

一般発表 | Track 5: 高度なデータ利活用・ドメイン応用 (医療情報, 教育, 地理情報等)

2026年3月1日(日) 13:00 ~ 15:10 | 会場

### [5L] 意思決定のモデリング

座長:山本 岳洋(兵庫県立大学) コメントータ:宇津呂 武仁(筑波大学) ジュニアコメントータ:伊藤 和浩(奈良先端科学技術大学院大学)

13:00 ~ 13:20

[5L-01] 奨学金制度文書からの適格性判定フローの自動抽出

\*金谷 一輝<sup>1</sup>、宮森 恒<sup>1</sup> (1. 京都産業大学)

---

13:20 ~ 13:40

[5L-02] 特徴量選択による日常的意思決定の個人化後悔予測モデル実装

\*乾 映之右<sup>1</sup>、石橋 直樹<sup>1</sup> (1. 武蔵野大学 データサイエンス学部)

---

13:40 ~ 14:05

[5L-03] クラウドソーシングにおけるワーカーの応募行動の推定

\*桑原 士門<sup>1</sup>、伊藤 寛祥<sup>1</sup>、森嶋 厚行<sup>1</sup> (1. 筑波大学)

---

14:05 ~ 14:30

[5L-04] 異種グラフを用いた知識モニタリング推定とメタ認知適応型学習支援

\*李 根<sup>1</sup>、陳 莉<sup>2</sup>、唐 成<sup>1</sup>、馬 博軒<sup>1</sup>、出口 大輔<sup>3</sup>、山下 隆義<sup>4</sup>、島田 敬士<sup>1</sup> (1. 九州大学、2. 大阪教育大学、3. 名古屋大学、4. 中部大学)

---

14:30 ~ 14:55

[5L-05] 社会選好を考慮したSNS上の誹謗中傷投稿の拡散を抑制する介入戦略

\*林 央祐<sup>1</sup>、山本 修平<sup>1</sup> (1. 筑波大学)

# 奨学金制度文書からの適格性判定フローの自動抽出

金谷 一輝<sup>†</sup> 宮森 恒<sup>†</sup>

<sup>†</sup> 京都産業大学情報理工学部情報理工学科 〒603-8555 京都府京都市北区上賀茂本山

E-mail: †{g2353351,miya}@cc.kyoto-su.ac.jp

**あらまし** 本研究では、奨学金制度文書から適格性判定フローを自動抽出における制度文書の適格性判定箇所を抽出する手法を提案する。各種支援制度における適格性判定は、現在も人手による文書解釈と判断に大きく依存しており、人手不足や判断の属人化といった課題がある。従来研究では、制度文書から条件要素を抽出する手法や、意思決定手順をルールベース表現へ変換する試みが行われてきたが、自然言語文書から完全な適格性判定フローを自動生成する手法は十分に検討されていない。本稿では、制度文書を対象として、受給資格要件を抽出し、表やフローチャートを構造化することに焦点を当てる。提案手法では、受給資格要件の抽出において VLM 単体を用いた場合と OCR を用いて文書のレイアウト情報を VLM に渡した場合を比較した結果、レイアウト情報を与えることで文書の抽出精度が向上することがわかった。

**キーワード** 制度文書, 適格性判定, 判断根拠, 意思決定支援, 大規模言語モデル, 行政手続き支援

## 1 はじめに

現在、我が国の行政サービスや社会保障制度の多くは、支援を必要とする個人自らが申し出ることによって初めて適用される申請主義に基づいている。しかし、これらの制度を規定する公的な募集要項やガイドラインは、法律用語などの専門的語彙や、複雑な条件分岐などを多用して記述されていることが多く、専門知識を持たない一般市民にとってその内容を正確に解読することは困難である場合も多い。その結果、本来は支援を受ける資格があるにもかかわらず、制度の複雑さゆえに自身の受給資格の有無を判断できずに申請を断念する事例や受給資格を満たす制度を自身で見つけることができないうえ、申請ができない事例などの申請漏れが課題となっている。

OpenFisca [1] などの Rules as Code (RaC) [4] の観点から制度の適格性判断を行うことのできるシステムは存在するものの、制度文書の適格性判定フローを手で構築するには、専門家による文書理解が必要であり、また作成に多大な時間を要する。加えて、受給要件などが変更された際にシステムの維持に課題が残る。そのため、適格性判定フローを自動構築できる手段が社会的に必要とされている。

本研究は、制度文書から適格性判定フローを自動構築するための基礎研究として位置づけられる。Nguyen らは [5] 手続き文書からフローチャートを自動生成するシステムを提案し、その構築過程が手順ステップの特定からフローチャートの組み立てまで 6 段階から構成されることを示している。本稿では、この過程の前提となる、制度文書からの受給資格要件の抽出に焦点を当てる。

制度文書は、フローチャートや表などの視覚的要素を含んでおり、視覚的に情報を処理することが必要である。従来、文書からの情報抽出には OCR (Optical Character Recognition) 技術が用いられてきた。近年、Vision-Language Model (VLM) が

画像とテキストを統合的に処理できる可能性を示している。しかし、日本語の制度文書における VLM 単体での文書処理と、レイアウト解析を経由した処理のどちらが効果的かは明らかでない。本稿では、VLM 単体による抽出と、レイアウト解析を経由した抽出を比較し、制度文書からの受給資格抽出におけるより効果的な手法を明らかにする。

本稿の構成は以下の通りである。第 2 節では関連研究について述べ、第 3 節では比較する 2 つの手法について詳述する。第 4 節では実験設定と結果を示し、第 5 節で本稿をまとめる。

## 2 関連研究

### 2.1 制度文書における情報処理

制度文書に含まれる受給資格要件には、条件分岐や場合分けなどの論理的構造があり、片山 [12] は、国民年金法の条文を述語論理により形式的に記述し、SMT ソルバー Z3Py を用いてその正しさを自動検証する手法を提案した。この研究は、被保険者の資格条件や年金の支給要件といった制度の規定が論理式として表現可能であることを示しており、制度文書に含まれる情報が構造的かつ形式的に扱えることを裏付けている。

また、法律文書を対象とした自然言語処理の取り組みとして、狩野ら [10] は国際コンペティション COLIEE を開催し、民法条文からの関連条文検索や含意関係認識といったタスクを設定している。参加システムには TF-IDF や BM25 などの情報検索手法に加え、BERT [2] などの深層学習手法が用いられており、法制度文書への NLP 技術の適用が進展していることを示している。

しかし、これらの研究はいずれもデジタル化済みのテキストデータを入力として前提としている。現実の制度文書は、行政機関から PDF 形式で配布されることが多く、受給資格の条件分岐をフローチャートで示したり、受給資格要件を表形式で整理したりするなど、視覚的なレイアウトによって情報を伝達す

る要素を含んでいる。このような視覚的要素に含まれる情報は、テキスト抽出のみでは分岐構造や要素間の関係性が失われるため、テキストベースの処理だけでは十分に捕捉できない。

本研究では、このような視覚的要素を含む制度文書を対象とし、文書画像からの受給資格要件の自動抽出を目的として、Vision-Language Model (VLM) および OCR を用いた手法の比較検討を行う。

## 2.2 Vision-Language Model を用いた文書理解

Shi ら [7] は GPT-4V [6] の OCR 能力を定量的に調査し、ラテン系以外の言語では認識精度が低下することを示している。また、緒方ら [11] は土木分野の日本語文書を対象にマルチモーダルモデルと OCR ツールの比較を行い、テキスト認識タスクにおいてマルチモーダルモデルが既存ツールに大きく劣ることを報告している。これらの知見は、日本語制度文書においても VLM 単体での処理に限界がある可能性を示唆しており、本研究で OCR を用いた場合との比較を行う動機となっている。また、文書のレイアウト情報とテキストを統合的に学習する LayoutLM [9] が提案されており、文書画像理解におけるレイアウト情報の有効性が示されている。しかし、日本語の制度文書への適用例は限定的であり、また VLM 単体では制度文書の文書構造の把握に限界がある可能性がある。

## 3 手 法

### 3.1 問題設定

本研究では、制度文書から受給資格要件を抽出するタスクを扱う。

入力: 制度文書の画像  $D = \{d_1, d_2, \dots, d_m\}$  ( $m$  はページ数)

制度文書は、大きく分けて、制度の説明、受給要件などの判断基準、申請手続きが記載されている。判断基準の記述形式は文書により異なり、文章のみで記述されている場合、フローチャートで記述されている場合、表を用いて記述されている場合などがある。本研究では、受給要件などの判断基準の抽出に焦点を当てる。

出力: 受給資格要件  $Y = (S, G)$

$S = \{s_1, s_2, \dots, s_m\}$  は文章および表形式で記述された受給資格要件の集合であり、 $G = (V, E)$  はフローチャートから抽出された構造である。 $V$  は要件ノードの集合、 $E \subseteq V \times V$  は分岐・遷移関係を表すエッジの集合である。

出力は、文書内の該当箇所を抜き出したものであり、表やフローチャートなどはその構造を反映した形式で記述する。

### 3.2 手法 A: VLM 単体による抽出

手法 A では、制度文書の画像を VLM に直接入力し、受給資格要件を抽出する。VLM は画像内のテキストおよび視覚的要素を統合的に処理し、受給資格要件に該当する箇所を選択的に抽出する。

#### 3.2.1 処理フロー

1. PDF 文書の各ページを PNG 画像に変換する
2. 各ページ画像とプロンプトを VLM に入力する

3. VLM が画像内から受給資格要件を抽出し、テキストとして出力する
4. 全ページの抽出結果を統合する

### 3.3 手法 B: レイアウト解析を利用した抽出

手法 B では、VLM への入力の前段階としてレイアウト解析を行い、文書の構造情報を明示的に付与した画像を VLM に入力する。これにより、VLM が文書のレイアウト構造を把握しやすくなることが期待される。

手法 A, B のシステム構造を図 2 に示す。

#### 3.3.1 処理フロー

1. PDF 文書を yomitoku [3] に入力し、レイアウト解析を行う
2. yomitoku がレイアウト情報（見出し、段落、表など）を抽出し、種別をアノテーションした画像を生成する
3. アノテーションした画像とプロンプトを VLM に入力する
4. VLM が画像内から受給資格要件を抽出し、テキストとして出力する
5. 全ページの抽出結果を統合する

図 1 に、教育訓練給付金の制度文書に対する yomitoku のレイアウト解析結果を示す。各構造要素に対してバウンディングボックスと要素種別が付与されていることが確認できる。

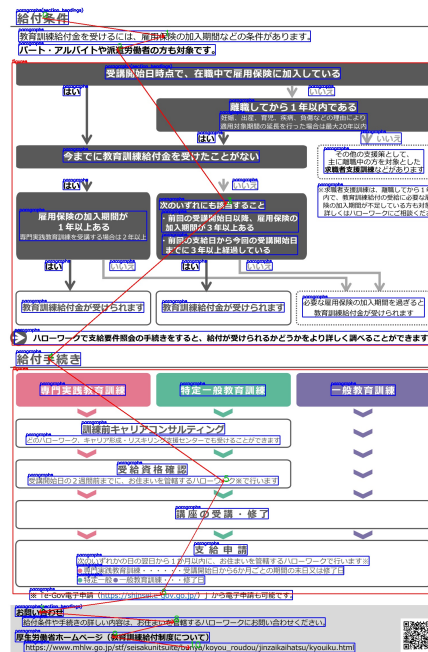


図 1 yomitoku によるレイアウト解析結果の例

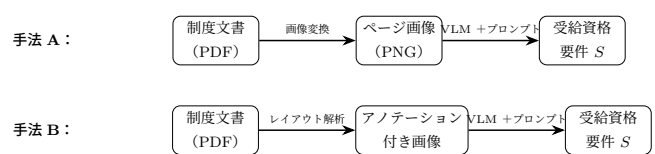


図 2 手法 A および手法 B の処理手順

## 4 実 験

### 4.1 実験設定

#### 4.1.1 データセット

本研究では、厚生労働省、日本年金機構、全国健康保険協会および JASSO が公開する pdf 形式の制度文書 9 件（1-40 ページ、中央値 3 ページ）を用いた。内訳はフローチャートを含むものが 2 件、表形式を含むものが 6 件、いずれも含まないものが 2 件である。

#### 4.1.2 実装環境

- VLM モデル: Qwen3-VL-30B [8]
- OCR ツール: yomitoku
- GPU: Quadro RTX 8000

### 4.2 評価指標

具体的には、以下の 2 点を評価対象として設定する。

- RQ1: 文書の該当箇所を正しく抽出できるか
- RQ2: フローチャートの構造が維持されるか

#### 4.2.1 RQ1 の評価指標

手法 A (VLM 単体) と手法 B (レイアウト解析+VLM) について、文章および表形式を含む文書からの抽出を以下の指標で評価した。ここで、正解としたものは適切に文章を抽出できたものに絞り、要約などの意味的な一致は不正解とする。

- **適合率 (Precision):** 抽出された内容が受給資格要件として正しいか
- **再現率 (Recall):** 受給資格要件の記述が漏れなく抽出されているか
- **F1 尺度**
- **表形式維持率:** 表が含まれる文書において、表の構造が保持されているか

#### 4.2.2 RQ2 の評価指標

手法 A (VLM 単体)、手法 B (レイアウト解析+VLM) について、フローチャートの構造維持を以下の指標で評価した。

- **フローチャートの構造維持:** ノードと分岐条件 (エッジ) が正しく抽出されているか
  - ノード抽出率: 正しく抽出できたノードの割合
  - エッジ抽出率: 正しく抽出できた遷移の割合

### 4.3 実験結果

#### 4.3.1 RQ1: 文章・表形式の抽出性能

表形式を含む 6 件を含めた 9 件の制度文書を用いて、手法 A と手法 B の抽出性能を比較した結果を表 1 に示す。

表 1 RQ1: 文章・表形式の抽出結果

手法	Precision	Recall	F1	表形式維持率
手法 A	29.69%	45.80%	36.03%	60.00%
手法 B	35.60%	47.24%	40.60%	71.56%

実験の結果、手法 A (VLM 単体) は、適合率が 29.69%、再現率が 45.80%、F1 尺度が 36.03%、表形式維持率が 60.00%であり、

表 2 RQ2: フローチャート構造の抽出結果

手法	ノード抽出率	エッジ抽出率
手法 A (VLM 単体)	77.78%	46.67%
手法 B (レイアウト解析+VLM)	58.30%	66.67%

手法 B (レイアウト解析+VLM) は、適合率が 35.60%、再現率が 47.24%、F1 尺度が 40.60%、表形式維持率が 71.56%であった。

この結果から、全ての評価指標において手法 B の方が上回り、特に抽出精度と表形式維持率においてその差は顕著であることがわかった。

#### 4.3.2 RQ2: フローチャート構造の抽出性能

フローチャートを含む 2 件の文書を用いて、2 つの手法の抽出性能を比較した結果を表 2 に示す。

実験の結果、手法 A (VLM 単体) は、ノード抽出率が 77.78%、エッジ抽出率が 46.67%であり、手法 B (レイアウト解析+VLM) は、ノード抽出率が 58.30%、エッジ抽出率が 66.67%、であった。

この結果からノード抽出率では手法 A の方が上回り、エッジ抽出率において手法 B の方が上回ることがわかった。

### 4.4 考 察

#### 4.4.1 レイアウト情報の付与による抽出精度の向上

RQ1 の結果から、レイアウト情報を付与した手法 B が全ての評価指標において手法 A を上回った。両手法の出力を比較すると、手法 A には以下の 2 つの誤りの傾向が確認された。

第一に、受給資格要件に該当しない箇所の誤抽出である。例えば、ある文書において手法 A はマイナンバーの提出手続きに関する説明文を「受給資格」「支給対象者」「支給条件」といった見出しとともに抽出した。これらは申請手続きの説明であり、受給資格要件には該当しない。第二に、同一の文章が異なる見出しのもとで繰り返し出力される現象が確認された。これらの誤りにより、手法 A では抽出精度が低下したと考えられる。

一方、手法 B ではこれらの誤抽出が抑制された。これは、レイアウト情報の付与により、VLM が文書内の各領域の役割（見出し、段落、表、手続き説明など）を区別できるようになり、受給資格要件に該当する箇所を選択的に抽出できたためと考えられる。また、表形式維持率においても手法 B が上回った（60.00% → 71.56%）のは、レイアウトアノテーションにより表領域が明示されたことで、VLM が表の構造を保持したまま抽出できたためと考えられる。

また、RQ2 の結果から、VLM にレイアウト情報を与えることでエッジ抽出率が上がった。これは VLM がフローチャートの内容を読み取る能力の向上を裏付けていると考えられる。一方、ノード抽出率においては手法 A が上回った。この結果は、レイアウトアノテーションがフローチャートの領域認識には寄与する一方で、VLM がフローチャートの内部構造を正確に読み取ることができないことを示唆している。レイアウト情報によりフローチャートの存在は明示されるが、ノード間の接続関係やテキスト内容の正確な抽出には、さらなる工夫が必要であると考えられる。

## 5 まとめと今後の課題

本研究では、制度文書からの受給資格要件の抽出において、VLM 単体による手法とレイアウト解析による構造情報を付与した手法の比較を行った。9 件の制度文書を用いた実験の結果、レイアウト情報の付与により適合率・再現率・表形式維持率のいずれにおいても改善が確認された。特に表形式の維持において効果が顕著であった。一方、フローチャートの抽出ではエッジ抽出率が改善した反面、ノード抽出率では逆の傾向が見られ、レイアウト情報の効果は抽出対象の性質によって異なることが示された。今後の課題として、データセットの拡充、受給資格要件およびフローチャート構造の抽出精度の改善、フローチャート以外の判断基準形式への対応が挙げられる。

## 謝 辞

本研究の一部は科研費 23K11342 の助成を受けたものである。

## 文 献

- [1] Openfisca: Introduction.
- [2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *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, 2019.
- [3] Kotaro Kinoshita. Yomitoku: Document ai engine specialized in japanese document image analysis.
- [4] James Mohun and Alex Roberts. Cracking the code: Rule-making for humans and machines. *OECD Working Papers on Public Governance*, No. 42, pp. 0\_1–109, 2020.
- [5] Nhat-Khiem Nguyen and Thanh-Tung Tran. Iuflowgen: An ai system for converting procedural texts into flowcharts. In *International Conference on Multi-disciplinary Trends in Artificial Intelligence*, pp. 222–234. Springer, 2025.
- [6] OpenAI. GPT-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.
- [7] Yongxin Shi, Dezhi Peng, Wenhui Liao, Zening Lin, Xinhong Chen, Chongyu Liu, Yuyi Zhang, and Lianwen Jin. Exploring ocr capabilities of gpt-4v (ision): A quantitative and in-depth evaluation. *arXiv preprint arXiv:2310.16809*, 2023.
- [8] Qwen Team. Qwen3 technical report, 2025.
- [9] Yiheng Xu, Minghao Li, Lei Cui, Shaohan Huang, Furu Wei, and Ming Zhou. Layoutlm: Pre-training of text and layout for document image understanding. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 1192–1200, 2020.
- [10] 狩野芳伸, 吉岡真治, 佐藤健ほか. Coliee: 法律文書の情報抽出および含意関係認識を行う国際コンペティション. 人工知能, Vol. 35, No. 3, pp. 377–384, 2020.
- [11] 緒方陸, 大久保順一, 藤井純一郎. 土木文書へのマルチモーダルモデル適用へ向けた調査. 人工知能学会全国大会論文集 第 38 回 (2024), pp. 3F1GS1001–3F1GS1001. 一般社団法人 人工知能学会, 2024.
- [12] 片山卓也. 国民年金法の述語論理による記述と検証 smt ソルバー z3py を用いたケーススタディ. コンピュータ ソフトウェア, Vol. 36, No. 3, pp. 3\_33–3\_46, 2019.

