$$

4.2.1 平均ペアワイズ距離 (APD)

APD は、利用者 u が読んだ記事の全ペア間の距離を平均したものであり、次のように定義される。

$$\text{APD}(u) = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n (1 - \cos(x_i, x_j)) \quad (2)$$

ここで、APD の値が大きいほど、利用者が読んだ記事間の意味的距離が大きく、トピックの多様性が高いことを意味する。

4.2.2 ジェネラリスト?スペシャリストスコア (GS-score)

GS-score は、利用者の閲覧履歴における内部的な一貫性を測定するものである。記事集合の重心ベクトル $\mu^{(u)}$ を

$$\mu^{(u)} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3)$$

と定義すると、GS-score は以下の式で与えられる。

$$\text{GS}(u) = \frac{1}{n} \sum_{i=1}^n \cos(x_i, \mu^{(u)}) \quad (4)$$

このスコアは、閲覧記事群がどの程度重心ベクトルに近いを示し、利用者の関心の集中度を反映する。GS-score が高いほど閲覧記事の主題的一貫性が強く、低いほど多様なトピックにまたがる傾向を示す。

4.2.3 補完的關係

APD と GS-score は、利用者行動の相補的な側面を捉える。すなわち、APD の高い利用者はより広範なトピックに触れており、GS-score の高い利用者は特定領域に集中している。両指標の併用により、読者の「広がり」と「集中度」を同時に評価できる。

謝辞

参考文献

- [1] First, A., Second, A.: Paper Title, *Journal Name*, Vol. xx, No. x, pp. xx-xx (2007)