一般発表 | Track 3: 情報検索・情報推薦・ソーシャルメディア

2026年3月2日(月) 9:30 ~ 11:40 | F会場

## [7F] SNS分析(社会/意見形成)

座長:佐々木 佑樹(富士通)　コメンテータ:児玉 直樹(新潟医療福祉大学)　ジュニアコメンテータ:寺本 優香(同志社大学)

9:30 ~ 9:55

[7F-01] SNSにおける利用者の記事拡散貢献度の定量的評価

*朝澤 颯[1]、吉川 正俊[1] (1. 大阪成蹊大学)

---

9:55 ~ 10:20

[7F-02] ソーシャルメディア断ちは幸福をもたらすか：利用目的と利用行動による影響の検証

*祖父江 智子[1]、林 純子[1]、伊藤 和浩[1]、久田 祥平[1]、若宮 翔子[1]、荒牧 英治[1] (1. 奈良先端科学技術大学院大学)

---

10:20 ~ 10:45

[7F-03] TelegramにおけるQAnon関連コミュニティとそのバックボーンネットワークの可視化

*吉田 真尋[1]、伊藤 貴之[1] (1. お茶の水女子大学大学院)

---

10:45 ~ 11:10

[7F-04] 認知科学に基づくユーザーの偶然性希求行動予測モデルの構築

*鷲見 優一郎[1]、中西 亮輔[1]、光田 英司[1]、二宮 由樹[2]、曽根 悠太郎[2]、三輪 和久[2] (1. トヨタ自動車株式会社 未来創生センター、2. 名古屋大学大学院 情報学研究科)

---

11:10 ~ 11:35

[7F-05] クエリ形式とランキング手法が検索結果のスタンス分布に与える影響の分析

*池元 太陽[1]、山本 岳洋[1] (1. 兵庫県立大学)

# SNS における利用者の記事拡散貢献度の定量的評価

朝澤　颯† 吉川　正俊†

† 大阪成蹊大学データサイエンス学部　〒 533-0007 大阪市東淀川区相川 1 丁目 3 番 7 号
E-mail: †{12370002,yoshikawa-mas}@g.osaka-seikei.ac.jp

**あらまし**　SNS において多数の利用者により記事が拡散された場合に，その拡散に対する各利用者の貢献度を定量化する方法を研究する．参加型ゲームにおけるプレイヤーの貢献度を定量評価するための手法であるバンザフ指数を利用することにより実際のツイートの伝搬データを用いた評価例を示し考察を行う．

**キーワード**　SNS，記事の拡散，情報の伝搬，貢献度評価，バンザフ指数

## 1　はじめに

SNS での情報拡散では，最初の投稿者，それを拡散したインフルエンサー，便乗した一般ユーザなど，複数の主体が関与する．直感的には，「最初の投稿者の貢献度が最も大きい」と理解できる一方で，その「貢献度」を客観的な数値で示すことは容易ではなく，定量化は各利用者にとり公平なものでなければならない．本稿は，この問題に対処するために，協力ゲーム理論の貢献度評価指数であるバンザフ指数 [4] を用いて SNS における各個人の情報拡散貢献度を定量化する枠組みを提案する．それにより，拡散に関わった各主体の寄与を数値として比較可能にし，議論の土台を与えることを目的とする．

真正かつ重要な情報が多くの利用者に広まった場合に，その貢献度を評価することは，利用者がそのような情報を拡散することの誘引となる．他方，拡散された情報がプライバシに関わる情報 [6] や誤情報/偽情報の場合は深刻な社会問題になる [26]．その対処のためには様々な対策を組み合わせる必要がある [3], [7]．このような社会的に問題がある情報が拡散された場合は，各利用者の拡散に対する貢献度は責任の程度と考えることができる．各利用者の責任程度の定量化は，拡散を抑止するための一つの基盤技術になり得ると考える．

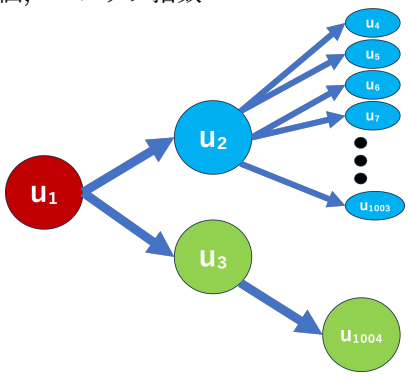


図 1　記事の伝搬グラフ．

## 2　伝搬グラフ

ある情報が SNS によって拡散される様子を表す非巡回有向グラフを定義する．

**定義 1** (伝搬グラフ)．**伝搬グラフ**は情報がノード間で伝搬される条件を表す非巡回有向グラフ $G(R, U, E)$ である．$U$ はノードの集合，$R$ $(\subset U)$ は最初に情報を生成したノードの集合でありソースノード集合と呼ぶ，また，ノード $r \in R$ をソースノードと呼ぶ．有向枝 $e(u_i, u_j) \in E$ はノード $u_i$ からノード $u_j$ へ情報が伝搬されることを表す．$R$ は入枝を持たないノードの集合である．

図 1 は，記事の伝搬グラフの例を示す．

表 1　協力の組合せによる価値（リーチ数）

| 組合せ | 影響力 |
|---|---|
| $u_1$ | 3 |
| $u_2$ | 0 |
| $u_3$ | 0 |
| $u_1, u_2$ | 1003 |
| $u_1, u_3$ | 4 |
| $u_2, u_3$ | 0 |
| $u_1, u_2, u_3$ | 1004 |

## 3　利用者の協力による記事の伝搬

ここでは，SNS での投稿拡散を単純化したモデルとして設定し，分析に用いる「価値」を定義する．

図 1 における次の 3 者に着目する．

$u_1$： 最初の投稿者（事象の起点）
$u_2$： インフルエンサー（拡散力が非常に高い）
$u_3$： 一般ユーザ（拡散力は低い）

本稿でいう「価値」は，投稿が届いた人数（リーチ数）で表す．以降，連携（誰が拡散に関与したか）を利用者の集合で表し，連携 $S$ によって最終リーチが変わるものとして特性関数 $v(S)$ を与える．誰が行動に関わったかによってリーチ数が変化する様子を，表 1 に示す．例えば，$u_1$ と $u_2$ が協力した場合，リーチ数は 1003 となる．

## 4　バンザフ指数

バンザフ指数（Banzhaf Index）は，協力ゲーム理論における貢献度を表す代表的な指数である．もともとは Banzhaf により提案され [4]，重み付き多数決ゲームにおける各プレイヤーの投票力（power）を測るために導入された．

### 4.1　投票ゲームにおける定義

$n$ 人のプレイヤー集合を $N = \{1, 2, \ldots, n\}$ とし，各プレイヤー $i$ に重み $w_i$ を与える．閾値 $q$ 以上で「可決」となる多数決ゲームを考え，任意の部分集合 $S \subseteq N$ に対して特性関数 $v(S)$ を以下のように定義する．

$$v(S) = \begin{cases} 1 & \text{もし } \sum_{i \in S} w_i \geqq q \text{ ならば（可決）} \\ 0 & \text{それ以外（否決）} \end{cases}$$

プレイヤー $i$ が含まれる連携 $S$ について，$v(S) = 1$ かつ $v(S \setminus \{i\}) = 0$ が成り立つとき，$i$ は連携 $S$ における**決定要因（critical player）**であるという．

プレイヤー $i$ が決定要因となる連携の数を $\eta_i$ とすると，非正規化バンザフ値は

$$B_i = \eta_i$$

であり，正規化バンザフ指数は

$$\beta_i = \frac{B_i}{\sum_{j \in N} B_j}$$

で与えられる [29]．

### 4.2　具体例：票数と影響力が一致しないケース

バンザフ指数は，票数（重み）が大きいことが影響力が大きいことと等しいとは限らないことを示すのにも有効である．以下に簡単な例を示す．

a）　例：学園祭の出し物を決める 3 クラス

A, B, C の 3 クラスが投票を行い，重みを

$$w_A = 5, \quad w_B = 4, \quad w_C = 2$$

とする．総票数は 11 票であり，可決ライン（閾値 $q$）を過半数の「6 票以上」とする．

b）「決定要因」の分析

可決となる連携（$v(S) = 1$）において，各クラスが決定要因（抜けると否決）になるかどうかを確認する．

連携 {A,B}:　$5 + 4 = 9$（可決）．A を除くと 4 で否決，B を除くと 5 で否決．よって A,B は決定要因．

連携 {A,C}:　$5 + 2 = 7$（可決）．A を除くと 2 で否決，C を除くと 5 で否決．よって A,C は決定要因．

連携 {B,C}:　$4 + 2 = 6$（可決）．B を除くと 2 で否決，C を除くと 4 で否決．よって B,C は決定要因．

連携 {A,B,C}:　$5 + 4 + 2 = 11$（可決）．A を除いても 6 で可決，B を除いても 7 で可決，C を除いても 9 で可決．よって誰も決定要因ではない．

c） 影響力の算出

決定要因となった回数は $B_A = 2, B_B = 2, B_C = 2$ であり，

$$\beta_A = \beta_B = \beta_C = \frac{2}{6} \approx 33.3\%$$

となる．このように「票数の大小」と「実質的な影響力」が一致しない場合があり得ることが分かる．

## 5 情報拡散における貢献度の定量的評価

本節では，前節で定義したバンザフ指数の数理的枠組みを，SNS 上の拡散現象へ適用し，各利用者の「構造的な責任」を評価する手法について述べる．本手法では，情報の維持および連鎖を協力ゲームの枠組みで捉え，拡散の結果得られた総リーチ数を価値関数 $v(S)$ として定義する．

利用者 $i$ が情報拡散プロセスに加わることで生じるリーチ数の増分は，限界貢献度

$$\Delta_i(S) = v(S) - v(S \setminus \{i\})$$

として計算される．これを全連携について集計したバンザフ値

$$B_i = \sum_{\substack{S \subseteq N \\ i \in S}} \{v(S) - v(S \setminus \{i\})\}$$

を，本研究における「影響力スコア」として定義し，拡散に関与した主体間の責任の重さを定量的に比較する．

### 5.1 理論モデルによる妥当性の確認

表 1 の定義に基づき，発信源（$u_1$），インフルエンサー（$u_2$），一般利用者（$u_3$）の 3 者によるトイモデルでのスコア算出を行った結果は以下の通りである．

$u_1$ （情報の起点）： 2014
$u_2$ （高影響力拡散者）： 2000
$u_3$ （一般利用者）： 2

算出されたスコアは，直感的な責任の重さの序列である $u_1 > u_2 > u_3$ と完全に整合している．特に，フォロワー数に基づく拡散能力が高い $u_2$ よりも，情報の起点である $u_1$ のスコアが上位となる点は重要である．これは，バンザフ指数が「その者がいなければ事象自体が成立しない」という情報の根源的な維持責任を正当に評価できていることを示唆している．

表 2 正規化バンザフ指数上位 10 位までの利用者．

| バンザフ指数順位 | 利用者仮名 | 正規化バンザフ指数 | フォロワー数 | 子孫数 |
|---|---|---|---|---|
| 1 | | 0.296884 | 291286 | 195 |
| 2 | $u_1$ | 0.257413 | -1 | 200 |
| 3 | | 0.221719 | 207674 | 195 |
| 4 | | 0.221719 | 134175 | 195 |
| 5 | $u_2$ | 0.000755 | -1 | 420 |
| 6 | | 0.000755 | -1 | 26 |
| 7 | | 0.000755 | -1 | 67 |
| 8 | | 0.000000 | 3322 | 1 |
| 9 | | 0.000000 | 730 | 1 |
| 10 | | 0.000000 | 9 | 1 |

表 3 フォロワー数上位 10 位までの利用者．

| フォロワー順位 | 利用者仮名 | 正規化バンザフ指数 | フォロワー数 | 子孫数 |
|---|---|---|---|---|
| 1 | | 0.0 | 20273671 | 8 |
| 2 | | 0.0 | 9967383 | 2 |
| 3 | | 0.0 | 2938197 | 13 |
| 4 | | 0.0 | 1684826 | 7 |
| 5 | | 0.0 | 1659278 | 28 |
| 6 | $u_3$ | 0.0 | 1602804 | 227 |
| 7 | | 0.0 | 909547 | 119 |
| 8 | | 0.0 | 617435 | 1 |
| 9 | | 0.0 | 419268 | 39 |
| 10 | | 0.0 | 409110 | 1 |

### 5.2 実際の投稿データへの適用と考察

実証解析として，2026 年 2 月 2 日から 3 日にかけて X に投稿された “Grammy” と “ICE” のキーワードを含む投稿群（5,348 件）を対象に，伝搬構造の解析とバンザフ指数の算出を行った．

表 2，表 3，表 4 は，それぞれ正規化バンザフ指数，フォロワー数，子孫数が上位 10 位までの利用者を示している．フォロワー数は，今回対象とした “Grammy” と “ICE” のキーワードを含む投稿とは独立の指標であるが，（正規化）バンザフ指数と子

表 4　子孫数上位 10 位までの利用者.

| 子孫数順位 | 利用者仮名 | 正規化バンザフ指数 | フォロワー数 | 子孫数 |
|---|---|---|---|---|
| 1 | | 0.000000 | -1 | 672 |
| 2 | | 0.000000 | 176209 | 475 |
| 3 | $u_2$ | 0.000755 | -1 | 420 |
| 4 | | 0.000000 | -1 | 396 |
| 5 | | 0.000000 | -1 | 367 |
| 6 | | 0.000000 | -1 | 258 |
| 7 | | 0.000000 | -1 | 255 |
| 8 | $u_3$ | 0.000000 | 1602804 | 227 |
| 9 | | 0.000000 | -1 | 209 |
| 10 | $u_1$ | 0.257413 | -1 | 200 |

孫数は，これらのキーワードを含む投稿に依存していることに注意されたい．二つ以上の表に現れる利用者には利用者仮名を与え，一つの表にしか現れない利用者の利用者仮名は空白としている．表 2，表 4 の中でフォロワー数が-1 と表示されている利用者は，情報の起点（Root）でありながらデータセット内にプロフィール情報が含まれていなかったアカウントである．

表 3 からは，フォロワー数が 2,000 万を超えるアカウントであっても，情報の連鎖を維持する役割を果たしていない場合，正規化バンザフ指数の値は非常に低くなるケースが確認された．これらの表からは，フォロワー数，（正規化）バンザフ指数，子孫数がそれぞれ異なる指標を与えることがわかる．この結果は，SNS における「知名度（フォロワー数）」と，情報の信頼性や維持に関わる「構造的責任」が乖離していることを示しており，本手法が従来の影響力評価とは異なる多角的な責任評価の軸を提供できることを示唆している．

## 6　関 連 研 究

### 6.1　ソーシャルネットワークにおける利用者の影響度計算

ソーシャルネットワークにおける利用者の影響度計算についてはいくつかの研究がある．

#### 6.1.1　PageRank を利用した Twitter における利用者の影響度計算

Weng ら [27] は，ソーシャルネットワークの構造，発信数，特定のトピックに関する利用者間の類似度をもとに利用者間の遷移確率行列をトピックごとに定義している．この行列を用いてトピック別 PageRank を計算し，トピックを意識した利用者の影響度 TwitterRank を提案した．さらにトピックごとの重み付を行うことにより利用者の全体的な影響力を定義している．我々の研究では，特定の記事が実際に拡散された場合の各利用者の貢献度（責任）を評価する点が異なる．

#### 6.1.2　シャープレイ値の利用

Narayanam ら [20] は，拡散モデルとして linear threshold model を用いている．拡散力の強い top-$k$ 利用者や全体に対して $\lambda$ の割合の利用者に情報を拡散させるための最少の利用者集合を求める問題に取り組み，各ノードの Shapley 値を近似的に計算するための SPINs (ShaPley value-based Influential Nodes) と呼ぶアルゴリズムを提案している．

Gaskó ら [12] は，Social network の拡散モデルとして Independent Cascade Model (ICM) を用い，与えられた大きさ $k$ 以下の初期シード集合 $S$ から最終的に活性化されるノード数を最大化する影響最大化問題（Influence Maximization Problem, IMP）に取り組んでいる．IMP は NP-hard であり，近似・ヒューリスティックが不可欠である．この論文では，協力ゲーム理論（Shapley 値）とメタヒューリスティクス（Extremal Optimization）を組み合わせて greedy 法の高計算コスト既存 Shapley 系手法（SPIN）のスケーラビリティ問題を克服する高精度かつ実用的に解く新手法を提案している．

Chen ら [8] は，social network における情報拡散に対する各ノードの貢献度を評価するために Shapley value を用いた Shapley centrality を提案している．この研究では拡散モデルとして the (random) triggering model を用いているため，ノード間の伝達可能性を確率として与える必要がある．我々は一般的な拡散モデルではなく，実際に投稿された記事を対象とした拡散の貢献度評価を対象としているためその点が異なる．

Becker ら [5] は，Chen ら [8] の Shapley centrality を各利用者から利用者集合に拡張した Group Shapley value を提案しその最大化問題の理論的限界と，小規模集合に対する近似可能性を示した．

**6.1.3** いくつかの centrality の比較

Molinero ら [18] は，ソーシャルネットワークの中心性として，Banzhaf, Shapley-Shubik という二つの指標および二つの新たな指標 effort, satisfaction を，従来からよく知られている degree, closeness, betweeness と比較した．

### 6.2 制約を持つ協力ゲーム

制約を持つ協力ゲーム [2] に関する研究は多くなされているが，その多くは公理系などの理論的性質を明らかにするものであり，本稿のような特定の応用シナリオにおける具体的な適用について研究したものは少ない．

Faigle と Kern [10] は参加順序に制約がある場合のシャープレイ値の理論的性質を明らかにしている．Michalak ら [17] は，ゲーム理論的なグラフ中心性を Shapley 値を用いて多項式時間で計算するアルゴリズムを開発した．Michalak らのモデルでは，効用関数が距離に応じて減衰することを仮定しているため，本論文で対象とした問題に適用することはできない．Myerson [19] は，TU ゲームにおいて，プレイヤーをノードとする無向グラフが与えられている場合に，プレイヤー集合が連結部分グラフを構成する場合にのみ協力可能であるゲームを設定した．この場合に，Myerson 値は，いわゆる Myerson 制限ゲームにおけるシャープレイ値として定義される．Herrings ら [16] は，無向 cycle-free グラフによって協力関係が制限される場合の TU ゲームについて考え，average tree soluction という解を与えている．プレイヤーを有向グラフのノードとし，先祖ノードが許可した場合にのみノードが連合に参加できる許可構造を持つ協力ゲーム [24] についてもいつかの研究がある．Gilles ら [13] や van den Brink ら [25] により導入された連言アプローチ (conjunctive approach) では，プレイヤーが他のプレイヤーと協力するためには，すべての祖先の許可を必要とすると仮定する．一方，階層的許可構造に対し，Gilles [14], van den Brink [22] によって導入された選言アプローチ (disjunctive approach) では，non-top プレイヤーが他のプレイヤーと協力するためには，少なくとも一つの親の許可を必要とすると仮定する．van den Brink [23] は，本稿で定義した伝搬グラフにおいてソースノードが唯一の場合のグラフ構造を階層的許可構造とする協力ゲームを考え公理的特徴づけを示した．Freixas ら [11] は，投票力が複数ありそれらの強弱が全順序でモデル化できる場合のバンザフ指数について研究している．

## 7 議論と今後の課題

### 7.1 評 価 指 標

**7.1.1** バンザフ指数とシャープレイ値

協力ゲームにおける各プレイヤーの貢献度を定量的に評価する方法として，バンザフ指数と並んでシャープレイ値 [21] がよく知られている．シャープレイ値では，プレイヤー $u$ の貢献度を計算する際，プレイヤーのあらゆる参加順序を列挙し，各順序における $u$ の限界貢献度の加重和を計算する．これに対し、バンザフ指数はプレイヤーの参加順序を考慮しない．バンザフ指数は半値 [9] であるため、線形性、対称性、単調性、ダミー変数性といった望ましい公理を満たす。これらの公理に加え、シャープレイ値は効率性（すなわち総ペイオフがプレイヤー間で完全に分配されること）を満たす．我々のシナリオでは、$u$ が $S$ に参加した際に情報が伝搬するノードは，プレイヤーが $S$ に参加する順序に依存しない．したがって，シャープレイ値ではなくバンザフ指数を用いることが適切であると考える．

**7.1.2** 価値関数の精緻化（リーチ数以外の指数）

本稿では「どれだけ届いたか（リーチ）」を価値として扱ったが，実データでは目的に応じて価値の定義を変えることができる．例えば，次のような指表を価値 $v(S)$ として用いる．

- **反応の量（エンゲージメント）**：いいね数・返信数・再投稿数など
- **広がる速さ（拡散速度）**：投稿後の一定時間（例：24 時間）に増えた再投稿数など
- **広がり方（拡散構造）**：どれだけ連鎖が続いた

か（深さ）／どれだけ分岐したか（広がり）
これにより，「どれだけ大きく広がったか」だけでなく，「どれだけ反応を生んだか」「どれだけ速く広がったか」「どういう形で広がったか」も含めて影響力を評価できるようにする．

**7.1.3** 他の指標との比較による妥当性検証

提案手法が妥当かどうかを確かめるため，次数中心性・媒介中心性・PageRank などの既存のネットワーク指標や，単純な再投稿数・フォロワー数と比較する．また，炎上・広告・流行など事例ごとに，「起点の人が強く出るのか」「インフルエンサーが支配的になるのか」といった傾向を整理し，バンザフ指数が**何を新しく説明できるのか**を明確にする．

### 7.2 伝搬グラフとバンザフ値の関係

トピックごとに拡散の形（例：一気に広がる／長く連鎖する）を比べ，バンザフ指数がどのタイプの拡散で大きく出るのかを整理する．

### 7.3 計算の効率性

バンザフ指数は，「考えられる協力パターン」をすべて検討する必要があるため，その時間計算複雑さは一般に参加者数の指数オーダになる．そのため大規模データでは，厳密計算ではなくランダムサンプリング（モンテカルロ法）などの近似計算を用いる．しかし，本稿で対象とするように参加者の参加順序が木構造により制約されている場合には，多項式オーダでの計算が可能となる [28].

### 7.4 手法の実用化

本論文の提案手法を実用化するためには，他のいくつかの要素技術と組み合わせることが必要になる．誤情報/偽情報の検知については多くの研究が行われている [15]．すべての誤情報/偽情報を確実に検知することは困難であるが，これらの技術を用いることにより社会的に影響が大きい誤情報/偽情報については検知することが重要である．

また，ソーシャルネットワークのアカウントと個人の確実な紐付け [1] も必要になる．すべてのソーシャルネットワークプラットフォームのアカウントに紐付けを課すことはプライバシー保護の点で困難であるが，個人とアカウントと紐付けられたプラットフォームの場合は，ここで提案した責任評価方法と組み合わせることにより，誤情報の流通が確実に減った情報流通空間を形成できることが期待できる．

以上より，モデルで示した考え方を実データへ拡張し，SNS 上の拡散における「責任の重さ」をより客観的に議論できる手法として整備していく．

## 8 おわりに

本稿では，SNS 上の拡散を単純化したモデルを用い，誤情報や偽情報が拡散した場合の利用者の責任程度の定量化を考察した．我々はそのためにバンザフ指数によって「誰の行動が拡散結果にどれだけ効いていたか（影響力）」を数値で比較できることを示した．バンザフ指数を用いることで，SNS 上の複数主体が関与する拡散に対し，各主体の「影響力（責任の重さの一側面）」を客観的に比較する一つの枠組みを与えられると考える．

一般に，情報の伝搬はグラフで表現できる．そのとき，その情報の価値，あるいはその情報を伝搬したことに対する貢献（または逆にペナルティ）を現実的な時間で定量的に計算することで社会における公正性を確保できる他の問題にも本研究の考え方を適用できると考える．本論文の手法をより効果的に適用できる応用課題の開拓も将来課題となる．

## 謝　辞

本研究は，JST CREST JPMJCR21M2 の支援を受けたものである

## 文　献

[1] Ireland pushes EU plan for ID-verified social media accounts | Digital Watch Observatory, December 2025. https://dig.watch/updates/ireland-pushes-eu-plan-for-id-verified-social-media-accounts.
[2] Encarnacion Algaba and René van den Brink. The Shapley Value and Games with Hierarchies. Technical Report TI 2019-064/II, Tinbergen Institute, 2019.
[3] Esma Aïmeur, Sabrine Amri, and Gilles Brassard. Fake news, disinformation and misinformation in social media: a review. *Social Network Analysis and Mining*, Vol. 13, No. 1, p. 30, February 2023.

[4] J.F. Banzhaf. Weighted voting doesn't work: A mathematical analysis. *Rutgers Law Review*, Vol. 19, No. 2, pp. 317–343, 1965.

[5] Ruben Becker, Gianlorenzo D'Angelo, and Hugo Gilbert. Maximizing Influence-Based Group Shapley Centrality. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '21, pp. 1461–1463, Richland, SC, May 2021. International Foundation for Autonomous Agents and Multiagent Systems.

[6] Ghazaleh Beigi and Huan Liu. A Survey on Privacy in Social Media: Identification, Mitigation, and Applications. *ACM/IMS Trans. Data Sci.*, Vol. 1, No. 1, pp. 7:1–7:38, March 2020.

[7] Neha Chaudhuri, Gaurav Gupta, Mehdi Bagherzadeh, Tugrul Daim, and Haydar Yalcin. Misinformation on social platforms: A review and research Agenda. *Technology in Society*, Vol. 78, p. 102654, September 2024.

[8] Wei Chen and Shang-Hua Teng. Interplay between social influence and network centrality: A comparative study on shapley centrality and single-node-influence centrality. In *Proceedings of the 26th International Conference on World Wide Web, WWW 2017*, pp. 967–976. ACM, 2017.

[9] Pradeep Dubey, Abraham Neyman, and Robert James Weber. Value theory without efficiency. *Mathematics of Operations Research*, Vol. 6, No. 1, pp. 122–128, 1981.

[10] U. Faigle and W. Kern. The Shapley value for cooperative games under precedence constraints. *International Journal of Game Theory*, Vol. 21, No. 3, pp. 249–266, September 1992.