# 特徴量選択による日常的意思決定の個人化後悔予測モデル実装

乾 映之右<sup>†</sup> 石橋 直樹<sup>†</sup>

<sup>†</sup> 武蔵野大学データサイエンス学部 〒135-0063 東京都江東区有明3丁目3-3

E-mail: †s2322007@stu.musashino-u.ac.jp, ††n-ishi@musashino-u.ac.jp

あらまし 後悔は「選ばなかった選択肢」との比較による反実仮想的思考から生じ、損失回避バイアスにより人々はこれを強く避けようとする。従来の後悔理論研究は投資・医療・キャリアなど大規模な意思決定を対象とし、意思決定支援も事後の振り返りに留まってきた。しかし、日常の小さな意思決定は頻度が高く累積的影響が大きいにもかかわらず、予測的支援は未開拓である。本研究では、意思決定履歴とコンテキスト情報から SHAP 値による適応的特徴量選択を行い、ユーザーごとに異なる後悔要因を抽出する個人化予測モデルを提案する。仮想ユーザーを用いた実験により、提案手法はルールベース手法と比較して MAE を 39%改善し、ユーザー間で平均 23%の特徴量が異なることを確認した。また、提案モデルを Web アプリケーションとして実装し、実用可能性を示した。

キーワード 機械学習, SHAP, 意思決定支援

## 1 はじめに

### 1.1 背景

日常生活において、我々は無数の小さな意思決定を行う。ランチで何を食べるか、衝動買いをするべきか、SNS を見続けるべきか。これらの決定は一つ一つは小さいが、毎日数十回繰り返されることで累積的に生活の質に影響する。しかし、疲労やストレスにより判断力が低下している時、我々はしばしば後悔する選択をしてしまう。

例えば、「ストレスが溜まっている時に高額のランチを食べて後悔した」「疲れている時に衝動買いをして後悔した」といった経験は誰にでもあるだろう。このような後悔は、その瞬間の気分や状況により判断力が歪められることで発生する。もし意思決定の瞬間に「あなたは今この選択を後悔する可能性が高い」と警告されれば、より良い選択ができる可能性がある。

### 1.2 既存研究の限界

後悔理論 (Regret Theory) に関する既存研究 [1] [2] は、主に投資判断、医療選択、キャリア決定など、「人生を変える大きな決定」に焦点を当ててきた。これらの研究は重要だが、頻度の観点では日常の小さな意思決定の方が圧倒的に多い。一方で、日常の小さな意思決定は研究対象として見過ごされてきた。

また、Decision Journal などの既存システム [3] は事後の振り返りに留まり、意思決定時点での予測的な支援を提供していない。これらのシステムは「なぜ後悔したか」を理解する助けにはなるが、「これから後悔するか」を予測することはできない。さらに、既存の推薦システムは「何を好むか」を予測するが、「何を後悔するか」という負の側面は扱わない。

### 1.3 本研究の貢献

本研究では、以下の3つの独自性を持つシステムを提案する:

(1) 日常的意思決定への特化: 頻度の高い小規模決定に焦点を当て、気分・ストレス・空腹度など日常的コンテキストを

18 次元で記録

(2) 予測型アプローチ: 過去データから意思決定時点で後悔リスクを予測し、事前に警告

(3) 個人化された学習モデル: SHAP 値による適応的特徴量選択により、「あなたが後悔しやすい状況」を特定

## 2 関連研究

### 2.1 後悔理論

後悔 (Regret) は意思決定研究において重要な概念である。Zeelenberg ら [1] は、後悔が「選ばなかった選択肢」との比較 (Counterfactual Thinking) で生じることを実証的に示した。彼らの研究によれば、後悔は単なる失望とは異なり、「別の選択をしていれば」という反実仮想的な思考を伴う。

Kahneman and Tversky [2] のプロスペクト理論は、人間が利得よりも損失に敏感であること (損失回避バイアス) を説明する。この理論は、なぜ人々が後悔を避けようとするのかを理論的に裏付ける。本研究はこれらの理論を日常的意思決定に応用し、個人の後悔パターンを機械学習で学習する点で新規性がある。

従来の後悔理論研究は、主に大規模な意思決定 (投資、医療、キャリア) を対象としてきた。しかし、日常の小さな意思決定は頻度が高く、累積的な影響が大きい。本研究は、この研究ギャップを埋めることを目指す。

### 2.2 推薦システム

推薦システムは、ユーザーの好みを予測し、適切なアイテムを提示する技術である。協調フィルタリング [4] は、ユーザー間の類似性から好みを予測する。コンテンツベース推薦 [5] は、アイテムの特徴から好みを予測する。

しかし、これらのシステムは「何を好むか」という正の側面のみを扱う。本研究は「何を後悔するか」という負の側面を予測する点で根本的に異なる。後悔予測は、推薦とは逆の視点か

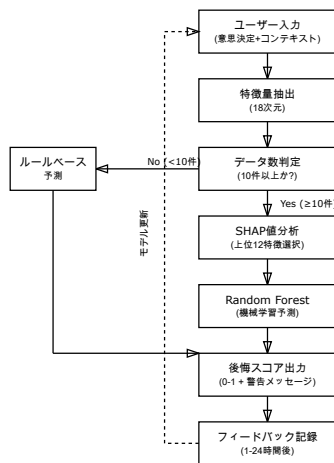


図 1 システム処理フロー

ら意思決定を支援する新しいアプローチである。

さらに、推薦システムは一般的にユーザーの状態（気分、ストレス）を考慮しない。本研究は、コンテキスト情報を重視し、「疲れている時は高額な選択を避ける」といった状態依存的な後悔パターンを捉える。

### 2.3 個人化機械学習

Few-shot learning [6] は、少数のサンプルから学習する手法である。メタ学習 [7] は、タスク間の共通構造を学習し、新しいタスクへの適応を高速化する。これらの手法は、データが少ない状況での個人化に有効である。

本研究は特徴量選択による個人化を採用する。SHAP 値を用いることで、ユーザーごとに重要な特徴量を自動的に発見し、解釈可能性を保ちながら個人化を実現する。このアプローチは、深層学習ベースの手法と比較して、少量データでも安定的に動作する。

## 3 提案システム

### 3.1 システムアーキテクチャ

RegretLens は以下の 3 層構造を持つ：

- フロントエンド (Jinja2)：UI 層
- バックエンド (Flask/Python)：ビジネスロジック層
- 機械学習エンジン：PostgreSQL + scikit-learn

図 1 にシステムの処理フローを示す。

### 3.2 データ収集

#### 3.2.1 意思決定記録

ユーザーは以下の情報を記録：

- 基本情報: カテゴリ（食事/買い物/娯楽/学習/仕事）、決定内容、代替案
- コンテキスト: 時刻, 曜日, 気分 (1-5), ストレスレベル (1-5), 空腹度 (1-5), 天気, 同伴者有無, 予算残高
- 決定要因: 価格, 期待度 (1-5), 健康価値 (1-5), 所要時間

#### 3.2.2 フィードバック

決定後、以下を記録：

- 後悔度 (1-5): 1=全く後悔なし, 5=とても後悔
- 満足度 (1-5)
- 後悔の理由（複数選択可）
- やり直したいか（はい/いいえ）

### 3.3 特徴量エンジニアリング

18 次元の特徴量を設計：

- 決定要因: price, taste\_expectation, health\_value, time\_required
- コンテキスト: mood\_score, stress\_level, hunger\_level, hour\_of\_day, is\_lunch\_time, is\_dinner\_time, day\_of\_week, weather\_encoded, with\_others, budget\_remaining
- 履歴特徴: user\_average\_regret\_this\_category, user\_regret\_variance, similar\_past\_decisions\_count, recent\_regret\_trend

### 3.4 ハイブリッド予測エンジン

データ量に応じて手法を切り替え：

- フィードバック < 10 件: ルールベース予測
- フィードバック ≥ 10 件: Random Forest 予測 (n\_estimators=50, max\_depth=10)

### 3.5 適応的特徴量選択

SHAP 値 [8] を用いて特徴量重要度を計算し、ユーザーごとに上位 12 特徴量を自動選択：

- (1) 訓練データで Random Forest を訓練
- (2) TreeExplainer で SHAP 値を計算
- (3) 重要度上位 12 特徴量を選択
- (4) 選択された特徴量のみで再訓練

### 3.6 説明可能性

予測結果に加えて以下を提示：

- 後悔スコア: 0-1 の範囲 (0=後悔なし, 1=確実に後悔)
- 警告メッセージ: 「あなたは疲れている時に高額な食事を選ぶと後悔しやすい (過去 5 回中 4 回)」
- 過去の類似ケース: 「2 週間前: 同じ店で大盛り → 後悔度 4」

図 2 にホーム画面, 図 3 に記録画面を示す。

## 4 実験

### 4.1 仮想ユーザーの設計

実ユーザーのバイアスを排除するため、心理学的に妥当な仮想ユーザーを生成した。4 つの典型的な後悔タイプを定義：ストレス敏感型 (30%, ストレスレベル ≥ 4 で後悔率急増), 価格敏感型 (30%, 価格 1000 円以上で後悔率上昇), 気分依存型 (20%, 気分スコア低下で後悔), ランダム型 (20%, 明確なパターンなし)。

### 4.2 実験 1: モデル性能比較

提案手法 (Random Forest + 適応的特徴量選択) が既存手

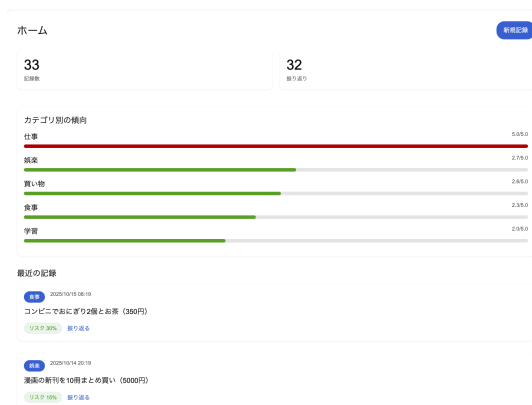


図 2 ホーム画面

The screenshot shows a form for recording a decision. It includes a 'カテゴリ' (Category) dropdown menu, a text area for '何をしましたか?' (What did you do?) with an example '例: ランチに1500円のパスタを注文した', and a '金額 (円)' (Amount) input field set to '0'. Below this is a section for '詳細情報を入力 (オプション)' (Enter detailed information (optional)) with sliders for '気分' (Mood) and 'ストレス' (Stress), and checkboxes for '曜日' (Day of the week) and '時刻' (Time of day).

図 3 意思決定記録画面

法より優れていることを示す。

データセット: 100 ユーザー × 100 決定 = 10,000 決定 (訓練 80 件/テスト 20 件, 時系列順守)

比較手法:

(1) ルールベースのみ: 既存手法 (後悔理論に基づく手動ルール)

(2) Random Forest (全 18 特徴量): ベースライン機械学習手法

(3) Random Forest + 適応的特徴量選択: 提案手法 (ユーザーごとに上位 12 特徴量を選択)

結果: 表 1 に各手法の MAE (Mean Absolute Error) を示す。

表 1 モデル性能比較 (実験 1)

手法	MAE	標準偏差	改善率
ルールベース	0.1365	0.0375	-
RF (全 18 特徴量)	0.0842	0.0191	38.3%
RF + 適応 (提案)	<b>0.0833</b>	0.0193	<b>39.0%</b>

提案手法が最も低い MAE (**0.0833**) を達成し、ルールベースと比較して **39%** の精度向上を実現した。また、全特徴量 RF と比較して 1.1% の改善を示し、統計的検定 ( $t$  統計量=3.1419,  $p=0.0022 < 0.01$ ) により 99% 信頼水準で有意差を確認した。適応的特徴量選択により、ユーザーごとに重要な特徴量が自動的に抽出され、予測精度が向上した。18 特徴量から 12 特徴量への削減 (33% 削減) でも性能を維持し、モデル解釈性の利点を示した。

#### 4.3 実験 2: 適応的特徴量選択の効果

個人化 (ユーザーごとに異なる特徴量を選択すること) の有効性を示す。

データセット: 50 ユーザー × 100 決定 = 5,000 決定 (訓練 80 件/テスト 20 件)

比較手法:

(1) 全 18 特徴量使用: すべてのユーザーで同じ 18 特徴量を使用

(2) 固定上位 12 特徴量: 全ユーザーの平均で重要度が高い 12 特徴量を使用 (個人化なし)

(3) 適応的 12 特徴量: ユーザーごとに SHAP 値で上位 12 特徴量を選択 (個人化あり, 提案手法)

結果: 表 2 に各手法の性能を示す。

表 2 特徴量選択の効果 (実験 2)

手法	MAE
全 18 特徴量	0.0860
固定 12 特徴量	0.0853
適応 12 特徴量 (提案)	<b>0.0857</b>

3 手法の性能差は小さく (MAE 差 0.0003~0.0007), いずれも実用的な精度を達成した。特徴量削減 (18 → 12) により性能をほぼ維持した。

個人化の実証として、平均 Jaccard 距離=0.2330 (ユーザー間で平均 23% の特徴量が異なる), ユニーク特徴量セット数=27/50 (54% が独自) を確認した。ユーザータイプ別の主要特徴量は、ストレス敏感型で stress\_level (15/15 ユーザー), 価格敏感型で price (15/15 ユーザー), 気分依存型で mood\_score (10/10 ユーザー) が選択された。適応的特徴量選択により、ユーザータイプに応じて異なる特徴量が自動的に選択されることを確認し、「万人に共通する後悔要因」ではなく「個人特有の後悔パターン」を捉える必要性を裏付けた。

#### 4.4 実験 3: ユーザータイプ別の予測精度

多様な後悔タイプのユーザーに対する適用可能性を検証する。100 ユーザー (ストレス敏感型 30 名, 価格敏感型 30 名, 気分依存型 20 名, ランダム型 20 名) × 100 決定で評価した。

明確なパターンを持つユーザー (ストレス敏感型・価格敏感型) では MAE 0.07~0.08 と非常に高精度を達成し、パターンのないランダム型でも MAE 0.1084 と実用的範囲に収まった。すべてのユーザータイプで実用的な精度 (MAE < 0.11)

表3 ユーザータイプ別の精度 (実験3)

ユーザータイプ	MAE	特徴
ストレス敏感型	<b>0.0712</b>	パターン明確
価格敏感型	<b>0.0770</b>	最も安定
気分依存型	0.0859	中程度
ランダム型	0.1084	パターンなし

を達成し、提案システムが多様なユーザーに適用可能であることを実証した。

#### 4.5 実験4: 時系列学習の効果

深層学習 (LSTM) が手動設計特徴量の Random Forest を上回るかを検証した。30 ユーザー × 200 決定 (シーケンス長 10) で評価した。

表4 時系列学習の比較 (実験4)

手法	MAE
<b>Random Forest</b>	<b>0.0782</b>
LSTM	0.1443
Simple RNN	0.1633

期待に反して、**Random Forest** が最良 (MAE 0.0782) を示し、LSTM は MAE 0.1443 と約 85% 悪化した。原因は、(1) データ量不足 (200 件は深層学習に不十分)、(2) 手動設計の集約特徴量 (recent\_regret\_trend 等) が時系列パターンを既に捉えている、(3) 過学習、である。本実験は、深層学習が常に優れているわけではなく、少量データでは適切な特徴量設計とシンプルなモデルが有効であることを示す重要な発見である。

## 5 考察

実験2により、ユーザーごとに重要な特徴量が異なることが確認された (Jaccard 距離 0.23)。例えば、ストレス敏感型では stress\_level、価格敏感型では price が最重要となり、「個人特有の後悔パターン」を捉える必要性を裏付けた。SHAP 値による解釈可能性により、「あなたはストレスレベルが4以上の時に後悔しやすい」といった具体的なフィードバックが可能となる。

実験4では、LSTM が Random Forest に劣る結果となったが、これは深層学習が常に最良とは限らないことを示す重要な発見である。データ量 200 件では不十分であり、手動設計の集約特徴量が時系列パターンを既に捉えていることが原因と考えられる。より大規模なデータが利用可能になれば、LSTM の優位性が現れる可能性がある。

## 6 おわりに

本研究では、日常的意思決定における後悔を予測するシステム RegretLens を提案した。SHAP 値による適応的特徴量選択により、ユーザーごとに異なる後悔要因を自動抽出する点が技術的な貢献である。

4つの実験により、以下を実証した：(1) 提案手法はルールベースと比較して MAE を 39% 改善 ( $p=0.0022$ )、(2) ユー

ザー間で 23% の特徴量が異なる (Jaccard 距離 0.23)、(3) 全ユーザータイプで実用的精度 (MAE 0.07-0.11)、(4) 深層学習 (LSTM) は少量データでは有効でない、ことを確認した。

学術的貢献として、(1) 日常的意思決定への後悔理論の応用、(2) 適応的特徴量選択による個人化手法、(3) 予測型意思決定支援の実現、がある。実用的には、モバイルアプリとしての展開が現実的であり、多様なユーザーに適用可能で、SHAP 値による解釈可能性で信頼性が高い。

今後の課題は、実ユーザーでの検証、より大規模データでの時系列学習の改善、RCT による介入効果の検証、他ドメインへの展開、長期的影響の評価である。

## 参考文献

- [1] M. Zeelenberg, et al., "The experience of regret and disappointment," *Cognition & Emotion*, Vol. 12, No. 2, pp. 221–230, 1998.
- [2] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," *Econometrica*, Vol. 47, No. 2, pp. 263–291, 1979.
- [3] Clearer Thinking, "Decision Journal," <https://www.clearerthinking.org/>
- [4] Y. Koren, et al., "Matrix factorization techniques for recommender systems," *Computer*, Vol. 42, No. 8, pp. 30–37, 2009.
- [5] M. J. Pazzani and D. Billsus, "Content-based recommendation systems," *The adaptive web*, pp. 325–341, 2007.
- [6] O. Vinyals, et al., "Matching networks for one shot learning," *Proc. NeurIPS*, 2016.
- [7] C. Finn, et al., "Model-agnostic meta-learning for fast adaptation," *Proc. ICML*, 2017.
- [8] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," *Proc. NeurIPS*, 2017.

# クラウドソーシングにおけるワーカーの応募行動の推定

桑原 士門<sup>†</sup> 伊藤 寛祥<sup>††</sup> 森嶋 厚行<sup>†††</sup>

<sup>†</sup> 筑波大学 情報学群 〒305-8550 茨城県つくば市春日 1-2

<sup>††</sup> 筑波大学 情報メディア系 〒305-8550 茨城県つくば市春日 1-2

E-mail: <sup>†</sup>s2211280@u.tsukuba.ac.jp, <sup>††</sup>ito@slis.tsukuba.ac.jp, <sup>†††</sup>morishima-office@ml.cc.tsukuba.ac.jp

**あらまし** 国内の労働人口減少に伴い、クラウドソーシングをはじめとしたプロジェクト型雇用の利用が急増している。しかしながら、既存の推薦システム研究は長期雇用を前提としたものが多く、プロジェクト型雇用特有のワーカーの応募行動は十分に解明されていない。本研究では、国内最大級のクラウドソーシングプラットフォームから提供された匿名化済み実データ（ジョブ数約7万件、ワーカー数約9,000名）を用い、ワーカーのジョブ応募行動を推定するモデルを構築し、どの属性が応募行動の予測に寄与するかを調査した。分析の結果、受注者であるワーカーの属性に関しては「最終提案からの経過日数」に代表される直近の活動頻度が最も強く応募行動に寄与し、発注者であるクライアントが投稿するジョブの属性に関しては「ジョブカテゴリ」および「公開からの経過日数（鮮度）」が重要であることがわかった。これらの知見は、プラットフォーム運営者やクライアント（発注者）が応募数を増加させるための実務的施策に資するものである。

**キーワード** クラウドソーシング, 応募行動推定, 機械学習, LightGBM, 特徴量重要度分析

## 1 序 論

### 1.1 研究背景

現在、日本国内においては少子高齢化に伴う構造的な労働人口不足が深刻な社会問題となっている [1]。この課題に対する解決策の一つとして、インターネットを通じて不特定多数の群衆（クラウド）に業務を委託する「クラウドソーシング」をはじめとした単発の仕事を受け負う新たな労働形態が急速に拡大している。

従来の雇用形態が企業への長期的な所属を前提としていたのに対し、これらの新しい働き方は、プロジェクト単位やタスク単位で労働契約が結ばれる「プロジェクト型雇用（ジョブ型雇用）」の性質を強く帯びている。この市場の拡大は、労働者に対しては時間や場所にとらわれない多様な働き方の選択肢を提供し、企業に対しては必要な時に必要な専門スキルを持つ人材を迅速に確保できるという、双方にとってのメリットをもたらしている。

### 1.2 研究の課題

しかしながら、クラウドソーシング市場の急成長に伴い、プラットフォーム上でのマッチングにおける非効率性が顕在化しつつある。膨大な数のジョブとワーカーが存在する中で、最適な組み合わせを効率的に見つけ出すことは容易ではない。

従来の HR 領域における推薦・マッチング技術の研究は蓄積されているが、その多くは長期雇用を前提とした求人サービスを対象としている。一方、クラウドソーシング領域、特にプロジェクト型雇用におけるワーカーの「応募行動」そのものを予測する研究は未だ発展途上である。

実際のプラットフォームにおいては、ジョブの閲覧数に対す

る応募率は必ずしも高くなく、特定の人気案件への応募の偏在や、逆に一件も応募が集まらない案件の発生など、プラットフォーム全体のマッチング効率を低下させる要因が散見される。

したがって、単にワーカーとジョブのスキル要件が一致しているかを見るだけでなく、ワーカーがそのジョブに対して実際に「応募する」という意思決定を行うか否かを高精度に推定する技術が求められている。