[11] Josep Freixas. The Banzhaf Value for Cooperative and Simple Multichoice Games. *Group Decision and Negotiation*, Vol. 29, No. 1, pp. 61–74, February 2020.

[12] Noémi Gaskó, Tamás Képes, Rodica Ioana Lung, and Mihai Suciu. Identification of influential nodes with Shapley Influence Maximization Extremal Optimization algorithm. *Applied Soft Computing*, Vol. 146, p. 110653, October 2023.

[13] R. P. Gilles, G. Owen, and R. van den Brink. Games with permission structures: The conjunctive approach. *International Journal of Game Theory*, Vol. 20, No. 3, pp. 277–293, September 1992.

[14] Robert P. Gilles. *The Cooperative Game Theory of Networks and Hierarchies*, Vol. 44 of *Theory and Decision Library C*. Springer, Berlin, Heidelberg, 2010.

[15] Vaishali U. Gongane, Mousami V. Munot, and Alwin D. Anuse. A survey of explainable AI techniques for detection of fake news and hate speech on social media platforms. *Journal of Computational Social Science*, Vol. 7, No. 1, pp. 587–623, April 2024.

[16] P. Jean Jacques Herings, Gerard van der Laan, and Dolf Talman. The average tree solution for cycle-free graph games. *Games and Economic Behavior*, Vol. 62, No. 1, pp. 77–92, January 2008.

[17] T. P. Michalak, K. V. Aadithya, P. L. Szczepanski, B. Ravindran, and N. R. Jennings. Efficient Computation of the Shapley Value for Game-Theoretic Network Centrality. *Journal of Artificial Intelligence Research*, Vol. 46, pp. 607–650, April 2013.

[18] Xavier Molinero, Fabián Riquelme, and Maria Serna. Power Indices of Influence Games and New Centrality Measures for Agent Societies and Social Networks. *Advances in Intelligent Systems and Computing*, pp. 23–30, 2014.

[19] Roger B. Myerson. Graphs and Cooperation in Games. *Mathematics of Operations Research*, Vol. 2, No. 3, pp. 225–229, August 1977.

[20] Ramasuri Narayanam and Yadati Narahari. A Shapley Value-Based Approach to Discover Influential Nodes in Social Networks. *IEEE Transactions on Automation Science and Engineering*, Vol. 8, No. 1, pp. 130–147, January 2011.

[21] Lloyd S. Shapley. A Value for N-Person Games. March 1952. 09488.

[22] René Van Den Brink. An axiomatization of the disjunctive permission value for games with a permission structure. *International Journal of Game Theory*, Vol. 26, No. 1, pp. 27–43, March 1997.

[23] René van den Brink. Axiomatizations of Banzhaf permission values for games with a permission structure. *International Journal of Game Theory*, Vol. 39, No. 3, pp. 445–466, July 2010.

[24] René van den Brink. Games with a permission structure - A survey on generalizations and applications. *TOP*, Vol. 25, No. 1, pp. 1–33, April 2017.

[25] René van den Brink and Robert P. Gilles. Axiomatizations of the Conjunctive Permission Value for Games with Permission Structures. *Games and Economic Behavior*, Vol. 12, No. 1, pp. 113–126, January 1996.

[26] Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. *Science*, Vol. 359, No. 6380, pp. 1146–1151, March 2018.

[27] Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He. TwitterRank: finding topic-sensitive influential twitterers. In *Proceedings of the third*

*ACM international conference on Web search and data mining*, WSDM '10, pp. 261–270, New York, NY, USA, February 2010. Association for Computing Machinery.

[28] Masatoshi Yoshikawa and Hayate Asazawa. Efficient quantification of responsibility for the spread of misinformation and disinformation using the banzhaf index. In *5th International Workshop on Data Science for Equality, Inclusion and Well-being (DS4EIW 2026)*. IEEE, May 2026. (to appear).

[29] 福田恵美子. 投票力指数. オペレーションズ・リサーチ, Vol. 55, No. 12, pp. 788–789, 2010.

# ソーシャルメディア断ちは幸福をもたらすか：利用目的と利用行動による影響の検証

祖父江智子† 林 純子† 伊藤 和浩† 久田 祥平† 若宮 翔子†
荒牧 英治†

† 奈良先端科学技術大学院大学 〒 630–0192 奈良県生駒市高山町 8916 番 5 号
E-mail: †sobue.tomoko.sr8@naist.ac.jp,
††{hayashi.junko.hh5,ito.kazuhiro.ih4,s-hisada,wakamiya,aramaki}@is.naist.jp

**あらまし** 近年，ソーシャルメディアユーザ数は大幅に増加しており，その利用がメンタルヘルスに与える影響を明らかにすることは重要な課題である．しかし，先行研究ではソーシャルメディア利用と幸福感の関連について一貫した知見は得られていない．その一因として，同じソーシャルメディア利用であっても，利用目的（例：逃避，暇つぶし，交流）や利用行動（例：投稿する，他者の投稿行動を閲覧する）によって幸福感への影響が異なる可能性が十分に考慮されてこなかった点が挙げられる．加えて，多くの先行研究は観察研究に依存しており，因果的な影響の推定は十分でない．こうした課題を克服するために，本研究では，利用目的および利用行動が異なる参加者を対象に，ソーシャルメディア断ちを伴う介入実験を実施した．その結果，普段の利用行動や利用目的の特性によって，断ちが幸福感に及ぼす影響が異なることが示された．例えば，同じ逃避目的であっても，投稿行動を主とするユーザと閲覧行動を主とするユーザとでは，ソーシャルメディア断ちが幸福感に及ぼす効果が逆方向となる場合が確認された．

**キーワード** ウェルビーイング，ソーシャルメディア，介入実験，ソーシャルメディア断ち，幸福感

## 1 はじめに

近年，ソーシャルメディアは情報通信インフラとして定着し，人々の日常生活やコミュニケーションの様式を大きく変容させている．その利便性の一方で，過度な利用が心身の健康に悪影響を及ぼす可能性が懸念されている．特に，青少年や若年成人においては，うつや睡眠障害などとの関連が指摘されており [1], [2]，一部の国や地域では，青少年のソーシャルメディア利用を制限する政策的対応も導入されている [3]．こうした懸念は世代を問わず，成人においてもソーシャルメディア利用と幸福感との間に負の関連が報告されている [4]．このような背景のもと，ソーシャルメディア利用が人々の幸福感にどのような影響を及ぼすのか，その因果関係を明らかにすることは重要な課題である．この課題に対し，ソーシャルメディア利用を一時的に停止させる「ソーシャルメディア断ち」などの介入実験がこれまでに複数実施されてきた．

しかし，先行研究の結論は必ずしも一貫していない．ソーシャルメディア断ちによる幸福感の向上 [5], [6] と低下 [7], [8] の双方が報告されており，依然として議論の余地が残されている．この結論の不一致の一因として，多くの研究が，利用時間や頻度といった単一の指標に依拠し，ソーシャルメディア上での多様な利用行動を十分に捉えられていない点が挙げられる [1]．メタ分析においても，具体的な利用行動や心理的要因に着目した検証の必要性が指摘されている [9], [10]．例えば，同じように長時間ソーシャルメディアを利用していた場合であっても，現実逃避を目的として他者の投稿を受動的に閲覧していたユーザと，同じく現実逃避を目的としながらも，自ら投稿を行っていたユーザとでは，ソーシャルメディア断ちによる幸福感の変化が異なる可能性がある．すなわち，ソーシャルメディア断ちの影響は，ユーザが「どのような行動」を「どのような目的」で行っていたかによって異なると考えられる．しかし，これらの要素を包括的に考慮した検討は十分に行われていない．


図 1: 実験概要．普段のソーシャルメディアの利用行動及び利用目的が異なる参加者を対象に，ソーシャルメディア断ちが幸福感に与える影響を検証する．

以上の背景を踏まえ，本研究では，次のリサーチクエスチョンを設定する．

**ソーシャルメディア断ちが幸福感に与える影響は，ユーザの利用行動や利用目的によってどのように変化するか？**

この問いに対し，本研究では，(i) 利用行動の違いが幸福感の変化に及ぼす影響，(ii) 利用行動と利用目的の組み合わせの違いが幸福感の変化に及ぼす影響の二点に着目し，分析を行う．

具体的には，普段のソーシャルメディア利用を利用行動（投稿・閲覧）および利用目的（娯楽，現実逃避，社会的交流，自己宣伝，情報収集，無目的）に分解し，ソーシャルメディア断ちが，幸福感に及ぼす影響を詳細に検討する．そのために，事前調査，1 週間の観察期間，そして 2 週間のソーシャルメディア断ち介入期間からなる実験を実施し，参加者の普段のソーシャルメディア利用行動や利用目的，ならびに実験期間中の幸福感に関するデータを収集した．収集したデータを用いて，まず利用行動の違いによる幸福感の変化を検討し，次に利用行動と利用目的の組み合わせが幸福感の変化に及ぼす影響を，階層ベイズモデルにより推定する．

本研究の主な貢献は以下の 3 点である．

- 利用時間のみに注目していた従来のソーシャルメディア断ち実験に対して，利用行動と利用目的を考慮した詳細な分析枠組みを導入した．
- 階層ベイズモデルを用いることで，個人差を考慮しつつ，利用行動と利用目的の組み合わせ効果を定量的に評価した．
- 特定の利用目的（例：逃避）において，投稿と閲覧が幸福感に対して逆方向の影響を与えることを明らかにし，利用行動と利用目的の組み合わせを分析することの重要性を実証的に示した．

## 2 関連研究

本節では，まず § 2.1 でソーシャルメディア利用と幸福感の因果関係を調べた介入実験研究の知見を概観する．次に，§ 2.2 で利用行動（能動的利用・受動的利用）に着目した先行研究をそれぞれ取り上げ，幸福感との関連についてどのような知見が得られてきたのかを検討する．さらに，利用行動と利用目的を組み合わせた観点から幸福感への影響を検討した研究がほとんど存在しない点を指摘し，本研究が取り組む研究ギャップを明確化する．

### 2.1 ソーシャルメディア利用と幸福感の因果関係

これまでに，ソーシャルメディア断ちを通じて幸福感への因果的影響を検討した介入実験研究が複数報告されているが，その結論は必ずしも一貫していない．例えば，1 日あたりのソーシャルメディア利用を 30 分に制限する 2 週間の介入実験では，幸福感の有意な向上が報告された [5]．一方で，ソーシャルメディア断ちを 1 週間実施した別の研究では，人生満足度の低下が示された [7]．ソーシャルメディア断ちによる幸福感への有意な影響は認められなかったとする研究も存在し [10], [11]，ソーシャルメディア断ちの効果の正負や大きさには依然として議論の余地がある．

このような結論の不一致の一因として，ソーシャルメディア利用の影響は，利用行動や利用目的によって異なることが考えられる．「ソーシャルメディア利用」と一括りにしても，実際にはユーザがどのような行動を行い，どのような目的で利用しているかは大きく異なる．これらの利用行動や利用目的の違いは，既存の断ち研究では十分に考慮されてこなかった．

### 2.2 ソーシャルメディア利用行動と幸福感

ソーシャルメディアの利用行動と幸福感との関係については，観察研究および介入研究の双方が実施されてきたが，その結論は必ずしも一致していない．先行研究では，利用行動は「能動的利用」と「受動的利用」に分類され，投稿やコメントといった能動的利用は幸福感と正の関連を示す一方で，閲覧中心の受動的利用は負の関連を示すとする見解が広く共有されてきた [12]．

こうした先行研究の知見に基づき，近年では利用行動の因果的影響を検討するための介入実験研究も報告されている．これらの研究では，閲覧行動が幸福感の低下と関連する可能性や，能動的利用が相対的に幸福感の向上と関連することが示唆されており，これまでの知見と整合的な結果が得られている [13], [14]．

一方で，能動的利用と受動的利用という単純な二分法に対しては，近年批判的な検討もなされている．Instagram において写真投稿を促す介入研究では，ポジティブ感情には正の効果が認められたものの，他の幸福感指標に有意な変化は見られなかった [15]．複数のレビュー論文では，能動的利用が一貫して幸福感に正の影響を及ぼし，受動的利用が負の影響を及ぼすとする仮説は必ずしも支持されておらず，利用行動と幸福感との関連は弱い，あるいは文脈依存的である可能性が指摘されている [16], [17]．

以上の先行研究を踏まえると，ソーシャルメディア断ちを用いた介入実験研究は一定数存在するものの，参加者の利用行動や利用目的の違いを考慮した上で介入効果を検討した研究はほとんど見られない．多くの研究では，ソーシャルメディア利用の多様性が十分に反映されておらず，介入による幸福感への影響について一貫した結論が得られていない可能性がある．そこで本研究では，利用行動および利用目的が異なる参加者を対象としたソーシャルメディア断ちの介入実験を実施し，これらの要素によって幸福感への因果的影響がどのように異なるかを検討する．

## 3 研究方法

### 3.1 実験デザイン

本研究では，普段のソーシャルメディア利用特性（利用行動・利用目的）が，ソーシャルメディア断ち期間中の幸福感の変化に及ぼす影響を検討するため，被験者内デザインを採用した．実験開始前に事前調査を実施し，観察期間（通常利用）1 週間と介入期間（ソーシャルメディア断ち）2 週間からなる計 21 日間の実験をオンライン環境下で実施し，参加者の日常生活の中でデータを収集した．本研究は，奈良先端科学技術大学院大学の倫理審査委員会の承認（承認番号: 2025-I-16）を得て実施した．

本研究における独立変数は，事前調査で測定された個人のソーシャルメディアにおける「利用行動（投稿・閲覧）」および「利用目的（娯楽，逃避，社会的交流，自己宣伝，情報収集，無目的）」である．従属変数は，実験期間を通じて繰り返し測定さ

表 1: 参加者の人口統計的属性 ($N = 77$)

| 属性 | 人数 | 割合 (%) |
|---|---|---|
| **性別** | | |
| 男性 | 30 | 39.0 |
| 女性 | 47 | 61.0 |
| **年齢** | | |
| 10 代 | 1 | 1.3 |
| 20 代 | 20 | 26.0 |
| 30 代 | 36 | 46.8 |
| 40 代 | 14 | 18.2 |
| 50 代 | 6 | 7.8 |

れた「現在の幸福感」である．

### 3.2 参 加 者

#### 3.2.1 事前スクリーニング

本研究では，ソーシャルメディアの利用行動（投稿・閲覧）が確立している参加者を選定するため，クラウドワークス [1]を通じて参加者を募集し，以下の条件に基づいて事前スクリーニングを実施した．

**X（旧 Twitter）の利用頻度**：プラットフォーム間の機能差を統制するため，主要なスクリーニング基準を X の利用状況に限定した．利用行動の習慣を確認するため，以下のいずれか一方，あるいは両方を満たすことを条件とした．

- **投稿行動**：1 日 1 回以上の投稿を行っていること．
- **閲覧行動**：1 日 1 時間以上，他者の投稿を閲覧していること．

ただし，他のソーシャルメディア（Instagram, Facebook 等）を併用しているユーザも募集対象に含めた．

**スマートフォンでの利用**：OS 標準のスクリーンタイム機能等を用いて，自己申告ではない客観的な利用状況（利用時間およびアプリ起動）の確認を行うため，スマートフォン（iOS または Android）で X を利用していることを条件とした．

**データ共有への同意**：アプリの利用時間データ（スクリーンショット等）の研究目的で共有することに同意できることを条件とした．

スクリーニング条件に同意した参加者は 94 名であった．このうち，実験期間中にドロップアウトした者および無効なアンケートが確認された者を除外した結果，最終的な参加者は **77 名**となった．

#### 3.2.2 事 前 調 査

選定された参加者に対し，実験開始前に事前調査を実施し，基本属性（性別，年齢，最終学歴）およびソーシャルメディアの利用実態を収集した．参加者の人口統計的属性を表 1 に，具体的な質問を付録の表 A·1 に示す．

利用実態の把握においては，X，Instagram，Facebook，TikTok，BeReal，Snapchat の主要 6 プラットフォームを対象に，以下の項目を測定した．これらの項目は，後の分析において利用行動や利用目的を分類するための独立変数として用いる．

- **投稿頻度**：リポストやコメントを除き，平均してどの程度の頻度で投稿を行うかを 6 件法（「3 日に 1 回未満」～「1 日に 11 回以上」）で測定した．
- **閲覧時間**：1 日に平均してどの程度の時間，他者の投稿を閲覧するかを 8 件法（「10 分未満」～「3 時間以上」）で測定した．
- **利用目的**：投稿および閲覧行動を行う目的（娯楽，逃避，交流，自己宣伝，情報収集など）について，その頻度を 5 段階リッカート尺度（1 =「全くしない」～5 =「ほとんど毎回」）で測定した．

参加者の利用実態を付録の表 A·2 に示す．なお，BeReal をインストールしている参加者はいなかったため，表 A·2 には同プラットフォームに関する記述統計量は含めていない．

### 3.3 実験手続き

本研究は，利用行動および利用目的ごとに，ソーシャルメディア利用の停止が幸福感に及ぼす影響を推定することを目的として，観察期間と介入期間からなる 21 日間（2025/11/5～2025/11/25）の実験を実施した．21 日間における回答状況を付録の図 A·1 に示す．

#### 3.3.1 観 察 期 間

最初の 1 週間（1 日目～7 日目）は観察期間とし，参加者には普段通りの生活およびソーシャルメディア利用を求めた．この期間中，参加者は 1 日 3 回配信されるアンケートに回答し，その時点における「現在の幸福感」を報告した（詳細は § 3.4 参照）．このデータは，介入前の状態（統制条件）として用いる．

#### 3.3.2 介 入 期 間

続く 2 週間（8 日目～21 日目）は介入期間とし，参加者に対して対象となるソーシャルメディアの利用を禁止した（ソーシャルメディア断ち）．禁止対象としたプラットフォームは，X, Instagram, Facebook, TikTok, BeReal, Snapchat, Threads, Pinterest，Bluesky である．なお，LINE や YouTube はソーシャルメディアに含まれる場合もあるが，本研究で対象とするコミュニケーションや情報発信・閲覧とは主目的が異なるため，利用禁止の対象から除外した．

参加者は介入期間中も観察期間と同様に，1 日 3 回の幸福感に関するアンケートへの回答を継続した．また，ソーシャルメディア断ちの遵守状況を把握するため，参加者に対してスマートフォンのスクリーンタイム機能のスクリーンショットの提出を求め，利用時間がゼロ（または意図しないバックグラウンド通信等の許容範囲内）であることを確認した．

### 3.4 幸福感の測定

幸福感の測定には，13 種類の感情（退屈感，孤独感，不安，嫌悪感，反発心，興奮，疲れ，ポジティブな感情，ネガティブな感情，悲しみ，恐れ，喜び，怒り）を用いた．これらの感情

1：https://crowdworks.jp/

は，Scale of Positive and Negative Experience (SPANE) [18] に基づく先行研究 [19] を参考に選定した．各感情について，その時点における程度を 5 段階リッカート尺度（1 =「全く感じない」〜5 =「非常に感じる」）で測定した．

本研究では，13 種類の測定項目のうち，感情価に基づく幸福感を捉えるため，ポジティブな感情および喜びの平均値から，ネガティブな感情，悲しみ，恐れ，および怒りの平均値を引いた値を**幸福感スコア**として用いた．以下では，記述の簡潔さのため，幸福感スコアを「幸福感」と表記する．

## 4 分析手法

本節では，事前調査および実験期間中に収集されたデータを整理し，仮説検証のための分析手続きについて説明する．§ 4.1 と § 4.2 では，利用行動および利用目的に関するアンケートの回答カテゴリを数値（スコア）に変換し，平均化することで，分析の独立変数を構築する手順を説明する．介入効果の分析は，大きく分けて二つの段階から構成される．まず，§ 4.3 では，ソーシャルメディア断ちによる幸福感の変化が，普段の利用行動の違いによってどのように異なるかを検討する．次に，§ 4.4 では，階層ベイズモデルを用い，利用行動と利用目的の組み合わせに着目し，幸福感の変化を説明する要因をより詳細に推定する．

### 4.1 利用行動の定量化

投稿頻度および閲覧時間として収集されたカテゴリカルデータを連続値（スコア）に変換し，参加者の利用行動を定量化した．

**投稿頻度スコア**：投稿頻度は，6 件法の回答カテゴリの階級値に基づき，1 日あたりの平均投稿回数として解釈可能な実数値へ変換した．具体的には，「3 日に 1 回未満」(0.2)，「2〜3 日に 1 回程度」(0.4)，「1 日に 1 回程度」(1.0)，「1 日に 2〜4 回程度」(3.0)（代表値），「1 日に 5〜10 回程度」(7.5)（中央値），「1 日に 11 回以上」(11.0)（最小値）として割り当てた．これにより，順序尺度であった回答を，比率尺度として扱う投稿頻度スコアへ変換した．

**閲覧時間スコア**：閲覧時間は，8 件法の回答カテゴリを時間（h）単位の実数値へ変換した．各カテゴリの範囲の中央値を代表値として採用し，具体的には以下のように換算した．「10 分未満」(0.1)，「10 分以上 30 分未満」(0.33)，「30 分以上 1 時間未満」(0.75)，「1 時間以上 1 時間 30 分未満」(1.25)，「1 時間 30 分以上 2 時間未満」(1.75)，「2 時間以上 2 時間 30 分未満」(2.25)，「2 時間 30 分以上 3 時間未満」(2.75)，「3 時間以上」(3.5)．なお，上限の設定がない「3 時間以上」については，保守的な推定値として 3.5 時間を割り当てた．これにより，日常的な閲覧行動を定量的な時間量として表現した．

**平均スコアによる集約**：各参加者のソーシャルメディアにおける利用行動をプラットフォーム横断的に表すため，スマートフォンにインストールされている全プラットフォームにおける投稿頻度スコアおよび閲覧時間スコアの平均値を算出した．具体的には，それぞれの平均値を，それぞれ平均投稿スコアおよび平均閲覧スコアと定義した．これらのスコアは，参加者のソーシャルメディア利用行動の頻度および量を集約した指標であり，§ 4.3 の利用行動タイプの分類に用いる．

### 4.2 利用行動・利用目的の定量化

利用目的として収集された自己報告データをスコアに変換し，投稿行動および閲覧行動ごとの利用目的を定量化した．

**利用目的スコア**：6 種類の目的（娯楽，逃避，社会的交流，自己宣伝，情報収集，無目的）に対する利用頻度について，5 段階リッカート尺度（1 =「全くしない」〜5 =「ほとんど毎回」）で測定した回答に基づき算出した．

**平均スコアによる集約**：各参加者のソーシャルメディアにおける利用目的をプラットフォーム横断的に表すため，スマートフォンにインストールされている全プラットフォームにおける利用目的スコアの平均値を算出した．具体的には，各目的 $k$ に対して，投稿行動および閲覧行動の平均値をそれぞれ $S_{post,k}$ および $S_{view,k}$ と定義した．例えば，娯楽目的（$k =$ 娯楽）での投稿行動について，X と Instagram におけるスコアがそれぞれ 2 と 4 である場合，その参加者の $S_{post,k}$ は 3.0 となる．これらのスコアは，参加者のソーシャルメディアにおける利用目的の傾向を，投稿行動および閲覧行動別に集約した指標であり，§ 4.4 の分析では個人特性 (Trait) を表す独立変数として用いる．

### 4.3 利用行動タイプ別の介入効果の検証：対応のある t 検定

本節では，ソーシャルメディア断ちが幸福感に与える影響が利用行動タイプによって異なるかを確認するため，対応のある $t$ 検定 (paired $t$-test) を用いた群別解析の手順を述べる．まず，§ 4.1 で定義した平均投稿スコアおよび平均閲覧スコアに基づき，参加者 ($N = 77$) を以下の 3 つの利用行動タイプに分類した．

- **投稿群** ($N = 19$)：能動的な発信行動が習慣化している群（平均投稿スコア $\geqq 1.0$）．
- **閲覧群** ($N = 30$)：投稿は少ないが，受動的な閲覧行動が習慣化している群（平均投稿スコア $< 1.0$ かつ平均閲覧スコア $\geqq 1.0$）．
- **低利用群** ($N = 28$)：投稿・閲覧ともに利用が少ない群（上記以外）．

これらの各群において観察期間（通常利用）と介入期間（ソーシャルメディア断ち）の幸福感スコアの平均値を比較し，その変化の有意性を検証した．なお，低利用群は，日常的な利用が少なく，介入の影響が相対的に小さいと考えられるため，分析における参照グループとして扱う．

### 4.4 利用行動・利用目的が介入効果に与える影響の推定：階層ベイズモデル

§ 4.3 に示した群別解析は，利用行動の主要な特性に基づく

傾向を把握する上で有効である．しかし，利用行動と利用目的の組み合わせといった詳細な条件を扱う場合，各条件におけるサンプルサイズが限定的となり，統計的な推定の不確実性が増大する懸念がある．また，連続的な利用傾向をカテゴリカルに分割することにより，境界付近の情報が失われ，閾値の設定に結果が左右される可能性がある．

そこで，各参加者の利用行動および利用目的のスコアを説明変数として直接モデルに組み込み，それらがソーシャルメディア断ちの介入効果にどのように影響するかを，交互作用として推定する．分析には，個人差（ランダム効果）と時系列相関（自己回帰項）を考慮した階層ベイズモデルを用いた．

本モデルでは，§ 4.2 で定義した利用行動（投稿・閲覧）と 6 種類の利用目的（娯楽，逃避，社会的交流，自己宣伝，情報収集，無目的）を組み合わせた計 12（$= 2 \times 6$）の特性変数をモデルに組み込んだ．例えば，「閲覧 × 娯楽」という特性変数は，娯楽目的での閲覧行動の頻度を表すスコア（$S_{\text{view, 娯楽}}$）に対応する．

モデルの基本式を以下に示す．

$$
\begin{aligned}
&WB_{t,i} \sim \mathcal{N}(\mu_{t,i}, \sigma_\varepsilon) \\
&\mu_{t,i} = \alpha + \alpha_i \\
&\qquad + \beta_{\text{phase}} \cdot Phase_{t,i} + \beta_{\text{time}} \cdot Time_{t,i} + \beta_{\text{lag}} \cdot WB_{t-1,i} \\
&\qquad + \sum_{k=1}^{K} \Big( \beta_{\text{trait},k} \cdot Trait_{k,i} + \beta_{\text{int},k} \cdot (Trait_{k,i} \times Phase_{t,i}) \Big)
\end{aligned}
$$