### 1.3 研究目的

以上の背景を踏まえ、本研究では、クラウドソーシングプラットフォームにおけるワーカーのジョブへの応募行動を予測するモデルを構築し、その決定要因を明らかにすることを目的とする。具体的には、以下の2つのリサーチクエスチョンを設定する。

RQ1: クラウドソーシングの実履歴データを用いてワーカー（応募者）の応募行動を推定するモデルを構築した場合、どのような属性（特徴量）がその予測に強く寄与するか？

RQ2: 得られた特徴量の重要度分析に基づき、プラットフォーム運営者やクライアント（発注者）の視点から、応募数を高めるためにどのような実務的示唆が得られるか？

なお、本研究は国内主要クラウドソーシングプラットフォームから提供された匿名化実データを用いた実証研究である。

### 1.4 本論文の構成

本論文は全6章から構成される。第1章では研究の背景と目的を述べた。第2章では関連研究を概説し、本研究の位置づけを示す。第3章では、提案する応募行動推定モデルの概要、使用するデータセット、特徴量エンジニアリングの手法について詳述する。第4章では評価実験の設定と結果を示し、特徴量重要度分析に基づいた考察を行う。第5章では本モデルの応用可能性と将来展望を論じ、第6章で結論と今後の課題をまとめる。

## 2 関連研究

### 2.1 クラウドソーシング市場とマッチング技術

クラウドソーシングは、Web を介して不特定多数の人材に業務をアウトソーシングする形態であり [2]、そのジョブの種類はシステム開発、デザイン、ライティングから単純作業まで多岐にわたる。Amazon Mechanical Turk に代表されるマイクロタスク型プラットフォームの登場以降、ワーカーの属性や行動パターンに関する実証研究が蓄積されてきた [3], [4]。多くのプラットフォームでは、依然としてキーワード検索やカテゴリ分類に基づくジョブ探索が主流 [6], [7] であり、個々のワーカーの過去の行動履歴や潜在的な選好を踏まえたパーソナライズされたマッチング最適化は、十分に体系化されているとは言い難い [5]。

オンライン労働市場に関する先行研究として、Horton [12] は Upwork などのフリーランスマーケットにおける市場設計と情報の非対称性について分析し、ワーカーの評判や過去の取引履歴が契約成立に与える影響を明らかにした。また、Kokkodis and Ipeirotis [13] は、フリーランサーの専門性の変遷と市場での成功要因を分析し、特定分野への特化と汎用スキルの獲得のトレードオフを示した。これらの研究は主に契約成立後の成果や評価に焦点を当てており、「応募するか否か」という意思決定プロセス自体を予測対象とした研究は限定的である。

### 2.2 HR 領域における推薦・予測システム

広義の HR 領域におけるデータ活用は進展しており、例えば求人票のテキスト情報から応募者数を推定する研究開発 [8] や、ユーザのクリック・応募などの行動ログを用いたリアルタイムな求人推薦システムの構築 [9] などが報告されている。また、クラウドソーシングに特化した研究として、ワーカーのスキルを推薦する試み [10] などもある。

Rendle et al. [16] が提案した BPR (Bayesian Personalized Ranking) は、ユーザの購買・クリックなどの暗黙的な行動データからユーザの選好を学習する手法であり、求人推薦にも応用可能である。また、協調フィルタリングに基づく行列分解 [14] や深層学習を用いた推薦手法 [15] も発展しているが、これらは主にユーザとアイテムの相互作用履歴が豊富な場合に有効である。クラウドソーシングでは同一ワーカーが短期間に多数の案件を比較検討するという特性があり、EC サイトの購買行動とは異なるダイナミクスを持つ。

しかしながら、従来の求人媒体を対象とした研究は、長期雇用を前提とすることが多く、クラウドソーシングのような短期・単発のプロジェクト型雇用における迅速な応募意思決定サイクルとは、その前提や決定要因が異なる可能性がある [11]。

### 2.3 本研究の位置づけ

本研究は、あるワーカーとあるジョブのペアに対して「ワーカーが応募するか否か」を直接推定する点に特徴がある。また、予測精度を追求した上で、機械学習モデルの特徴量重要度に基づいて応募行動の主要因を解釈し、実務に資する具体的な示唆(案件内容の改善指針や提示順の最適化など)へと落とし込む

ことを狙う。これは、学術的な理解を深めると同時に、プラットフォームの実運用における課題解決に直結するアプローチである。

## 3 応募行動推定モデルの提案

### 3.1 提案手法の概要

本研究では、クラウドソーシングにおけるマッチングの問題を、「あるワーカー(応募者)  $u$  が、あるクライアント(発注者)が投稿したジョブ  $i$  に対して応募するか否か」を予測する二値分類問題として定式化する。以降、本論文では仕事を受注する側を「ワーカー(応募者)」、仕事を発注する側を「クライアント(発注者)」と呼ぶ。

提案システムの概要を図 1 に示す。システムは、ワーカー(応募者)情報、ジョブ(発注情報)、および両者の過去の相互作用履歴を入力とする。

- 入力：ワーカー特徴量  $\mathbf{x}_u$  (応募者情報)、ジョブ特徴量  $\mathbf{x}_i$  (発注者情報・発注情報)、および相互作用特徴量  $\mathbf{x}_{ui}$
- 出力：応募予測  $\hat{y}_{ui} \in \{0, 1\}$  (0: 応募しない, 1: 応募する)
- ラベル：  $y_{ui} = 1$  (応募あり),  $y_{ui} = 0$  (応募なし)

本研究の学習アルゴリズムでは、ワーカー属性・ジョブ属性・テキスト由来特徴量などからなる多様なデータを学習できる柔軟性、学習・推論の高速性、および特徴量重要度に基づく寄与分析を通じたモデル判断の説明可能性を重視した。これらの要件を踏まえ、線形モデル(ロジスティック回帰)やアンサンブル学習(ランダムフォレスト [17])を含む複数の学習手法を比較検討した結果、勾配ブースティング決定木 [18] の高速実装である LightGBM [19] を採用した。各手法の詳細な性能比較については第 4.2.1 節で述べる。

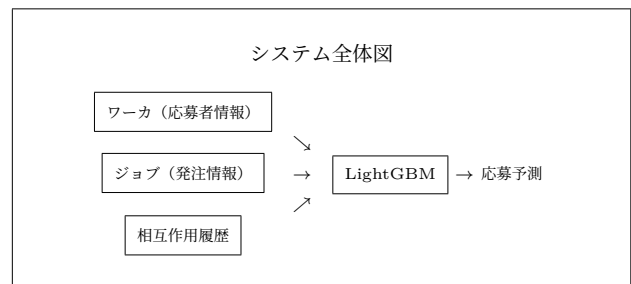


図 1 提案システムの全体図

### 3.2 データセット

本研究では、国内主要クラウドソーシングプラットフォームより提供された匿名化済みの実データを使用する。個人情報保護の観点から、特定の個人や特定の案件を識別し得る情報は一切扱わず、集計・分析結果は一般化した形で記述する。

対象期間： 2022 年 12 月～2023 年 3 月 (提供データ範囲)

データ規模： ジョブ数 約 70,000 件、ワーカー数 約 9,000 名

含まれる主な情報：

- ジョブ情報：カテゴリ、予算、公開日時、募集文のタイトル・本文等

- ワーカ情報：過去の応募・契約履歴，自己申告スキル等
- 応募情報：ワーカによる応募日時，応募結果（契約完了，辞退等）

### 3.3 データ前処理

提供されたデータは，基本的に「誰がどの案件に応募したか」という正例 ( $y = 1$ ) のペアが中心である．二値分類モデルの学習には，「応募しなかった」という負例 ( $y = 0$ ) のデータが必要となる．そこで，本研究では以下の手順で負例を生成した．

1. あるジョブ  $i$  に実際に応募したワーカ集合  $U_i^+$  を特定する．
2. 当該ジョブの募集期間中にプラットフォーム上で活動履歴（ログインや他案件への応募など）があるワーカ集合から， $U_i^+$  に含まれないワーカをランダムに選択し，（ジョブ  $i$ ，ワーカ  $u$ ）のペアを負例として追加する．
3. 負例の数は，ジョブごとに「正例数  $\times$  一定比率」で決定する．本研究では，正例 1 件に対し負例 1 件を生成する比率 1.0 を基本設定とした．この比率の選択については，第 4.2.2 節で複数の比率（0.5, 1.0, 2.0）を比較検証し，その妥当性を確認している．

この負例サンプリング手法は，観測されていないペアをすべて「非応募」と断定するのではなく，「当該期間に活動していたにもかかわらず応募しなかった」ペアを非応募の近似として扱うものである．ただし，この手法では「ジョブを閲覧した上で応募しなかった」のか，「そもそもジョブを閲覧していない（閲覧されていない）」のかを区別できないという限界がある．この点については第 4 章の考察で触れる．

### 3.4 特徴量エンジニアリング

ワーカの応募行動を決定づける要因を多角的に捉えるため，本研究では「ワーカ（応募者情報）」「ジョブ（発注者情報・発注情報）」「相互作用」の観点から特徴量を設計した．主な特徴量の一覧を表 1 に示す．

表 1 本研究で使用した主な特徴量一覧

カテゴリ	特徴量名	説明
ワーカ	worker_total_proposals	総提案数
ワーカ	worker_contract_rate	契約成功率
ワーカ	worker_total_earnings	総獲得金額
ワーカ	worker_days_since_last_prop.	最終提案からの経過日数
ワーカ	worker_proposals_last_90d	直近 90 日間の提案数
ワーカ	worker_contracts_last_90d	直近 90 日間の契約数
ワーカ	worker_skill_tfidf_sim.	スキルのコサイン類似度
ジョブ	client_total_jobs	発注者の過去ジョブ件数
ジョブ	client_median_budget	発注者の予算中央値
ジョブ	job_category	ジョブカテゴリ
ジョブ	job_budget_avg	平均予算
ジョブ	job_description_length	説明文の文字数
ジョブ	job_open_days	公開期間（日）
ジョブ	job_days_since_release	公開からの経過日数
ジョブ	job_title_lda_[0-15]	タイトルの LDA トピック
ジョブ	job_desc_lda_[0-31]	説明文の LDA トピック
相互作用	budget_fit_score	予算適合度
相互作用	client_repeat_worker_ratio	過去リピート率

ワーカ（応募者）の属性としては，過去の総提案数や契約率

などの実績だけでなく，直近の活動状況を示す指標として「最終提案からの経過日数（worker\_days\_since\_last\_proposal）」などを導入した．スキルに関しては，ジョブが求める必須スキルとの完全一致数に加え，表記ゆれや類似スキルを考慮するため，スキル記述テキストを TF-IDF [22] で算出したコサイン類似度（worker\_skill\_tfidf\_similarity）も特徴量とした．

ジョブ（発注者情報）としては，クライアントの過去のジョブ件数や予算のメディアンといったクライアント自身の特性を表す指標を設計した．また，ジョブ（発注情報）として，カテゴリや予算といった構造化データに加え，ジョブタイトルや説明文のテキストデータから Latent Dirichlet Allocation (LDA) [23] を用いてトピックモデリングを行い，各トピックへの所属確率（トピック分布）を特徴量（job\_title\_lda/job\_desc\_lda）として利用した．ジョブの鮮度を示す「公開からの経過日数（job\_days\_since\_release）」も重要な指標として加えた．

相互作用特徴量としては，ワーカ（応募者）とクライアント（発注者）の間の関係性を表すものであり，ワーカの過去の平均契約金額と当該ジョブの予算帯との適合度（budget\_fit\_score）などを設計した．

### 3.5 予測モデルの構築と評価方法

前述の通り，学習モデルとして LightGBM を採用した．主要なハイパーパラメータは，学習率 0.05，葉の数 64，サブサンプル比率 0.8，特徴量サンプル比率 0.8 とした．また，クラス不均衡に対処するため class\_weight を balanced に設定した．学習時には early stopping を適用し，検証用データでの AUC が 100 ラウンド連続で改善しない場合に学習を停止することで過学習を防いだ．

評価においては，モデルの識別性能を示す AUC [20] だけでなく，予測確率の正確性を示す Brier score，および上位推薦時の実用性を示す Precision@10 や Recall@50 も併せて用いる．なお，LightGBM の選定理由および他の学習手法との比較については，第 4.2.1 節で詳述する．

## 4 評価実験

### 4.1 実験設定

モデルの学習と評価において，未来の情報が学習データに漏洩すること（リーク）を防ぐため，応募日時に基づく時系列分割を行った．データ分割の概要を図 2 に示す．



図 2 時系列に基づくデータ分割の概要

訓練データ（Training）： 2023 年 1 月～2 月の応募データ

検証データ（Validation）： 2023 年 3 月前半の応募データ

テストデータ（Test）： 2023 年 3 月後半の応募データ

評価指標は以下の通りである．

AUC： 正例と負例を正しく分離・順位付けできるかの全体的な性能指標。

Precision@10： 各ワーカーに対して予測スコア上位 10 件のジョブを推薦した場合の平均応募率。推薦システムの品質を測る実用的な指標。

Recall@50： ワーカーが実際に応募した案件のうち、予測スコア上位 50 件に含まれていた割合。取りこぼしの少なさを評価する。

Brier score： 予測確率と実際のラベル (0 または 1) との平均二乗誤差。値が小さいほど確率予測の整合性が高いことを示す。

## 4.2 実験結果

各データセットにおける性能評価指標を表 2 に示す。

表 2 各データセットにおける性能評価指標

データセット	AUC	P@10	R@50	Brier
Training	0.914	0.377	0.997	0.137
Validation	0.845	0.446	0.999	0.154
Test	0.833	0.493	0.999	0.166

### 4.2.1 学習手法の比較

提案手法として採用した LightGBM の妥当性を検証するため、他の代表的な機械学習手法との比較を行った。比較対象として、線形モデルの代表であるロジスティック回帰、およびアンサンブル学習の代表であるランダムフォレストを選定した。各手法のハイパーパラメータ設定を表 3 に、テストデータにおける性能比較を表 4 に示す。

表 3 各学習手法のハイパーパラメータ設定

手法	主要なハイパーパラメータ
ロジスティック回帰	正則化強度 $C = 1.0$ , 正則化タイプ $L2$ , ソルバー $lbfgs$ , 最大反復回数 1000
ランダムフォレスト	決定木数 100, 最大深度 16, 最小分割サンプル数 2, 最小リーフサンプル数 1
LightGBM	学習率 0.05, 葉数 64, サブサンプル比 0.8, 特徴量サンプル比 0.8

表 4 学習手法間の性能比較 (Test データ)

手法	AUC	P@10	R@50	Brier
ロジスティック回帰	0.798	0.491	0.999	0.202
ランダムフォレスト	0.833	0.493	0.999	0.182
LightGBM	0.833	0.493	0.999	0.183

表 4 より、LightGBM はランダムフォレストと同等の AUC (0.833) を達成し、ロジスティック回帰 (0.798) を有意に上回った。この結果は、本タスクにおいて非線形な特徴量間の相互作用が重要であることを示唆している。LightGBM とランダムフォレストの性能が同等である一方、LightGBM は学習速度が高速であり、かつ特徴量重要度の解釈が容易であるため、本研究では LightGBM を採用した。

### 4.2.2 負例サンプリング比率の影響

第 3.3 節で述べた負例サンプリングにおいて、正例に対する負例の比率が予測性能に与える影響を検証した。比率 0.5, 1.0,

表 6 ワーカー (応募者) 特徴量の重要度ランキング (上位 7 項目)

特徴量名	重要度
worker_days_since_last_proposal	458
worker_total_proposals	201
worker_proposals_last_90days	88
worker_total_earnings	32
worker_contract_rate	32
worker_contracts_last_90days	12
worker_skill_tfidf_similarity	7

2.0 の 3 条件で実験を行い、結果を表 5 に示す。

表 5 負例サンプリング比率の検証 (Test データ)

負例比率	訓練データ数	正例率	AUC	P@10	Brier
0.5	884,475	0.656	0.834	0.669	0.197
1.0	1,208,796	0.480	0.833	0.493	0.183
2.0	1,837,382	0.316	0.833	0.326	0.157

表 5 より、AUC については負例比率によらずほぼ一定 (0.833 ~ 0.834) であり、モデルの識別性能は安定していることが確認された。一方、Precision@10 は比率 0.5 で最も高く (0.669)、比率の増加に伴い低下する傾向が見られた。これは、負例比率が大きいほど全体に占める負例の割合が増え、上位 10 件に正例が含まれる確率が相対的に減少するためと考えられる。Brier score は比率 2.0 で最も良好 (0.157) であり、比率が大きいほど確率予測の校正精度が向上する傾向が示された。

本研究では、識別性能 (AUC) と確率校正精度 (Brier) のバランスを考慮し、負例比率 1.0 を基本設定として採用した。ただし、実際の推薦システムへの適用時には、ユースケースに応じて比率を調整することで、Precision@10 (比率 0.5) や Brier score (比率 2.0) を優先した最適化も可能である。

表 2 より、Test データにおける AUC は 0.833 となり、未知のデータに対しても一定の高い識別性能を有していることが確認された。

実務的な推薦品質を示す Precision@10 は Test データで 0.493 であった。これは、モデルが「応募可能性が高い」と予測して上位 10 件に提示したジョブのうち、平均して約半数に対して実際にワーカーが応募したことを意味しており、推薦システムとして実用上十分な水準に達していると考えられる。Recall@50 も 0.999 と非常に高く、ワーカーが関心を持つ可能性のある案件のほとんどを上位 50 件以内に捕捉できていることがわかる。

一方で、Brier score は Test データで 0.166 と、Training (0.137) と比較して若干の悪化が見られた。これは、未知データに対して確率の過小評価や過大評価が一部残存している可能性を示唆しており、確率校正手法のさらなる改善が今後の課題となる。

### 4.3 特徴量重要度の分析と考察 (RQ1)

RQ1「どの属性が寄与するか」に答えるため、学習済み LightGBM モデルの特徴量重要度 (split importance: 決定木の分岐に寄与した回数に基づく重要度) を算出し、分析を行った。ワーカーに関する特徴量とジョブに関する特徴量の重要度ランキングをそれぞれ表 6, 表 7 に示す。

表7 ジョブ（発注情報）特徴量の重要度ランキング（上位10項目）

特徴量名	重要度
job_category	341
job_days_since_release	305
job_desc_lda_10	53
job_description_length	45
job_desc_lda_15	27
job_description_unique_ratio	22
job_desc_lda_6	21
job_desc_lda_11	18
job_desc_lda_25	18
job_open_days	15

表8 重要度上位のLDAトピックの解釈と実務的示唆

特徴量名	主要キーワード	解釈	効果的カテゴリ例
job_title_lda_11	継続依頼あり、初心者歓迎、在宅	安心感・継続性訴求型	データ入力、映像編集、ECサイト
job_desc_lda_10	報酬、応募条件、応募方法	応募手順明示型	データ入力、翻訳
job_desc_lda_15	納品、修正、デザイン	クリエイティブ案件型	英語翻訳、各種代行
job_desc_lda_6	仕事内容、主婦、学生	ターゲット層明示型	英語翻訳、発送作業
job_title_lda_1	簡単作業、データ入力、ボーナス	低ハードル型	データ入力、ウェブデザイン

表6、表7の結果から、以下のような知見が得られた。

第一に、ワーカー（応募者）の動的な活動状況が極めて強い影響力を持っている点である。表6において最も重要度が高かったのは「最終提案からの経過日数（worker\_days\_since\_last\_proposal, 重要度458）」であり、次いで「総提案数（worker\_total\_proposals, 201）」「直近90日間の提案数（worker\_proposals\_last\_90days, 88）」などが続いた。これは、ワーカーの静的なスキルセットよりも、「今現在、活発に仕事を探している状態にあるか」という直近の活動度合いが、応募行動を規定する最大の要因であることを示唆している。

第二に、ジョブ（発注情報）側の要因としては、ジョブのカテゴリと鮮度が重要である。表7によれば、「ジョブカテゴリ（job\_category, 341）」が最も高く、次いで「公開からの経過日数（job\_days\_since\_release, 305）」が続いた。ワーカーは自身の専門領域に合致するカテゴリを重視しており、かつ、公開されて間もない「新規案件」に優先的に応募する傾向が見て取れる。

第三に、テキスト情報の有用性が確認された点である。表7に示されるように、ジョブ説明文のLDAトピック特徴量（job\_desc\_ldaの各成分）が複数上位にランクインしている。本研究では、テキストデータの特徴量化にLatent Dirichlet Allocation (LDA) [23]を用いてトピックモデリングを行った。特徴量重要度上位にランクインしたLDAトピックの解釈を表8に示す。

これは、カテゴリや予算といった構造化データだけでは表現しきれない仕事の具体的な内容や文脈が、応募判断に影響を与

えていることを裏付けている。

一方で、TF-IDFを用いた「スキル類似度」は重要度7と一定の寄与を示した。これは、ワーカーが自己申告しているスキルとジョブのマッチ度よりも前述した項目の方が、応募のシグナルとして有効に機能している可能性を示唆している。