ここで，$WB_{t,i}$ は時点 $t$ における参加者 $i$ の幸福感，$\mu_{t,i}$ は $WB_{t,i}$ の期待値，$\sigma_\varepsilon$ は観測誤差の標準偏差，$\alpha$ は全参加者に共通する切片，$\alpha_i$ は参加者 $i$ ごとのランダム切片，$\beta_{\text{phase}}$ は介入フェーズ（ソーシャルメディア断ち）の主効果，$\beta_{\text{time}}$ は時間経過が幸福感に与える主効果，$\beta_{\text{lag}}$ は直前の幸福感が現在の幸福感に与える自己回帰効果，$\beta_{\text{trait},k}$ は $k$ 番目の特性の主効果，$\beta_{\text{int},k}$ は $k$ 番目の特性と介入フェーズとの交互作用効果，$Phase_{t,i}$ は観察期間 (0) と介入期間 (1) を表すダミー変数，$Time_{t,i}$ 時間指標（観察期間および介入期間開始からの経過日数），$WB_{t-1,i}$ は直前時点の幸福感，$K$ は特性の総数，$Trait_{k,i}$ は参加者 $i$ の $k$ 番目の特性スコアを表す．

推定には，4 本のチェーンを用いたマルコフ連鎖モンテカルロ（Markov chain Monte Carlo: MCMC）法を適用し，各チェーンから 2,000 サンプルを生成した．得られた事後分布については，事後平均および両側 94% 最高密度区間（highest density interval: HDI）を指標として算出する．判断基準として，HDI が完全に 0 を下回る項目は，当該利用の停止と幸福感との間に負の関連（利用停止に伴う幸福感の低下）があるものとし，逆に HDI が完全に 0 を上回る項目は，正の関連（利用停止に伴う幸福感の上昇）があるものとして解釈した．これにより，利用行動と利用目的の交互作用を考慮したソーシャルメディア断ちが幸福感に与える影響を推定した．

## 5 結　　果

本節では，ソーシャルメディア断ち実験から得られた分析結果を報告する．まず § 5.1 では，利用行動別（投稿・閲覧・低利用）における幸福感の変化について，基礎的な統計検定の結果を示す．続いて § 5.2 では，階層ベイズモデルによる結果に基づき，利用行動および利用目的が幸福感の変動に与える影響を詳細に検討する．

### 5.1 利用行動タイプ別の幸福感変化

§ 4.3 で定義した 3 つの利用行動タイプ（投稿群・閲覧群・低利用群）について，観察期間（通常利用）と介入期間（ソーシャルメディア断ち）の幸福感スコアを比較した結果を表 2 に示す．表より，日常的に投稿行動を行う**投稿群**においてのみ，観察期間（平均 0.021）から介入期間（平均 −0.304）にかけて，幸福感の有意な低下が確認された（変化量 $-0.325$, $p < 0.05$）．一方で，主に閲覧行動を行う**閲覧群**においては，幸福感の低下傾向が見られたものの，有意な差ではなかった（変化量 $-0.185$, $p > 0.05$）．また，日常的な利用頻度が低い**低利用群**では，期間による幸福感の変化はほとんど認められなかった（変化量 $-0.028$, $p > 0.05$）．

以上の結果より，ソーシャルメディア断ちが幸福感に与える影響は一様ではなく，特に投稿群において幸福感の低下が確認された．

### 5.2 利用行動・利用目的別の幸福感変化

次に，ソーシャルメディア断ちが幸福感に及ぼす影響が，利用行動（投稿・閲覧）と利用目的の組み合わせによってどのように異なるかを検討した（表 3）．

投稿行動においては，日常的に逃避または情報収集を目的とする傾向が強い参加者ほど，介入期間中に幸福感が大きく低下することが示された（投稿 × 逃避：$\beta = -0.176$，投稿 × 情報収集：$\beta = -0.167$, いずれも 94% HDI は 0 を含まない）．これは，これらの目的による投稿行動が，日常的な幸福感の維持に寄与していた可能性を示唆する．一方，社会的交流を目的とした投稿行動をする傾向が強い参加者では，ソーシャルメディア断ちによって幸福感が上昇する傾向が認められた（$\beta = 0.146$, 94% HDI は 0 を含まない）．これは，日常の社会的交流を目的でとした発信行動が，幸福感を抑制する要因となっていた可能性を示している．

閲覧行動に関しては，利用目的によって介入効果が対照的な結果となった．日常的に娯楽目的で閲覧する参加者ほど，介入に伴う幸福感の低下が大きかった（$\beta = -0.118$，94% HDI は 0 を含まない）．一方で，逃避や情報収集を目的とする参加者では，ソーシャルメディア断ちにより幸福感の上昇がみられた（閲覧 × 逃避：$\beta = 0.119$，閲覧 × 情報収集：$\beta = 0.068$，いずれも 94% HDI は 0 を含まない）．これは，日常的な娯楽目的での閲覧は幸福感の維持に質する一方で，逃避や情報収集を目的とした閲覧は，幸福感を阻害する機能を果たしていた可能性を示唆する．

以上の結果を視覚的に補足するため，介入効果と顕著な関連を示した特定の組み合わせ（正の関連：**投稿 × 社会的交流**，負の関連：**投稿 × 逃避**）に着目する．これらの特性スコア（$S_{\text{post, 社会的交流}}$ または $S_{\text{post, 逃避}}$）が 4.0 以上の参加者をそ

表 2: 利用行動タイプ別の幸福感の推移と変化の要約．変化量は「介入期間 − 観察期間」の差を示す．† は対応のある $t$ 検定において $p < 0.05$ で有意であることを示す．

| 利用行動タイプ | 人数 | 観察期間 | 介入期間 | 変化量 |
|---|---|---|---|---|
| 投稿群 | 19 | 0.021 | -0.304 | -0.325† |
| 閲覧群 | 30 | 0.500 | 0.315 | -0.185 |
| 低利用群 | 28 | 0.694 | 0.666 | -0.028 |

表 3: 階層ベイズモデルによる介入効果の推定結果．$\beta$ は各特性と介入フェーズの交互作用効果を示す．† は正または負の効果が認められた（94 % HDI が 0 を含まなかった）項目を示す．利用行動と利用目的の組み合わせごとに異なる幸福感への影響が示唆された．

| 利用行動・利用目的 | 推定値 ($\beta$) |
|---|---|
| 投稿 × 娯楽 | -0.051 |
| 投稿 × 逃避 | -0.176† |
| 投稿 × 社会的交流 | 0.146† |
| 投稿 × 自己宣伝 | -0.027 |
| 投稿 × 情報収集 | -0.167† |
| 投稿 × 無目的 | 0.055 |
| 閲覧 × 娯楽 | -0.118† |
| 閲覧 × 逃避 | 0.119† |
| 閲覧 × 社会的交流 | 0.009 |
| 閲覧 × 自己宣伝 | -0.036 |
| 閲覧 × 情報収集 | 0.068† |
| 閲覧 × 無目的 | 0.034 |

れぞれ抽出し，各群における幸福感の平均値の推移を図 2 に示す．なお，その他の利用行動・利用目的の組み合わせに関する推移は付録の図 A·2 に示す．

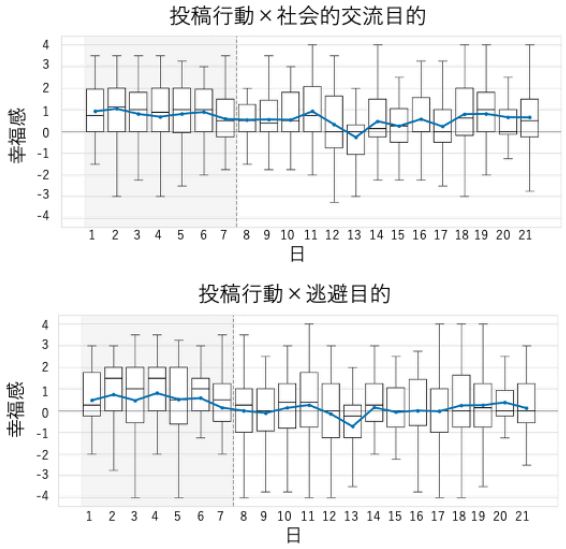


図 2: 特性別の幸福感スコアの推移比較．日常的に「投稿 × 社会的交流」の傾向が強い群（上）と，「投稿 × 逃避」の傾向が強い群（下）の幸福感平均値を示す．網掛けは観察期間（1–7 日目），白地は介入期間（8–21 日目）を表す．社会的交流目的の投稿が多い群では，実験期間を通じて幸福感が概ね正の値で推移しているのに対し，逃避目的の投稿が多い群では，介入期間において幸福感が低下し，負の値を示す日が相対的に多く認められる．

## 6 考　　察

### 6.1 利用行動タイプ別の幸福感変化

利用行動別の幸福感変化を分析した結果，日常的に投稿を行う参加者がソーシャルメディア断ちをすると，幸福感が低下することが示された．これは，日常的な投稿行動がユーザの幸福感を一定程度支える機能を果たしていた可能性を示唆する．これは，投稿行動をはじめとする能動的なソーシャルメディア利用が幸福感と正の関連を持つことを示唆する点で先行研究 [12], [14], [15] と一致する．一方で，本研究では，投稿行動の中でも利用目的によっては介入期間中に幸福感の低下と関連する場合があることも示されており，能動的利用が一様に幸福感に正の影響を及ぼすわけではない点にも注意を払う必要がある．

### 6.2 利用行動・利用目的別の幸福感変化

利用行動と利用目的の交互作用に着目した先行研究は，我々の知る限り存在せず，本研究はこの領域に体系的な知見を提供するものである．特に，投稿行動と閲覧行動それぞれに対して目的別の効果が異なることが明らかになった点は，両者を独立に扱う既存研究を補完する新規性の高い知見といえる．

本研究では，同一の利用目的であっても，投稿行動と閲覧行動では，介入期間中の幸福感への影響が逆方向となる結果が観察された．例えば，投稿行動における逃避目的および情報収集目的では，介入期間中に幸福感の低下がみられたことから，これらの投稿行動が日常的には幸福感の維持に一定程度寄与していた可能性が考えられる．逃避目的での投稿は，ストレスや疲労感を伴う状態において，感情を言語化し外在化する手段として機能していた可能性がある．このようなネガティブな内容の開示は幸福感の改善と関連していることが示されている [20]．また，情報収集目的での投稿は，単なる情報取得行動ではなく，他者からの反応を通じて理解を深める探索的行動として機能していた可能性がある．他者からのフィードバックや社会的サポートの知覚が促進されたこと [21] が，幸福感を支える要因となっていた可能性が示唆される．

一方で，閲覧行動における逃避および情報収集目的では，介入期間中に幸福感の上昇が観察された．このことは，日常的な受動的閲覧が幸福感を抑制していた可能性を示唆している．逃避目的での閲覧行動は，感情を処理したり外在化したりすることなく，受動的に注意を逸らす行為に留まる可能性がある．その結果，過剰な情報接触や他者との比較を通じて，日常的な幸福感を低下させていた可能性が考えられる [22], [23]．情報収集目的での閲覧行動は，多量の情報に受動的に曝露される形になりやすく，認知的負荷や情報疲労を高める可能性がある．この

ような情報過多による疲労が幸福感に負の関連を持つことが考えられる [22]．そのため，日常的には幸福感を阻害する側面を有していた可能性が示唆される．

以上の解釈は介入による幸福感の変化から導かれた推論に基づくものであり，各行動が持つ心理的機能を直接測定したものではない．今後は，認知的負荷や自己効力感，注意の消耗といった媒介要因を含めた検討が求められる．

### 6.3 今後の課題

本研究にはいくつかの限界が存在する．1 点目は，サンプルサイズが比較的限定的であるため，推定性能および外的妥当性に一定の制約があることである．特に，介入研究では効果量の検出力がサンプルサイズに依存するため，本研究の推定結果が小規模サンプル特有のばらつきの影響を受けている可能性がある．

2 点目は，本研究の参加者がソーシャルメディア断ちに前向きな態度を持つ者に偏っている可能性である．このような選択バイアスが存在する場合，ソーシャルメディア利用に強い動機づけを持つユーザやソーシャルメディア断ちに抵抗感のあるユーザが過小代表され，得られた効果推定が母集団全体の平均的反応を必ずしも反映していない懸念がある．

3 点目は，利用行動および利用目的の指標が自己報告に基づいている点である．これらの指標は想起バイアスや社会的望ましさバイアスの影響を受ける可能性があり，利用ログデータやアプリ使用時間といった客観的指標を併用できなかったことによる測定誤差が，結果の解釈に影響している可能性がある．

4 点目は，本研究の参加者が主に 30 代から 40 代で構成されている点である．ソーシャルメディア利用の心理的影響については，青年期や若年成人を中心に懸念が示されることが多い [2], [3]．年齢層によってソーシャルメディアの利用目的や日常生活における位置づけは異なる可能性があるため，本研究の結果を青年期や若年成人にそのまま一般化することには注意が必要である．今後の研究では，20〜30 代の年齢層の参加者を対象とした介入実験を通じて，本研究の知見の外的妥当性を検討することが求められる．

以上を踏まえると，本研究はソーシャルメディアの利用行動および利用目的と幸福感の関係について一定の知見を提供するものである一方，その解釈にあたってはサンプル特性，測定方法，および研究デザインに由来する制約を考慮する必要がある．

## 7 おわりに

本研究では，ソーシャルメディア断ちという介入を通じて，日常的な利用行動や利用目的が幸福感の変化に及ぼす影響を検討した．分析の結果，まず利用行動の観点では，日常的に投稿行動を行う個人ほど，ソーシャルメディア断ちによって幸福感が大きく低下する傾向が確認された．これは，ソーシャルメディアにおける投稿行動が，日常生活における幸福感の維持に一定の機能を果たしている可能性を示唆している．さらに，利用行動と利用目的の組み合わせに着目すると，逃避や情報収集を目的とした投稿行動は，介入後の幸福感低下と関連しており，これらの目的に基づく投稿行動が日常の幸福感維持に寄与していたことが推察された．一方で，社会的交流を目的とした投稿行動は，介入によって幸福感が向上する傾向が認められ，日常的な交流目的の発信がむしろ幸福感を抑制する要因となっていた可能性が示された．また，閲覧行動に関しては，娯楽目的の閲覧が幸福感の維持に資する一方で，逃避および情報収集目的での閲覧は，介入後に幸福感が上昇したことから，日常的には幸福感を阻害する側面を持つことが明らかとなった．

本研究の新規性は，利用行動や多様な利用目的の個人差を踏まえたうえでソーシャルメディア利用と幸福感の関係性を因果的観点から検討した点にある．本研究で得られた知見は，ソーシャルメディア利用の是非を画一的に論じるのではなく，個々人の利用形態に応じた健康的なソーシャルメディア利用のあり方を検討する上で，重要な指針や示唆を提供するものと考えられる．

## 謝　辞

本研究の一部は，「戦略的イノベーション創造プログラム (SIP)」「統合型ヘルスケアシステムの構築」JPJ012425 の補助を受けて行った．

## 文　献

[1] Chirag Gupta, Sangita Jogdand, Mayank Kumar, CHIRAG GUPTA, and Sangita D Jogdand. Reviewing the impact of social media on the mental health of adolescents and young adults. *Cureus*, Vol. 14, No. 10, 2022.

[2] Patti M. Valkenburg, Adrian Meier, and Ine Beyens. Social media use and its impact on adolescent mental health: An umbrella review of the evidence. *Current Opinion in Psychology*, Vol. 44, pp. 58–68, 2022.

[3] Rachel Fieldhouse and Mohana Basu. Australia's world-first social media ban is a 'natural experiment' for scientists. *Nature*, December 2025.

[4] Holly B. Shakya and Nicholas A. Christakis. Association of Facebook Use With Compromised Well-Being: A Longitudinal Study. *American Journal of Epidemiology*, Vol. 185, No. 3, pp. 203–211, February 2017.

[5] Paige Coyne and Sarah J Woodruff. Taking a break: the effects of partaking in a two-week social media digital detox on problematic smartphone and social media use, and other health-related outcomes among young adults. *Behavioral Sciences*, Vol. 13, No. 12, p. 1004, 2023.

[6] Hunt Allcott, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. The welfare effects of social media. *American economic review*, Vol. 110, No. 3, pp. 629–676, 2020.

[7] Zahir Vally and Caroline G. D'Souza. Abstinence from social media use, subjective well-being, stress, and loneliness. *Perspectives in Psychiatric Care*, Vol. 55, No. 4, pp. 752–759, 2019.

[8] Lisa C. Walsh, Annie Regan, Karynna Okabe-Miyamoto, and Sonja Lyubomirsky. Does putting down your smartphone make you happier? the effects of restricting digital media on well-being. *PLOS ONE*, Vol. 19, No. 10, pp. 1–25, October 2024.

[9] Kaitlyn Burnell, Diana J. Meter, Fernanda C. Andrade, Ashley N. Slocum, and Madeleine J. George. The effects of social media restriction: Meta-analytic evidence from randomized controlled trials. *SSM - Mental Health*, Vol. 7, p.

100459, 2025.

[10] Laura Lemahieu, Yannick Vander Zwalmen, Marthe Mennes, Ernst Koster, Mariek Abeele, and Karolien Poels. The effects of social media abstinence on affective well-being and life satisfaction: a systematic review and meta-analysis. *Scientific Reports*, Vol. 15, p. 7581, March 2025.

[11] Michael Wadsley and Niklas Ihssen. Restricting social networking site use for one week produces varied effects on mood but does not increase explicit or implicit desires to use SNSs: Findings from an ecological momentary assessment study. *PLOS ONE*, Vol. 18, No. 11, pp. 1–20, November 2023.

[12] Philippe Verduyn, Oscar Ybarra, Maxime Résibois, John Jonides, and Ethan Kross. Do Social Network Sites Enhance or Undermine Subjective Well-Being? A Critical Review: Do Social Network Sites Enhance or Undermine Subjective Well-Being? *Social Issues and Policy Review*, Vol. 11, pp. 274–302, January 2017.

[13] Philippe Verduyn, David Seungjae Lee, Jiyoung Park, Holly Shablack, Ariana Orvell, Joseph Bayer, Oscar Ybarra, John Jonides, and Ethan Kross. Passive Facebook usage undermines affective well-being: Experimental and longitudinal evidence. *Journal of Experimental Psychology: General*, Vol. 144, No. 2, p. 480, 2015.

[14] Sarah M. Hanley, Susan E. Watt, and William Coventry. Taking a break: The effect of taking a vacation from Facebook and Instagram on subjective well-being. *PLOS ONE*, Vol. 14, No. 6, pp. 1–13, June 2019.

[15] Hannes-Vincent Krause, Fenne große Deters, Annika Baumann, and Hanna Krasnova. Active social media use and its impact on well-being—an experimental study on the effects of posting pictures on instagram. *Journal of Computer-Mediated Communication*, Vol. 28, No. 1, p. zmac037, 2023.

[16] Patti M Valkenburg, Irene I van Driel, and Ine Beyens. The associations of active and passive social media use with well-being: A critical scoping review. *New media & society*, Vol. 24, No. 2, pp. 530–549, 2022.

[17] Rebecca Godard and Susan Holtzman. Are active and passive social media use related to mental health, wellbeing, and social support outcomes? a meta-analysis of 141 studies. *Journal of Computer-Mediated Communication*, Vol. 29, No. 1, p. zmad055, 2024.

[18] Ed Diener, Derrick Wirtz, Robert Biswas-Diener, William Tov, Chu Kim-Prieto, Dong-Won Choi, and Shigehiro Oishi. New Measures of Well-Being. *Social Indicators Research Series*, Vol. 39, pp. 247–266, April 2009.

[19] Victoria Oldemburgo de Mello, Felix Cheung, and Michael Inzlicht. Twitter (X) use predicts substantial changes in well-being, polarization, sense of belonging, and outrage. *Communications Psychology*, Vol. 2, No. 1, February 2024.

[20] Koustuv Saha, Dong Whi Yoo, Vedant Das Swain, and Munmun Choudhury. Mental wellbeing effects of disclosing life events on social media. *Scientific Reports*, Vol. 15, p. 23519, July 2025.

[21] Jack Lipei Tang. Are You Getting Likes as Anticipated? Untangling the Relationship between Received Likes, Social Support from Friends, and Mental Health via Expectancy Violation Theory. *Journal of Broadcasting & Electronic Media*, Vol. 66, No. 2, pp. 340–360, June 2022.

[22] Jörg Matthes, Kathrin Karsay, Desirée Schmuck, and Anja Stevic. "Too much to handle": Impact of mobile social networking sites on information overload, depressive symptoms, and well-being. *Computers in Human Behavior*, Vol. 105, p. 106217, 2020.

[23] Ziyu Liu and Liyao Xiao. Is It Just About Scrolling? The Correlation of Passive Social Media Use with College Students' Subjective Well-Being Based on Social Comparison Experiences and Orientation Assessed Using a Two-Stage Hybrid Structural Equation Modeling–Artificial Neural Network Method. *Behavioral Sciences*, Vol. 14, No. 12, 2024.

## 付　　録

表 A·1: 事前調査の質問項目の概要および質問文

(a) 質問項目の概要

| **質問項目** | **内容（回答形式）** |
|---|---|
| 基本属性 | 性別（男性／女性／無回答），年齢（10 代〜90 代以上），最終学歴（該当選択肢） |
| 利用行動 | 投稿頻度（3 日に 1 回未満〜1 日に 11 回以上），閲覧時間（10 分未満〜3 時間以上） |
| 利用目的 | 娯楽，逃避，社会的交流，自己宣伝，情報収集，無目的（5 件法リッカート尺度） |

(b) 質問文

| **指標** | **質問文** |
|---|---|
| 投稿頻度 | 平均してどの程度の頻度で投稿を行いますか（リポストやコメントを除く） |
| 閲覧時間 | 平均してどの程度の時間，SNS 上で他の人の投稿を閲覧しますか |
| 投稿 × 娯楽 | 楽しんだりリラックスしたりするために投稿しますか |
| 投稿 × 逃避 | 日常を忘れたり，ストレスを紛らわせたりするために投稿しますか |
| 投稿 × 社会的交流 | 同じ関心を持つ人と交流したり，自分の考えや反応を共有したりするために投稿しますか |
| 投稿 × 自己宣伝 | 自分の活動やビジネスをアピールしたり，収益化したりするために投稿しますか |
| 投稿 × 情報収集 | 他の人の反応や意見を通してインスピレーションを得たり，情報を探したりするために投稿しますか |
| 投稿 × 無目的 | 特に理由はなく，習慣的または無意識に投稿しますか |
| 閲覧 × 娯楽 | 楽しんだりリラックスするために，他の人の投稿を閲覧しますか |
| 閲覧 × 逃避 | 日常を忘れたり，ストレスから気を逸らしたりするために，他の人の投稿を閲覧しますか |
| 閲覧 × 社会的交流 | 同じ関心を持つ人と交流したり，自分の考えを共有したりするために，他の人の投稿を閲覧しますか |
| 閲覧 × 自己宣伝 | 自分の活動やビジネスのアピールや収益化に役立てるために，他の人の投稿を閲覧しますか |
| 閲覧 × 情報収集 | インスピレーションを得たり，企業や製品に関する情報を探したりするために，他の人の投稿を閲覧しますか |
| 閲覧 × 無目的 | 特に理由はなく，習慣的または無意識に，他の人の投稿を閲覧しますか |

注：利用行動に関する質問は，X，Instagram，Facebook，TikTok，BeReal，Snapchat の各プラットフォームを対象として実施した．各プラットフォームにおいて，回答者が行った利用行動（投稿・閲覧）に対応する利用目的を尋ねた．投稿頻度および閲覧時間は順序尺度に基づく質問項目である．利用目的は 5 段階リッカート尺度（1=全くしない〜5=ほとんど毎回）で回答を求めた．

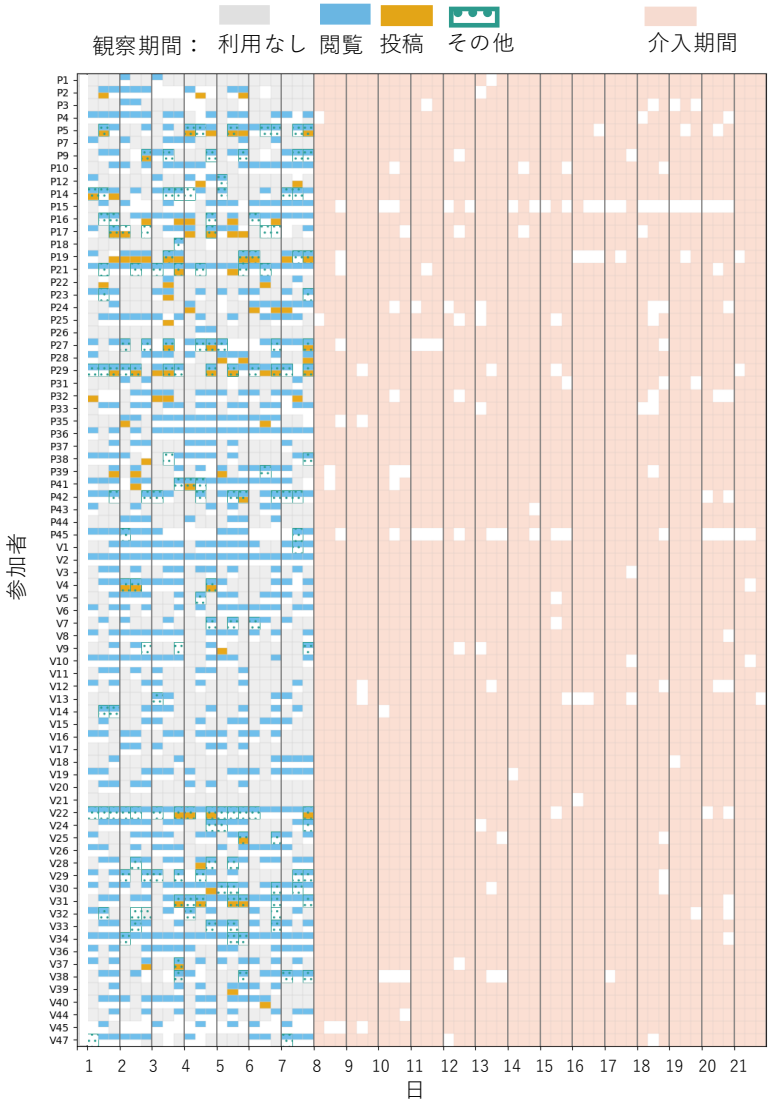


図 A·1: 21 日間の調査における参加者の回答状況およびソーシャルメディア利用形態．各行は個々の参加者を，各列は調査日（1–21 日目）を示す．1–7 日目は観察期間 (通常利用)，8–21 日目は介入期間（ソーシャルメディア断ち）である．各セルの色は，当該日の回答の有無（未回答の場合は白）およびソーシャルメディアの利用形態（利用なし，閲覧，投稿，その他）を区別している．介入期間中のセルの着色は，介入の遵守状況に関わらず，当該日の回答が行われたことを示している．

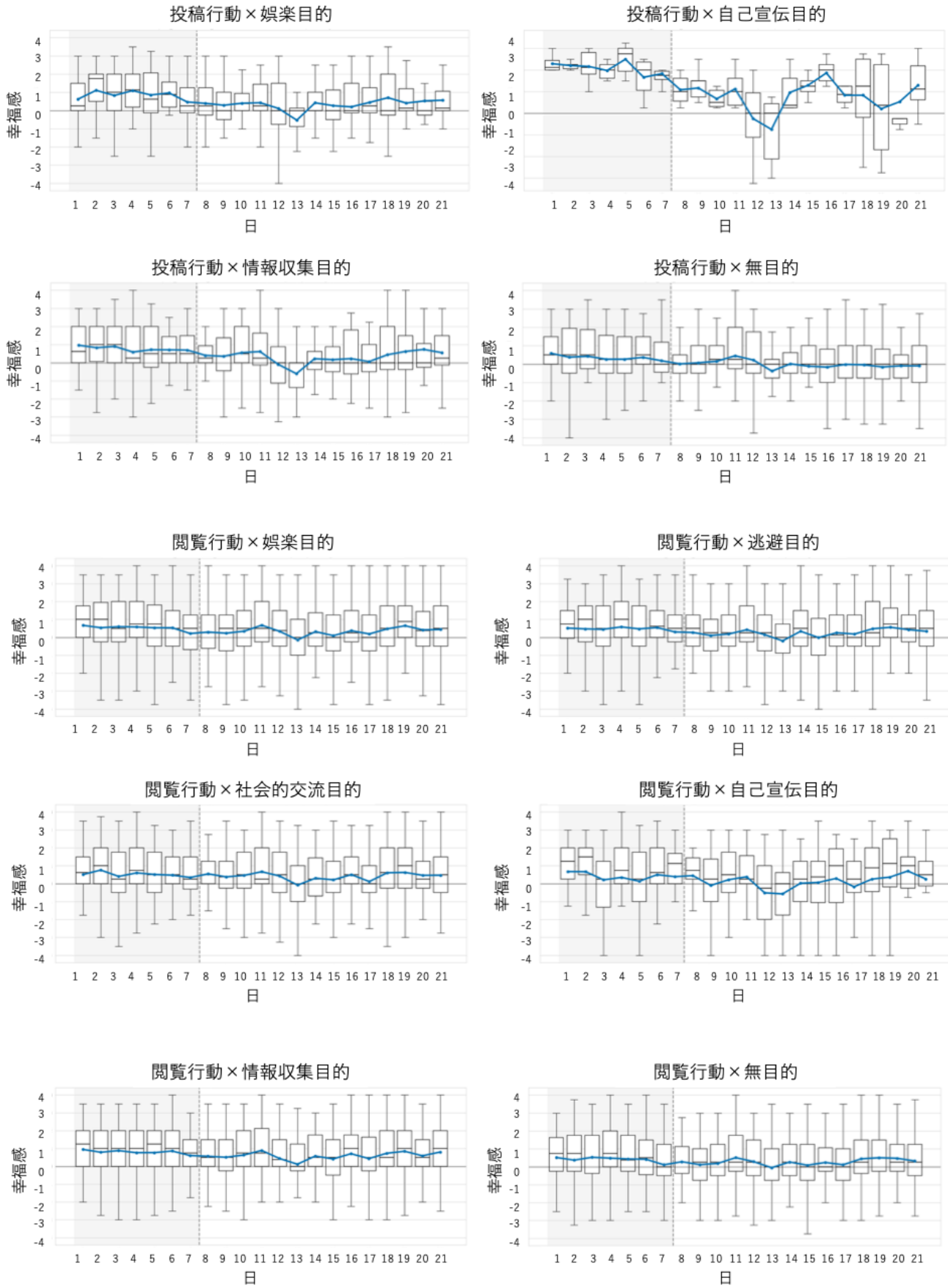


図 A·2: 特性別の幸福感スコアの推移．娯楽，自己宣伝，情報収集および無目的での投稿行動を日常的に行う群と，娯楽，逃避，社会的交流，自己宣伝，情報収集および無目的での閲覧行動を日常的に行う群の幸福感平均値を示す．網掛けは観察期間（1–7 日目），白地は介入期間（8–21 日目）を表す．

表 A·2: 各プラットフォームにおける利用実態の記述統計量

(a) 投稿頻度の分布

| プラットフォーム | 3 日に 1 回未満 | 2〜3 日に 1 回 | 1 日に 1 回 | 1 日に 2〜4 回 | 1 日に 5〜10 回 | 1 日に 11 回以上 |
|---|---|---|---|---|---|---|
| X | 35 | 9 | 18 | 7 | 6 | 1 |
| Instagram | 44 | 3 | 1 | 1 | 0 | 0 |
| Facebook | 7 | 0 | 0 | 0 | 0 | 0 |
| TikTok | 16 | 1 | 0 | 0 | 0 | 0 |
| Snapchat | 1 | 0 | 0 | 0 | 0 | 0 |

(b) 閲覧時間の分布

| プラットフォーム | 10 分未満 | 10〜30 分 | 30 分〜1 時間 | 1〜1.5 時間 | 1.5〜2 時間 | 2〜2.5 時間 | 2.5〜3 時間 | 3 時間以上 |
|---|---|---|---|---|---|---|---|---|
| X | 2 | 4 | 9 | 28 | 13 | 7 | 5 | 8 |
| Instagram | 23 | 6 | 11 | 3 | 0 | 2 | 2 | 2 |
| Facebook | 6 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| TikTok | 10 | 2 | 1 | 0 | 0 | 0 | 0 | 3 |
| Snapchat | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

(c) 投稿行動 × 利用目的

| プラットフォーム | 娯楽 | 逃避 | 社会的交流 | 自己宣伝 | 情報収集 | 無目的 |
|---|---|---|---|---|---|---|
| X | 2.618 (1.119) | 2.816 (1.314) | 3.053 (1.574) | 1.461 (0.807) | 2.382 (1.404) | 2.566 (1.370) |
| Instagram | 2.184 (1.495) | 1.878 (1.333) | 2.143 (1.514) | 1.327 (0.774) | 1.939 (1.329) | 1.939 (1.391) |
| Facebook | 1.857 (1.464) | 1.286 (0.488) | 1.714 (0.951) | 1.429 (0.787) | 1.714 (0.951) | 1.857 (1.215) |
| TikTok | 1.688 (1.352) | 1.688 (1.401) | 1.812 (1.223) | 1.500 (0.966) | 1.688 (1.302) | 1.750 (1.390) |
| Snapchat | 4.000 (–) | 4.000 (–) | 2.000 (–) | 2.000 (–) | 2.000 (–) | 2.000 (–) |

(d) 閲覧行動 × 利用目的

| プラットフォーム | 娯楽 | 逃避 | 社会的交流 | 自己宣伝 | 情報収集 | 無目的 |
|---|---|---|---|---|---|---|
| X | 3.895 (1.040) | 3.934 (0.957) | 3.671 (1.269) | 1.763 (1.165) | 3.250 (1.338) | 4.224 (0.903) |
| Instagram | 3.367 (1.410) | 2.898 (1.475) | 2.551 (1.487) | 1.510 (1.003) | 2.939 (1.435) | 3.184 (1.564) |
| Facebook | 3.286 (1.254) | 2.571 (1.618) | 3.000 (1.528) | 1.571 (0.787) | 1.857 (1.215) | 2.714 (1.704) |
| TikTok | 3.375 (1.746) | 3.375 (1.746) | 2.688 (1.621) | 1.875 (1.455) | 2.062 (1.289) | 3.125 (1.708) |
| Snapchat | 2.000 (–) | 2.000 (–) | 2.000 (–) | 2.000 (–) | 2.000 (–) | 2.000 (–) |

注：(a) と (b) は，投稿頻度および閲覧時間の回答分布，(c) と (d) は，投稿行動および閲覧行動に対応する利用目的の平均値（標準偏差）を示す．投稿頻度（「3 日に 1 回未満」: 0.2〜「1 日に 11 回以上」: 11.0）および閲覧時間（「10 分未満」: 0.1〜「3 時間以上」: 3.5）は，回答カテゴリに順序変数を割り当てて算出した．利用目的は 5 段階リッカート尺度（1：全くしない〜5：ほとんど毎日）の平均値である．なお，Snapchat 利用者は 1 名のみであったため，標準偏差は算出していない．また，BeReal 利用者は存在しなかったため，同プラットフォームの統計量は除外している．

# Telegram における QAnon 関連コミュニティとそのバックボーンネットワークの可視化

吉田　真尋† 　伊藤　貴之†

† お茶の水女子大学大学院人間文化創成科学研究科　〒 112–8610 東京都文京区大塚 2-1-1
E-mail: †{g2020542,itot}@is.ocha.ac.jp

**あらまし**　QAnon に代表される陰謀論・過激派運動は，オンライン空間を基盤として国際的に拡散し現実社会に深刻な影響を及ぼしている．本研究では，QAnon 支持者の主要な活動拠点となっている Telegram に着目し，チャンネルをノード，引用メッセージをエッジとするネットワークとして大規模データセットを構築した．まず，ネットワーク全体の可視化を通じて，QAnon 関連チャンネルが政治，誤情報，ウェルネス等の多様なトピック領域と結びついている構造を俯瞰的に把握した．さらに可視化システムで発見できた興味深いサブネットワークに有向グラフに拡張した Disparity Filter を適用し，バックボーンネットワークを抽出した．これにより，情報拡散において本質的な役割を果たすチャンネル間の関係を抽出し，QAnon 関連コミュニティと他トピック領域との結節点を明確化した．本研究は，オンライン過激主義の拡散構造をネットワーク可視化とバックボーン分析の両面から明らかにする点に特徴がある．

**キーワード**　ソーシャルメディア，Telegram，情報可視化

## 1　はじめに

インターネットを起点とする過激派運動や陰謀論の拡散は，オンライン空間にとどまらず，現実社会における暴力行為や政治的不安定化を引き起こす要因として深刻な問題となっている．特に，匿名性と高い拡散性を特徴とするオンライン・プラットフォームは，過激な思想や陰謀論が形成・共有される温床となっている．2016 年のコメット・ピンポン銃撃事件 [1] や，2021 年のアメリカ合衆国議会議事堂襲撃事件に代表されるように，オンライン上で醸成された世界観が集団的暴力として顕在化する事例が確認されている．これらの事例の中心に位置するのが，アメリカ合衆国発の陰謀論運動である QAnon である．QAnon は，2017 年に匿名掲示板 4chan および 8chan（8kun）への投稿を起点とし，「政財界のエリートが悪魔崇拝的な犯罪に関与している」という主張を核に支持を拡大してきた．当初は特定の匿名掲示板に限定されていたが，主要 SNS におけるデプラットフォーミングの進展により，支持者は規制の緩やかな Telegram などのプラットフォームへと活動の場を移行している．

2024 年時点において，QAnon はアメリカ国内にとどまらず国際的に拡散し，各国の社会的・政治的文脈と結びつきながら独自の展開を見せている．ドイツにおける国家転覆未遂事件や，日本における QAnon 的言説の拡散は，この運動が国境を越えて市民社会に影響を及ぼしていることを示している．さらに QAnon 内部には，政治的トピックに加え，ネットミーム，健康，スピリチュアリティ，代替医療といった多様な関心領域に基づく分派が存在し，従来の政治的過激主義とは異なる経路で支持者を獲得している点も特徴的である．一方で，SNS 上の QAnon に関する先行研究の多くは，投稿内容や言説の分析に主眼を置いており，チャンネルやユーザ間の関係性が形成するネットワーク構造に十分な注意を払っていない．また，Telegram を対象とした研究においても，QAnon 関連チャンネルのみを抽出する分析にとどまり，それらが他の非 QAnon チャンネルとどのように接続し，情報を流通させているのかを包括的に可視化した研究は限られている．

そこで本研究では，QAnon 支持者の主要な活動拠点となっている Telegram に着目し，QAnon 関連チャンネルとそれ以外のチャンネルを含む大規模データセットを構築する．そして，チャンネルをノード，引用メッセージをエッジとするネットワークとして可視化する．本研究により，オンライン上に形成されるコミュニティ構造と情報拡散のパターンを直感的かつ定量的に把握し，QAnon 運動がいかに他のトピック領域と結びつきながら維持・拡散されているのかを明らかにする．

## 2　関連研究

極右運動およびその分派に関する研究は，これまで政治学・社会学・メディア研究の分野を中心に蓄積されてきた [2], [3]．その中でも QAnon は，オンライン空間を基盤とした新しい形態の陰謀論的・過激派運動として注目を集めており，思想内容のみならず，その拡散経路や他領域との接合に関する研究が進められている．

### 2.1　QAnon とその分派に関する関連研究

Baker [4] は，COVID-19 パンデミック期に影響力を持った代替医療系インフルエンサーの Instagram 投稿を対象にその投稿テーマを分析し，ウェルネス文化と陰謀論が容易に結びつく構造を明らかにした．この研究では，QAnon 的言説が健康増進や自己啓発といった語彙に包摂されることで，政治的過激性を弱めた形で拡散されている点を指摘している．また，red pill や awakening といった QAnon 特有のコードを用いたゲーミフィケーションが，支持者の参加と没入を促進していることも示された．

Conner [5] は，QAnon インフルエンサー 100 名の投稿や相互参照関係を質的に分析し，ヨガやスピリチュアリティといった一見 QAnon と無関係なサブカルチャーが，陰謀論への入口として機能していることを明らかにした．これらの領域では，政治的主張が前面に出ることなく，精神的共同体や健康志向といった動機を通じて支持者が取り込まれる点に特徴があり，QAnon が自己を擬態させながら浸透していると論じられている．

さらに，ネットミームを介した極右言説の拡散も重要な研究テーマである．Pepe the Frog に代表されるネットミームは，極右思想の象徴として機能するだけでなく [6], [7]，医療誤情報や陰謀論の拡散手段としても利用されていることが報告されている [8]．

これらの研究は，QAnon が単一の政治運動ではなく，複数の文化的領域と接合した複合的現象であることを示している．

### 2.2 SNS における QAnon コミュニティについての関連研究

Hoseini ら [9] は，Telegram 上の多言語 QAnon チャンネルを対象に大規模データを分析し，言語圏ごとに主流となるトピックが異なることを明らかにした．この研究では BERT を用いたトピック分類が採用されており，QAnon の国際的拡散を言語の観点から捉えている．Angermaier ら [10] は，スノーボールサンプリングと Human-in-the-Loop を組み合わせることで，QAnon 関連チャンネルのみからなる大規模データセットを構築し，その研究利用価値を示した．また，Thomas ら [11] は，Telegram 上の引用関係を用いたスノーボールサンプリングにより QAnon コミュニティを収集・分析している．しかし，これらの関連研究の多くは，QAnon 関連チャンネルのみを分析対象としており，それらが他の非 QAnon チャンネルとどのように接続しているのかをネットワーク構造として可視化した研究は限られている．

以上の関連研究を踏まえ，本研究では QAnon に直接関連するチャンネルに加え，ウェルネスやネットミームなど間接的に関与するチャンネルも含むデータセットを構築し，引用関係に基づくネットワーク可視化を行うことで，QAnon 運動の拡散構造と他領域との接合点を明らかにする．

## 3 提案手法

### 3.1 データセットの収集手法

本研究ではスノーボールサンプリングによってデータセットを収集した．具体的には，QAnon に関連するニュースを発信するチャンネルから，多数の最新メッセージ（本稿では 1000 件）を走査し，その中に他のチャンネルから引用されたメッセージがあったら引用元のチャンネルを同様に走査する，という処理を（本稿では 7 回）繰り返した．以上の処理により，米国大統領候補討論会開催直後である 2024 年 9 月 11 日から 2024 年 9 月 12 日を対象として，チャンネル 4912 件と引用メッセージ 93567 件を抽出したデータセットを生成した．このデータセットの基本的な統計情報を表 1 に示す．この統計情報から，本データセットが多言語に対応していることがわかる．

表 1　データセットの基本統計情報

| 項目 | 値 |
|---|---|
| 最古引用日 | 2015-11-04 16:35:06 |
| 最新引用日 | 2024-09-10 20:56:46 |
| チャンネル数 | 4913 |
| メッセージ数 | 93,567 |
| メディア付きメッセージ数 | 86,745 |
| メディア付きメッセージの割合（%） | 92.71 |
| チャンネル概要文使用言語数 | 40 |
| メッセージ使用言語数 | 40 |

さらに，チャンネルの説明文をもとに各チャンネルに BERTopic [12] を適用して，（本稿では 30 種類の）トピックに分類した．本研究は多言語にわたるチャンネルの分析を目的とすることから，トピック分類には事前学習済み多言語 BERT モデル [13] を使用した．さらに 30 トピックの中から，Sharma ら [14] の研究結果をもとに，QAnon に直接関連するトピック，QAnon に直接関連しないが誤情報や陰謀論と関係の深いトピック，ネットミームに関連するトピック，ロシア大使館に関連するトピックを 8 個抜粋した．これにより，QAnon に直接関連しないトピックもデータセットに含まれることがわかる．表 2 は 8 個のトピックとその代表的単語及びテキストの表である．

また，チャンネルの説明文とメッセージが書かれた言語をもとに，チャンネルとメッセージの言語を特定した．表 3 は可視化に使用したチャンネルのメタデータ，表 4 は可視化に使用したメッセージのメタデータを表す．

表 3　可視化に使用したチャンネルのメタデータ

| 変数名 | メタデータ |
|---|---|
| id | チャンネルの ID |
| label | チャンネル名 |
| value | チャンネルの参加者数 |
| description | チャンネルの説明文 |
| group | description から決定されたトピック |
| language | description の言語 |

表 4　可視化に使用したメッセージのメタデータ

| 変数名 | メタデータ |
|---|---|
| source | メッセージの引用元チャンネルの ID |
| target | メッセージの引用先チャンネルの ID |
| date | メッセージが引用された日時 |
| forwards | メッセージが引用された回数 |
| views | メッセージが閲覧された回数 |
| message | メッセージの内容 |
| language | メッセージが書かれた言語 |
| url | メッセージに添付された URL |

表 2　可視化に使用したトピック例

| トピック名 | チャンネル数 | 代表的単語 |
|---|---|---|
| 1_канал_россии_официальный_russian | 530 | ["канал", "россии", "официальный","russian","новости","ru","на","по","российской","для"] |
| 2_news_and_independent_the | 333 | ["news","and","independent","the","world","on","media","from","journalist","geopolitics"] |
| 5_catholic_and_god_orthodox | 138 | ["catholic","and","god","orthodox","the","christ","to","is","of","church"] |
| 7_health_covid_and_vaccine | 126 | ["health","covid","and","vaccine","19","of","the","vaccines","medical","to"] |
| 9_israel_news_islamic_palestine | 100 | ["israel","news","islamic","palestine","resistance","the","middle","east","gaza","and"] |
| 10_memes_meme_and_me | 99 | ["memes","meme","and","me","of","for","all","the","content","your"] |
| 14_white_pro_nationalist_our | 48 | ["white","pro","nationalist","our","and","we","aktivistmann","anti","com","racist"] |
| 15_trump_of_president_the | 46 | ["trump","of","president","the","donald","congresswoman","america","patriot","united","house"] |
| 18_cats_images_ai_photos | 36 | ["cats","images","ai","photos","or","other","art","cat","random","and"] |

### 3.2　可視化手法

以上の手法で得られたデータセットに cosmograph [15] を適用し，チャンネルをノード，引用したメッセージをエッジとして可視化した．図 1 は可視化システムの使用例であり，ノードの大きさは参加者数，エッジの太さは引用回数に比例する．

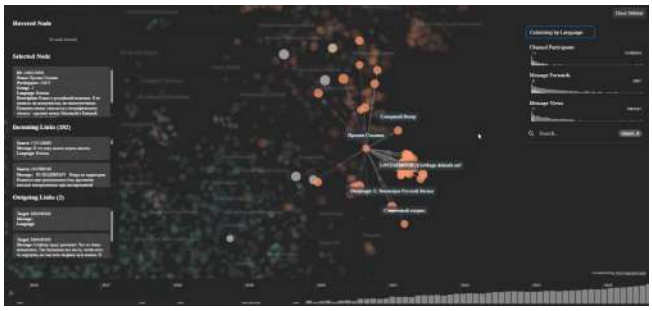

図 1　可視化システムの使用例

可視化システムは大きく分けて以下の 4 つの機能で構成される．

**ネットワーク表示 (図 1 の画面全体)**

データセットから構成されるネットワーク全体を表示する．コントロールパネルからハイライトしたエッジとノードを決定することで，特定のネットワークのみを表示することもできる．またノードをクリックすることで，そのノードに直接接続されるエッジとノードをハイライトすることができる．

**コントロールパネル (図 1 の画面右)**

上部のドロップダウンリストを操作することで，ノードに直接するエッジの数，チャンネルの言語，チャンネルのトピックをもとにノードの色付けを変更することができる．また中部のチャンネル参加者数，メッセージの引用回数，メッセージの閲覧回数のヒストグラムを操作することで，特定の条件に該当するノードとエッジをハイライトできる．ここで複数の条件を組み合わせることが可能である．例として，「引用回数 100 回以上かつ閲覧回数 50 回以下のエッジとそれに直接接続するノード」という条件でノードをハイライトすることもできる．さらに下部の検索欄から，チャンネル ID，チャンネル名，チャンネルの言語，チャンネルの説明文をもとにチャンネルを検索できる．

**選択したチャンネル，メッセージの情報 (図 1 の画面左)**

上から，ホバーしたノードのメタデータ，クリックしたノードのメタデータ，クリックしたノードが引用したエッジ，クリックしたノードから引用されたエッジが表示される．

**タイムライン (図 1 の画面下)**

メッセージが引用された日時をもとに構成されたヒストグラムを操作することで，ノードおよびエッジをハイライトできる．引用日時をもとにネットワークをアニメーションとして表示することも可能である．

## 4　実行結果と考察

本章では収集したデータセットの可視化結果について議論する．

### 4.1　開発システムを使用した可視化結果

代替医療，反ワクチン陰謀論で有名な Web サイト Natural News [16] の距離 1 のエゴネットワークをシステム上で可視化した結果を図 2 に示す．これにより，エゴネットワークの規模自体は小さいが，Dr Naomi Wolf [17]，Ron Johanson 上院議員など現実世界での有名人との繋がりが多いことを確認できた．

さらに，Natural News，Dr Naomi Wolf，Real Time Daily News などのチャンネルを起点としてエゴネットワークを段階的に追跡した結果，最終的に「/CIG/ Telegram | Counter Intelligence Global」と称する極右系チャンネルへと接続していることが確認された．/CIG/ Telegram | Counter Intelligence Global の距離 1 のネットワークをシステム上で可視化した結果を図 3 に示す．CIG は Natural News のように現実世界において高い知名度を有する媒体ではないものの，Telegram 上における参加者数や活動規模といった指標に基づけば，相対的に大きな影響力を持つチャンネルであることが確認された．また，Disclose.tv [18] に代表される誤情報拡散で知られる他のウェブサイト由来チャンネルとも接続しており，これらが相互に結び付くことで複雑なネットワーク構造を形成していることが明らかとなった．

### 4.2　Disparity Filter を利用した可視化結果

前項で発見されたウェルネス・代替医療関係のチャンネルと QAnon 関連チャンネル，ネットミームに関連するチャンネルを含むサブネットワークに，有向グラフに対応した Disparity Filter を適用し，バックボーンネットワークを抽出した．メッセージの引用回数の合計数を重みとしバックボーンネットワークを抽出した結果を図 4 に，平均数を重みとしバックボーンネットワークを抽出した結果を図 5 に，2 つの抽出結果を重ね合わせたものを図 4 に示す．エッジの色と意味の対応は表 5 の通りである．