#### 4.4 実務的示唆についての考察（RQ2）

RQ2「応募数を高めるための示唆」について、重要度分析の結果に基づき、プラットフォーム運営者およびクライアントに対して以下の点が提案できる。

##### a) ジョブの鮮度維持と露出の最大化

「公開からの経過日数」の重要度が高いことから、ジョブは公開直後が最も応募を集めやすい「旬」の時期であるといえる。したがって、クライアントは募集開始直後の期間に、プラットフォーム上の広告機能などを活用して露出を最大化することが効果的である。また、長期間掲載しても応募が集まらない場合は、内容を見直した上で再投稿し、定期的な情報の更新をすることも一案である。

##### b) 適切なカテゴリ設定とターゲットへのアプローチ

「ジョブカテゴリ」の重要度が示す通り、ワーカーは自身の得意領域の案件を探している。クライアントは、依頼内容に最も合致するカテゴリを正確に選択する必要がある。プラットフォーム側は、活発なワーカー（直近の提案・活動履歴がある層）に対して、そのワーカーが好むカテゴリの新着案件を優先的に通知・推薦する仕組みを強化すべきである。

##### c) ジョブ説明文の具体化による情報補完

テキスト特徴量の寄与は、説明文の記述内容が応募判断を左右することを示している。クライアントは、具体的な作業内容、必要なスキルレベル、報酬体系などを明確かつ詳細に記述することで、ワーカーの不安を払拭し、応募を後押しできる可能性がある。

#### 4.4.1 テキスト改善による応募増加シミュレーション

上述のテキスト特徴量の重要性をより具体的に検証するため、LDAトピックの導入効果に関するシミュレーション分析を実施した。本シミュレーションでは、各ジョブカテゴリ内において、特定のLDAトピック比率が高いジョブ群（上位20%）と低いジョブ群（下位20%）の間で、1日あたりの平均応募数を比較した。

この分析により、当該トピックの特徴が強く表れた表現に改善した場合に期待される応募増加率を推定した。主要なカテゴリにおけるシミュレーション結果を表9に示す。

表9 LDAトピック導入による応募増加シミュレーション（カテゴリ別上位）

カテゴリ	トピック	主要キーワード	増加量	増加率
ECサイト制作	desc_lda_18	契約金額、ポートフォリオ、納期	+0.96 件/日	582%
HTML・CSS	desc_lda_18	契約金額、応募方法、納期	+3.04 件/日	197%
YouTube編集	desc_lda_5	質問歓迎、応募待ち	+0.56 件/日	132%
WordPress	desc_lda_18	依頼工程、提供素材	+1.54 件/日	81%
その他SNS	desc_lda_26	契約金額、受取金額	+0.30 件/日	75%

表9の結果から、以下の実務的示唆が得られた。

第一に、応募方法や契約条件の明示が効果的である。複数のカテゴリで高い効果を示した `job_desc_lda_18` は、「契約金額」「応募方法」「納期」「ポートフォリオ提示依頼」などのキーワードで特徴づけられる。これは、ワーカが応募を検討する際に、報酬条件や応募手順が明確に記載されているジョブを選好することを示唆している。クライアントは説明文において、これらの情報を漏れなく具体的に記述することで、応募数の増加が期待できる。

第二に、カテゴリによって効果的なトピックが異なる。例えば、YouTube 動画編集カテゴリでは `job_desc_lda_5`（「ご質問がありましたら気軽にお問い合わせください」「応募をお待ちしております」等）が最も効果的であり、ワーカへの歓迎姿勢を示す表現が重要である。一方、技術系カテゴリ（HTML・CSS, WordPress 等）では、依頼工程や提供素材といった案件の具体性を示すトピックが有効である。

第三に、シミュレーション上の増加率は最大で約6倍（ECサイト制作カテゴリ）に達する。ただし、これは相関関係に基づく推定値であり、テキスト改善のみで同等の効果が得られることを保証するものではない。実際の効果検証には、A/Bテスト等による因果推論的アプローチが必要である。

なお、本研究の限界として、第3.3節で述べた負例サンプリングの問題が挙げられる。本モデルは「閲覧した上での非応募」と「未閲覧による非応募」を区別できていない。そのため、例えば「重要度が高い特徴量（例：カテゴリ）」が適合していても応募に至らないケースの中には、単にそのジョブがワーカの目に触れていなかった（閲覧機会の不足）というケースが含まれている可能性がある。より厳密な分析を行うためには、ワーカのジョブ閲覧ログを取得し、閲覧を明示的に考慮したモデリングが必要となる。

## 5 応募行動モデルの応用と展望

### 5.1 クライアント（発注者）支援への応用

本研究で構築した応募行動推定モデルは、クライアントがジョブを作成する段階での支援に応用可能である。例えば、ジョブの下書き段階で入力された情報（カテゴリ、予算、タイトル、説明文など）に基づき、モデルが「この内容で公開した場合、どの程度の応募が見込めるか」をリアルタイムに推定・提示する機能が考えられる。もし推定された応募可能性が低い場合には、重要度分析の知見に基づき、「説明文をより具体的に記述する」「適切なカテゴリに変更する」「予算を上げる」といった改善案をシステムが自動的にフィードバックすることで、マッチングの成功率向上に寄与できる。

### 5.2 ワーカ（応募者）支援への応用

ワーカに対しては、パーソナライズされたジョブ推薦への直接的な応用が可能である。現在公開されている多数のジョブの中から、個々のワーカにとって応募する可能性が高いと予測される順に案件をランキング表示することで、ワーカが自分に適

した案件を探すための探索コストを大幅に削減できる。本実験の Test データにおける Precision@10 が約50%であったことは、上位に提示された案件の2件に1件は実際に応募に結びつく可能性を示しており、ワーカの機会損失を減らし、効率的な受注活動を支援する上で有効なアプローチとなると期待される。

### 5.3 プラットフォーム全体への波及効果

これらの応用を通じて、プラットフォーム全体で応募予測に基づくマッチングの最適化が進めば、クライアントは適切なスキルを持つワーカとより早期に出会うことができ、ワーカはより多くの適切な報酬機会を得ることができる。結果として、ミスマッチの減少、平均成約までの期間短縮、成約率の向上といったプラットフォーム全体の健全化と、全ユーザの体験向上という正の波及効果が期待される。

## 6 結論

### 6.1 本研究のまとめ

本研究では、クラウドソーシングプラットフォームにおけるマッチング効率化を目的として、実履歴データに基づきワーカのジョブへの応募行動を予測するモデルの構築を試みた。問題を二値分類として定式化し、LightGBMを用いて応募行動を推定した結果、時系列分割したテストデータにおいて AUC 0.833 という良好な識別性能を確認した。また、Precision@10 は 0.493 に達し、実用的な推薦システムとしての適用可能性が示された。

さらに、特徴量重要度分析を通じて応募行動の決定要因を考察し、ワーカの「直近の活動度」、ジョブの「カテゴリ適合性」および「鮮度（新規性）」が、静的なスキルマッチング以上に重要な要因であることを明らかにした。

### 6.2 本研究の貢献

本研究の学術的貢献は、従来の研究では十分に扱われてこなかった短期・プロジェクト型のクラウドソーシング領域において、「応募の有無」という具体的な行動を直接的な予測対象とし、大規模実データを用いた実証分析を通じてその実現可能性と主要な決定要因を提示した点にある。

実務的貢献としては、構築したモデルがレコメンデーションエンジンのコアとして利用可能であることを示すとともに、重要度分析の結果から、ジョブの鮮度管理やテキスト記述の改善といった、クライアントやプラットフォーム運営者が応募数を増やすために取り得る具体的な施策への示唆を与えた点が挙げられる。

### 6.3 今後の課題

今後の課題としては、以下の点が挙げられる。

第一に、閲覧情報の導入である。前述の通り、現在の負例サンプリングは閲覧を考慮できていない。閲覧ログを用いて「ジョブを詳細画面で閲覧したが、応募しなかった」ケースを明示的に負例として扱うことで、モデルの学習データとしての妥当性を高め、より精緻な予測と分析が可能になると考えられる。

第二に、多様なワーカ層への一般化可能性の検証である。本研究のデータは特定の4ヶ月間に活動したワーカを対象としている。新規登録直後のワーカや、長期休眠後に復帰したワーカなど、異なる活動パターンを持つ層に対しても同様の予測性能が得られるかは未検証である。より長期間にわたるデータを用いた検証や、ワーカのライフサイクルステージを考慮したモデルの拡張が求められる。また、近年の大規模言語モデル(LLM)の普及により、クライアントがジョブ文を作成する際のテキスト品質が向上している可能性がある。本研究のデータはLLM普及前の時期(2022年末~2023年初)のものであり、最新のジョブデータではテキスト特徴量の寄与パターンが変化している可能性があり、継続的なモデル更新が重要となる。

第三に、因果推論に基づく施策効果の定量化である。本研究では特徴量重要度に基づく要因分析を行ったが、これは相関関係の分析に留まる。例えば「ジョブの鮮度が重要」という知見から、実際に「早期露出を強化すれば応募が増える」という因果関係を立証するには、因果推論手法の適用が必要である。具体的には、傾向スコアマッチング(施策を受けたグループと受けていないグループの背景要因を揃えて比較する手法)や、差分の差分法(DID: 施策導入前後の変化を、施策対象群と非対象群で比較し、施策固有の効果を抽出する手法)などが考えられる。これらの手法を用いることで、プラットフォーム運営者が施策の投資対効果をより正確に見積もることが可能になる。

最後に、オンライン環境での評価である。本研究は過去データを用いたオフライン評価に留まっている。構築したモデルを実際のプラットフォームに適用し、A/Bテストなどを通じて、実際の応募数、成約率、ユーザの探索時間といった事業KPIに与える影響を検証することが、実用化に向けた最終的なステップとなる。

## 謝 辞

本研究を遂行するにあたり、多大なるご指導を賜りました森嶋厚行教授に深く感謝申し上げます。また、伊藤寛祥助教には、研究の進め方や実験手法について多くの助言をいただきました。さらに、本研究で使用したデータをご提供いただいたクラウドソーシングプラットフォーム運営企業様に心より御礼申し上げます。最後に、研究室のメンバーをはじめ、支えてくださった全ての方々に感謝いたします。

本研究の一部はJSPS科研費(22H00508)とJSTCREST(Grant Number JPMJCR22M2)の支援を受けました。ここに謝意を表します。

## 文 献

- [1] 総務省統計局: 労働力調査(基本集計)2024年(令和6年)平均結果, 2024.
- [2] J. Howe: The Rise of Crowdsourcing, *Wired Magazine*, Vol. 14, No. 6, pp. 176–183, 2006.
- [3] P. G. Ipeirotis: Demographics of Mechanical Turk, *CeDER Working Paper*, New York University, 2010.
- [4] D. Difallah, E. Filatova, and P. Ipeirotis: Demographics and Dynamics of Mechanical Turk Workers, in *Proceedings*

- of the 11th ACM International Conference on Web Search and Data Mining (WSDM '18), pp. 135–143, 2018.
- [5] A. Kittur, J. V. Nickerson, M. Bernstein, E. Gerber, A. Shaw, J. Zimmerman, M. Lease, and J. Horton: The Future of Crowd Work, in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13)*, pp. 1301–1318, 2013.
- [6] CrowdWorks 公式サイト, <https://crowdworks.jp/> (参照 2023-10-27).
- [7] Upwork 公式サイト, <https://www.upwork.com/> (参照 2023-10-27).
- [8] 産業技術総合研究所: 求人票から応募者数を高精度に予測するAIを開発, 産総研プレスリリース, 2021-03-02.
- [9] S. Toma, S. Koga, Y. Matsuda, T. Hara, and S. Nishio: A Practical Real-Time Job Recommendation System with Compact Models, in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM '18)*, pp. 1959–1962, 2018.
- [10] 大石啓登, 伊藤寛祥, 豊田正史, 森嶋厚行: プロジェクト型クラウドソーシングにおけるワーカのスキル推薦, 第16回データ工学と情報マネジメントに関するフォーラム (DEIM2024), 2024.
- [11] V. De Stefano: The Rise of the “Just-in-Time Workforce”: On-Demand Work, Crowdwork, and Labor Protection in the “Gig-Economy”, *Comparative Labor Law & Policy Journal*, Vol. 37, No. 3, pp. 471–504, 2016.
- [12] J. J. Horton: Online Labor Markets, in *Internet and Network Economics (WINE 2010)*, Lecture Notes in Computer Science, Vol. 6484, pp. 515–522, 2010.
- [13] M. Kokkodis and P. G. Ipeirotis: Reputation Transferability in Online Labor Markets, *Management Science*, Vol. 62, No. 6, pp. 1687–1706, 2016.
- [14] Y. Koren, R. Bell, and C. Volinsky: Matrix Factorization Techniques for Recommender Systems, *Computer*, Vol. 42, No. 8, pp. 30–37, 2009.
- [15] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua: Neural Collaborative Filtering, in *Proceedings of the 26th International Conference on World Wide Web (WWW '17)*, pp. 173–182, 2017.
- [16] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme: BPR: Bayesian Personalized Ranking from Implicit Feedback, in *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI '09)*, pp. 452–461, 2009.
- [17] L. Breiman: Random Forests, *Machine Learning*, Vol. 45, No. 1, pp. 5–32, 2001.
- [18] J. H. Friedman: Greedy Function Approximation: A Gradient Boosting Machine, *Annals of Statistics*, Vol. 29, No. 5, pp. 1189–1232, 2001.
- [19] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu: LightGBM: A Highly Efficient Gradient Boosting Decision Tree, in *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 30, pp. 3146–3154, 2017.
- [20] T. Fawcett: An Introduction to ROC Analysis, *Pattern Recognition Letters*, Vol. 27, No. 8, pp. 861–874, 2006.
- [21] S. M. Lundberg and S.-I. Lee: A Unified Approach to Interpreting Model Predictions, in *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 30, pp. 4765–4774, 2017.
- [22] J. Ramos: Using TF-IDF to Determine Word Relevance in Document Queries, in *Proceedings of the First Instructional Conference on Machine Learning*, pp. 133–142, 2003.
- [23] D. M. Blei, A. Y. Ng, and M. I. Jordan: Latent Dirichlet Allocation, *Journal of Machine Learning Research*, Vol. 3, pp. 993–1022, 2003.

# 異種グラフを用いた知識モニタリング推定とメタ認知適応型学習支援

李 根<sup>†</sup> 陳 莉<sup>††</sup> 唐 成<sup>†</sup> 馬 博軒<sup>†</sup> 出口 大輔<sup>†††</sup>

山下 隆義<sup>††††</sup> 島田 敬士<sup>†</sup>

<sup>†</sup>九州大学 〒819-0395 福岡県福岡市西区元岡 744 番地

<sup>††</sup>大阪教育大学 〒582-8582 大阪府柏原市旭ヶ丘 4 丁目 698 番 1 号

<sup>†††</sup>名古屋大学 〒464-8601 愛知県名古屋市千種区不老町

<sup>††††</sup>中部大学 〒487-8501 愛知県春日井市松本町 1200 番地

E-mail: <sup>†</sup> {gen.li,tang,atsushi}@limu.ait.kyushu-u.ac.jp, boxuan@artsci.kyushu-u.ac.jp, <sup>††</sup> chen-l68@cc.osaka-kyoiku.ac.jp, <sup>†††</sup> ddeguchi@nagoya-u.ac.jp, <sup>††††</sup> takayoshi@isc.chubu.ac.jp

**あらまし** 従来の学習支援システムは主にテストログから推定される知識レベルに基づいて支援を行うが、学習者が自身の理解度を適切に判断できるかという知識モニタリング能力の観点には十分に考慮されていない。知識モニタリングとは自身の理解状態を客観的に判断する能力であり、学習の質や効率に大きく影響するため、知識レベルと併せて考慮した支援が重要である。本研究では、テストログと学習リフレクションを統合し、学習者の潜在的な知識モニタリング状態を推定する手法を提案する。学習リフレクションでは全ての知識概念が言及されるとは限らないため、学習者・知識概念・概念間関係を異種グラフとして構築し、欠損する学習者と概念関係の補完をリンク予測問題として定式化する。さらに、推定結果に基づき学習者を知識レベルと知識モニタリング能力の二軸で類型化し、Feed Up・Feed Back・Feed Forward の三段階に従ったメタ認知適応型フィードバックを提示するシステムを実装した。実データを用いた推定精度の検証およびユーザ実験によるシステムの評価を通じて、提案手法の有効性と教育実践への適用可能性を示す。

**キーワード** 異種グラフ, 知識モニタリング, 学習支援, LLM

## 1 はじめに

効果的な学習支援の実現には、学習者が自身の理解状態を正確に把握し、適切な学習方略を選択する能力が不可欠である [31]。近年、学習管理システム (LMS) や知的学習支援システム (ITS) の普及により、テストログや行動ログから学習者の知識レベルを推定し、個別化された支援を提供することが可能になっている [17], [23]。これらのシステムは主に学習者が「何を知っているか」を評価し、不足領域に対する学習教材の推薦などにより一定の成果を上げている [4]。

しかしながら、知識レベルの評価のみでは効果的な学習支援には不十分な場合がある。教育現場では、実際には理解していないにも関わらず「わかったつもり」になる過信や、十分な知識があるにも関わらず自信を持ってない過小評価が頻繁に観察される [14], [18]。このような現象は、学習者の知識モニタリング能力、すなわち自身の理解状態を正確に判断するメタ認知能力の不足に起因する [3]。したがって、学習支援システムは知識レベルだけでなく、知識モニタリング能力も考慮した支援を提供することが重要である。

従来、知識モニタリング能力の評価には、各問題に対する自信度を回答させる構造化質問紙が用いられてきた [25]。しかし、この手法は評価項目を事前に固定する必要があり、多様なカリキュラムへの柔軟な適用が困難であった。一方、自由記述式の

学習リフレクションは、自己調整学習 (SRL) における自己省察フェーズの中核的实践として多くの教育機関で実施されている [26]。学習者が能動的に自らの理解状態を言語化するこの形式は、事前定義された選択肢への受動的な回答と比較して、より真正なメタ認知プロセスを反映できる。また、評価項目が固定されないため、カリキュラム変更にも柔軟に対応可能である。しかし、学習者が評価対象となる全ての知識概念について言及するとは限らないため、言及されなかった概念に対する認知状態の推論が課題となる。

本研究では、テストログと学習リフレクションを統合し、学習者の知識レベルと知識モニタリング能力を同時に評価するフレームワークを提案する。大規模言語モデル (LLM) を用いて教材と自己報告テキストから知識エンティティを抽出し、学習者・知識概念・概念間関係からなる異種グラフを構築する。学習者が言及しなかった概念に対する認知状態の推定は、異種グラフニューラルネットワーク (HGNN) によるリンク予測問題として定式化する。さらに、推定結果に基づき学習者を知識レベルと知識モニタリング能力の二軸で5つのパターンに類型化し、Hattie & Timperley のフィードバックモデル [11] に基づく Feed Up・Feed Back・Feed Forward の三段階適応型フィードバックを提供するシステムを実装した。

本研究では、以下の研究課題に取り組む：

- **RQ1:** 異種グラフに基づくリンク予測は、学習者の潜在的な知識モニタリング状態をどの程度正確に推定できるか？

- **RQ2:** 知識モニタリング能力に基づく適応型フィードバックは、学習者と教育者にどのように評価されるか？

本論文の貢献は以下の通りである：(1) 自由記述式の学習リフレクションから知識モニタリング能力を推定する異種グラフベースの手法、(2) 知識レベルと知識モニタリング能力を統合した学習者類型化に基づく適応型フィードバックシステムの設計と実装、(3) 実データによる推定精度の検証およびユーザ実験によるシステム評価

## 2 関連研究

### 2.1 知識モニタリング評価

知識モニタリングとは、自身の知識状態を正確に判断するメタ認知能力であり、効果的な学習に不可欠な要素である。初期の研究では、学習者の自信判断と実際のパフォーマンスを比較することでメタ認知的校正を評価する構造化パラダイムが確立された [27]。これらの研究は、自己調整学習における校正の役割を強調し、メタ認知的制御やタスク解釈を含む理論的拡張を提案した [24]。

知識モニタリング能力の定量化には、信号検出理論に基づく指標が広く用いられている [2], [20]。学習者の認識（理解している/していない）と実際のパフォーマンス（正答/誤答）の対応関係から、誤警報率、見逃し率、および弁別力  $d'$  などの指標が算出される。これらの指標により、単なる正答率では捉えられない学習者の自己認識の正確さを評価できる。しかし、これらの手法は事前定義された閉鎖的な回答形式に依存しており、学習者の表現の柔軟性が制限されるという課題がある。