図 6(左下) の最も大きいネットワークはニュース関連のチャ

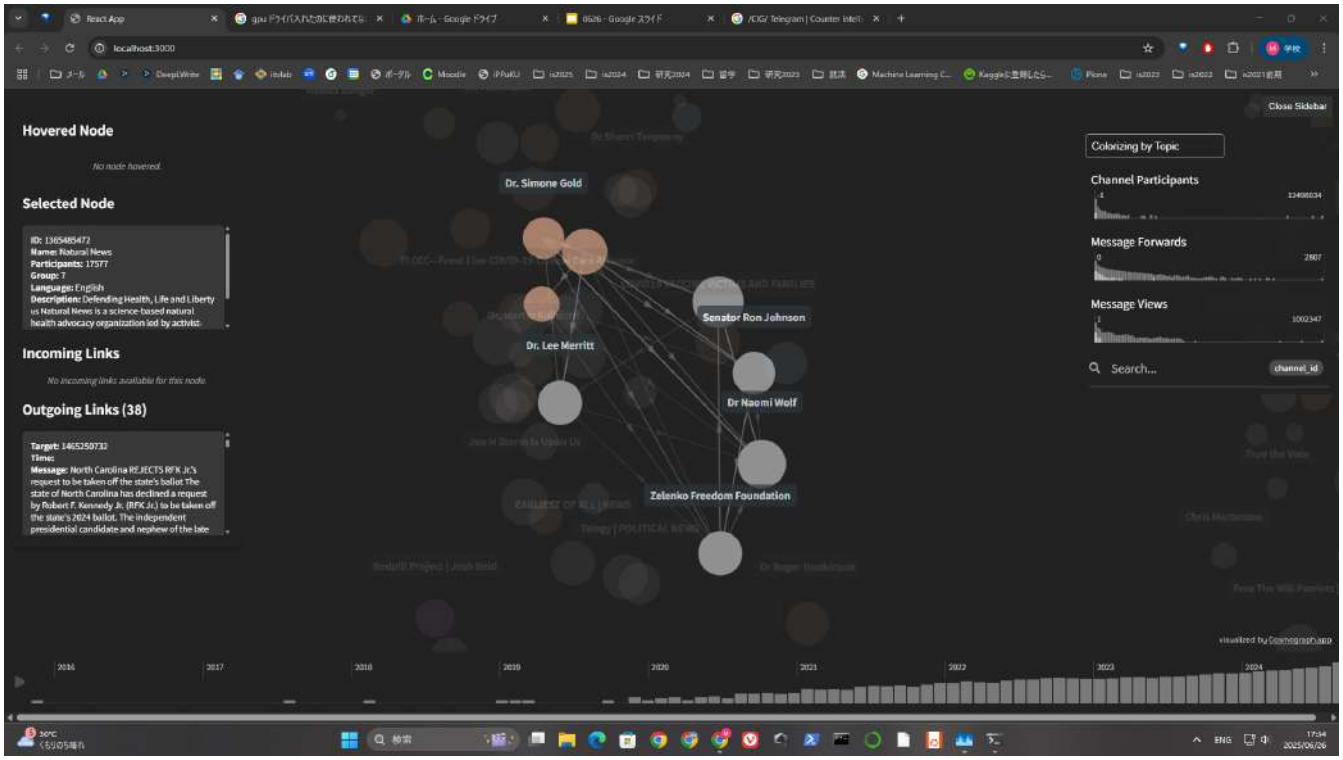


図 2　トピック別で色付けした Natural News のエゴネットワーク

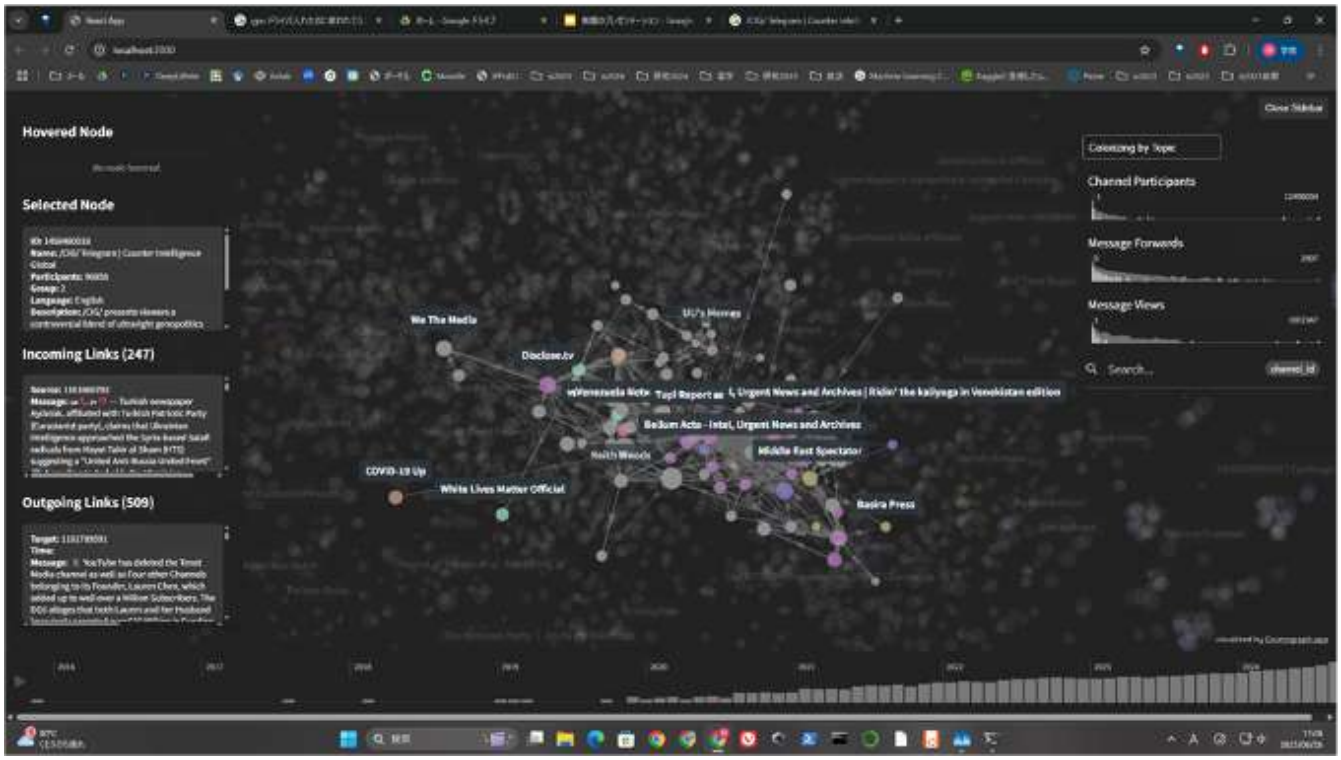


図 3　トピック別で色付けした/CIG/ Telegram | Counter Intelligence Global のエゴネットワーク

図 4　メッセージの引用回数の総和を重みとした場合のバックボーンネットワークの可視化例

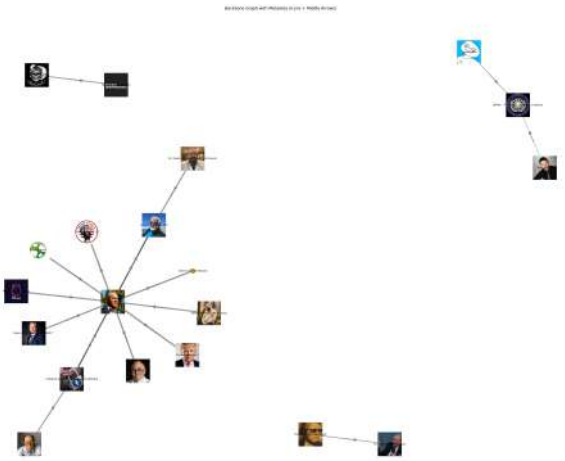

図 5　メッセージの引用回数の平均を重みとした場合のバックボーンネットワークの可視化例

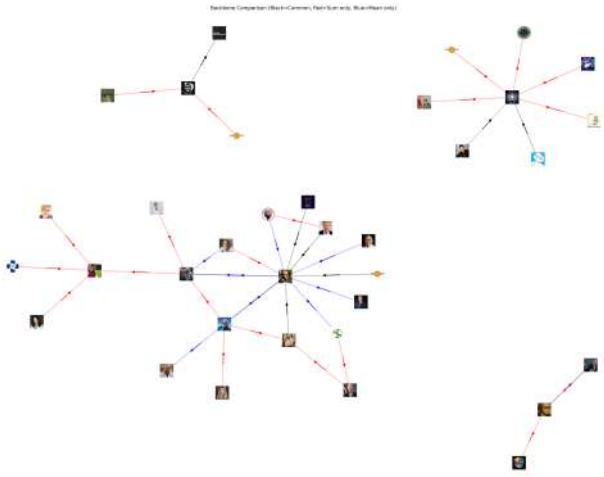

図 6　2 つのバックボーンネットワークを重ね合わせた場合の可視化例

表 5　バックボーン抽出条件とエッジの意味

| 色 | 条件 | エッジの意味 |
|---|---|---|
| 黒 | 両方のバックボーンに残る | 大規模ハブと局所的な結合の双方において有意と判定されたエッジであり，ネットワーク全体における中核的な引用関係を示す． |
| 赤 | 総和重みのみで残る | 引用総数が多いチャンネル間に形成されるハブ中心の強いエッジであり，ネットワーク全体における影響力の伝播経路を表す． |
| 青 | 平均重みのみで残る | 小規模チャンネル間で局所的に強い引用関係を持つエッジであり，コミュニティ内部の密な相互作用の兆候を示す． |

ンネル，ドナルド・トランプに関連するチャンネル，上院議員のチャンネル，代替医療に関連するチャンネルから構成される．その一方でその他の小さなネットワークはネットミーム関連

チャンネル，極右系チャンネルが主である．これにより，ウェルネス・代替医療領域と QAnon 関連チャンネルが，ニュース系・政治系チャンネルを介して強く結びついている一方で，ネットミームや極右系チャンネルは比較的独立した小規模なクラスターを形成していることを確認できた．

## 5 まとめ・今後の課題

本報告では，オンライン発の過激派運動である QAnon を対象に，Telegram 上のチャンネル間の引用関係をネットワークとして可視化し，その構造と情報拡散の特徴を明らかにした．スノーボールサンプリングにより QAnon 関連および非関連チャンネルを含むデータセットを構築し，チャンネルをノード，引用関係をエッジとする可視化システムを開発した．さらに，可視化システムで発見された興味深いエゴネットワークに対し，有向グラフ対応の Disparity Filter を用いてバックボーンネットワークを抽出した結果，ウェルネス・代替医療領域と QAnon はニュース・政治系チャンネルを介して強く結合している一方で，ネットミームや極右系は比較的独立したクラスターを形成していることを確認できた．以上により，QAnon 運動は話題領域ごとに異なる接続様式を持つ複合的なネットワーク構造として維持・拡散されていると結論づけられる．

本研究で開発した可視化システムは探索的分析に有効である一方，分析結果の定量的比較や再現性の確保という点では改善の余地がある．ネットワーク指標の統計的検定によって本研究の枠組みをより一般化することが今後の重要な課題である．

## 文　献


[1] Susan Svrluga and Faiz Siddiqui. N.c. man told police he went to d.c. pizzeria with gun to investigate conspiracy theory. https://www.washingtonpost.com/news/local/wp/2016/12/04/d-c-police-respond-to-report-of-a-man-with-a-gun-at-comet-ping-pong-restaurant/, 2016. 2025 年 12 月 13 日に参照.

[2] Nathan Katz. Do-it-yourself white supremacy: Linking together punk rock and white power. *Poetics*, Vol. 82, p. 101476, 2020.

[3] Consumption, wellness, and the far right. Vol. 25, .

[4] Stephanie Alice Baker. Alt. health influencers: how wellness culture and web culture have been weaponised to promote conspiracy theories and far-right extremism during the covid-19 pandemic. *European Journal of Cultural Studies*, Vol. 25, No. 1, pp. 3–24, 2022.

[5] Christopher T. Conner. Qanon, authoritarianism, and conspiracy within american alternative spiritual spaces. *Frontiers in Sociology*, Vol. Volume 8 - 2023, , 2023.

[6] Pepe the frog meme branded a 'hate symbol'. https://www.bbc.com/news/world-us-canada-37493165, 2016. 2024 年 12 月 29 日に参照.

[7] Laura Glitsos and James Hall. The pepe the frog meme: an examination of social, political, and cultural implications through the tradition of the darwinian absurd. *Journal for Cultural Research*, Vol. 23, No. 4, pp. 381–395, 2019.

[8] Stephanie Alice Baker and Michael Walsh. How memes transformed from pics of cute cats to health disinformation super-spreaders, February 2024.

[9] Mohamad Hoseini, Philipe Melo, Fabricio Benevenuto, Anja Feldmann, and Savvas Zannettou. On the globalization of the qanon conspiracy theory through telegram, 2021.

[10] Mathias Angermaier, Elisabeth Hoeldrich, Jana Lasser, and Joao Pinheiro Neto. The schwurbelarchiv: a german language telegram dataset for the study of conspiracy theories, 2025.

[11] W.F. Thomas. German QAnon Telegram Dataset. 11 2021.

[12] Maarten Grootendorst. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*, 2022.

[13] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 11 2019.

[14] Karishma Sharma, Emilio Ferrara, and Yan Liu. Characterizing online engagement with disinformation and conspiracies in the 2020 us presidential election. In *Proceedings of the international AAAI conference on web and social media*, Vol. 16, pp. 908–919, 2022.

[15] Nikita Rokotyan, Olga Stukova, and Denis Ovsyannikov. Cosmograph: GPU-accelerated Force Graph Layout and Rendering, December 2024.

[16] Daniel Funke. Health misinformation website rebrands as pro-trump outlet to get around ban from facebook. hhttps://www.politifact.com/article/2020/jun/05/health-misinformation-website-rebrands-pro-trump-o/, 2020.

[17] David Connet. Naomi wolf banned from twitter for spreading vaccine myths. https://www.theguardian.com/books/2021/jun/05/naomi-wolf-banned-twitter-spreading-vaccine-myths, 2021.

[18] W.F Thomas. Disclose.tv: Conspiracy forum turned disinformation factory. https://web.archive.org/web/20220112153617/https://www.logically.ai/articles/disclose.tv-conspiracy-forum-turned-disinformation-factory, 2022. 2024 年 1 月 7 日に参照.

# 認知科学に基づくユーザーの偶然性希求行動予測モデルの構築

鷲見優一郎† 中西 亮輔† 光田 英司† 二宮 由樹†† 曽根悠太郎††
三輪 和久††

† トヨタ自動車株式会社 未来創生センター 〒 471–8572 愛知県豊田市トヨタ町 1 番地
†† 名古屋大学大学院 情報学研究科 〒 464–8603 愛知県名古屋市千種区不老町
E-mail: † {yuichiro_sumi,ryosuke_nakanishi,eiji_mitsuda}@mail.toyota.co.jp,
††{ninomiya.yuki.t1,miwa.kazuhisa.m6}@f.mail.nagoya-u.ac.jp, sone.yutaro.n2@s.mail.nagoya-u.ac.jp

**あらまし** 近年，推薦システム分野において，セレンディピティが注目されている．認知科学の先行研究では，ユーザーの「偶然性希求行動」（偶然が起こりやすい場を作ろうとする行動）に注目し，その時々の偶然性希求行動に応じて，推薦システムが提供する偶然の度合いを調整する重要性が指摘されている．本研究では，一般的な推薦システムでの実現を目指し，偶然性希求行動を検索クエリから予測するモデルを提案する．その際，偶然性希求行動に影響を与える因子も検索クエリから予測し，中間変数として導入することでモデルの解釈性を高めた．段階的な予測は一般に予測性能の低下を招くが，本モデルは検索クエリから直接予測するモデルよりも高い予測性能を示した．その要因についても考察する．

**キーワード** 推薦システム，セレンディピティ，ユーザ理解

## 1 はじめに

近年，推薦システムにおいて，類似したアイテムが繰り返し推薦されてしまう「過剰最適化 [1]」が問題視されている．過剰最適化は，ユーザー満足度の低下につながり得るだけでなく [2]，ユーザーが自身とは異なる価値観に触れる機会を失う「フィルターバブル [3]」状態を招く可能性が指摘されている．

これらの問題を解決するため，近年セレンディピティが注目されている [4]．セレンディピティのある推薦は，ユーザー満足度や購買意向と相関があることが示されている [5]．従来研究の多くは，セレンディピティを推薦アイテムが持つ属性として捉え，条件として新規性，関連性，意外性を定義し，ユーザーの行動履歴やフィードバックからそれらを抽出・制御するアルゴリズムの設計に注力してきた [2]．

一方，記事推薦システムにおける大規模フィールド調査 [6] では，こうしたアイテム属性に基づく定義だけでは，多くのセレンディピティ事例の取りこぼしがあると報告されている．このことは，セレンディピティのある推薦システムの設計において，アイテム属性や行動履歴などの客観指標のみを重視する推薦手法には限界があり，ユーザーの心理特性や心理状態などに関連する主観指標を考慮する必要性を示唆している．

認知科学分野において，オンラインショッピングの利用場面を想定し，ユーザーが偶然性をどのように捉え，どの程度求めるのかに関する心理学実験が報告されている．従来研究 [7], [8] では，ユーザーの「偶然性希求行動」―ユーザーが偶然性を期待して行う（方略的）行動― に着目し，それに影響を与える心理的および状況要因が特定されている．さらに，その時々の偶然性希求行動に応じて，推薦システムが提供する偶然の度合いを調整することの重要性も指摘されている．

しかし，ユーザーの偶然性希求行動とその影響因子の取得は，現状ではアンケート以外の方法では困難である．ユーザーエクスペリエンス（UX）を損なわないためには，推薦システムの一般的な利用過程で自然に得られるデータから予測する方法が求められる．また，機械学習モデル等による予測結果に基づいてユーザーへの情報提供を行う場合，モデルの解釈性が高いと有用である．解釈性の高いモデルにより，サービス提供者がなぜそのような予測結果となったのかを理解し，提供情報の内容の検討が可能となることが期待される．

以上より，本研究では偶然性希求行動を検索クエリから予測するモデルを提案する．提案手法では，偶然性希求行動に影響を与える因子も検索クエリから予測し，中間変数として導入することでモデルの解釈性を高める．このように解釈性向上を目的とした段階的予測を行う偶然性希求行動予測モデルと，検索クエリから直接予測するモデルの予測性能を比較し，有効性を検証する．

## 2 関連研究

### 2.1 「Beyond-Accuracy」の推薦システム

推薦システムは，オンラインショッピングサイトやニュースサイトなどで広く利用されている．これまで，推薦システムは，Recall@$k$ のような，推薦リスト上位 $k$ 件のうちに含まれる正解数に基づいて算出される指標などを用いて，ユーザーの行動履歴をできるだけ再現することを主眼に設計・評価されてきた．その結果，行動履歴に関連するアイテムを過度に推薦し，「過剰最適化」を招き得ることが指摘されている [1], [9]．この課題を解決するため，推薦リストの予測性能だけでなく，推薦の多様性や公平性などの指標を加えてシステムを評価する枠組みである「Beyond-Accuracy」が注目されている．多様性のある推薦

は，UX を豊かにし，ユーザーの視野を広げるのに寄与するとされている [10]．さらに，公平性のある推薦では，高齢者など特定の属性を有するユーザー群において性能が低下する，あるいは人気の低いアイテムが推薦されにくいといった偏りを防ぎ，より公正なユーザー体験の実現に寄与するとされている [11]．セレンディピティは，Beyond-Accuracy 指標の 1 つとして提案されており [4]，ユーザーが自力では発見しにくかった，興味対象を見つける手助けになると主張されている [12]．

推薦システム分野では，セレンディピティを推薦アイテムが備える属性として捉え，新規性，関連性，意外性のいずれか 1 つ，またはそれらの組み合わせとする定義が広く用いられてきた [2], [13]．新規性は，一般にユーザーに依存しない指標とされ，推薦時点で広く未知である可能性や想定外の程度を表し，評価数などから推定される人気度の低さに基づき算出される [14]．関連性は，ユーザーの興味・関心への適合度を表し，嗜好の予測値としてスコア化されることが多い [15]．意外性は，ユーザーが普段消費するアイテムとの非類似性や，予期していなかったアイテムを消費した際の満足度など，複数の観点から定義される．例えば，過去履歴に含まれるアイテムおよびそれらの類似アイテムから成る集合と，推薦候補アイテムとの距離に基づき算出する定義がある [16]．

しかし，これらの指標のうちどれがセレンディピティに不可欠であるか，また各指標をどのように測定すべきかについては，理論や手法が未だ確立されていない [6]．加えて，記事推薦を対象とした大規模フィールド調査では，これらの指標と，ユーザーが感じるセレンディピティとの間に乖離が存在することを報告している [6]．このことは，アイテム属性に加え，ユーザーの状況や主観に関わる要素を考慮した指標が必要であることを示唆する．

### 2.2 セレンディピティにおけるユーザーの心理的要因

認知科学を用いた研究 [7] では，オンラインショッピングの場面でユーザーがどの程度偶然性を求めるか，および，その傾向に影響を与える心理的要因に着目している．どの程度偶然性を求めて行動するかは「ユーザーが偶然の出会いを期待して行う（方略的）行動」を意味する，偶然性希求行動で表される [7]．この指標が高いユーザーは，セレンディピティを強く求めていると考えられる．これまで偶然性希求行動に影響を与える心理的要因と製品属性として，それぞれ目標具体性と快楽次元が報告されている [8]．目標具体性は，ユーザーが検索時に「欲しい商品」をどの程度具体的に想起しているかを示す指標である．例えば車の購入を目的とした検索において，特定の車種名まで具体的に想起している場合は目標具体性が高く，コンパクトカーなどの大まかなカテゴリのみを想起している場合は目標具体性が低い．快楽次元とは，製品（またはブランド）に対する消費者態度を構成する次元の一つであり，使用経験に伴う感覚的・情緒的反応に基づく快楽的な評価を表す概念である．測定には快楽尺度を用いる [17]．

実験の結果，目標具体性が高いほど偶然性希求行動が低くなり，探索対象の快楽次元が高いほど偶然性希求行動が高くなる傾向が報告されている．一方で，ユーザーの目標具体性が低いほど，快楽次元が偶然性希求行動に与える影響がより大きくなることも確認されており，これらの変数は偶然性希求行動と密接に関連していることが報告されている．

## 3 提 案 手 法

提案モデルの全体像を図 1 に示す．本研究では，実サービスでの利用を想定し，入力は検索クエリのみとする．提案モデルでは，モデルの解釈性を向上させるため，中間変数として，文献 [7], [8] により偶然性希求行動に影響することが確認されている，目標具体性と快楽次元（ 2.2 節参照），および，目標具体性の予測に有効と考えられる検索目標と検索クエリの間の関係性を表す「上位下位関係」および「属性関係」（ 4.1 節参照）を導入する．

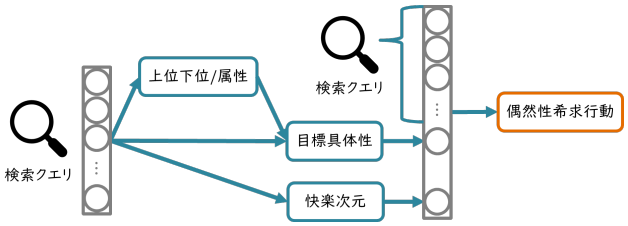


図 1 提案モデルの全体像

### 3.1 中間変数予測

目標具体性の予測では，入力特徴量として検索クエリ，上位下位関係ラベル，属性関係ラベルを用いる．このとき，検索クエリのベクトルと各関係ラベルの予測値を連結して 1 つのベクトルを作成し，予測に用いる．具体的には，検索クエリは事前学習済み言語モデルでベクトル化し，上位下位関係ラベルと属性関係ラベルは，その検索クエリのベクトルのみで学習されたモデルの予測値を利用する．

快楽次元の予測に用いる入力特徴量は，関係ラベルの予測と同様に，検索クエリのベクトルのみである．全ての予測モデルは教師あり学習手法を用いて学習させる．

### 3.2 偶然性希求行動予測

偶然性希求行動の予測モデルは，図 1 で示す通り，検索クエリの埋め込みベクトル，目標具体性の予測値，快楽次元の予測値を入力特徴量として，教師あり学習手法を用いて学習させる．入力特徴量は，検索クエリのベクトルに対して，目標具体性と快楽次元の予測値を連結して作成する．

## 4 実　　験

### 4.1 データセット

本実験では，クラウドソーシング [8] により収集された 2,200 件の検索クエリデータを用いた（参加者は成人 200 名．平均年齢 42.1 歳，SD = 8.86，女性 93 名，男性 107 名）．このうち，検索クエリとして成立していない 4 件（入力なし，または「なし」と入力）を除外し，2,196 件を提案手法の学習・検証・

テストに用いた．各検索クエリには，15 種の関連情報が含まれており，本研究では以下に示すような検索クエリと 3 種類の関連情報を利用した．

**検索クエリ**：参加者に「あなたが今実際にオンラインショッピングを利用して購入したいと思っている製品を思い浮かべてください.」と質問し，想起した製品を検索するために入力するクエリを記述させた．

**快楽次元**：快楽尺度 [17] [18] を用いて，参加者が想起した製品に対する態度（快楽次元）を測定した．快楽尺度は意味的差異尺度であり，5 組の形容詞対を 7 件法（例：「1：非常に面白くない」–「7：非常に面白い」）で評価させた．

**目標具体性**：参加者が想起した製品（質問文中では“X”として提示）について,「あなたが 1 つ目の製品 "X" を思い浮かべたとき，買いたいものは具体的に決まっていましたか？」と質問し，5 件法（「1：全く決まっていなかった」–「5：非常に決まっていた」）で回答を得た．

**偶然性希求行動**：参加者に「これから，先ほどあなたが入力した検索語による検索結果に，どの程度「偶然性」を取り入れるかを設定していただきます．スライダーを用いて 0 から 100 まで，段階的に偶然性が高くなっていきます，あなたはどの程度，検索結果に偶然性を持たせたいですか？」と質問し，回答させた．

さらに本研究では，参加者から直接収集したデータに対して，訓練された第三者がアノテーションした上位下位関係ラベルおよび属性関係ラベルも用いた．各ラベルの定義は以下の通りである [19].

- 上位下位関係ラベル： 検索目標と検索クエリの関係性が概念的に上下関係であることを示す．具体的には,「単語 A が単語 B の上位語（Hypernym）である」は，"B is a (kind of) A" が成立することに等しい.「単語 B が単語 A の下位語（Hyponym）である」を考える場合も同様である．なお,「スポーツカー」などの複合名詞の場合も同様に，上位語もしくは下位語に相当すれば，上位下位関係とみなした．
- 属性関係ラベル： 検索クエリが検索目標の属性（性質，特徴，構成要素など）に相当することを示す．例えば，検索目標が「車」，検索クエリが「車 かっこいい」である場合，「かっこいい」という単語が属性関係に該当する．

### 4.2 前 処 理

**検索クエリ**：日本語事前学習済みの BERT [20](tohoku-nlp/bert-base-japanese) [21] および RoBERTa [22](nlp-waseda/roberta-base-japanese) [23] を用いて，クエリの埋め込みベクトルを取得した．具体的には，言語モデルによりクエリ中の各単語をそれぞれベクトル化し，pooling 手法を利用することで，検索クエリ全体を表すベクトルを算出した．pooling 手法には，[24] で高い性能が報告されている，CLS pooling と Mean pooling を採用した．各 pooling 手法の詳細は次の通り [25], [26].

CLS　BERT の事前学習時に，次文予測に用いられる，[CLS] トークンの埋め込みを検索クエリ全体の埋め込みとして用いる．RoBERTa の場合，[CLS] トークンが存在しないため，代替として，文頭トークンである <s> トークンの埋め込みを用いる．

Mean　クエリ中の各単語の埋め込みベクトルを要素ごとに平均することにより，検索クエリ全体の埋め込みを算出する．

本研究で扱うサンプルサイズは，2,196 件であり，クラス数に依存して 1 クラスあたりのサンプル数が小さくなり，学習が困難になる．そこで，快楽次元，目標具体性，偶然性希求行動は，学習時にクラス間のサンプル数が均衡するように二値化を行い，1 クラスあたりのサンプル数を確保した．詳細は以下の通りである．

**快楽次元・目標具体性**：快楽次元と目標具体性は，それぞれ 7 件法と 5 件法で参加者の回答を取得している．二値化するためにそれぞれの頻度分布を確認したところ，全 2,196 件のサンプルにおいて，快楽次元では，4 以下の値が 48.7%（1,070 件）と 4 より大きい値が 51.3%（1,126 件）であり，目標具体性では，2 以下の値が 48.4%（1,062 件）と 2 より大きい値が 51.6%（1,134 件）である．これらの統計量に基づいて，本研究では，快楽次元の 4 以下の値を「低」，4 より大きい値を「高」とし，また目標具体性の 2 以下の値を「低」，2 より大きい値を「高」と定義した．

**偶然性希求行動**：0 から 100 の整数値で取得された回答を，回答値が 55 以下を「低」，56 以上を「高」として二値化した．各クラスの割合は,「低」が 50.6%（1,111 件）と「高」が 49.4%（1,085 件）である．

本データ取得は，名古屋大学倫理審査委員会およびトヨタ自動車株式会社の倫理審査（2024TMC246）において厳正な審査を受け，倫理的配慮が適切に確保されたうえで実施した．

### 4.3 実 験 設 定

本研究では， 3 章で述べた提案手法の有効性を検証するため，ベースライン手法と比較する．中間変数の導入による段階的な予測に対して，ベースライン手法では，検索クエリから直接，偶然性希求行動を予測した．

提案手法の学習は， 3 章で述べた通りに実行した．それぞれの予測モデルには，全て LightGBM [27] を用いて，二値分類問題として学習した．LightGBM は，表形式データを入力とする二値分類タスクで広く用いられ，高い予測性能と計算効率が報告されているため，採用した．中間変数を用いて学習する予測モデルでは，陽性ラベルの予測確率を用いており，入力特徴量の該当要素は 0 から 1 の間の実数値である．また，モデルのハイパーパラメータは Optuna [28] によりチューニングした．二値分類問題であることを考慮して，性能指標には，二値交差エントロピーを使用した．

ハイパーパラメータチューニングとモデルの性能評価は，二重交差検証 [29], [30] により実施した. 本検証法は，汎化性能とモデル選択の評価を厳密に行うため，外側と内側の検証ループを組合せ，二重に交差検証を行う手法である．本研究では，外側・内側ともに $K = 5$ の Stratified K-Fold Cross Validation を用い，クラス比率を維持した分割を行った．具体的な学習・

評価フローを次に示す.

1. 全データを外側の交差検証用に 5 個のセットに分割した. 各セットは，1 つを外側テスト用セット，残りを外側学習用セットとした. 外側テスト用セットは最終的な性能評価のみに用い，モデル選択には用いない
2. 外側学習用セットをさらに 5 個のセットに分割し，内側検証用セットと，内側学習用セットを作成した. 内側ループでは，候補となるハイパーパラメータ群に対して性能指標（二値交差エントロピー）を算出し，Optuna により最適なハイパーパラメータを選択した
3. 内側ループで選択されたハイパーパラメータを用いて，外側学習セット全体でモデルを再学習した
4. 外側テストセットに対して予測を行い，各評価指標のスコアを算出した. これを外側ループの各分割で繰り返すことで，モデルの予測性能を評価した.

### 4.4 実 験 結 果

モデルの性能は，Accuracy，Precision，Recall，F1-score を用いて評価し，それぞれについて 5-Fold の平均値と標準偏差を算出した. 結果を表 1 に示す. 表中の括弧内は標準偏差，太字は同一言語モデルおよび Pooling 手法内でベースラインと提案手法を比較した際の，各指標の最良値を示す.

まず，表 1 の太字で示すように，各条件に対してベースライン手法と提案手法を比較すると，提案手法が優れていることがわかる. 具体的には，4 つの評価指標，2 つの言語モデル，2 つの Pooling 手法の組合せによる 16（$4 \times 2 \times 2$）通りの比較のうち，14 通りで提案手法が優位な結果である. また，全ての評価指標において，最良値を示したのは，CLS pooling と RoBERTa モデルを用いた提案手法（Accuracy=0.598，Precision=0.593，Recall=0.594，F1-Score=0.593）であった. このことから，中間変数の予測値を導入した段階的予測が，予測性能の向上に寄与することが確認された.

## 5 考　　察

### 5.1 モデル構造

提案手法はベースライン手法と比較して性能が向上しており，この結果は，本研究で導入した中間変数が偶然性希求行動の予測に有用であることを示している. 段階的なモデル構造の有用性は，複雑な推論を中間ステップへ分解することで推論を助けるという観点で，大規模言語モデルにおけるプロンプトエンジニアリング手法の 1 つである Chain-of-Thought [31] 研究の主張と類似している. 実際，本研究でも事前学習済みの大規模言語モデルを用いており，段階的な推論フローが有効に働いた可能性がある.

一方で，提案手法では中間変数に真値ではなく予測値を用いた. そのため，中間変数の予測性能が後段の予測性能を制約し得る. 提案手法の限界性能を確認するため，中間変数に真値を与える条件に変更し，同様に実験した. 実験結果を表 2 に示す. 真値を用いた場合は予測値を用いた場合に比べて Accuracy，Precision，F1-Score では約 1%，Recall では 2% の改善にとどまった. つまり，中間変数が予測値であっても提案手法は限界性能に近い水準の性能を示していると解釈できる.

次に，その理由を検討するため，中間変数そのものの予測性能を確認する. 提案手法において最良の結果を示した条件（CLS pooling を用いた RoBERTa）における Accuracy を表 3 に示す. 本実験では，クラス分布の偏りが小さいため，チャンスレベル（50%）との比較が直観的である Accuracy を用いた. 目標具体性および快楽次元はいずれも 65% を下回り，二値分類問題として十分な予測性能とは言い難い. にもかかわらず，偶然性希求行動予測における，中間変数に予測値を用いた場合と真値を用いた場合との性能差は 1.1% 程度にとどまる. このことは，提案手法の限界性能に近い性能は，中間変数の予測性能そのものよりも，段階的予測を行うモデル構造に起因している可能性を示唆している.

### 5.2 言語モデルと pooling 手法

表 1 より，提案手法における F1-Score の最良値は，Mean pooling を用いた BERT で 0.577，CLS pooling を用いた RoBERTa で 0.593 であり，RoBERTa を用いた条件の方が高い性能を示した. このことは，RoBERTa が BERT と同一構造でありながらも，Next Sentence Prediction の除去や動的マスキング等で事前学習レシピを最適化している点と整合的である. すなわち，検索クエリのような短い入力に対してもロバストな表現を得やすい性質が，提案手法の性能に有利に働いた可能性がある.

次に Pooling 手法の影響を確認する. 表 1 より，Pooling 手法以外の条件を同一にして比較したとき（例えば，BERT を用いたベースライン），一貫性を確認できなかった（RoBERTa を用いた提案手法では CLS が良く，それ以外の条件では Mean の方が優れていた）. Mean pooling は，ベクトルの単純な要素平均による情報圧縮手法であるため，検索クエリの長さに依存せず安定した情報の集約が可能であるが，重要語も非重要語も等しく集約する. 一方で，CLS pooling は，次文予測のためにトークンがもつ意味単位で情報を圧縮する手法であり，Mean pooling と比較して，短文における重要語を強調して集約する. 検索クエリのような短文において，CLS pooling が優位でなかったのは，検索クエリの質が低かったことが考えられる. 例えば，商品名が「ベビー綿棒」で，検索クエリが「抗菌」のみ，となっているサンプルがあり，ユーザーの意図を正しく反映する検索クエリを取得できていないことが示唆される.

### 5.3 解　釈　性

本研究では，検索クエリから偶然性希求行動を予測するにあたり，解釈性の向上を目的として中間変数を導入したモデルを提案した. そこで，提案手法で学習したモデルが，実際にどの特徴量に基づいて予測しているのかを分析し，解釈性の観点から妥当性を検討する. 解釈性の検証には，Shapley Additive Explanations（SHAP）[32] による，特徴量が予測値へ与える寄与度（SHAP 値）を用いる. SHAP 値は，一般的に，絶対

表 1　偶然性希求行動の予測結果

| 言語モデル | Pooling | 手法 | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|---|---|
| BERT | CLS | ベースライン | 0.528 (0.025) | 0.523 (0.025) | 0.545 (0.008) | 0.533 (0.016) |
| | | 提案手法 | **0.536 (0.027)** | **0.530 (0.027)** | **0.547 (0.023)** | **0.538 (0.024)** |
| | Mean | ベースライン | 0.567 (0.024) | 0.562 (0.025) | 0.571 (0.031) | 0.566 (0.023) |
| | | 提案手法 | **0.572 (0.025)** | **0.564 (0.025)** | **0.592 (0.025)** | **0.577 (0.022)** |
| RoBERTa | CLS | ベースライン | 0.554 (0.013) | 0.547 (0.013) | 0.569 (0.008) | 0.558 (0.009) |
| | | 提案手法 | **0.598 (0.025)** | **0.593 (0.025)** | **0.594 (0.036)** | **0.593 (0.027)** |
| | Mean | ベースライン | 0.563 (0.032) | 0.554 (0.030) | **0.587 (0.035)** | **0.570 (0.032)** |
| | | 提案手法 | **0.566 (0.027)** | **0.560 (0.028)** | 0.575 (0.034) | 0.567 (0.026) |

表 2　提案手法における中間変数の条件を変えた際の偶然性希求行動の予測結果

| 言語モデル | Pooling | 中間変数 | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|---|---|
| BERT | CLS | 予測値 | 0.536 (0.027) | 0.530 (0.027) | 0.547 (0.023) | 0.538 (0.024) |
| | | 真値 | **0.593 (0.027)** | **0.585 (0.027)** | **0.609 (0.037)** | **0.596 (0.029)** |
| | Mean | 予測値 | 0.572 (0.025) | 0.564 (0.025) | 0.592 (0.025) | 0.577 (0.022) |
| | | 真値 | **0.602 (0.025)** | **0.595 (0.027)** | **0.614 (0.037)** | **0.604 (0.025)** |
| RoBERTa | CLS | 予測値 | 0.598 (0.025) | 0.593 (0.025) | 0.594 (0.036) | 0.593 (0.027) |
| | | 真値 | **0.609 (0.024)** | **0.604 (0.026)** | **0.610 (0.029)** | **0.607 (0.022)** |
| | Mean | 予測値 | 0.566 (0.027) | 0.560 (0.028) | 0.575 (0.034) | 0.567 (0.026) |
| | | 真値 | **0.595 (0.023)** | **0.590 (0.022)** | **0.592 (0.033)** | **0.591 (0.025)** |

表 3　RoBERTa かつ CLS pooling 条件における中間変数の Accuracy

| 目標具体性 | 快楽次元 |
|---|---|
| 0.636 (0.016) | 0.641 (0.017) |

値が大きいほど予測値への影響が大きいことを意味する．

全サンプルに対して特徴量（770 次元）ごとに SHAP 値を算出し，その絶対値を平均した値を算出した．図 2 に，その値が大きい上位 10 個の特徴量を降順に示す．ここで，数字で表記された特徴量は，検索クエリベクトルにおける該当次元番号である．

まず，中間変数のうち目標具体性（図中，purpose）は予測への影響が最も大きく，2 番目に大きい特徴量（図中，714；検索クエリベクトルの 714 番目の要素）と比較して，約 4 倍の影響度であった．一方で、もう 1 つの中間変数である快楽変数は，上位 10 個に含まれず，影響度は全体の 513 位であった．つまり，提案手法は，偶然性希求行動を主に目標具体性で説明するモデルと言える．この傾向は，[7] の結果とも整合する．

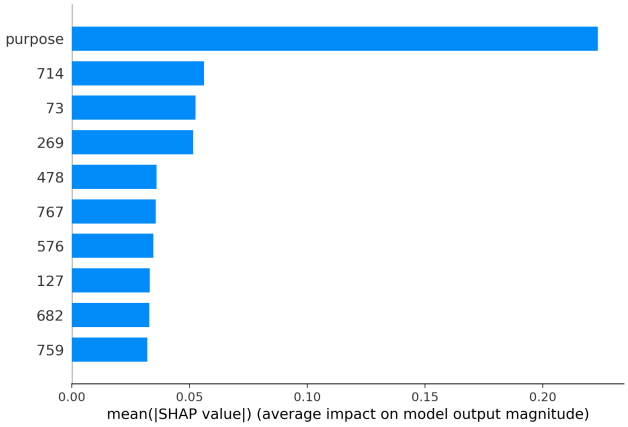


図 2　RoBERTa（CLS）を用いた提案手法における Shapley 値の Summary Plot（縦軸中の数字はベクトルの要素番号）

## 6　おわりに

本研究では，偶然性希求行動を検索クエリから予測する解釈性の高いモデルを提案した．提案モデルは，検索クエリのみを入力とし，目標具体性と快楽次元を中間変数とする，人の思考過程に基づくモデル構造である．検索クエリから直接予測を行うベースラインと比較して，多くの条件において評価指標の 4 項目すべてで同等または優れた性能を示した．この結果は，中間変数の導入がもたらした効果によるものであると考えられる．5.1 節の考察で示した通り，中間変数を組み込み，段階的に予測を行うモデル構造は，大規模言語モデル分野におけるプロンプトエンジニアリング手法と類似した推論フローを有しており，その構造自体が中間変数の予測性能よりも重要であることが確認された．

今後の展望は 2 つある．1 つは被験者実験の実施である．本研究では，オンラインショッピングを想定した仮想的な設定のもとでアンケートを通じて収集されたデータを使用した．今後は，実サービスへの応用を見据え，提案手法による予測結果に基づいて実験的に推薦システムを制御し，使用時における行動データやユーザー満足度により効果を検証する必要がある．

2 つ目は学習データの拡充と質向上である．今回使用した約 2,200 件のデータは，アンケートにより，検索クエリと同時に検索対象も取得したが，その際，両者の対応関係が不明瞭で（例えば，検索対象に「ベビー綿棒」と記し，検索クエリには「抗菌」のみを記すなど），実験意図の理解が十分でない可能性があるサンプルも含まれていた．今後はデータ量を増やすと共

に，不適切なサンプルを除去する前処理を行い，より信頼性の高いデータを用いて学習することで，予測性能のさらなる向上を図る．

## 文　献


[1] Panagiotis Adamopoulos and Alexander Tuzhilin. On over-specialization and concentration bias of recommendations. *Proceedings of the 8th ACM Conference on Recommender systems*, pp. 153–160, 10 2014.

[2] Denis Kotkov, Shuaiqiang Wang, and Jari Veijalainen. A survey of serendipity in recommender systems. *Knowledge-Based Systems*, Vol. 111, pp. 180–192, 2016.

[3] Tien T. Nguyen, Pik-Mai Hui, F. Maxwell Harper, Loren Terveen, and Joseph A. Konstan. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, p. 677–686, New York, NY, USA, 2014. Association for Computing Machinery.

[4] Mouzhi Ge, Carla Delgado-Battenfeld, and Dietmar Jannach. Beyond accuracy: evaluating recommender systems by coverage and serendipity. In *Proceedings of the Fourth ACM Conference on Recommender Systems*, RecSys '10, p. 257–260, New York, NY, USA, 2010. Association for Computing Machinery.

[5] Li Chen, Yonghua Yang, Ningxia Wang, Keping Yang, and Quan Yuan. How serendipity improves user satisfaction with recommendations? a large-scale user evaluation. In *The World Wide Web Conference*, WWW '19, p. 240–250, New York, NY, USA, 2019. Association for Computing Machinery.

[6] Denis Kotkov, Alan Medlar, Triin Kask, and Dorota Glowacka. The dark matter of serendipity in recommender systems. In *Proceedings of the 2024 Conference on Human Information Interaction and Retrieval*, CHIIR '24, p. 108–118, New York, NY, USA, 2024. Association for Computing Machinery.

[7] Yuki Ninomiya, Yutaro Sone, Kazuhisa Miwa, Yuichiro Sumi, Ryosuke Nakanishi, Eiji Mitsuda, Koji Sato, and Tadashi Odashima. Determinants of users' chance-seeking behavior in search-based recommendation. In *Proceedings of the Nineteenth ACM Conference on Recommender Systems*, RecSys '25, p. 564–569, New York, NY, USA, 2025. Association for Computing Machinery.

[8] 曽根悠太郎, 二宮由樹, 三輪和久, 鷲見優一郎, 中西亮輔, 光田英司, 佐藤浩司, 小田島正. 製品の快楽・功利次元と偶然性希求行動の関連. 2025 年度日本認知科学会第 42 回大会, 2025.

[9] Zeinab Abbassi, Sihem Amer-Yahia, Laks V.S. Lakshmanan, Sergei Vassilvitskii, and Cong Yu. Getting recommender systems to think outside the box. In *Proceedings of the Third ACM Conference on Recommender Systems*, RecSys '09, p. 285–288, New York, NY, USA, 2009. Association for Computing Machinery.

[10] Pablo Castells, Neil J. Hurley, and Saul Vargas. *Novelty and Diversity in Recommender Systems*, pp. 881–918. Springer US, Boston, MA, 2015.

[11] Yifan Wang, Weizhi Ma, Min Zhang, Yiqun Liu, and Shaoping Ma. A survey on the fairness of recommender systems. *ACM Trans. Inf. Syst.*, Vol. 41, No. 3, February 2023.

[12] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.*, Vol. 22, No. 1, p. 5–53, January 2004.

[13] Denis Kotkov, Joseph A. Konstan, Qian Zhao, and Jari Veijalainen. Investigating serendipity in recommender systems based on real user feedback. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, SAC '18, p. 1341–1350, New York, NY, USA, 2018. Association for Computing Machinery.

[14] Marius Kaminskas and Derek Bridge. Diversity, serendipity, novelty, and coverage: A survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Trans. Interact. Intell. Syst.*, Vol. 7, No. 1, December 2016.

[15] Michael D. Ekstrand, John T. Riedl, and Joseph A. Konstan. Collaborative filtering recommender systems. *Foundations and Trends® in Human–Computer Interaction*, Vol. 4, No. 2, pp. 81–173, 2011.

[16] Panagiotis Adamopoulos and Alexander Tuzhilin. On unexpectedness in recommender systems: Or how to better expect the unexpected. *ACM Trans. Intell. Syst. Technol.*, Vol. 5, No. 4, December 2014.

[17] Kevin E. Voss, Eric R. Spangenberg, and Bianca Grohmann. Measuring the hedonic and utilitarian dimensions of consumer attitude. *Journal of Marketing Research*, Vol. 40, No. 3, pp. 310–320, 2003.

[18] 岡田庄生, 西川英彦. 消費者の功利主義的・快楽主義的モノづくり動機と，製品成果・公開. マーケティングジャーナル, Vol. 39, No. 1, pp. 75–87, 2019.

[19] 中西亮輔, 鈴木結友, 鷲見優一郎, 光田英司, 二宮由樹, 曽根悠太郎, 三輪和久. プロンプト最適化を用いた検索クエリと検索目標の関係性アノテーション. 第 266 回 自然言語処理研究発表会, 2025.

[20] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[21] 東北大学自然言語処理研究グループ. tohoku-nlp/bert-base-japanese, 2020. `https://huggingface.co/tohoku-nlp/bert-base-japanese` [アクセス日： (2025 年 10 月 20 日)].

[22] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach, 2019.

[23] 早稲田大学河原研究室. nlp-waseda/roberta-base-japanese, 2022. `https://huggingface.co/nlp-waseda/roberta-base-japanese` [アクセス日： (2025 年 10 月 20 日)].

[24] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. *CoRR*, Vol. abs/1908.10084, , 2019.

[25] 塚越駿, 笹野遼平, 武田浩一. 定義文を用いた文埋め込み構成法. 言語処理学会 第 27 回年次大会 発表論文集, 2021.

[26] 原知正, 栗田宙人, 横井祥, 乾健太郎. 平均プーリングによる文埋め込みの再検討: 平均は点群の要約として十分か? 言語処理学会 第 30 回年次大会 発表論文集, 2024.

[27] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: a highly efficient gradient boosting decision tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, p. 3149–3157, Red Hook, NY, USA, 2017. Curran Associates Inc.

[28] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

[29] Sudhir Varma and Richard Simon. Bias in error estimation when using cross-validation for model selection. *BMC bioinformatics*, Vol. 7, No. 1, p. 91, 2006.

[30] Gavin C Cawley and Nicola LC Talbot. On over-fitting in model selection and subsequent selection bias in perfor-

mance evaluation. *The Journal of Machine Learning Research*, Vol. 11, pp. 2079–2107, 2010.

[31] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, Vol. 35, pp. 24824–24837, 2022.

[32] Scott M. Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, p. 4768–4777, Red Hook, NY, USA, 2017. Curran Associates Inc.

## 付　　録

### 1　中間変数予測結果

快楽次元の予測結果を表 A·1 に，上位下位および属性関係ラベルの予測結果を表 A·2 に，目標具体性の予測結果を表 A·3 示す.

表 A·1　快楽次元の予測結果

| 言語モデル | Pooling | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|---|
| BERT | CLS | 0.619 (0.011) | 0.634 (0.017) | 0.611 (0.031) | 0.622 (0.014) |
| | Mean | 0.637 (0.011) | 0.650(0.013) | **0.655 (0.032)** | **0.657 (0.018)** |
| RoBERTa | CLS | 0.641 (0.017) | 0.653 (0.019) | 0.641 (0.011) | 0.647 (0.014) |
| | Mean | **0.653 (0.032)** | **0.665 (0.029)** | 0.647 (0.043) | 0.656 (0.035) |

表 A·2　上位下位/属性関係ラベルの予測結果

| | 言語モデル | Pooling | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|---|---|
| 上位下位 | BERT | CLS | 0.634 (0.062) | 0.328 (0.043) | 0.414 (0.134) | 0.344 (0.059) |
| | | Mean | 0.638 (0.055) | 0.337 (0.042) | 0.444 (0.124) | 0.367 (0.051) |
| | RoBERTa | CLS | 0.623 (0.060) | 0.332 (0.045) | **0.475 (0.146)** | 0.371 (0.054) |
| | | Mean | **0.657 (0.056)** | **0.363 (0.058)** | 0.444 (0.140) | **0.375 (0.065)** |
| 属性 | BERT | CLS | 0.885 (0.011) | **0.970 (0.007)** | 0.899 (0.010) | 0.933 (0.007) |
| | | Mean | 0.895 (0.013) | 0.966 (0.005) | 0.915 (0.015) | 0.940 (0.008) |
| | RoBERTa | CLS | **0.920 (0.013)** | 0.969 (0.006) | **0.940 (0.014)** | **0.954 (0.008)** |
| | | Mean | 0.903 (0.006) | 0.966 (0.012) | 0.925 (0.007) | 0.945 (0.004) |

表 A·3　目標具体性の予測結果

| | 言語モデル | Pooling | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|---|---|
| 提案手法 | BERT | CLS | 0.619 (0.015) | 0.643 (0.019) | 0.592 (0.047) | 0.615 (0.026) |
| | | Mean | **0.637 (0.011)** | **0.657 (0.026)** | 0.628 (0.038) | 0.641 (0.011) |
| | RoBERTa | CLS | 0.636 (0.016) | 0.649 (0.020) | **0.647 (0.026)** | **0.647 (0.015)** |
| | | Mean | 0.617 (0.010) | 0.637 (0.010) | 0.599 (0.033) | 0.617 (0.018) |
| 中間変数が真値 | BERT | CLS | 0.613 (0.017) | 0.635 (0.018) | 0.591 (0.044) | 0.611 (0.026) |
| | | Mean | **0.643 (0.018)** | **0.659 (0.025)** | **0.642 (0.011)** | **0.650 (0.011)** |
| | RoBERTa | CLS | 0.622 (0.012) | 0.644 (0.018) | 0.601 (0.024) | 0.622 (0.011) |
| | | Mean | 0.625 (0.012) | 0.644 (0.015) | 0.615 (0.035) | 0.628 (0.018) |
| ベースライン | BERT | CLS | 0.616 (0.019) | 0.638 (0.020) | 0.594 (0.039) | 0.615 (0.025) |
| | | Mean | **0.638 (0.022)** | **0.661 (0.025)** | 0.617 (0.030) | 0.638 (0.022) |
| | RoBERTa | CLS | 0.624 (0.024) | 0.643 (0.027) | 0.615 (0.023) | 0.628 (0.022) |
| | | Mean | 0.637 (0.017) | 0.656 (0.018) | **0.624 (0.029)** | **0.640 (0.019)** |

# クエリ形式とランキング手法が検索結果のスタンス分布に与える影響の分析

池元　太陽†　　山本　岳洋††

† 兵庫県立大学 社会情報科学部　〒 651–2197 兵庫県神戸市西区学園西町 8–2–1
†† 兵庫県立大学 大学院情報科学研究科　〒 651–2197 兵庫県神戸市西区学園西町 8–2–1
E-mail: †fa22t006@guh.u-hyogo.ac.jp, ††t.yamamoto@sis.u-hyogo.ac.jp

**あらまし**　本研究では，賛否両論あるトピックに関するウェブ検索において，クエリ形式の違いとランキング手法が検索結果の偏りに与える影響を分析した．具体的には，「学校に制服は必要か」といった多様な意見が存在するトピックを使用し，質問クエリとキーワードクエリに分けられるクエリ形式と，特性の異なる 3 つのランキング手法（BM25，e5，PRP）が，検索結果の偏りを示すスタンス分布にどのような影響を及ぼすかについて分析した．分析の結果，クエリ形式の違いによって検索結果におけるスタンス分布が変化することを示したが，その傾向はランキング手法によって異なることが確認された．また，検索有効性が高いランキング手法は，クエリ形式の違いによる影響を受けにくい一方で，検索結果の偏りを強める可能性があることを示した．本研究で得られた知見は，検索システムの評価において，検索有効性のみならず，クエリの表現の違いやランキング手法の選択によって生じる検索結果の偏りを考慮することの重要性を提示するものである．

**キーワード**　情報検索，スタンス分布，検索モデル，クエリ形式

## 1　はじめに

賛否両論あるトピックにおいて，ウェブ検索は，ユーザが多角的な視点から情報を収集し，責任ある意見形成を行うための重要な手段として機能している．例えば，「学校に制服は必要か」，「死刑制度は廃止すべきか」といったトピックには，1 つの明確な正解が存在せず，多様な価値観や意見が対立し，継続的に議論されている．このようなトピックに対して，ユーザはウェブ検索を通じて，様々な主張や根拠に触れ，それらを批判的に評価することで自身の意見を形成していくことが望まれる [19].