### 2.2 教育分野における LLM 活用

教育分野において、LLM の重要な応用の一つは、教育コンテンツからの知識構造抽出である。LLM は教材から知識概念を抽出し [22]、知識グラフと組み合わせることで非構造化コンテンツを構造化表現に変換できる [5], [6], [15], [21]。このグラフベースの文脈モデリングは、教育応用における概念間の依存関係を捉える機会を提供する。また、プログラミングバグへの対応 [9]、自動評価 [12] などのタスクにも活用されている。さらに、学習者へのフィードバック生成にも LLM が活用されており [13], [28]、適応型学習支援の実現に貢献している。

### 2.3 グラフベースの学習モデリング

グラフベースのモデリングは、教育コンテンツと学習者行動の関係性を表現する有効な手法である。教育知識グラフは前提関係や概念階層を符号化し [19]、その構築手法はニューラルシーケンスラベリングから確率的アソシエーションマイニングまで多岐にわたる [7]。近年では、強化学習を用いた解釈可能な推薦パスの生成 [10] や、RDF ベースの表現 [1] も提案されている。

グラフニューラルネットワーク (GNN) は、演習問題と概念間の潜在的依存関係を捉えることができる [16], [30]。GKT-CD [29] のようなアプローチは、学習者-概念インタラクションを通じて習熟度を推論する。また、教育理論に基づくグラフベースフレームワークにより認知プロセスのモデル化も試みら

れている [8]。しかし、既存のグラフベースモデルは知識モニタリング評価を十分に扱っていない。

## 3 提案手法

本章では、テストログと学習リフレクションを統合し、学習者の知識モニタリング状態を推定するフレームワークを提案する。図 1 に提案フレームワークの概要を示す。

### 3.1 問題設定

学習者集合  $S = \{s_1, \dots, s_n\}$  がコースに参加している教育環境を考える。 $\mathcal{K}$  をコースで扱われる知識概念（概念、理論、原理など）の集合とする。学習者は以下の 2 つの活動に従事する：(1) 自由記述式の学習リフレクションにおいて、自身の理解状態を振り返り自由に記述する中で、理解していると認識する知識概念 ( $\mathcal{K}_i^+$ ) と理解していないと認識する知識概念 ( $\mathcal{K}_i^-$ ) に言及する；(2) 特定の知識概念を評価するための評価項目に回答する。各評価項目  $q_j \in \mathcal{Q}$  は主に知識概念  $k_j \in \mathcal{K}$  を評価し、学習者のパフォーマンスは正誤の二値として記録される。

しかしながら、課題となるのは評価される知識概念の集合  $\mathcal{K}_Q = \{k_j | q_j \in \mathcal{Q}\}$  が、学習者が明示的に言及した知識概念  $\mathcal{K}_i^+ \cup \mathcal{K}_i^-$  に完全には含まれない場合がある。すなわち、一部の学習者において  $\mathcal{K}_Q \not\subseteq (\mathcal{K}_i^+ \cup \mathcal{K}_i^-)$  となる。このような「言及されなかった概念」に対する認識状態を潜在的認識状態 (Latent Perceived State) と呼ぶ。潜在的認識状態が存在する場合、テスト結果との対応付けが不完全となり、知識モニタリング能力の正確な評価が困難となる。本研究の目標は：(1) 潜在的認識状態を推定し、各学習者の知識モニタリング状態を補完すること；(2) 推定結果に基づく学習者類型化と適応型フィードバックの有効性を検証すること。

### 3.2 知識グラフの構築

#### 3.2.1 知識概念の抽出と関係同定

LLM を用いて、コース教材から知識概念とその間の階層的・意味的關係を抽出し、知識グラフ  $G = (\mathcal{K}, \mathcal{E}_{\mathcal{K}-\mathcal{K}})$  を構築する。ここで  $\mathcal{K}$  は知識概念の集合、 $\mathcal{E}_{\mathcal{K}-\mathcal{K}}$  は概念間の関係（前提関係、包含関係など）を表すエッジ集合である。

#### 3.2.2 評価項目と知識概念のマッピング

各評価項目  $q_j$  とそれが主に評価する知識概念  $k_j$  を関連付けるため、LLM を用いてマッピング  $M_Q : \mathcal{Q} \rightarrow \mathcal{K}$  を構築する。これにより、各学習者について正答した概念集合と誤答した概念集合を特定できる。

#### 3.2.3 明示的認識概念の抽出

各学習者  $s_i$  の学習リフレクションに対して LLM を適用し、言及された知識概念を抽出する。抽出された概念は、記述の文脈に基づき  $\mathcal{K}_i^+$ （「理解している」と認識）または  $\mathcal{K}_i^-$ （「理解していない」と認識）に分類される。

### 3.3 HGNN に基づくリンク予測

#### 3.3.1 異種学習者グラフの構築

構築したコンテキスト知識グラフから、異種学習者グラフ

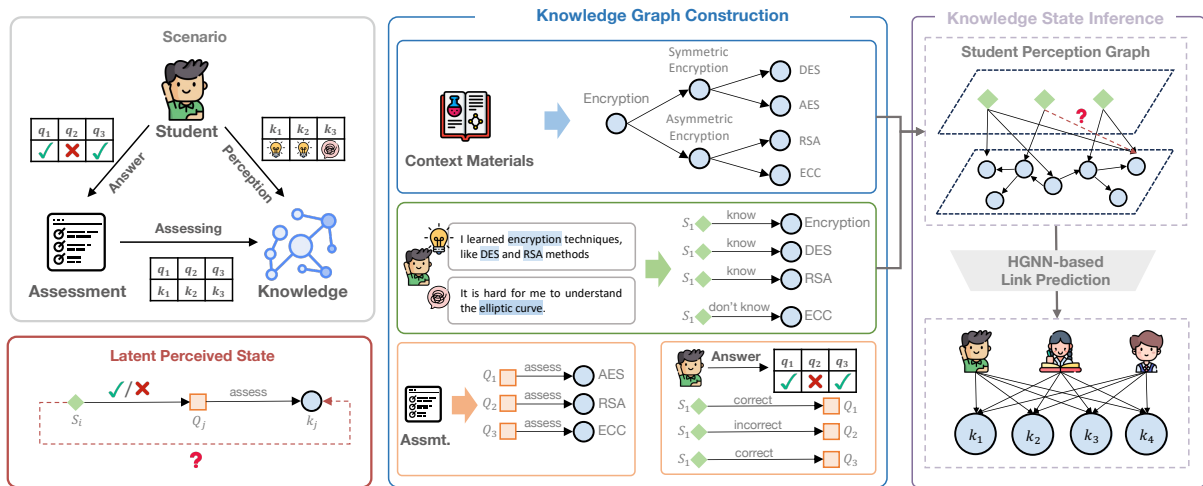


図 1: 提案フレームワークの概要. 左部のシナリオに示すように, 学習者はテストへの回答と学習リフレクションの記述という 2 つの活動に従事する. 中央部の知識グラフ構築では, LLM を用いて教材から知識概念の階層構造を抽出し, 学習リフレクションから学習者の認識状態(「理解している」/「理解していない」)を抽出する. 右部の知識状態推論では, 学習者が明示的に言及しなかった概念(図中の?マーク)に対する認識状態を, HGNN によるリンク予測として推定する.

$G_H = (V, E)$  を構築する. このグラフは学習者ノード  $V_S = S$  と知識概念ノード  $V_K = \mathcal{K}$  を含む. エッジ集合  $E$  は以下の 3 種類から構成される:

- **学習者-概念エッジ  $\mathcal{E}_{S-K}$** : 学習者と, その学習者が「理解している」と認識した知識概念 ( $\mathcal{K}_i^+$ ) を接続する.
- **概念-概念エッジ  $\mathcal{E}_{K-K}$** : 知識概念間の意味的・階層的關係を表現する.
- **概念-学習者エッジ  $\mathcal{E}_{K-S}$** : 双方向メッセージパッシングのための逆エッジ.

重要な点として, 「理解していない」と認識した概念 ( $\mathcal{K}_i^-$ ) へのエッジはこの段階ではグラフに含めない. この設計により, 潜在的認識状態の推定は「学習者-概念間のエッジが存在するか否か」を予測するリンク予測問題として定式化される. すなわち, エッジの存在は「理解している」という認識を, エッジの不在は「理解していない」という認識を表す.

### 3.3.2 異種グラフニューラルネットワークによるリンク予測

潜在的認識状態を推論するため, HGNN を用いたリンク予測を行う. HGNN は異種情報をタイプ別メッセージパッシング機構により処理する:

$$h_v^{(l+1)} = \text{Aggregate}_{\tau(v)}^{(l)} \left( h_v^{(l)}, \{h_u^{(l)}, r : u \in N_r(v), \forall r \in R\} \right) \quad (1)$$

ここで,  $\tau(v)$  はノード  $v$  の種類,  $N_r(v)$  は関係タイプ  $r$  で接続された  $v$  の近傍ノード,  $R = \{\mathcal{E}_{S-K}, \mathcal{E}_{K-K}, \mathcal{E}_{K-S}\}$  は全関係タイプ集合を表す.

ノード埋め込み  $h_s$  (学習者) と  $h_k$  (知識概念) を計算後, 学習者  $s$  が知識概念  $k$  を「理解している」と認識している確率を  $\hat{y}_{s,k} = \sigma(f(h_s, h_k))$  として予測する. モデルは二値交差エントロピー損失を最小化するように学習される.

### 3.3.3 明示的情報に基づく負例サンプリング (EINS)

効果的なリンク予測モデルの学習のため, 明示的情報に基づ

く負例サンプリング (EINS: Explicit-Informed Negative Sampling) 方式を設計する. 前述の通り,  $\mathcal{K}_i^-$  (「理解していない」と認識した概念) はグラフのエッジとして含まれていない. EINS はこの情報を負例として積極的に活用する:

1. **明示的負例**:  $\mathcal{K}_i^-$  から直接サンプリング. 学習者が明示的に「理解していない」と報告した概念である.
2. **暗黙的負例**: 潜在的認識状態(言及されなかった概念)からランダムサンプリング(比率パラメータ  $\rho$  で制御).

学習者の明示的な「理解していない」という報告を負例として優先的に活用することで, 学習者の自己認識の境界をより正確に学習できる.

### 3.3.4 潜在的認識状態の推論

学習済みモデルを用いて, 明示的に言及されなかった知識概念  $k_j \notin (\mathcal{K}_i^+ \cup \mathcal{K}_i^-)$  に対する学習者  $s_i$  の潜在的認識状態を, 予測リンク確率に基づき推論する:

$$k_j \in \begin{cases} \hat{\mathcal{K}}_i^+ & \text{if } \hat{y}_{s_i, k_j} \geq \theta \\ \hat{\mathcal{K}}_i^- & \text{if } \hat{y}_{s_i, k_j} < \theta \end{cases} \quad (2)$$

ここで  $\theta$  は閾値パラメータである. 推論の結果, 学習者の認識状態は  $\mathcal{K}_i^+ \cup \hat{\mathcal{K}}_i^+$  (理解認識) と  $\mathcal{K}_i^- \cup \hat{\mathcal{K}}_i^-$  (非理解認識) として補充される.

## 3.4 学習者類型化

### 3.4.1 知識モニタリング評価のためのカテゴリ分類

推論により補充された認識状態と実際のテスト結果を比較し, 知識モニタリング能力を評価する. 各知識概念を, 認識状態(「理解している」/「理解していない」)と実際の正誤の組み合わせにより, 以下の 4 カテゴリに分類する(表 1):

各カテゴリの解釈は以下の通りである:

- **ヒット (A)**: 理解していると認識し, 実際に正答
- **フォルスアラーム (B)**: 理解していると認識するが, 実際は誤答

表 1: 知識モニタリング評価のためのカテゴリ分類

	正答	誤答
「理解している」と認識	ヒット (A)	フォルスアラーム (B)
「理解していない」と認識	ミス (C)	正棄却 (D)

- ミス (C) : 理解していないと認識するが、実際は正答
- 正棄却 (D) : 理解していないと認識し、実際に誤答

### 3.4.2 知識モニタリング能力の定量化

信号検出理論に基づき、各学習者の知識モニタリング能力を定量化する。表 1 の 4 カテゴリから、以下の指標を導出する。

#### a) 弁別力 ( $d'$ )

学習者が「理解している」状態と「理解していない」状態を正確に区別できる能力を表す総合指標である：

$$d'_i = z\left(\frac{|A_i|}{|A_i| + |C_i|}\right) - z\left(\frac{|B_i|}{|B_i| + |D_i|}\right) \quad (3)$$

ここで  $z(\cdot)$  は標準正規分布の逆累積分布関数である。 $d'$  が高いほど、学習者は自身の理解状態を正確に弁別できている。

#### b) 誤警報率 (False Alarm Rate)

誤答した概念のうち、「理解している」と誤認した割合を表す：

$$FAR_i = \frac{|B_i|}{|B_i| + |D_i|} \quad (4)$$

この値が高い学習者は、実際には理解していない概念を「理解している」と認識する過信傾向を示す。

#### c) 見逃し率 (Miss Rate)

正答した概念のうち、「理解していない」と誤認した割合を表す：

$$Miss_i = \frac{|C_i|}{|A_i| + |C_i|} \quad (5)$$

この値が高い学習者は、実際には理解している概念を「理解していない」と認識する過小評価傾向を示す。

これらの指標により、知識モニタリング能力の全体的な弁別力 ( $d'$ ) だけでなく、その内訳として過信傾向 (FAR) と過小評価傾向 (Miss) を把握できる。

### 3.4.3 学習者の 5 パターン類型化

学習者を知識レベル (テスト正答率) と知識モニタリング能力に基づき、以下の 5 パターンに分類する (表 2)：

表 2: 学習者の 5 パターン類型化

パターン	知識レベル	弁別力	追加条件	特徴
校正良好	高	高	—	理想的な状態
限界認識	低	高	—	不足を正確に認識
自信不足	高	低	見逃し率 高	能力を過小評価
自信過剰	低	低	誤警報率 高	理解したつもり
イエス傾向	高	低	見逃し率 低	常に『理解している』と回答

## 3.5 適応型フィードバック

学習者類型化の結果に基づき、個別最適化されたフィードバックを提供するための方略を設計する。

### 3.5.1 3段階フィードバックモデル

本研究では、Hattie & Timperley [11] が提唱した 3 段階フィードバックモデルを採用する。このモデルは、効果的なフィードバックが以下の 3 つの問いに答えるべきことを示している：

- **Feed Up** : 目標は何か (Where am I going?)
- **Feed Back** : 現状はどうか (How am I going?)
- **Feed Forward** : 次に何をすべきか (Where to next?)

本研究では、このモデルを知識レベルと知識モニタリング能力の二軸フレームワークに適用し、各学習者類型に応じた方略を定義する。

### 3.5.2 類型別フィードバック方針

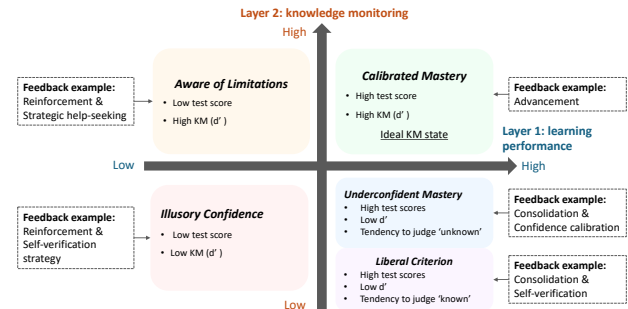


図 2: 学習者類型とフィードバック方略の対応

図 2 に、各類型に対応するフィードバック方略の概要を示す。知識レベルの改善に対しては、低成績の学習者には基礎概念の復習と定着を促す**強化学習**を、高成績の学習者には習得済み知識の発展・応用を促す**定着学習**を提供する。知識モニタリングの改善に対しては、類型別の方略を提供する。自信過剰の学習者には、理解を「既知」と判断する前に自己説明や回答確認などの**自己検証方略**を促す。自信不足の学習者には、正答した概念を振り返り成功体験を認識させる**自信校正**を促す。限界認識の学習者には、不明な概念に対する質問や教材参照などの**戦略的援助要請**を促す。

## 4 学習支援ダッシュボードの実装

### 4.1 概要

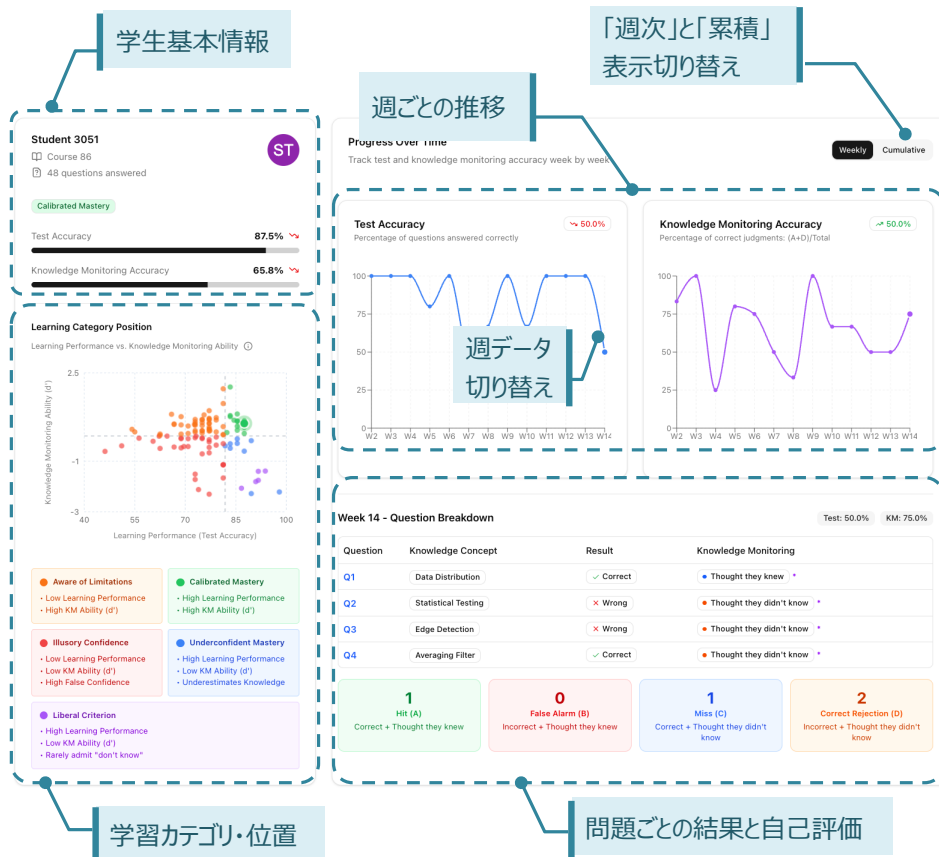
提案ダッシュボードは、学習者の知識レベルと知識モニタリング能力を可視化し、第 3 章で設計した適応型フィードバック方略に基づく支援を提供する Web アプリケーションである。図 3 に全体構成を示す。ダッシュボードは上部の**結果可視化部** (図 3a) と下部の**フィードバック部** (図 3b) から構成される。

### 4.2 結果可視化部

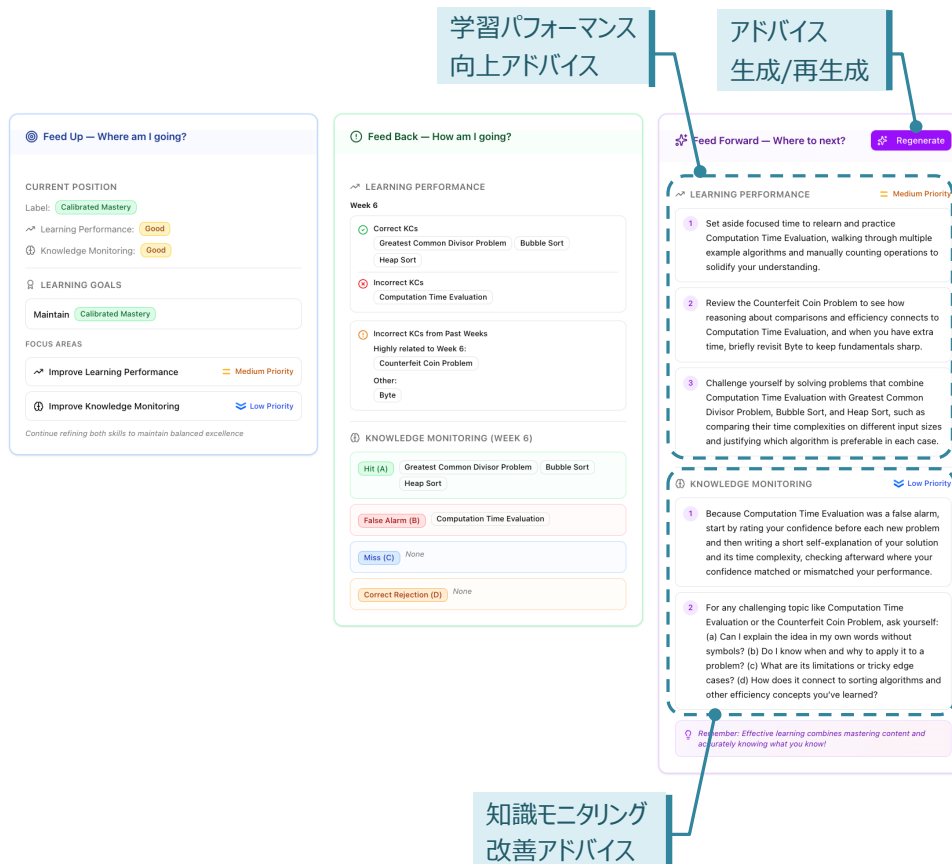
結果可視化部では、学習者が自身の学習状態を客観的に把握できるように、4 つのコンポーネントを配置する (図 3a)。