一方で，検索結果の偏りがユーザの意見形成や態度変容に強い影響を与えることが，これまでの研究により指摘されている [8,9,17]．検索結果の上位に提示された文書の賛否のスタンスに偏りがある場合，それを受け取ったユーザは偏った意見を形成する可能性がある．本研究では，このような検索結果の偏りを捉える指標として，スタンス分布に着目する．スタンス分布とは，検索結果として提示された個々の文書が，そのトピックに対してどのような立場をとっているか（賛成・中立・反対など）の割合を示すものである．この分布が特定のスタンスに偏っている場合，ユーザが接触する視点が限定され，多角的な情報収集が阻害される可能性がある．したがって，スタンス分布は，検索結果がユーザの意見形成に与える影響を定量的に評価するための有効な指標であると考えられる．

検索結果のスタンス分布は，ユーザの検索行動によって異なる可能性がある．既存研究では，同じ情報要求を持つユーザ間でも，入力するクエリの表現は異なること [1,2,22,27]，及びその表現の違いが検索結果やランキングに大きな変化をもたらすことが示されている [1,2,12,27]．また，ユーザが利用する検索エンジンやシステムによって検索結果を提示するためのランキング手法も異なることが想定され，それらが検索性能および検索結果を変化させることが示されている [23,28]．しかし，クエリの表現の違いやランキング手法の選択が検索結果のスタンス分布に与える影響については検証されていない．

そこで，本研究では，クエリの表現方法の違いの 1 つであるクエリ形式と，アドホック検索において代表的な 3 つのランキング手法に注目し，それらが検索結果のスタンス分布に与える影響を調査する．これにより，賛否両論あるトピックに関する情報収集におけるバイアスへの理解を深め，ユーザの責任ある意見形成を支援するための，偏りの少ない検索結果の提示に寄与することを目指す．

上記の目的を達成するため，以下の 2 つのリサーチクエスチョン（RQ）を設定する：

- **RQ1**：同じ情報要求でも，クエリ形式が違えばスタンス分布は異なるのか？
- **RQ2**：同じクエリでも，ランキング手法が違えばスタンス分布は異なるのか？

これらのリサーチクエスチョンに答えるため，本研究では以下の実験を実施した．まず，賛否両論あるトピックを扱った既存のデータセットを用いて，質問形式のクエリとキーワード形式のクエリを，賛成・中立・反対のスタンス別に大規模言語モデル（LLM）を用いて生成した．次に，生成したクエリを用いて，3 つのランキング手法（BM25，e5，LLM を用いたペアワイズランキングプロンプティング（PRP））それぞれで検索結果を求めた．そして，検索結果からスタンス分布を求め，クエリ形式とランキング手法が検索結果のスタンス分布に与える影

響を定量的に評価した．

実験の結果，以下の知見が得られた：

1. 同じ情報要求でも，クエリ形式の違いによってスタンス分布は異なることを示した．また，ランキング手法によって，クエリ形式の違いによるスタンス分布の差異の大きさと影響の性質は異なることを示した．
2. 同じクエリでも，ランキング手法の違いによって検索結果のスタンス分布は異なることを示した．特に検索有効性の高いランキング手法で，クエリのスタンスが検索結果に強く反映される傾向が見られた．
3. クエリ形式やランキング手法に関わらず，文書コーパスの偏りは検索結果のスタンス分布に影響を与えることを示した．特に BM25 は，文書コーパスのスタンス分布に近似しやすく，コーパスの偏りの影響を受けやすいランキング手法であることを示した．

本研究の主な貢献は，同一の情報要求であっても，クエリ形式の違いとランキング手法の選択によって，検索結果のスタンス分布が変化することを定量的に明らかにしたことである．

## 2 関 連 研 究

### 2.1 検索結果とユーザの意見形成

検索結果はユーザの意見形成および態度変容に強い影響力を持つことが多くの研究で示されている [8,9,17]．Epstein ら [9] は，ウェブ検索を用いた有権者の投票に関する意思決定において，偏った検索結果がユーザの態度や信念を変化させる検索エンジン操作効果（Search Engine Manipulation Effect，SEME）と呼ばれる現象が確認され，検索結果の偏りがユーザの意見形成に影響を及ぼすことを示した．そして，検索エンジン操作効果の要因として，Epstein らは，上位にランキングされているページの意見をより重視する順序効果が影響していると分析したが，後の Draws ら [6] の研究で，特定の視点（一方のスタンス）を支持する文書を閲覧するほど，その視点を受け入れやすくなるという接触効果に起因する可能性が高いことが示唆されている．

### 2.2 賛否両論あるトピックにおける検索行動と検索結果のバイアス

賛否両論あるトピックに関する検索結果において，検索行動および検索結果の提示過程に内在する複数のバイアスが，間接的にユーザの意見形成に影響を及ぼすことが示されている．具体的には，ユーザが自身の事前信念に合致する情報を優先的に選択する確証バイアスに加え，ユーザの満足度最大化を目的として設計された検索エンジンのランキングアルゴリズムによるランキングバイアス，さらにユーザが検索結果の上位項目を優先的に選択するポジションバイアスが相互に作用することで，検索エンジン操作効果を引き起こし，ユーザの多角的な視点に基づいた意見形成を阻害する可能性が指摘されている [19]．

特に，ユーザは自身の政治的信条などの事前信念に合致する検索クエリを選択する傾向があり，そのようなクエリを検索エンジンに入力することで，事前信念に偏った検索結果が提示されやすくなる．この現象は，ユーザの事前信念をさらに強化する方向に作用することが報告されている [7,22]．同様に，ユーザの事前信念は入力するクエリの文言に反映されることが知られており，クエリ自体に事前信念との明確な意味的関連性が見られない場合であっても，検索結果としてはユーザの信念や態度に合致した情報が提示されることが明らかにされている．その結果，対立する態度を持つ個人間で受け取る情報が大きく異なることが報告されている [11]．

さらに，賛否両論あるトピックを対象とした議論検索においては，クエリへの適合性を重視した高い検索有効性を持つランキング手法が，情報の公平性を犠牲にし，検索結果におけるスタンス分布の不均衡を助長する可能性があることが示されている [16]．このようなランキングの偏りと確証バイアスが結びつくことで，ユーザは検索結果を十分に精査しなくなり [14]，事前信念と一致する情報のみを選択的に閲覧するため，信念が補強される傾向が強まることが示唆されている [25]．

### 2.3 クエリ形式が検索有効性に与える影響

同じ情報要求であっても様々なクエリの表現方法があり，ユーザによって入力されるクエリは異なるが [1,2,22,27]，本研究では，異なるクエリの表現としてクエリ形式に注目する．クエリ形式は質問クエリとキーワードクエリの 2 種類に大別される [26]. キーワードクエリは，従来から主要なクエリ形式として広く用いられてきたが [26]，近年では質問クエリの利用が増加傾向にある [5]．既存研究では，一般的な検索タスクにおいて，質問クエリとキーワードクエリの検索有効性の差異に統計的な有意差は確認されなかったと報告している [26]．

しかし，特定のタスクにおいて，クエリ形式が検索有効性に影響を与えることが示されている．松田ら [28] は，多言語検索において，クエリ形式の違いが検索有効性に与える影響について分析した．語彙ベースの検索モデルである BM25 では，その影響は限定的であったが，DPR [13] などの意味ベースの検索モデルにおいて，質問クエリの方がキーワードクエリよりも高い検索有効性を示す傾向にあることを明らかにした．また，Wang ら [24] は複数の情報を統合する必要がある複雑な検索タスクにおいて，キーワードクエリの方が質問クエリよりも検索有効性が高く，単一の情報を要求する単純な検索タスクにおいては，質問クエリの方がキーワードクエリよりも検索有効性が高い傾向にあることを示し，タスクの複雑性がクエリ設計の有効性に影響を及ぼすことを報告した．

### 2.4 ランキング手法が検索有効性に与える影響

既存研究では，語彙ベースの検索から意味ベースの検索，さらに LLM を用いたリランキングへと至るランキング手法の進展が，検索有効性に与える影響について多角的に検討されてきた．BM25 は，語彙の一致に基づくランキング手法として，高い効率性と解釈性を備え [20]，多様な検索タスクにおけるベースラインとして広く利用されている．一方で，意味ベースの密検索モデルは，Transformer に基づく埋め込み表現を用いるこ

とで，語彙の不一致により関連文書が検索されない問題を克服し，BM25 を上回る検索性能を発揮する可能性があることが示唆されている [23].

さらに，BM25 や密検索モデルによって取得された文書候補に対し，LLM を直接リランカーとして適用する手法は，クロスエンコーダやハイブリッド検索などの高性能な手法を上回る検索有効性を達成しており，LLM の高度な文脈理解能力がリランキングにおいて有効であることが示唆されている [4].

また，デジタル化されていない歴史的文書のように，ノイズや複雑な文書構造を含む多数の文書を対象とした検索タスクにおいては，意味ベースの検索モデルよりも BM25 が高い性能を示す傾向が確認されており，検索の対象となるデータの特性に応じて適切なランキング手法を選択する重要性が指摘されている [10].

そこで，本研究では，語彙ベースの手法として BM25，意味ベースの手法として e5，および LLM を直接ランキングモデルに適用する手法としてペアワイズランキングプロンプティング（PRP） [18] の 3 つのランキング手法を用いる．

### 2.5 本研究の位置付け

このように，クエリ形式やランキング手法が検索有効性にどう影響するかについては様々なタスクで検証されているが，それらが賛否両論あるトピックにおける検索結果のスタンス分布に与える影響については検証されていない．そこで，本研究では質問クエリとキーワードクエリという 2 つのクエリ形式がスタンス分布に与える影響と，検索モデルとして一般的に用いられ，高い検索有効性を持つ 3 つのランキング手法がスタンス分布に与える影響を定量的に評価して分析する．

## 3 実験方法

### 3.1 実験の概要

リサーチクエスチョンを明らかにするため，本研究では以下の手順で実験を行った．まず，賛否両論あるトピックを扱った既存のデータセットを用いて，質問クエリとキーワードクエリからなるクエリペアを，賛成・中立・反対のスタンス別に 30 件ずつ LLM を用いて生成した．次に，生成したクエリペアを用いて，データセットの文書コーパスを対象に，3 つの手法で文書をランキングした．そして，得られたランキング上位 $k$ 件のスタンス分布を求め，クエリ形式とランキング手法による影響を複数の指標で評価した．

### 3.2 データセット

本研究の実験には，Draws ら [6] の研究において，検索結果のスタンスの偏りがユーザの態度にもたらす影響を分析するために構築された，賛否両論あるトピックに関するデータセットを用いた．表 1 にデータセットの統計情報を示す．このデータセットでは，ユーザの意見が特定の立場に偏っていないことと，確信度の低い意見が多数を占め，論争の余地があることをトピックの選定基準として，“Is obesity a disease?” といった健康課題に関するトピックや “Should bottled water be banned?” といった環境問題に関するトピックなど，学術的かつ賛否両論あるトピックが 5 つ選定されている．

表 1 データセットの全トピックと各トピックにおけるスタンス別文書数（表 2 に示すスタンス評価値が+1 から+3 の文書を賛成文書，0 の文書を中立文書，-1 から-3 の文書を反対文書として集計）．

| トピック | 文書数 | | |
|---|---|---|---|
| | 賛成 | 中立 | 反対 |
| Is obesity a disease? | 24 | 13 | 19 |
| Is cell phone radiation safe? | 12 | 11 | 33 |
| Should bottled water be banned? | 26 | 9 | 21 |
| Should zoos exist? | 22 | 6 | 28 |
| Are social networking sites good for our society? | 20 | 14 | 22 |

表 2 7 段階のリッカート尺度によるスタンスの評価値とラベル，及びラベルを具体的に表す例文（文献 [6] の表 1 を引用）．

| Viewpoint Label | | Example(Topic:“Should Zoos Exist?”) |
|---|---|---|
| +3 | strongly supporting | “There is nothing wrong with zoos!” |
| +2 | supporting | “I‘m in favor of zoos, let’s keep them.” |
| +1 | somewhat supporting | “Zoos are not great, but they benefit society.” |
| 0 | neutral | “We present arguments for and against zoos.” |
| -1 | somewhat opposing | “Despite some benefits, I’m against zoos.” |
| -2 | opposing | “We should strive towards closing all zoos.” |
| -3 | strongly opposing | “Horrible places! All zoos should be closed.” |

各トピックに 56 件の文書からなるコーパスがあり，各文書にはクラウドワーカによって二値の適合性と，トピックに対する見解（スタンス）が「強く支持」から「強く反対」までの 7 段階のリッカート尺度で与えられている．文献 [6] に記載されているスタンスの評価基準とその例を表 2 に示す．また，不適合文書におけるスタンスの評価値は 0 である．しかし，不適合文書は各トピックのコーパスで 0 から 2 件と非常に少ないため，本研究の実験において不適合文書が検索結果に与える影響は限定的であり，主に適合文書間の順序関係に着目したリランキングの問題に近い性質を持つ．

表 1 から，大半のトピックのコーパスが，約 20 件の賛成文書と反対文書，及び約 10 件の中立文書で構成されていることが分かる．しかし，“Is Cell Phone Radiation Safe?” というトピックのみ，反対文書数が賛成文書数より多く，他のトピックと相対的に比較しても，偏りのあるコーパスであることが分かる．このような文書コーパスの偏りが検索結果のスタンス分布に及ぼす影響について，4 節の実験結果で議論する．

### 3.3 クエリの生成

本研究では，データセットのトピックから，質問クエリとキーワードクエリで構成されるクエリペアを LLM で生成した．表 3 に生成したクエリペアの一例を示す．例は使用するデータセットにおけるトピック “Should bottled water be banned?” に対する生成クエリペアである．

既存研究において，ユーザは事前信念に沿った検索クエリを選択する傾向にあり，選択したクエリを検索エンジンで使用することで，事前信念を補強するような検索結果が表示される傾向にあることが確認されているため [7, 11, 22]，ユーザの検索行動を考慮して，クエリペアを賛成・中立・反対のスタンス別

表 3 GPT-4o を用いて生成したクエリペアの例（トピック：“Should bottled water be banned?”）.

| | 質問クエリ | キーワードクエリ |
|---|---|---|
| 賛成クエリ | What are the environmental benefits of banning bottled water?<br>How does banning bottled water reduce plastic waste? | Environmental benefits of banning bottled water<br>Banning bottled water and plastic waste reduction |
| 中立クエリ | What are the environmental impacts of bottled water?<br>What are the health benefits and risks of bottled water? | Environmental impacts of bottled water<br>Health benefits and risks of bottled water |
| 反対クエリ | Why shouldn't bottled water be banned?<br>What are the benefits of bottled water? | Reasons to keep bottled water<br>Advantages of bottled water |

に複数生成した．本節では，クエリペアの生成方法と生成されたクエリペアの品質について説明する．

### 3.3.1 クエリペアの生成

クエリペアの生成には，OpenAI の GPT-4o (2024-11-20)[1] を用いた．生成する際の temperature は 0 に設定した．そして，データセットの各トピックについて，賛成・中立・反対の 3 つのスタンス別に，質問クエリとキーワードクエリからなるクエリペアを 30 件ずつ生成した．

例として，クエリペアを生成する際に与えるプロンプトを以下に示す．プロンプトは，クエリの例を数件提示する Few-shot プロンプトであり，与えるクエリの例は生成するクエリのスタンスによって異なる．また，赤字の stance には，賛成クエリを生成する際は supportive，中立クエリを生成する際は neutral，反対クエリを生成する際は opposing を指定する．

> You are an AI assistant that generates diverse search queries based on users' information needs. For the given query, please infer the user's information need and then generate 30 **pairs** of search queries that a user with a **stance** stance might formulate.
> Each pair must consist of one **Question Query** (e.g., "Should we adopt X?", "What is the best way to achieve Y?") and one related **Keyword Query** (e.g., "X benefits," "Y advantages," "reasons to support Z"). The two queries in each pair must seek **exactly the same information**.
>
> ---
>
> Query: {original_query}
>
> ---
>
> ### Output Format
> Please return the generated queries as a comma-separated list of 30 pairs. Each pair should be formatted as: **["Question Query", "Keyword Query"]**.
> Example:
> [["Question Query 1", "Keyword Query 1"], ["Question Query 2", "Keyword Query 2"], ..., ["Question Query 30", "Keyword Query 30"]]

Bailey らの研究 [2] では，クラウドワーカを利用した多様なクエリの収集の際に，事前に作成した情報要求を与え，その情報要求に応えるために検索エンジンで使用するクエリを入力することを求めた．本研究で使用するプロンプトではそれらのクエリ収集プロセスを参考にして，与えられたトピックに対する情報要求を考え，各スタンスを支持するユーザが作成，入力すると考えられる検索クエリを生成するように指示した．そして，生成するクエリペアが，それぞれのペアで同じ情報要求を持つことを強調した．

表 4 生成したクエリペアの意味的類似度の平均（標準偏差）.

| | 平均 | 最大 | 最小 |
|---|---|---|---|
| 賛成クエリペア | 0.880(0.050) | 0.972 | 0.743 |
| 中立クエリペア | 0.914(0.041) | 0.978 | 0.775 |
| 反対クエリペア | 0.887(0.038) | 0.967 | 0.782 |

### 3.3.2 クエリペアの意図保持性の評価

生成したクエリペアが同じ情報要求を満たすことを確認するため，英語テキストの高性能な埋め込みモデル `BAAI/bge-large-en-v1.5`[2]を用いて，質問クエリとキーワードクエリのコサイン類似度を算出した．既存研究では，形式の異なるクエリが同じ情報要求を満たしているかを測る意図保持性の評価に Universal Sentence Encoder [3] や OpenAI の text-embedding-3-large[3] などのモデルが用いられているが [12, 28]，本研究では検索タスク向けに最適化された埋め込みモデルであり，情報検索分野において高い性能が報告されている点[4]を考慮し，bge-large-en-v1.5 を採用した．表 4 に算出したコサイン類似度の値を示す．

生成したクエリペアは，どのスタンスにおいてもコサイン類似度の平均値が 0.8 を超えており，最小値は 0.7 以上であった．このコサイン類似度の目安として，Iovine らの研究 [12] で，意図保持性を満たすコサイン類似度の閾値が定義されている．Iovine らは，質問クエリとキーワードクエリの間で双方向の書き換えを可能にする教師なしモデルの提案において，モデルの学習に使用するクエリペアの類似度の閾値を 0.6 としている．これは，ペアとなる質問クエリとキーワードクエリが類似したドメインを共有していることを満たす基準として適用された．また，モデルを評価するためのテストデータとして使用するクエリペアの類似度の閾値を 0.8 としている．これは，同じ情報

1：https://platform.openai.com/docs/models/gpt-4o?snapshot=gpt-4o-2024-11-20

2：https://huggingface.co/BAAI/bge-large-en-v1.5

3：https://platform.openai.com/docs/models/text-embedding-3-large

4：https://huggingface.co/spaces/mteb/leaderboard(2026 年 2 月 2 日閲覧)

要求を満たし，意味的にほとんど等しいことを示す，より厳格な基準として適用された．本研究で生成したクエリペアは，コサイン類似度の最小値が 0.6 より大きく，平均値も 0.8 より大きいため，既存研究で定義されている意図保持性の基準を概ね満たしているといえる．

### 3.4 ランキング手法

本研究では，BM25，e5，PRP の 3 つのランキング手法を使用した．

#### 3.4.1 BM25

BM25 [20] は，クエリと文書の単語の一致に基づくランキング手法であり，文書内の単語頻度（TF）と逆文書頻度（IDF）を用いて適合性を評価する．本研究では，語彙ベースの疎検索モデルとして BM25 を用いる．

#### 3.4.2 e5

e5 [23] は，情報検索のタスクに特化して学習されたテキスト埋め込みモデルであり，入力されたクエリや文書のテキストを意味的な特徴を捉えた高次元のベクトルに変換する．本モデルは対照学習を用いて，クエリと適合する文書のベクトルは近くに，適合しない文書のベクトルは遠くなるような埋め込み空間を構築する．本研究では，語彙一致に依存しない意味ベースの密検索モデルとして `intfloat/e5-base-v2`[5]を用いる．

#### 3.4.3 ペアワイズランキングプロンプティング（PRP）

PRP [18] は，LLM に 2 つの文書を提示し，どちらがクエリに対してより適合性が高いかを判断させる手法である．この手法では，LLM に与えるプロンプトを工夫することで，モデルが文書間の相対的な優劣を比較し，ランキングに必要なスコアを生成する．本研究では，LLM を直接的にランキングモデルとして利用するためのアプローチとして，PRP を用いる．以下に PRP で使用するペアワイズランキングのプロンプトを示す．プロンプトは既存研究 [18] を参考にし，GPT-4o-mini (2024-06-01)[6]を LLM として使用した．

> Given a query {query}, which of the following two passages is more relevant to the query?
> Passage A: {document1}
> Passage B: {document2}
> Output Passage A or Passage B:

LLM はプロンプト内のテキスト順序に敏感であるため [15]，PRP では，文書の順序を入れ替えて，2 回問い合わせを行うことで，入力順序に対するバイアスを低減させている．各クエリにおいて，コーパス内の全ての組み合わせで PRP を行ってスコアを割り当て，合計したスコアが高い順にソートしてランキングする．

### 3.5 評価指標

#### 3.5.1 スタンス分布

本研究では，検索結果の偏りを定量的に評価するために，検索結果における賛成文書・中立文書・反対文書の割合を示すスタンス分布を求める．具体的には，ランキング上位 $k$ 件において，スタンス評価値が $+1$ から $+3$ の文書を賛成文書，0 の文書を中立文書，$-1$ から $-3$ の文書を反対文書として，賛成・中立・反対のスタンス別文書数の割合を計算する．

#### 3.5.2 Root Normalized Order-aware Divergence(RNOD)

本研究では，質問クエリによって得られたスタンス分布とキーワードクエリによって得られたスタンス分布の差異を評価する指標として RNOD を用いる．RNOD [21] は，カテゴリ間に定義された順序構造を考慮しながら 2 つの確率分布間の差異を測定する距離指標である．RNOD の値が小さいほど，2 つの分布がより類似していることを示す．

RNOD の特徴は，カテゴリ間の距離を考慮して分布間の差異を評価できる点にある．たとえば，上位 5 件が全て賛成文書である検索結果 A，上位 5 件が全て中立文書である検索結果 B，上位 5 件が全て反対文書である検索結果 C を比較する．このとき，賛成文書が全て中立文書に置き換わっている A と B よりも，賛成文書がすべて反対文書に置き換わっている A と C の方を，分布間の差異として大きいものとして評価できる．検索結果間のスタンス分布の差異を測定するため，本研究では，賛成，中立，反対を区別して評価可能な RNOD を採用した．

本研究では，賛成，中立，反対という 3 つの要素と順序からなるカテゴリ集合 $C = \{\text{賛成}, \text{中立}, \text{反対}\}$ を考える．質問クエリによって得られた上位 $k$ 件のスタンス分布を $P = (p_{\text{賛成}}, p_{\text{中立}}, p_{\text{反対}})$ と表す．例えば，質問クエリによる上位 5 件の検索結果において，賛成文書が 3 件，中立文書が 1 件，反対文書が 1 件出現していれば，$P = (p_{\text{賛成}} = 0.6,\ p_{\text{中立}} = 0.2,\ p_{\text{反対}} = 0.2)$ となる．同様に，キーワードクエリによって得られた上位 $k$ 件のスタンス分布を $P^* = (p^*_{\text{賛成}}, p^*_{\text{中立}}, p^*_{\text{反対}})$ と表す．このとき，2 つのスタンス分布 $P$ と $P^*$ の RNOD は以下の式で計算される：

$$\mathrm{RNOD}(P \parallel P^*) = \sqrt{\frac{\mathrm{OD}(P \parallel P^*)}{|C| - 1}} \tag{1}$$

ここで，$\mathrm{OD}(P \parallel P^*)$ は以下の式で定義される：

$$\mathrm{OD}(P \parallel P^*) = \frac{1}{|C|} \sum_{i \in C} \mathrm{DW}_i \tag{2}$$

$$\mathrm{DW}_i = \sum_{j \in C} \delta_{ij} (p_j - p^*_j)^2, \qquad \delta_{ij} = |i - j| \tag{3}$$

ただし，$\delta_{ij}$ はカテゴリ間の距離を表し，本研究では 賛成$= 0$, 中立$= 1$, 反対$= 2$ とした．例えば，$\delta_{\text{賛成反対}} = |0 - 2| = 2$ となる．

## 4 実験結果

本節では，本研究の実験結果として，まず，RQ1 に関する結果を示し，次に，RQ2 に関する結果を示す．

### 4.1 RQ1：同じ情報要求でも，クエリ形式が違えばスタンス分布は異なるのか？

本節では，クエリ形式の違いがスタンス分布に与える影響を

5：https://huggingface.co/intfloat/e5-base-v2
6：https://platform.openai.com/docs/models/gpt-4o-mini

表 5　ランキング上位 5 件のスタンス分布（%）.

| | | BM25 | | | e5 | | | PRP | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 賛成文書 | 中立文書 | 反対文書 | 賛成文書 | 中立文書 | 反対文書 | 賛成文書 | 中立文書 | 反対文書 |
| 質問クエリ | 賛成クエリ | 37.9 | 18.3 | 43.9 | 39.2 | 25.6 | 35.2 | 63.9 | 16.0 | 20.1 |
| | 中立クエリ | 38.7 | 16.1 | 45.2 | 31.9 | 23.6 | 44.5 | 45.7 | 18.3 | 36.0 |
| | 反対クエリ | 36.9 | 13.5 | 49.6 | 22.8 | 23.2 | 54.0 | 15.7 | 13.1 | 71.2 |
| キーワードクエリ | 賛成クエリ | 39.2 | 20.7 | 40.1 | 42.8 | 24.4 | 32.8 | 63.3 | 16.1 | 20.5 |
| | 中立クエリ | 34.4 | 24.3 | 41.3 | 33.5 | 21.3 | 45.2 | 43.9 | 19.2 | 36.9 |
| | 反対クエリ | 35.5 | 20.9 | 43.6 | 24.5 | 17.3 | 58.1 | 16.7 | 13.9 | 69.5 |

表 6　クエリペアにおける質問クエリ上位 5 件のスタンス分布とキーワードクエリ上位 5 件のスタンス分布の RNOD の平均（標準偏差）.

| | BM25 | e5 | PRP |
|---|---|---|---|
| 賛成クエリ | 0.236(0.162) | 0.213(0.149) | 0.096(0.115) |
| 中立クエリ | 0.214(0.138) | 0.159(0.130) | 0.109(0.124) |
| 反対クエリ | 0.242(0.134) | 0.174(0.138) | 0.124(0.172) |
| 平均 | 0.231(0.145) | 0.182(0.139) | 0.110(0.137) |

ランキング手法別に分析する．表 5 に各ランキング手法における上位 5 件のスタンス分布を示す．そして，表 6 に各ランキング手法の，質問クエリにおける上位 5 件のスタンス分布とキーワードクエリにおける上位 5 件のスタンス分布の RNOD の平均値を示す．また，実験はランキング結果上位 3 件，上位 5 件，上位 10 件でスタンス分布を求めて評価したが，どの件数においても同様の傾向が得られた．

**4.1.1**　BM25 におけるクエリ形式の影響

表 5 に示す BM25 の結果において，質問クエリでは中立文書の割合が低く (賛成クエリ：18.3%，中立クエリ：16.1%，反対クエリ：13.5%)，キーワードクエリで高い（賛成クエリ：20.7%，中立クエリ：24.3%，反対クエリ：20.9%) 傾向が観察された．この結果は，BM25 が語彙の一致に基づいたランキング手法であることに起因すると考えられる．質問クエリには機能語や文の構造を含む表現が多く含まれるため，それらと語彙的に強く対応する文書が上位にランク付けされやすくなる．その結果，賛成・反対といった立場を示す語彙を多く含む文書が上位にランキングされたと考えられる．一方，キーワードクエリでは話題や対象を表す語が中心となるため，特定の立場を示さずに事実や背景を説明する文書とも一致しやすく，中立的な文書が相対的に多く検索されたと考えられる．

また，表 6 の RNOD の結果から，BM25 は本研究で用いた 3 つのランキング手法の中で RNOD の平均値が最も大きく（平均：0.231），クエリ形式の違いによるスタンス分布の差異が大きいことが確認された．これは，語彙ベースのランキング手法がクエリ表現の表層的な違いに敏感であるという既存研究の知見 [1,2,28] とも一致しており，BM25 においては同一の情報要求であっても，クエリ形式の違いが検索結果のスタンス分布に影響を及ぼす可能性が高いことを示している．

**4.1.2**　e5 におけるクエリ形式の影響

表 5 に示す e5 の結果において，質問クエリでは中立文書の割合が高く（賛成クエリ：25.6%，中立クエリ：23.6%，反対クエリ：23.2%），キーワードクエリでは中立文書の割合が低い（賛成クエリ：24.4%，中立クエリ：21.3%，反対クエリ：17.3%）傾向が観察された．この傾向は BM25 における影響とは異なるものである．また，特徴的な傾向として，賛成クエリにおいては賛成文書の割合が高く（質問クエリ：39.2%，キーワードクエリ：42.8%），反対クエリにおいては反対文書の割合が高い（質問クエリ：54.0%，キーワードクエリ：58.1%）というクエリのスタンスに応じた偏りが，キーワードクエリで強まる傾向が観察された．この結果は，e5 ではキーワードクエリが，クエリの潜在的なスタンスをより強調し，スタンスの偏りを増幅させる可能性があることを示唆している．

また，表 6 に示す RNOD の結果から，e5 の RNOD の平均値は BM25 よりも小さい値を示し（平均：0.182），クエリ形式の違いによるスタンス分布の差異が BM25 よりも小さいことが確認された．これは e5 が語彙の一致に依存せず意味的類似性に基づいて検索を行うため，クエリの表現の違いによる影響を緩和していることを示す．

**4.1.3**　PRP におけるクエリ形式の影響

表 5 に示す PRP の結果において，質問クエリによるスタンス分布とキーワードクエリによるスタンス分布で同様の傾向を示した．また，表 6 に示す RNOD の結果から，PRP は本研究で用いた 3 つのランキング手法の中で RNOD の平均値が最も小さく（平均：0.110），クエリ形式の違いによるスタンス分布の差異が最も小さい手法であることが示された．これらの結果は，PRP がクエリと文書を同時に入力するペアワイズ比較に基づく手法であり，クエリ形式の差異よりも，クエリが示す本質的な立場や意図を優先して評価していることを示唆する．

### 4.2　RQ2：同じクエリでも，ランキング手法が違えばスタンス分布は異なるのか？

本節では，表 5 の結果から，ランキング手法の違いがスタンス分布に与える影響を分析する．まず，BM25 においては，e5 や PRP と比べて，スタンス分布に大きな偏りは見られなかった．質問形式の反対クエリにおいては反対文書の割合がやや高い傾向にあったが（質問・反対クエリ：0.496），全体として，賛成文書と反対文書が比較的均等に分布している．これは，BM25

が単語の一致に基づく手法であり，文脈やスタンスといった意味的特徴を十分に捉えることができないため，クエリのスタンスに対する感度が低いことに起因すると考えられる．この特性は，BM25 が多様な視点を含む検索結果を比較的維持しやすい一方で，特定の立場に基づく情報探索には適さない可能性を示している．

一方で，e5 ではクエリのスタンスに応じた分布の偏りが観察された．具体的には，賛成クエリでは賛成文書の割合が高く（質問クエリ：0.392，キーワードクエリ：0.428），反対クエリでは反対文書の割合が高まる傾向を示した（質問クエリ：0.540，キーワードクエリ：0.581）．e5 は，高次元の埋め込み空間を用いて，クエリと文書の意味的特徴を捉えることが可能なため，クエリが持つ潜在的なスタンスに感応し，類似したスタンスを持つ文書を上位にランク付けしたと考えられる．この結果は，意味ベースの検索モデルが検索有効性を向上させる一方で，ユーザの立場を強化する方向に検索結果のスタンスが偏る可能性を持つことを示唆している．

さらに PRP では，その傾向が顕著であり，賛成クエリでは賛成文書の割合が 6 割以上を占め（質問クエリ：0.639，キーワードクエリ：0.633），反対クエリでは反対文書の割合が 7 割程を占めた（質問クエリ：0.712，キーワードクエリ：0.695）．これは，LLM が単なる意味的類似性ではなく，クエリと文書の間のスタンスの整合性や論調の一致を重視してランキングを行っている可能性を示している．既存研究 [18] が指摘するように，LLM は高度な文脈理解能力を持つ一方で，クエリに含まれる前提や価値観を強く反映した検索結果を生成・選択する傾向があり，本研究の結果はその特性が検索結果のスタンス分布にも表れていることを示している．

### 4.3 文書コーパスの偏りがスタンス分布に与える影響

本節では，3.3 節で言及した文書コーパスの偏りが検索結果のスタンス分布に与える影響について検討する．表 7 にランキング上位 5 件におけるトピック別のスタンス分布を示す．また，表 8 に各トピックにおけるデータセット全体の文書コーパスのスタンス分布とランキング上位 5 件のスタンス分布の RNOD の平均を示す．RNOD の値が小さいほど，ランキング結果から得られたスタンス分布がコーパスのスタンス分布に類似していることを示す．まず，表 7 の結果において，コーパスにおける反対文書の割合が高い “Is cell phone radiation safe?” というトピックのみ，クエリのスタンスやランキング手法に関わらず，検索結果の反対文書の割合が高い傾向が確認された．これは，文書コーパスの偏りがクエリの表現やランキング手法に関わらず，検索結果の偏りを増強することを示唆する．

また，文書コーパスのスタンス分布と各ランキング手法で上位 5 件のスタンス分布の差異を定量的に評価した表 8 の結果において，3 つのランキング手法のうち，BM25 における RNOD の平均が最も低い（BM25：0.218，e5：0.246，PRP：0.308）ことが確認された．これは BM25 を用いてランキングしたスタンス分布は，コーパスのスタンス分布に近似することを示し，BM25 がコーパスの偏りの影響を受けやすい手法であることを示唆する．

## 5 議　　論

本研究は，クエリ形式およびランキング手法の違いが検索結果のスタンス分布に与える影響を明らかにすることを目的として実施した．実験結果から，クエリ形式とランキング手法の双方が検索結果のスタンス分布を変化させることを示した．これにより，同一の情報要求であっても，クエリの形式やランキング手法によって，意図せず異なる検索結果が提示される可能性があることを明らかにした．ユーザが入力するクエリの表現や検索システムが採用するランキング手法によって，検索結果に偏りが生じる可能性を考慮する必要性があることを指摘した点は，本研究の重要な意義である．

RQ1 に関して，同一の情報要求であっても，質問クエリとキーワードクエリというクエリ形式の違いが，検索結果のスタンス分布に影響を与えることを明らかにした．BM25 はクエリの表層的な違いに敏感に反応し，クエリ形式の違いによるスタンス分布の差異が大きいことが示された．一方で，クエリと文書の意味的類似性を考慮できる e5 と PRP では，スタンス分布の差異が小さく，クエリの表現の揺らぎに対して頑健な手法であることが示された．また，クエリ形式の違いがスタンス分布に与える具体的な影響については，ランキング手法によって異なることが確認された．これらの結果から，ユーザの些細なクエリ表現の違いが，意図しない検索結果の変化を生む可能性があるといえる．

次に，RQ2 に関して，ランキング手法の違いがスタンス分布に与える影響は明確であった．クエリと文書の意味的類似性を考慮できる e5 や PRP でランキングしたスタンス分布は，クエリのスタンスに応じた検索結果の偏りが強まる傾向が確認された．これは，意味ベースや LLM ベースのランキング手法は，検索有効性は高い一方で，クエリに内在するスタンスを強く反映し，検索結果の偏りを強める可能性があるため，ユーザが受け取る情報の多様性を損なう危険性があることを示す．

さらに，本研究では，文書コーパスの偏りが検索結果のスタンス分布に与える影響について検証した．分析の結果から，検索結果のスタンス分布が，コーパスに内在する偏りの影響を強く受けることが確認された．実際に，コーパスにおける反対文書の割合が高いトピックでは，クエリ形式やランキング手法に関わらず，検索結果のスタンス分布における反対文書の割合が高まる傾向が確認された．また，実験で使用した 3 つのランキング手法のうち，BM25 を用いてランキングしたスタンス分布が，コーパスのスタンス分布に最も近似していたことから，BM25 のような単語の一致に基づく語彙ベースの検索手法は，コーパスの偏りの影響を受けやすい手法である可能性が示唆された．

以上の結果から，本研究は，検索結果のスタンス分布に生じる偏りが，クエリ形式，ランキング手法，及び文書コーパスの偏りという複数の要因の相互作用によって形成されることを示したといえる．

表 7　ランキング上位 5 件のトピック別スタンス分布（%）.

| | | | BM25 | | | e5 | | | PRP | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 賛成文書 | 中立文書 | 反対文書 | 賛成文書 | 中立文書 | 反対文書 | 賛成文書 | 中立文書 | 反対文書 |
| 質問クエリ | 賛成クエリ | Is obesity a disease? | 48.0 | 24.0 | 28.0 | 46.0 | 16.7 | 37.3 | 78.0 | 21.3 | 0.7 |
| | | Is cell phone radiation safe? | 8.7 | 11.3 | 80.0 | 8.7 | 25.3 | 66.0 | 17.3 | 16.7 | 66.0 |
| | | Should bottled water be benned? | 26.0 | 19.3 | 54.7 | 26.0 | 34.0 | 40.0 | 52.7 | 34.0 | 13.3 |
| | | Should zooz exist? | 56.7 | 14.0 | 29.3 | 74.0 | 8.7 | 17.3 | 94.7 | 5.3 | 0.0 |
| | | Are social networking sites good for our society? | 50.0 | 22.7 | 27.3 | 41.3 | 43.3 | 15.3 | 76.7 | 2.7 | 20.7 |
| | 中立クエリ | Is obesity a disease? | 57.3 | 15.3 | 27.3 | 38.7 | 26.7 | 34.7 | 78.0 | 16.0 | 6.0 |
| | | Is cell phone radiation safe? | 8.0 | 8.7 | 83.3 | 12.7 | 25.3 | 62.0 | 11.3 | 16.0 | 72.7 |
| | | Should bottled water be benned? | 37.3 | 18.0 | 44.7 | 32.7 | 30.7 | 36.7 | 43.3 | 34.7 | 22.0 |
| | | Should zooz exist? | 47.3 | 17.3 | 35.3 | 53.3 | 6.7 | 40.0 | 59.3 | 10.7 | 30.0 |
| | | Are social networking sites good for our society? | 43.3 | 21.3 | 35.3 | 22.0 | 28.7 | 49.3 | 36.7 | 14.0 | 49.3 |
| | 反対クエリ | Is obesity a disease? | 53.3 | 17.3 | 29.3 | 28.0 | 18.7 | 53.3 | 44.0 | 11.3 | 44.7 |
| | | Is cell phone radiation safe? | 10.0 | 10.0 | 80.0 | 10.0 | 22.7 | 67.3 | 0.7 | 9.3 | 90.0 |
| | | Should bottled water be benned? | 40.0 | 14.0 | 46.0 | 32.0 | 32.7 | 35.3 | 26.0 | 32.0 | 42.0 |
| | | Should zooz exist? | 42.0 | 7.3 | 50.7 | 30.0 | 6.0 | 64.0 | 2.7 | 0.0 | 97.3 |
| | | Are social networking sites good for our society? | 39.3 | 18.7 | 42.0 | 14.0 | 36.0 | 50.0 | 5.3 | 12.7 | 82.0 |
| キーワードクエリ | 賛成クエリ | Is obesity a disease? | 39.3 | 42.7 | 18.0 | 35.3 | 39.3 | 25.3 | 82.7 | 17.3 | 0.0 |
| | | Is cell phone radiation safe? | 7.3 | 13.3 | 79.3 | 5.3 | 22.0 | 72.7 | 10.0 | 14.0 | 76.0 |
| | | Should bottled water be benned? | 44.0 | 12.0 | 44.0 | 50.0 | 13.3 | 36.7 | 54.0 | 39.3 | 6.7 |
| | | Should zooz exist? | 55.3 | 12.0 | 32.7 | 77.3 | 6.7 | 16.0 | 97.3 | 2.0 | 0.7 |
| | | Are social networking sites good for our society? | 50.0 | 23.3 | 26.7 | 46.0 | 40.7 | 13.3 | 72.7 | 8.0 | 19.3 |
| | 中立クエリ | Is obesity a disease? | 36.7 | 40.0 | 23.3 | 31.3 | 32.7 | 36.0 | 76.0 | 20.7 | 3.3 |
| | | Is cell phone radiation safe? | 5.3 | 16.0 | 78.7 | 5.3 | 21.3 | 73.3 | 4.7 | 15.3 | 80.0 |
| | | Should bottled water be benned? | 44.0 | 27.3 | 28.7 | 45.3 | 17.3 | 37.3 | 44.7 | 36.0 | 19.3 |
| | | Should zooz exist? | 50.0 | 13.3 | 36.7 | 60.0 | 6.0 | 34.0 | 61.3 | 5.3 | 33.3 |
| | | Are social networking sites good for our society? | 36.0 | 24.7 | 39.3 | 25.3 | 29.3 | 45.3 | 32.7 | 18.7 | 48.7 |
| | 反対クエリ | Is obesity a disease? | 42.7 | 29.3 | 28.0 | 26.7 | 11.3 | 62.0 | 39.3 | 9.3 | 51.3 |
| | | Is cell phone radiation safe? | 6.0 | 14.0 | 80.0 | 6.0 | 16.0 | 78.0 | 0.0 | 6.0 | 94.0 |
| | | Should bottled water be benned? | 50.0 | 30.0 | 20.0 | 50.0 | 14.7 | 35.3 | 25.3 | 38.7 | 36.0 |
| | | Should zooz exist? | 50.0 | 9.3 | 40.7 | 29.3 | 5.3 | 65.3 | 15.3 | 0.0 | 84.7 |
| | | Are social networking sites good for our society? | 28.7 | 22.0 | 49.3 | 10.7 | 39.3 | 50.0 | 3.3 | 15.3 | 81.3 |

表 8　各トピックにおけるデータセットのコーパスのスタンス分布とランキング上位 5 件のスタンス分布の RNOD の平均（標準偏差）.

| | | BM25（標準偏差） | e5（標準偏差） | PRP（標準偏差） |
|---|---|---|---|---|
| 質問クエリ | 賛成クエリ | 0.226（0.117） | 0.264（0.097） | 0.329（0.069） |
| | 中立クエリ | 0.216（0.102） | 0.238（0.104） | 0.279（0.093） |
| | 反対クエリ | 0.194（0.101） | 0.231（0.098） | 0.284（0.074） |
| キーワードクエリ | 賛成クエリ | 0.230（0.112） | 0.268（0.082） | 0.345（0.054） |
| | 中立クエリ | 0.216（0.100） | 0.237（0.129） | 0.298（0.088） |
| | 反対クエリ | 0.225（0.093） | 0.235（0.098） | 0.314（0.068） |
| 平均 | | 0.218（0.104） | 0.246（0.101） | 0.308（0.074） |

しかし，本研究は，英語のトピックに限定された小規模なデータセットを対象としている点に限界がある．加えて，検索結果の評価は上位件数に基づいており，実際のユーザの検索行動や意見形成への影響までは直接的に検証していない．今後は，多言語および多文化環境における分析や，大規模なデータセットを対象とした検索実験，さらにはユーザ実験との統合を行っていく必要がある．

## 6 ま と め

本研究では，クエリ形式およびランキング手法の違いが，検索結果の偏りを示すスタンス分布に与える影響を調査した．実験の結果，検索有効性が高い意味ベースの検索モデルや LLM ベースのランキング手法で，クエリ形式の違いによる影響を抑制する傾向が確認されたが，クエリのスタンスに応じて検索結果の偏りが強まる傾向が観察された．また，文書コーパスの偏りが，クエリ形式やランキング手法に関わらず，検索結果に影響を与えることが確認され，特に BM25 などの語彙ベースの検索手法がその影響を受けやすい手法である可能性を示した．本研究は，検索有効性のみならず，スタンス分布を用いた検索結果の偏りの評価の重要性を示した．

## 謝　　辞

本研究は，JSPS 科研費 JP24K03228，JP25K03229 の助成を受けたものです．ここに記して謝意を表します．

## 文　　献

[1] Marwah Alaofi, Luke Gallagher, Dana Mckay, Lauren L. Saling, Mark Sanderson, Falk Scholer, Damiano Spina, and Ryen W. White. Where do queries come from? In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 2850–2862, 2022.

[2] Peter Bailey, Alistair Moffat, Falk Scholer, and Paul Thomas. UQV100: A test collection with query variability. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 725–728, 2016.

[3] Daniel Cer, Yinfei Yang, Sheng yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St. John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, Yun-Hsuan Sung, Brian Strope, and Ray Kurzweil. Universal sentence encoder. *arXiv preprint arXiv:1803.11175*, 2018.

[4] Murilo Cunha, Marilia Silveira, Brenda Santana, Larissa Freitas, and Ulisses Corrêa. Optimizing and evaluating a retrieval-augmented generation system for normative document retrieval in hospital settings. In *Proceedings of the 31st Brazilian Symposium on Multimedia and the Web*, pp. 385–393, 2025.

[5] Brian Dean. We analyzed 306m keywords. here's what we learned about google searches. `https://backlinko.com/go`

ogle-keyword-study.

[6] Tim Draws, Nava Tintarev, Ujwal Gadiraju, Alessandro Bozzon, and Benjamin Timmermans. This is not what we ordered: Exploring why biased search result rankings affect user attitudes on debated topics. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 295–305, 2021.

[7] Axel G Ekström, Guy Madison, Erik J Olsson, and Melina Tsapos. The search query filter bubble: Effect of user ideology on political leaning of search results through query selection. *Information Communication and Society*, Vol. 27, No. 5, pp. 878–894, 2024.

[8] Robert Epstein and Ji Li. Can biased search results change people's opinions about anything at all? A close replication of the search engine manipulation effect (SEME). *Plos one*, Vol. 19, No. 3, e0300727, 2024.

[9] Robert Epstein and Ronald E. Robertson. The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences*, Vol. 112, No. 33, pp. E4512–E4521, 2015.

[10] Haruki Fujimaki and Makoto P. Kato. KASYS at the NTCIR-18 SUSHI task. In *Proceedings of the 18th NTCIR Conference on Evaluation of Information Access Technologies*, pp. 422–455, 2025.

[11] Hussam Habib, Ryan Stoldt, Andrew High, Brian Ekdale, Ashley Peterson, Katy Biddle, Javie Ssozi, and Rishab Nithyanand. Algorithmic amplification of biases on google search. *arXiv preprint arXiv:2401.09044*, 2024.

[12] Andrea Iovine, Anjie Fang, Besnik Fetahu, Jie Zhao, Oleg Rokhlenko, and Shervin Malmasi. CycleKQR: Unsupervised bidirectional keyword-question rewriting. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 11875–11886, 2022.

[13] Vladimir Karpukhin, Barlas Oguz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wentau Yih. Dense passage retrieval for open-domain question answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pp. 6769–6781, 2020.

[14] Suzuki Masaki and Yusuke Yamamoto. Characterizing the influence of confirmation bias on web search behavior. *Frontiers in Psychology*, Vol. 12, No. 771948, pp. 1–11, 2021.

[15] Anna Neumann, Elisabeth Kirsten, Muhammad Bilal Zafar, and Jatinder Singh. Position is power: System prompts as a mechanism of bias in large language models (LLMs). In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, pp. 573–598, 2025.

[16] Sachin Pathiyan Cherumanal, Damiano Spina, Falk Scholer, and W. Bruce Croft. Evaluating fairness in argument retrieval. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pp. 3363–3367, 2021.

[17] Frances A. Pogacar, Amira Ghenai, Mark D. Smucker, and Charles L.A. Clarke. The positive and negative influence of search results on people's decisions about the efficacy of medical treatments. In *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval*, pp. 209–216, 2017.

[18] Zhen Qin, Rolf Jagerman, Kai Hui, Honglei Zhuang, Junru Wu, Le Yan, Jiaming Shen, Tianqi Liu, Jialu Liu, Donald Metzler, Xuanhui Wang, and Michael Bendersky. Large language models are effective text rankers with pairwise ranking prompting. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pp. 1504–1518, 2024.

[19] Alisa Rieger, Tim Draws, Nicolas Mattis, David Maxwell, David Elsweiler, Ujwal Gadiraju, Dana McKay, Alessandro Bozzon, and Maria Soledad Pera. Responsible opinion formation on debated topics in web search. In *Advances in Information Retrieval*, pp. 437–465, 2024.

[20] Stephen Robertson and Hugo Zaragoza. The probabilistic relevance framework: BM25 and beyond. *Foundations and Trends in Information Retrieval*, Vol. 3, No. 4, pp. 333–389, 2009.

[21] Tetsuya Sakai. On variants of root normalised order-aware divergence and a divergence based on Kendall's Tau. *arXiv preprint arXiv:2204.07304*, 2022.

[22] Marieke van Hoof, Corine S Meppelink, Judith Moeller, and Damian Trilling. Searching differently? How political attitudes impact search queries about political issues. *New Media & Society*, Vol. 26, No. 7, pp. 3728–3750, 2024.

[23] Liang Wang, Nan Yang, Xiaolong Huang, Binxing Jiao, Linjun Yang, Daxin Jiang, Rangan Majumder, and Furu Wei. Text embeddings by weakly-supervised contrastive pre-training. *arXiv preprint arXiv:2212.03533*, 2024.

[24] QianYing Wang, Clifford Nass, and Jiang Hu. Natural language query vs. keyword search: Effects of task complexity on search performance, participant perceptions, and preferences. In *Proceedings of the 2005 IFIP TC13 International Conference on Human-Computer Interaction*, pp. 106–116, 2005.

[25] Ryen W. White and Horvitz Eric. Belief dynamics and biases in web search. *ACM Transactions on Information Systems*, Vol. 33, No. 4, pp. 1–46, 2015.

[26] Ryen W. White, Matthew Richardson, and Wen-tau Yih. Questions vs. queries in informational search tasks. In *Proceedings of the 24th International Conference on World Wide Web*, pp. 135–136, 2015.

[27] Guido Zuccon, Joao Palotti, and Allan Hanbury. Query variations and their effect on comparing information retrieval systems. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pp. 691–700, 2016.

[28] 松田明梨, 加藤誠. 多言語検索における質問クエリとキーワードクエリの性能評価. 第 17 回データ工学と情報マネジメントに関するフォーラム, 6L–04, 2025.