#### 4.2.1 学習者基本情報

学習者 ID、累積解答数、現在の学習者類型 (5 パターンのいずれか)、テスト正答率、および知識モニタリング能力 ( $d'$ ) を表示する。



(a) 結果可視化部



(b) フィードバック部

図 3: 学習支援ダッシュボードの全体構成

#### 4.2.2 学習者ポジションマップ

知識レベル（横軸）と知識モニタリング能力（縦軸）の二次元空間上に、学習者の現在位置を表示する散布図である。5つの学習者類型に対応する領域を色分けして示すことで、学習者が自身の位置づけと目指すべき方向を直感的に把握できるようにする。

#### 4.2.3 週次推移グラフ

テスト正答率と知識モニタリング正確度の時系列変化を折れ線グラフで表示する。「週次」モードでは各週の値を、「累積」モードでは学期開始からの累積値を表示する。グラフ上の任意の週を選択することで、当該週の詳細データを表示できる。表示モードは「週次」（各週の値）と「累積」（学期開始からの累積値）を切り替え可能である。

#### 4.2.4 問題別詳細テーブル

選択された週の各評価問題について、対応する知識概念、正誤結果、および学習者の自己評価（理解認識/非理解認識）を一覧表示する。これにより、知識モニタリングの4カテゴリ（ヒット、フォルスアラーム、ミス、正棄却）が問題単位で確認できる。

### 4.3 フィードバック部

フィードバック部では、第3章で設計した3段階フィードバックモデルに基づき、Feed Up・Feed Back・Feed Forwardの3セクションを配置する（図3b）。

#### 4.3.1 Feed Up：目標提示

学習者の現在の類型、目指すべき目標状態、および改善の優先順位を提示する。図2に基づき、知識レベルとモニタリング能力のどちらを優先的に改善すべきかを明示する。

#### 4.3.2 Feed Back：現状提示

学習者の現状を知識レベルとモニタリング能力の2観点から詳細に提示する。

知識レベルの観点では、当該週で正答した知識概念と誤答した知識概念をリスト表示する。さらに、知識グラフの関係性に基づき、過去に誤答した概念のうち当該週の内容と意味的に関連の高いものを抽出し、「関連する過去の誤答概念」として提示する。モニタリングの観点では、4カテゴリに分類された知識概念をそれぞれ提示する。

#### 4.3.3 Feed Forward：アドバイス生成

学習者類型と現状に応じた個別最適化されたアドバイスを提示する。アドバイスは知識レベル向上とモニタリング改善の2種類から構成される。アドバイス生成にはLLMを使用する。学習者の類型、正答・誤答履歴、および類型別フィードバック方針に基づき、学習者の具体的な状況に応じたアドバイスを生成する。

### 4.4 実装詳細

ダッシュボードはWebアプリケーションとして実装した。フロントエンドはReact.jsを用いて実装し、バックエンドはPython (FastAPI)を用いた。第3章で述べたHGNNによる潜在的認識状態の推論機能、およびLLMベースのアドバイス

生成機能を統合した。学習者類型化における知識レベルおよび知識モニタリング能力の高低判定には、各クラスの中央値を閾値として使用した。アドバイス生成にはOpenAI GPT-5.2 APIを使用し、学習者の状況に応じたプロンプトエンジニアリングにより個別最適化を実現した。

## 5 評価実験

### 5.1 潜在的認識状態の推定精度 (RQ1)

#### 5.1.1 実験設定

情報科学コースの5クラス、計684名の学生を対象としたデータセットを使用した。コースは14週間にわたり、情報理論、暗号、アルゴリズム、データ構造、機械学習などの基礎概念を扱う。各学生は各講義後に(1)自由記述式の学習リフレクション、(2)評価テストに従事した。表3にデータセットの統計情報を示す。#LPSは潜在的認識状態の数。

表3: データセットの統計情報

	FC <sub>A</sub>	FC <sub>B</sub>	FC <sub>C</sub>	FC <sub>D</sub>	FC <sub>E</sub>
学生数	109	181	101	148	145
評価項目数	48	38	43	42	50
# $\mathcal{E}_{S-K}$	6,349	10,128	6,908	9,396	6,918
# LPS	2,187	3,002	1,727	2,634	3,807

全クラス共通：知識概念数 211, #  $\mathcal{E}_{K-K} = 226$

比較手法として、Random Guesser (RG), GCN, GAT, Label Propagation (LP) を用いた。RGはリンクをランダムに予測するベースライン、GCNとGATは同種グラフ ( $\mathcal{E}_{S-K}$ のみ) 上のグラフニューラルネットワーク、LPは異種グラフ ( $\mathcal{E}_{S-K}$ と $\mathcal{E}_{K-K}$ ) 上のラベル伝播法である。提案手法はHANベースの異種グラフニューラルネットワークにEINSを組み合わせたものである。評価指標にはROC-AUCを使用し、各クラスについて80%を訓練データ、20%をテストデータとして分割し、30回の試行の平均値を報告する。

#### 5.1.2 結果

表4に各手法の性能比較を示す。提案手法は全クラスにおいて最高のAUC値を達成し、平均85.21%のAUCで潜在的認識状態を推定できた。

表4: 推定精度 (AUC%) の比較

手法	FC <sub>A</sub>	FC <sub>B</sub>	FC <sub>C</sub>	FC <sub>D</sub>	FC <sub>E</sub>	平均
RG	50.00	50.00	50.00	50.00	50.00	50.00
GCN	77.91	76.97	79.18	79.53	76.45	78.01
GAT	75.44	74.71	75.81	77.18	73.89	75.41
LP	81.30	81.85	80.68	82.85	81.47	81.63
提案手法	<b>86.52</b>	<b>85.17</b>	<b>84.81</b>	<b>87.23</b>	<b>82.33</b>	<b>85.21</b>
w/o EINS	76.52	76.49	78.47	79.49	77.58	77.71

提案手法は同種グラフを用いるGCN (78.01%) およびGAT

(75.41%)を大幅に上回った。この結果は、知識概念間の意味的・階層的関係( $\mathcal{E}_{K-K}$ )を活用することの有効性を示している。また、異種グラフを用いる LP (81.63%)と比較しても 3.58 ポイント高い精度を達成した。EINS の導入により 7.50 ポイントの向上が確認され、学習者が明示的に報告した非理解認識を負例として活用することの重要性が示された。

### 5.1.3 考察

提案手法は 85.21% の AUC を達成し、ベースライン手法を上回った。特に EINS による 7.50 ポイントの向上は、学習者の明示的な自己報告を負例として活用することの有効性を示している。これは、「理解していない」という学習者の認識が、モデルの学習において重要な境界情報となることを示唆している。

また、同種グラフ手法 (GCN, GAT) と比較して異種グラフ手法 (LP, 提案手法) が高い性能を示したことから、知識概念間の意味的・階層的関係が潜在的認識状態の推定に有効であることが確認された。

## 5.2 ユーザ実験 (RQ2)

### 5.2.1 実験設定

システムの受容性を評価するため、大学生 36 名および教員 11 名、計 47 名を対象としたユーザ実験を実施した。

実験手順は以下の通りである：(1) システム使用 (ダッシュボードの操作と学習状況の確認)、(2) 評価アンケート (フィードバック品質 16 項目, SUS 10 項目, 自由記述 2 項目)。

### 5.2.2 結果

#### a) フィードバック品質

三段階フィードバックの評価結果を表 5 に示す。学生群 ( $M=4.19$ )、教員群 ( $M=4.52$ ) ともに高い評価を得た。特に Feed Back (現状提示) の評価が高く、誤答した知識概念、関連する過去に誤答した知識概念の提示が有用と評価された。

表 5: フィードバック品質評価 (5 段階)

セクション	学生		教員	
	M	SD	M	SD
Feed Up	3.95	1.01	4.41	0.70
Feed Back	4.39	0.81	4.59	0.61
Feed Forward	4.06	0.93	4.51	0.61
全体	4.19	0.30	4.52	0.15

#### b) システムユーザビリティ

System Usability Scale (SUS) による評価結果を表 6 に示す。教員群は 70.68 点を達成した一方、学生群は 63.26 点と平均をやや下回った。

表 6: SUS スコア

	平均	SD
学生 ( $n=36$ )	63.26	14.37
教員 ( $n=11$ )	70.68	17.45

### c) 定性評価

自由記述の分析から、最も多く言及されたのは「知識モニタリングの 4 カテゴリ可視化」(15 件)であり、「単なる正解・不正解だけでなく、自己評価との組み合わせで復習の優先順位をつけやすい」という意見が得られた。次いで、「過去に誤答した知識概念との関連性提示」(12 件)、「学習者類型とポジションマップ」(7 件)が有用と評価された。

改善要望としては、「UI/UX」に関する指摘 (10 件) が最も多く、文字サイズや週の切り替え操作に関する意見があった。また、「ヒット」「フォルスアラーム」等の専門用語の分かりにくさ (5 件) が指摘された。

### 5.2.3 考察

フィードバック品質は学生・教員ともに高い評価を得た ( $M>4.0$ )。特に知識モニタリングの 4 カテゴリ可視化と過去の誤答との関連性提示が有用と評価され、提案システムの核心機能が受容されたことを示している。

一方、学生群の SUS スコアは平均を下回った。教員群のスコアが高かった要因として、教員は学習分析ダッシュボードの使用経験があり、類似システムへの親和性が高かったことが考えられる。今後の改善点として、(1) UI の操作性改善、(2) 専門用語のより直感的な表現への変更が挙げられる。

## 6 まとめ

本研究では、学習者の知識モニタリング能力を評価するための異種グラフベースのフレームワークと、その結果に基づく適応型フィードバックシステムを提案した。LLM による知識概念抽出と異種グラフ構築により、学習者の潜在的認識状態をリンク予測問題として推定する手法を開発した。また、知識レベルと知識モニタリング能力の二軸による学習者類型化と、Feed Up・Feed Back・Feed Forward の三段階フィードバックを提供するダッシュボードを実装した。評価実験により提案手法の有効性を確認し、ユーザ実験により知識モニタリングの可視化と過去の誤答との関連性提示が有用であることを示した。今後は、他分野への適用、UI の改善、長期的な学習効果の検証に取り組む。

## 謝辞

本研究は、JST CREST (JPMJCR22D1)、JST AIP チャレンジ、ならびに JSPS 科研費 (JP22H00551, JP25K21360, JP24K16759) の支援を受けて実施された。

## 文献

- [1] Ismail Aliyu, A Kana, Salisu Aliyu, et al. Development of knowledge graph for university courses management. *International Journal of Education and Management Engineering*, Vol. 10, No. 2, p. 1, 2020.
- [2] Adam B Barrett, Zoltan Dienes, and Anil K Seth. Measures of metacognition on signal-detection theoretic models. *Psychological methods*, Vol. 18, No. 4, p. 535, 2013.
- [3] Tony Blackwood. Business undergraduates' knowledge monitoring accuracy: How much do they know about how

- much they know? *Teaching in Higher Education*, Vol. 18, No. 1, pp. 65–77, 2013.
- [4] Dejun Cai, Yuan Zhang, and Bintao Dai. Learning path recommendation based on knowledge tracing model and reinforcement learning. In *2019 IEEE 5th international conference on computer and communications (ICCC)*, pp. 1881–1885. IEEE, 2019.
  - [5] Li Chen, Gen Li, Boxuan Ma, Cheng Tang, and Masanori Yamada. A three-step knowledge graph approach using llms in collaborative problem solving-based stem education. *International Association for Development of the Information Society*, 2024.
  - [6] Li Chen, Gen Li, Buxuan Ma, Cheng Tang, Masanori Yamada, and Atsushi Shimada. Classifying knowledge nodes and analyzing activation features: An integrated knowledge graph approach for collaborative problem-solving. In *2025 IEEE International Conference on Advanced Learning Technologies (ICALT)*, pp. 107–111, 2025.
  - [7] Penghe Chen, Yu Lu, Vincent W Zheng, Xiyang Chen, and Boda Yang. Knowedu: A system to construct knowledge graph for education. *Ieee Access*, Vol. 6, pp. 31553–31563, 2018.
  - [8] Jiajun Cui, Hong Qian, Bo Jiang, and Wei Zhang. Leveraging pedagogical theories to understand student learning process with graph-based reasonable knowledge tracing. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 502–513, 2024.
  - [9] Tung Do Viet and Konstantin Markov. Using large language models for bug localization and fixing. In *2023 12th International Conference on Awareness Science and Technology (iCAST)*, pp. 192–197. IEEE, 2023.
  - [10] Jibril Frej, Neel Shah, Marta Knezevic, Tanya Nazaretsky, and Tanja Käser. Finding paths for explainable mooc recommendation: A learner perspective. In *Proceedings of the 14th Learning Analytics and Knowledge Conference, LAK '24*, p. 426–437, New York, NY, USA, 2024. Association for Computing Machinery.
  - [11] John Hattie and Helen Timperley. The power of feedback. *Review of educational research*, Vol. 77, No. 1, pp. 81–112, 2007.
  - [12] Owen Henkel, Libby Hills, Adam Boxer, Bill Roberts, and Zach Levonian. Can large language models make the grade? an empirical study evaluating llms ability to mark short answer questions in k-12 education. In *Proceedings of the Eleventh ACM Conference on Learning@ Scale*, pp. 300–304, 2024.
  - [13] Stephen Hutt, Allison DePiro, Joann Wang, Sam Rhodes, Ryan S Baker, Grayson Hieb, Sheela Sethuraman, Jaclyn Ocumpaugh, and Caitlin Mills. Feedback on feedback: Comparing classic natural language processing and generative ai to evaluate peer feedback. In *Proceedings of the 14th Learning Analytics and Knowledge Conference*, pp. 55–65, 2024.
  - [14] Priscilla Lee and Soohyun Nam Liao. Targeting metacognition by incorporating student-reported confidence estimates on self-assessment quizzes. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education, SIGCSE '21*, p. 431–437, New York, NY, USA, 2021. Association for Computing Machinery.
  - [15] Gen Li, Li Chen, Cheng Tang, Daisuke Deguchi, Takayoshi Yamashita, and Atsushi Shimada. From reflections to motifs: A graph-based analysis of learners' knowledge construction. In *International Conference on Artificial Intelligence in Education*, pp. 299–307. Springer, 2025.
  - [16] Gen Li, Cheng Tang, Li Chen, Daisuke Deguchi, Takayoshi Yamashita, and Atsushi Shimada. Llm-driven ontology learning to augment student performance analysis in higher education. In *International Conference on Knowledge Science, Engineering and Management*, pp. 57–68. Springer, 2024.
  - [17] Chien-Chang Lin, Anna YQ Huang, and Owen HT Lu. Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic review. *Smart learning environments*, Vol. 10, No. 1, p. 41, 2023.
  - [18] Klaus Lingel, Jan Lenhart, and Wolfgang Schneider. Metacognition in mathematics: Do different metacognitive monitoring measures make a difference? *ZDM*, Vol. 51, No. 4, pp. 587–600, 2019.
  - [19] Hanxiao Liu, Wanli Ma, Yiming Yang, and Jaime Carbonell. Learning concept graphs from online educational data. *Journal of Artificial Intelligence Research*, Vol. 55, pp. 1059–1090, 2016.
  - [20] Brian Maniscalco and Hakwan Lau. A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and cognition*, Vol. 21, No. 1, pp. 422–430, 2012.
  - [21] Lars-Peter Meyer, Claus Stadler, Johannes Frey, Norman Radtke, Kurt Junghanns, Roy Meissner, Gordian Dziwis, Kirill Bulert, and Michael Martin. Llm-assisted knowledge graph engineering: Experiments with chatgpt. In *Working conference on Artificial Intelligence Development for a Resilient and Sustainable Tomorrow*, pp. 103–115. Springer Fachmedien Wiesbaden Wiesbaden, 2023.
  - [22] Lin Ni, Sijie Wang, Zeyu Zhang, Xiaoxuan Li, Xianda Zheng, Paul Denny, and Jiamou Liu. Enhancing student performance prediction on learnersourced questions with sgnn-llm synergy. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38, pp. 23232–23240, 2024.
  - [23] Zilong Pan, Lauren Biegley, Allen Taylor, and Hua Zheng. A systematic review of learning analytics: Incorporated instructional interventions on learning management systems. *Journal of Learning Analytics*, Vol. 11, No. 2, pp. 52–72, 2024.
  - [24] Stephanie Pieschl. Metacognitive calibration—an extended conceptualization and potential applications. *Metacognition and Learning*, Vol. 4, pp. 3–31, 2009.
  - [25] Tricia R Prokop. Calibration and academic performance in students of health sciences. *Health Professions Education*, Vol. 6, No. 4, pp. 564–573, 2020.
  - [26] T. Ullmann. Automated analysis of reflection in writing: Validating machine learning approaches. *International Journal of Artificial Intelligence in Education*, Vol. 29, pp. 217–257, 2019.
  - [27] Christopher A Was. Discrimination in measures of knowledge monitoring accuracy. *Advances in Cognitive Psychology*, Vol. 10, No. 3, p. 104, 2014.
  - [28] Kevin P Yancey, Geoffrey Laflair, Anthony Verardi, and Jill Burstein. Rating short l2 essays on the cefr scale with gpt-4. In *Proceedings of the 18th workshop on innovative use of NLP for building educational applications (BEA 2023)*, pp. 576–584, 2023.
  - [29] Junrui Zhang, Yun Mo, Changzhi Chen, and Xiaofeng He. Gkt-cd: Make cognitive diagnosis model enhanced by graph-based knowledge tracing. In *2021 International joint conference on neural networks (IJCNN)*, pp. 1–8. IEEE, 2021.
  - [30] Yifen Zhou and Xian Yu. Multi-graph spatial-temporal synchronous network for student performance prediction. *IEEE Access*, 2024.
  - [31] Barry J Zimmerman. Becoming a self-regulated learner: An overview. *Theory into practice*, Vol. 41, No. 2, pp. 64–70, 2002.

# 社会選好を考慮した SNS 上の誹謗中傷投稿の拡散を抑制する介入戦略

林 央祐<sup>†</sup> 山本 修平<sup>††</sup>

<sup>†</sup> 筑波大学 情報学群 知識情報・図書館学類 〒 305-8550 茨城県つくば市春日 1 丁目 2 番地

<sup>††</sup> 筑波大学 図書館情報メディア系 〒 305-8550 茨城県つくば市春日 1 丁目 2 番地

E-mail: <sup>†</sup>s2210263@u.tsukuba.ac.jp, <sup>††</sup>syamamoto@slis.tsukuba.ac.jp

**あらまし** SNS 上の誹謗中傷拡散が深刻な社会問題となる中、法や技術的手法による対策が数多く行われているが、SNS 上の誹謗中傷の本質的な原因はその同調や拡散であるため、同調や拡散を抑制するアプローチも重要である。本研究は、行動経済学に基づき、SNS 上での誹謗中傷投稿への拡散を抑制するため、社会選好に応じて個別化された介入戦略の有効性を検討した。SNS 利用者には、4 種類の介入メッセージを提示する実験を行い、行動意向の変化を分析した。その結果、特定の社会選好タイプと介入メッセージの組み合わせによる誹謗中傷の有意な抑制効果はあまり確認されなかったが、特定の条件下では通常投稿の拡散まで抑制してしまう副作用が観察された。また、利得局面よりも損失局面の方が、介入メッセージの個別化に適している可能性が示されたため、社会選好は個別化の有力な手段となり得るが、メッセージとの関連性を考慮すべきであり、今後より効果的な分析をおこなうためには社会選好以外の個人の属性や傾向を取り入れる必要がある。

**キーワード** 誹謗中傷, SNS, ナッジ, 行動変容, 社会選好 (Social Value Orientation)

## 1 はじめに

**SNS での誹謗中傷:** 近年、SNS (ソーシャル・ネットワーキング・サービス) の普及により、時間や場所を問わず誰もがコミュニケーションを取ることが可能となった。一方で、SNS の持つ匿名性や拡散性を悪用した誹謗中傷が増加し、深刻な社会問題となっている。誹謗中傷に関する相談件数は年々増加しており、2024 年度には違法・有害情報相談センターへの相談件数が 6,403 件に達した [1]。また、インターネット利用者の 60.6% が誹謗中傷を目撃した経験があると報告されており [2]、問題の深刻さが窺える。誹謗中傷の被害は一般ユーザにも及ぶが、特に著名人に対するものは、メディアでの注目度が高く、被害者への心理的影響も大きい [3]。このような被害を防ぐための対策が急務となっている。炎上において実際に書き込みに参加する人は、インターネット利用者のうちごく一部にすぎない。むしろ、「安易に再投稿・拡散する人」の増加により誹謗中傷が集積し集団攻撃となることが示されている [4]。したがって、効果的な対策には、主体的な投稿者だけでなく、誹謗中傷に同調する人を減らすことが重要である。

**効果的な誹謗中傷対策の難しさ:** 現在、誹謗中傷対策として法的手法と技術的手法が用いられている。法的手法では、プロバイダ責任制限法の改正により発信者情報開示請求の手続きが簡略化され、加害者特定が迅速化した [5]。技術的手法では、自然言語処理や機械学習による誹謗中傷の自動検出システムの開発が進められている [6]。しかし、法的手法では被害者の負担が依然として大きく、匿名性の高いプラットフォームでの加害者特定が困難になることもある。また、自動検出では文脈依存の表現や間接的な攻撃の判別精度が不十分であり、検出できな

かったり、誤検出により正当な批判まで削除したりするリスクが指摘されている。このように、法律や技術のみに依存した対策には限界がある。

**ナッジ:** 本研究ではこの問題に対し、行動経済学で注目されているナッジ理論に着目する。ナッジ理論は、個人の自由を損なわずに望ましい行動を促進する手法であり、意思決定に影響を与える環境や情報提示を工夫することで行動を自然に誘導する。ナッジは法律や罰則と異なり強制力を伴わないため、心理的抵抗が少ない点が特徴である。近年、オンラインハラスメントの文脈でもナッジの有効性が実証されつつあり、正木ら [7] の研究で検証されている。誹謗中傷の抑制においても、損失回避を意識させるメッセージ (例: 「あなたの投稿が法的責任を問われる可能性があります」) や利他性に基づくメッセージ (例: 「この投稿は相手を深く傷つけます」) の有効性が示唆されている [8]。このように介入メッセージにより誹謗中傷への同調行動を抑制できる可能性がある一方で、その効果は受け手の個人特性によって大きく異なることが指摘されている [9]。

**社会選好:** こうした個人差は、利他性や社会規範への感受性といった心理的特性に関連していると考えられる。先行研究でも、このような個人差が攻撃性や同調行為に影響することが明らかにされている。Hilbig ら [10] は、社会的価値志向性 (Social Value Orientation: SVO) が攻撃性や反社会的行動と密接に関連することを実証的に示し、SNS での誹謗中傷やその同調への影響も例外ではないと考えられる。こうした特性は、行動経済学において社会選好として概念化されている。社会選好とは、自己の物質的利益だけでなく他者や集団の利益、あるいは分配の公平性を考慮する選好を指し、伝統的な経済学の自己利益最大化の仮定を超えて人間行動を説明する重要な概念である。社

会選好の測定には、SVO スライダー (Social Value Orientation Slider) が広く用いられている [11]。この手法は、自己と他者への資源配分課題を通じて社会選好を定量的に評価し、利他主義、協調主義、個人主義、競争主義などにタイプ分けすることができる。利他主義は他者の利益の最大化を、協調主義は自己と他者の利益の合計の最大化を、個人主義は自己の利益のみを、競争主義は他者との利益差の最大化を志向する。このように社会選好には多様な傾向が存在し、それぞれが攻撃性や社会規範への反応において異なる特徴を持つ。したがって、誹謗中傷への同調を抑制する介入メッセージの効果も、個人々の社会選好の傾向を考慮することでより高まると考えられる。

**本研究:** 本研究は、SNS 上での誹謗中傷への同調や拡散を抑制するため、ナッジ理論を活用し、社会選好に応じて個別化された介入戦略の有効性を検討する。具体的には、実験参加者へのメッセージ提示実験を通じて、利己性と利他性、社会規範を強調するメッセージの効果を評価する。社会規範を提示するメッセージについては、多数派への協調を促すものと少数派からの離脱を促すものに区別し、比較分析を行う。実験参加者の社会選好を測定するため、資源配分課題を提示し、自己と他者への配分パターンから個々の社会選好のタイプを特定する。その上で、社会選好の傾向ごとに各介入メッセージを提示し、誹謗中傷への同調をどの程度抑制できるかを評価する。加えて、先行研究 [8] で問題とされた、介入策が誹謗中傷以外の投稿に対しても同調を抑制する副作用をどの程度低減できるかについても評価する。以上により、異なるメッセージが各傾向の受け手に与える影響を明らかにし、SNS 上での誹謗中傷を効果的に抑制するアプローチを構築する。

## 2 関連研究

本研究は、社会選好に応じた介入メッセージを提示することで、誹謗中傷への同調を効果的に抑制できるかを明らかにすることを目的としている。この章では、法的・技術的手法による SNS 上の誹謗中傷対策に関する研究、ナッジによる介入メッセージの効果に関する研究、社会選好の測定方法とその活用に関する研究を概観する。最後に、本研究の位置付けについて述べる。

### 2.1 法的・技術的手法での SNS 上の誹謗中傷対策に関する研究

SNS 上での誹謗中傷対策として、法的手法と技術的手法が発展してきている。法的手法では、プロバイダ責任制限法の改正により発信者情報開示請求の手続きが簡略化され、加害者特定の迅速化が実現されている [5]。従来、被害者は裁判所を通じてプロバイダと投稿者の両方に情報開示請求を行う必要があり、複数回の訴訟手続きを要していた。2021 年の法改正により一度の非訟手続きで発信者情報の開示が可能となり、手続きの簡略化と迅速化が達成された。また、侮辱罪の法定刑引き上げにより誹謗中傷に対する抑止力が強化され、法的枠組みの整備が進展している。これらの法的整備により、被害者が加害者に対し

て責任を追及する環境が改善されつつある。技術的手法では、自然言語処理や機械学習を用いた誹謗中傷の自動検出システムの開発が進展している。西谷ら [6] は、BERT 等の深層学習モデルを用いて日本語の誹謗中傷表現を高精度に検出する手法を提案し、実用的な精度を達成した。この手法では、大規模なテキストデータセットを学習することで文脈を考慮した誹謗中傷の判別が可能となっている。Zhang ら [12] は、畳み込みニューラルネットワーク (CNN) と注意機構を組み合わせたモデルを提案し、X (旧 Twitter) 上の攻撃的な言語を高精度で検出することに成功した。Davidson ら [13] は、ヘイトスピーチと単なる攻撃的な言語を区別するための分類器を開発し、両者の違いを機械学習で識別する手法を確立した。これらの技術により、大量の投稿から有害コンテンツを迅速に識別することが可能になり、プラットフォーム運営者による効率的なモデレーションが実現されつつある。さらに、深層学習の進展により、文脈依存の表現や新たなスラング表現への対応も継続的に改善が進められている。

### 2.2 ナッジによる介入メッセージの効果に関する研究

ナッジ理論は、個人の自由を損なわずに望ましい行動を促進する手法として、オンライン上の行動変容にも応用されている。Thaler ら [14] が提唱したナッジ理論の枠組みでは、損失回避や社会規範の活用が行動変容に効果的であるとされている。この理論は、法律や罰則と異なり強制力を伴わないため、心理的抵抗が少ない点が特徴である。正木ら [7] は、SNS 上での攻撃的コメント投稿を抑制するために、投稿前に注意喚起メッセージを表示するナッジ介入を実施した。その結果、注意喚起メッセージを受け取ったユーザは、攻撃的コメントの投稿率が有意に低下することが示された。また、Munger ら [15] は、Twitter 上で人種差別的な発言をしたユーザに対し、影響力のあるアカウントから警告メッセージを送信する実験を行った。その結果、警告を受けたユーザは、その後の差別的発言を有意に減少させることが実証された。この研究は、社会的影響力を持つ存在からのメッセージが行動変容に効果的であることを示している。誹謗中傷への同調抑制に関して、村田ら [8] は、利己性メッセージ (例:「あなたの投稿が法的責任を問われる可能性があります」) や利他性に基づくメッセージ (例:「この投稿は相手を深く傷つけます」)、また社会規範メッセージ (例:「たった 1.5% の人しか誹謗中傷に加担していません」) の 3 つのメッセージの有効性を検証した。実験の結果、利己性・利他性のメッセージが誹謗中傷への同調を一定程度抑制する効果があることが示されたが、社会規範メッセージには効果がなかったことが示された。また、介入メッセージが誹謗中傷以外の投稿に対しても同調を抑制してしまう副作用も確認された。これらの研究により、ナッジを用いた介入メッセージは誹謗中傷自体の抑制や、誹謗中傷への同調に対して一定の効果を持つものの、その効果を最大化し、副作用を最小化するためには、副作用の原因の特定や効果を高める更なる工夫を見出すことが求められる。

---

1: X. <https://x.com/?lang=ja>

### 2.3 社会選好の測定方法とその活用に関する研究

社会選好は、自己の物質的利益だけでなく他者や集団の利益、あるいは分配の公平性を考慮する選好として、行動経済学において重要な概念として確立されている。社会選好の測定には複数の手法が開発されてきたが、中でも SVO スライダー (Social Value Orientation Slider) が広く用いられている。Murphy ら [11] は、SVO スライダーを開発し、自己と他者への資源配分課題を通じて個人の社会選好を定量的に評価する手法を確立した。SVO スライダーは、簡便かつ高い信頼性を持つ測定手法として多くの研究で採用されている。SVO スライダーを用いた研究では、社会選好が様々な社会的行動の予測に有効であることが示されている。Murphy ら [11] の研究では、SVO スライダーで測定された社会選好が協力行動や向社会的行動を予測することが実証された。特に、協調的志向性を持つ個人は、資源配分ゲームにおいてより公平な分配を選択し、他者との協力を重視する傾向が確認された。さらに、Pletzer ら [16] は、SVO スライダーを用いたメタ分析により、社会選好が社会的ジレンマにおける協力行動の強力な予測因子であることを明らかにした。分析の結果、協調的な社会選好を持つ個人は、個人主義的または競争的な社会選好を持つ個人と比較して、有意に高い協力行動を示すことが確認された。一方で、社会選好は協力行動だけでなく、攻撃性や反社会的行動とも関連することが示されている。Hilbig ら [10] は、SVO スライダーで測定した個人主義的・競争的な社会選好を持つ個人が攻撃的・反社会的な行動傾向を示しやすいことを実証した。彼らの研究では、社会選好と攻撃性の間に有意な関連が見られ、特に個人主義的・競争主義的志向性を持つ個人は、他者を犠牲にしても相対的優位を得ようとする行動パターンを示すことが明らかになった。これらの研究により、SVO スライダーで測定される社会選好は協力行動から攻撃性まで幅広い社会的行動を予測する重要な個人特性であることが確認されている。

### 2.4 本研究の位置付け

誹謗中傷投稿そのものを技術的に防ぐ試みは、今後の技術の進歩に伴いより効果的なものになるだろう。一方で、SNS 上の誹謗中傷被害は、誹謗中傷投稿が多くの人間に拡散され、被害者の目に届いてしまうことが原因であり、この同調等による拡散行動を抑制することも効果的である。また、ナッジ理論を用いた介入メッセージがオンライン上の誹謗中傷、それらの同調に対して有効であることが様々な研究で示されている [8][15]、一方で、従来のナッジは一般的に効果があるメッセージを全員に一律で提示するものであり、個人の特性に応じてメッセージを出し分けることで効果を高める研究は十分に行われていない。先行研究から、他者との協力行動や社会的行動が社会選好に応じて異なることが示されており [10][11][16]、誹謗中傷への同調行動も社会選好に応じて異なる可能性が考えられる。本研究は、攻撃性や同調行動と関連する社会選好に着目し、個人の社会選好の傾向に応じた介入メッセージの効果を検証することで、より効果的な誹謗中傷対策の実現を目指す。

## 3 社会選好を考慮したメッセージ提示実験

### 3.1 実験の概要

本研究では、社会選好に応じて個別化された介入戦略が、SNS 上での誹謗中傷への同調や拡散をどの程度抑制できるか評価する。本研究では 2025 年 10 月 27 日、28 日、11 月 10 日の 3 日間にわたり質問紙調査を実施した。実験参加者には、SNS 投稿を閲覧し介入メッセージを見る前の反応と、介入メッセージを見た後の反応として適切な選択肢を回答してもらい、両者を前後比較することで介入効果を測定した。本章の構成は以下の通りである。3.2 節で実験参加者の募集方法と参加条件、及び 4 つの実験グループへの割り当てについて説明する。3.3 節では実験で提示した介入メッセージを述べる。3.4 節では実験の具体的な流れを説明し、3.5 節で収集したデータの概要をまとめる。本研究は、筑波大学図書館情報メディア系研究倫理委員会の承認を受けて実施された (承認番号: 第 25-93 号)。

### 3.2 実験参加者とグループ分け

#### 3.2.1 実験参加希望者の募集

本研究の実験参加者は、オンラインアンケートサイト「Freeasy」<sup>2</sup>を通じて募集した。Freeasy は、調査対象者のリクルーティングとアンケート配信を行うマーケティングリサーチプラットフォームである。募集に際して設定した条件は、(1) 18 歳以上、(2) 直近 1 か月以内に SNS (実験で使用する X) で「いいね」や「リポスト・引用ポスト」を利用した経験がある、(3) 実験で提示する X 上の投稿内容 (野球・サッカー) に興味・関心がある、の 3 つである。

条件 (2)、(3) は、SNS やテーマへの関心が低い参加者が含まれることで適切な分析結果が得られないことを防ぐために設定した。特に条件 (3) でスポーツ、具体的には野球とサッカーをテーマに指定した理由は以下の 2 点である。第一に、Williams らの研究 [17] において、スポーツに関する誹謗中傷は他分野の誹謗中傷と比べて試合等のイベント直後に急速に拡散されやすいことが示されており、誹謗中傷の拡散抑制を目的とする本研究に適したテーマであると判断したためである。第二に、総務省の調査結果 [18] から、野球とサッカーは日本における観戦者割合が最も高い上位 2 つのスポーツであり、多くの実験参加者が関心を持つテーマであると考えたためである。

これらの条件を満たす参加者を絞り込むため、Freeasy 上でスクリーニング調査を事前に実施した。その結果、計 15,000 名の応募者の中から条件を満たす 400 名の参加希望者を確保した。

#### 3.2.2 個人の社会選好の測定

実験参加者の個人の社会選好を測定するために、上記のスクリーニング条件を満たした 400 人に社会選好に関するアンケートを実施した。本実験で実際に参加希望者に回答してもらった社会選好測定法を表 1、2 に示す。本実験では SVO スライダー [11] を参考にした。スクリーニングで提示する 6 問と本実験中に提示する 6 問でそれぞれ異なるシナリオを提示する。前

2: Freeasy, <https://freeasy-survey.com/>

半では、実験参加者自身と見ず知らずの第三者を想定し、利得を得られる金額配分の選択肢の中から実験参加者が適切だと思うものを選択する。後半では、想定する人物は変更せず、前半の6問の選択肢すべてから一律10,000円を引いた損失を被る金額配分の選択肢の中から、適切だと思うものを選択する。本来のSVOスライダーでは、社会選好を測定する際には利得の金額配分のみを用いるが、本実験で使用する介入メッセージは損失回避に働きかけるものであるため、利得局面で考えるSVOスライダーよりも損失の金額配分を用いたSVOスライダーを使用することで、より介入メッセージが効果的に作用する個人の傾向を測定できると考えたため、二種類のSVOスライダーで測定する。スクリーニング時点で利得局面に関する社会選好測定を行うのは、個人の社会選好の傾向を測定する目的と、実験参加者の社会選好に偏りが無いことを確認する目的があるためである。

また、表から実験参加者が選択した金額配分から、個人の社会選好の傾向を測定できるSVOスコアを算出する。自己への配分の平均を $A_s$ 、他者への平均を $A_o$ として、SVOスコアは、以下の式で計算される。

$$SVO_p = \arctan\left(\frac{A_o - 5000}{A_s - 5000}\right), SVO_n = \arctan\left(\frac{A_o + 5000}{A_s + 5000}\right)$$

また、 $SVO_p$ と $SVO_n$ は、それぞれ利得局面、損失局面によるスコアになっている。このスコアは、 $-16.26^\circ \sim 61.39^\circ$ の範囲で算出される。

### 3.3 実験グループの割り当て

本実験では、実験参加者を4つの介入群にランダムに割り当てた。各介入群には異なるタイプの介入メッセージを提示し、個人の社会選好の傾向と介入メッセージの組み合わせによって誹謗中傷の拡散行動がどのように変化するかを分析した。本研究では個人の社会選好に応じて効果的な誹謗中傷を抑制する効果を持つ介入メッセージがどのようなものか明らかにすることを目的としているため、実験参加者の社会選好に極端な偏りが無いことを確認したうえで、提示する介入メッセージの種類ごとに社会選好を考慮せずランダムな100人に振り分けた。

### 3.4 メッセージ提示実験

#### 3.4.1 実験参加者に提示した投稿内容

本実験では、介入群に関わらず、実験参加者全員に対してX上で実際に投稿された野球・サッカーのテーマそれぞれの誹謗中傷を含む投稿と、誹謗中傷を含まない投稿の計4種類を提示した。誹謗中傷を含む投稿に対しては、介入メッセージによって拡散行動を抑制したい狙いがあるが、誹謗中傷を含まない投稿に対しては、介入メッセージによって拡散行動を抑制してしまう副作用の効果を検証するために用いた。図1に実験参加者に提示した各投稿を示す。野球についての投稿は2025年8月に、サッカーについての投稿は2024年10月にそれぞれX上で実際に投稿されたニュース記事に返信された誹謗中傷を含む投稿と含まない投稿を選択した。誹謗中傷の定義として、国際大学グローバル・コミュニケーション・センターが定めた9

つの定義[19]を用いた。中でも、経験率が高かった「侮辱・攻撃」要素を含みかついいねやリポストが、対応する誹謗中傷なし投稿とできるだけ同数程度になっていることに注意して設定した。選択した計4つの投稿を、SNSを使用する際実際に閲覧するであろうXの画面を模した形式で提示した。

#### 3.4.2 実験参加者に提示した介入メッセージの内容

本節では、実験に使用した介入メッセージの作成方法について説明する。先行研究[8]では、自分自身の損失を強調する利己性メッセージや、他人の損失を強調する利他性メッセージが、誹謗中傷の拡散行動を効果的に抑制できることが報告されている。一方で、先行研究では社会規範メッセージは有意な抑制効果が確認されていないことも報告されている。本研究で改めて介入メッセージを提示する理由は2つある。まず、利己性メッセージと利他性メッセージは、誹謗中傷拡散を効果的に抑制できる一方で、誹謗中傷がない通常の投稿の拡散まで抑制してしまうことが分かっている。この原因として、個人の特性を考えずにメッセージを提示したことにより効果に個人差が出たと考えられる。よって、利己性メッセージと利他性メッセージにあたる利己損失強調メッセージと利他損失強調メッセージは本研究でも採用する。次に、社会規範メッセージは有意な抑制効果がなかったことが分かっている。この原因として、先行研究でも課題として挙げられていたが、実験参加者に社会規範を意識させられなかったことが考えられる。つまり社会規範メッセージを効果のある介入メッセージにするためには、より社会規範を意識させる工夫が必要である。よって、本研究では先行研究の社会規範メッセージを参考にしつつ、より多様な人に社会規範を意識させるために2つのメッセージを作成した。自身が多数派に属することを強く意識させる多数派帰属強調メッセージと、自身が少数派でありそこから脱却することを強く意識させる少数派脱却メッセージに細分化した。以上の知見を踏まえて、本研究では社会選好を考慮した効果的な介入戦略を構築するために、以下4種類のメッセージ介入群の効果を検証する。

**利己損失強調群：**自分自身の損失を強調するメッセージを提示する群。自身の利得を最大化／損失を最小化しようとする利己的な傾向が強い個人に効果的であると考えられる。よって、本研究では次のようなメッセージを作成した。「SNSでの誹謗中傷への関与は、**あなたの評判を下げ、法的リスクを高める可能性があります。**」太字のように、個人の損失に関する具体的な内容を盛り込むことで、損失回避傾向が強い個人に効果的に働きかけることを狙っている。

**利他損失強調群：**他者の損失を強調するメッセージを提示する群。他者の利得を最大化／損失を最小化しようとする利他的な傾向が強い個人に効果的であると考えられる。よって、本研究では次のようなメッセージを作成した。「あなたの誹謗中傷への関与は、**アスリートの心理的健康に影響を与える可能性があります。**」太字のように、第三者の損失に関する具体的な内容を盛り込むことで、損失回避傾向が強い個人に効果的に働きかけることを狙っている。

**多数派帰属強調群：**社会規範メッセージを細分化したうちの一つで、自身が多数派に属することを強く意識させるメッセー

	選択肢 1	選択肢 2	選択肢 3	選択肢 4	選択肢 5	選択肢 6	選択肢 7	選択肢 8	選択肢 9
Q1	10,000	9,400	8,800	8,100	7,500	6,900	6,300	5,600	5,000
	5,000	5,600	6,300	6,900	7,500	8,100	8,800	9,400	10,000
Q2	8,500	8,700	8,900	9,100	9,300	9,400	9,600	9,800	10,000
	1,500	1,900	2,400	2,800	3,300	3,700	4,100	4,600	5,000
Q3	5,000	5,400	5,900	6,300	6,800	7,200	7,600	8,100	8,500
	10,000	9,800	9,600	9,400	9,300	9,100	8,900	8,700	8,500
Q4	5,000	5,400	5,900	6,300	6,800	7,200	7,600	8,100	8,500
	10,000	8,900	7,900	6,800	5,800	4,700	3,600	2,600	1,500
Q5	8,500	8,500	8,500	8,500	8,500	8,500	8,500	8,500	8,500
	8,500	7,600	6,800	5,900	5,000	4,100	3,300	2,400	1,500
Q6	10,000	9,800	9,600	9,400	9,300	9,100	8,900	8,700	8,500
	5,000	5,400	5,900	6,300	6,800	7,200	7,600	8,100	8,500

表 1: SVO スライダー測定法の質問と選択肢 (利得局面)。各セルの上の行が自分の利得, 下の行が他者の利得を示す。実験参加者は見ず知らずの第三者を想定し、金額配分 (Q1~Q6) の選択肢 (1~9) のの中から実験参加者が最も好ましい組み合わせを選択する。

	選択肢 1	選択肢 2	選択肢 3	選択肢 4	選択肢 5	選択肢 6	選択肢 7	選択肢 8	選択肢 9
Q1	0	-600	-1,200	-1,900	-2,500	-3,100	-3,700	-4,400	-5,000
	-5,000	-4,400	-3,700	-3,100	-2,500	-1,900	-1,200	-600	0
Q2	-1,500	-1,300	-1,100	-900	-700	-600	-400	-200	0
	-8,500	-8,100	-7,600	-7,200	-6,700	-6,300	-5,900	-5,400	-5,000
Q3	-5,000	-4,600	-4,100	-3,700	-3,200	-2,800	-2,400	-1,900	-1,500
	0	-200	-400	-600	-700	-900	-1,100	-1,300	-1,500
Q4	-5,000	-4,600	-4,100	-3,700	-3,200	-2,800	-2,400	-1,900	-1,500
	0	-1,100	-2,100	-3,200	-4,200	-5,300	-6,400	-7,400	-8,500
Q5	-1,500	-1,500	-1,500	-1,500	-1,500	-1,500	-1,500	-1,500	-1,500
	-1,500	-2,400	-3,200	-4,100	-5,000	-5,900	-6,700	-7,600	-8,500
Q6	0	-200	-400	-600	-700	-900	-1,100	-1,300	-1,500
	-5,000	-4,600	-4,100	-3,700	-3,200	-2,800	-2,400	-1,900	-1,500

表 2: SVO スライダー測定法の質問と選択肢 (損失局面)。各セルの上の行が自分の損失, 下の行が他者の損失を示す。実験参加者は見ず知らずの第三者を想定し、金額配分 (Q1~Q6) の選択肢 (1~9) のの中から実験参加者が最も好ましい組み合わせを選択する。



(a) テーマ：野球，誹謗中傷あり (b) テーマ：野球，誹謗中傷なし (c) テーマ：サッカー，誹謗中傷あり (d) テーマ：サッカー，誹謗中傷なし  
図 1: 実験で提示した 4 種類の投稿。「野球」か「サッカー」をテーマとする投稿に対して、誹謗中傷が含まれる投稿と、含まれない投稿の合計 4 つの組み合わせを用意した。

ジを提示する群。社会的同調傾向が強い個人に効果的であると  
考えられる。よって、本研究では次のようなメッセージを作成

した。「98.5%の人は誹謗中傷に関与していないという研究報告があり、あなたの言動が社会から逸脱していないかを考慮す

る必要があります。」太字のように、多数派の行動に関する具体的な内容を盛り込むことで、社会的同調傾向が強く、社会への帰属意識が高い個人に効果的に働きかけることを狙っている。

**少数派脱却強調群：**社会規範メッセージを細分化したうちのひとつで、自身が少数派でありそこから脱却することを強く意識させるメッセージを提示する群。社会的同調傾向が強い個人に効果的であると考えられる。よって、本研究では次のようなメッセージを作成した。「**たった1.5%の人しか誹謗中傷に関与していない**という研究報告があり、あなたの言動がその一部の少数派になってしまっていないですか。」太字のように、少数派の行動に関する具体的な内容を盛り込むことで、社会的同調傾向が強く、自身が少数派であることを避けようとする個人に効果的に働きかけることを狙っている。

これら4種類の介入メッセージを用いて、個人の社会選好の傾向に合わせた誹謗中傷拡散を効果的に抑制できる介入メッセージがどのようなものかを検証した。

### 3.5 実験の流れ

本実験は以下の手順で実施した。Freeasy では年齢や性別などの基本的な属性情報を収集できるため、本実験の質問項目ではこれらの情報は収集せず、以下の6つのステップで実施した。

1. 同意確認： Freeasy 上で募集した参加希望者の中から選ばれた、実験の分析効果を高めるためのスクリーニングを通過した実験参加者に対して、社会選好の傾向に応じた介入メッセージの効果の検証という本研究の目的についてや、60分程度の所要時間、いつでも休憩や中断を行えることを説明し、同意をした方のみ次に手順に進んだ。
2. 社会選好の測定： SVO スライダーを参考にした社会選好に関するアンケートを実施する。スクリーニング時点で社会選好に偏りが無いことを確認するために一度利得分配の6問は測定しているため、本実験中では上記で説明した損失の分配を行う6問を提示した。
3. 誹謗中傷を含む／含まない投稿の提示： 野球とサッカーの各テーマにおいて、誹謗中傷を含む投稿と含まない投稿の計4種類の投稿を表示した。投稿内容は図1の通りである。
4. 反応計測（1回目）： 提示された各投稿に対する行動意向を測定した。ここでは、いいね、リポスト、引用ポストの3種類の拡散行動への行動意向を0%～100%の範囲で、20%刻みで回答してもらった。
5. 介入メッセージの提示： 上記の説明の通り、実験参加者が所属する介入群に応じた介入メッセージを提示した。
6. 反応計測（2回目）： 介入メッセージの提示後、1回目の反応計測と同様に、いいね、リポスト、引用ポストの3種類の拡散行動への行動意向を0%～100%の範囲で、20%刻みで回答してもらった。

表3: 介入群ごとの分析対象となる実験参加者統計

統計量	利己損失	利他損失	多数派帰属	少数派脱却
男性人数	41	43	31	33
女性人数	10	16	15	14
平均年齢	51.1	48.2	43.5	45.2
標準偏差	9.6	12.5	9.0	10.3
参加者数	51	59	46	47

## 4 分析結果

### 4.1 実験データの概要

#### 4.1.1 介入メッセージ提示前後の行動意向に関する分布

介入メッセージ効果を示すヒートマップを図2に示す。行動意向すべてにおいて、介入メッセージ提示前後で行動意向の変化がある程度見られるものの、介入前後でともに「0%」を選択する参加者が多かったことが分かる。全てのアンケート項目に対して「0%」と回答している実験参加者は、行動意向の変化が計測できないことから、今回の実験参加者の対象外であるため、除外して以降の分析を行った。

#### 4.1.2 実験参加者の性別／年齢分布

分析対象となる実験参加者の年齢、性別の分布を図3に示す。分析対象となる人数は合計203人で、その内男性が148人、女性が55人であった。表3に各介入群ごとの実験参加者数と平均年齢、年齢の標準偏差を示す。各介入ごとで実験参加者数は概ね50人程度であった。

#### 4.1.3 SVOスコアの分布

SVOスコアの分布を図4に示す。この図では、値全体を25分割したヒストグラムとなっており、 $SVO_p$ は刻み幅 $3.11^\circ$ 、 $SVO_n$ は刻み幅 $2.82^\circ$ となっている。横軸はSVOスコア、縦軸はその値をとった参加者数を示しており、値が大きいほど利他性を強く示す。 $SVO_p$ と $SVO_n$ の中央値がそれぞれ $32.45^\circ$ 、 $28.44^\circ$ と、 $SVO_n$ の方がやや低くなった。これは、損失回避傾向の方が利得配分傾向よりも利己的な傾向を示す参加者が多いことを示している。グラフ内の縦線は中央値を示しており、中央値を基準に参加者の社会選好を利己的（PROSELF）と利他的（PROSOCIAL）に分類した。

### 4.2 メッセージ提示実験の分析

#### 4.2.1 重回帰分析の式と各変数の説明

本研究では、個人の社会選好に対して介入メッセージがどのような介入効果をもたらすかを分析するため、重回帰分析モデルを構築した。重回帰分析の式は以下の通りである。

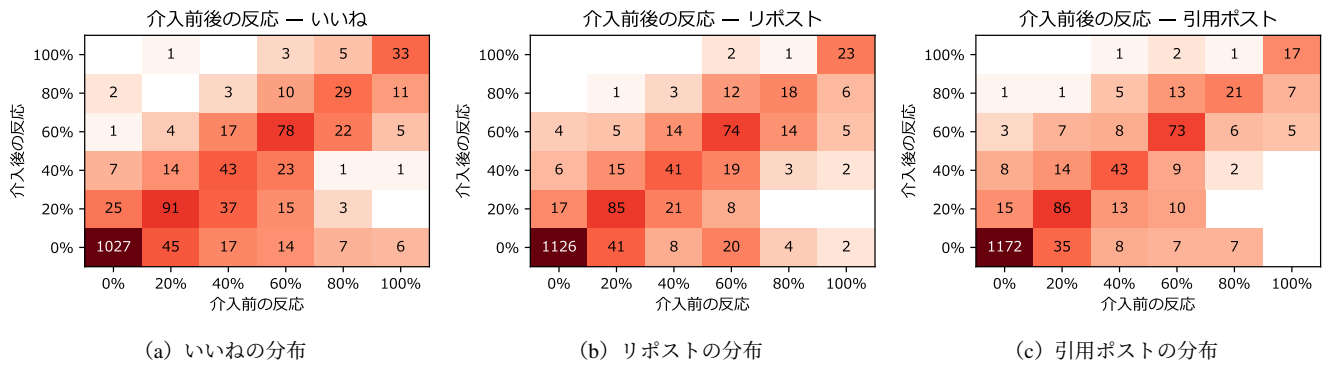


図2: 介入前後の行動意向別の人数。行方向に介入前の行動意向を、列方向に介入後の行動意向を示し、対応する人数をセル内に示している。

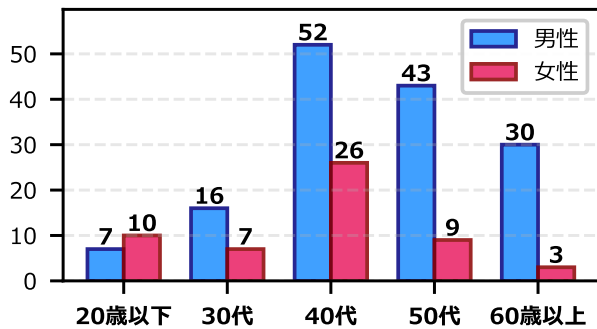


図3: 実験参加者の性別・年齢分布

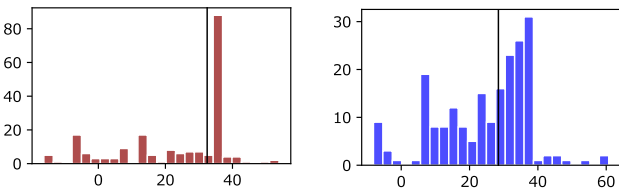


図4: 除外後のSVOスライダーによって測定した $SVO_p$ と $SVO_n$ の分布。それぞれの中央値を、縦軸と平行に引かれている黒の実線で示している。

$$y_{i,j} = \beta_0$$

$$\begin{aligned}
 & + \beta_1 \cdot \text{PROSELF}_x \\
 & + \beta_2 \cdot \text{PROSOCIAL}_x \\
 & + \beta_3 \cdot \text{利己損失強調} \\
 & + \beta_4 \cdot \text{利他損失強調} \\
 & + \beta_5 \cdot \text{多数派帰属強調} \\
 & + \beta_6 \cdot \text{少数派脱却強調} \\
 & + \beta_7 \cdot \text{PROSELF}_x \times \text{利己損失強調} \\
 & + \beta_8 \cdot \text{PROSELF}_x \times \text{利他損失強調} \\
 & + \beta_9 \cdot \text{PROSELF}_x \times \text{多数派帰属強調} \\
 & + \beta_{10} \cdot \text{PROSELF}_x \times \text{少数派脱却強調} \\
 & + \beta_{11} \cdot \text{PROSOCIAL}_x \times \text{利己損失強調} \\
 & + \beta_{12} \cdot \text{PROSOCIAL}_x \times \text{利他損失強調} \\
 & + \beta_{13} \cdot \text{PROSOCIAL}_x \times \text{多数派帰属強調} \\
 & + \beta_{14} \cdot \text{PROSOCIAL}_x \times \text{少数派脱却強調} \\
 & + \epsilon
 \end{aligned}
 \tag{1}$$

**目的変数:** 本研究の目的変数  $y_{i,j}$  は、介入前後の反応の差分を離散化して定義する。具体的には、まず以下の式で結果の差分を算出する。

$$y_{i,j} = \begin{cases} 1 & \text{介入前の行動意向} > \text{介入後の行動意向} \\ 0 & \text{介入前の行動意向} = \text{介入後の行動意向} \\ -1 & \text{介入前の行動意向} < \text{介入後の行動意向} \end{cases} \tag{2}$$

ここで、 $i$ は投稿テーマ(野球・誹謗中傷あり/なし, サッカー・誹謗中傷あり/なし)種類を表す。また、 $j$ は反応の種類(いいね, リポスト, 引用ポスト)を表す。 $y_{i,j}$ は $\{-1, 0, 1\}$ の3種類の値をとることになり、1は介入メッセージによって介入後の行動意向が介入前に比べて低下したことを示し、-1は行動意向が上昇したことを示す。0は行動意向に介入前後で変化がなかったことを示す。本研究では計12種類(投稿テーマ4種類×行動意向3種類)の目的変数 $y_{i,j}$ に対して重回帰分析を行う。誹謗中傷を含む投稿( $y_{1,1} \sim y_{1,3}, y_{3,1} \sim y_{3,3}$ )は説明変数の回帰係数が正の値となっている場合に、誹謗中傷拡散の抑制効果があることを示す。一方で、誹謗中傷を含まない投稿( $y_{2,1} \sim y_{2,3}, y_{4,1} \sim y_{4,3}$ )の回帰係数が正の値となっている場合に、副作用があることを示す。

**説明変数:** 各説明変数については以下の通り。

- **PROSELF<sub>x</sub>, PROSOCIAL<sub>x</sub>:** 個人の社会選好に基づく分類を示すダミー変数であり、利己的(PROSELF)な場合は1, 利他的(PROSOCIAL)な場合は1, それ以外の場合は0として定義する。社会選好の分類は、前節で示したSVOスコアの中央値を基準に行う。 $x$ は利得配分局面 $p$ または損失回避局面 $n$ を表す。
- **利己損失強調, 利他損失強調, 多数派帰属強調, 少数派脱却強調:** 各介入群を示すダミー変数であり、介入群に属する場合は1, それ以外の場合は0として定義する。
- **交互作用項:** 個人の社会選好に基づく分類と各介入群の交互作用を示す変数であり、例えばPROSELF<sub>x</sub>×利己損失強調は、利己的な参加者が利己損失強調介入群に属する場合に1, それ以外の場合は0として定義する。

以上により、本研究では各目的変数 $y_{i,j}$ に対して、個人の社会選好、各介入群、および個人の社会選好と各介入群の交互作用が行動意向の差分に与える影響を分析した。

#### 4.2.2 重回帰分析の結果

重回帰分析の結果を表4～7に示す。表は投稿テーマごとに  $SVO_p$ ,  $SVO_n$  をそれぞれ説明変数に用いた場合の分析結果である。各表には、説明変数ごとに回帰係数、標準誤差を示している。

表4に示すように、テーマが野球で誹謗中傷を含む投稿に対して  $SVO_p$  を説明変数に用いた場合、「いいね」において  $PROSOCIAL_p$  が回帰係数 0.1023 であり、 $p < 0.01$  で有意な抑制効果が見られた。また、利他損失強調  $\times$   $PROSOCIAL_p$  の交互作用項で回帰係数が 0.1748 であり、 $p < 0.05$  で有意な抑制効果が見られた。

表5に示すように、テーマがサッカーで誹謗中傷を含む投稿に対して  $SVO_p$  を説明変数に用いた場合、「引用ポスト」において利他損失強調が回帰係数 0.0729 であり、 $p < 0.05$  で有意な抑制効果が見られた。また、誹謗中傷を含まない投稿に対しても、「引用ポスト」において利他損失強調が回帰係数 0.1015 であり、 $p < 0.01$  で有意な副作用が確認された。さらに、誹謗中傷を含まない投稿の「引用ポスト」において、利己損失強調  $\times$   $PROSOCIAL_p$  の交互作用項が回帰係数 -0.1155 であり、 $p < 0.05$  で有意な効果が見られた。

表6に示すように、テーマが野球で誹謗中傷を含む投稿に対して  $SVO_n$  を説明変数に用いた場合、「いいね」において  $PROSOCIAL_n$  が回帰係数 0.1161 であり、 $p < 0.01$  で有意な抑制効果が見られた。また、「リポスト」においても  $PROSOCIAL_n$  が回帰係数 0.0693 であり、 $p < 0.05$  で有意な抑制効果が確認された。さらに、「引用ポスト」においても  $PROSOCIAL_n$  が回帰係数 0.1056 であり、 $p < 0.01$  で有意な抑制効果が見られた。一方で、誹謗中傷を含まない投稿の「いいね」において、利他損失強調  $\times$   $PROSOCIAL_n$  の交互作用項が回帰係数 0.2101 であり、 $p < 0.01$  で有意な副作用が確認された。

表7に示すように、テーマがサッカーで誹謗中傷を含む投稿に対して  $SVO_n$  を説明変数に用いた場合、「いいね」において  $PROSOCIAL_n$  が回帰係数 0.0721 であり、 $p < 0.05$  で有意な抑制効果が見られた。また、「リポスト」において  $PROSOCIAL_n$  が回帰係数 0.0849 であり、 $p < 0.01$  で有意な抑制効果が確認された。一方で、誹謗中傷を含まない投稿に対しては、「リポスト」において  $PROSOCIAL_n$  が回帰係数 0.0791 であり、 $p < 0.01$  で有意な副作用が見られた。さらに、誹謗中傷を含まない投稿の「引用ポスト」において、利他損失強調  $\times$   $PROSOCIAL_n$  の交互作用項が回帰係数 0.1166 であり、 $p < 0.05$  で有意な副作用が確認された。

## 5 考 察

### 5.1 社会選好が持つ誹謗中傷拡散の抑制効果

本研究では、SNS 上の誹謗中傷投稿への同調・拡散を抑制するために、個人の社会選好 (Social Value Orientation; SVO) を考慮した介入戦略の有効性を検証した。特に、利得局面で測定される  $SVO_p$  と損失局面で測定される  $SVO_n$  の差異に着目し、社会選好が誹謗中傷拡散行動に与える影響を分析した。

分析結果では、利得局面の社会選好タイプ ( $PROSOCIAL_p$ ) において、テーマが野球で誹謗中傷を含む投稿に対する「いいね」に有意な抑制効果が確認された。具体的には、 $PROSOCIAL_p$  が回帰係数 0.1023 ( $p < 0.01$ ) となり、利他的傾向を持つ参加者ほど誹謗中傷投稿への拡散意向が低下する傾向が示された。さらに、利他損失強調メッセージと  $PROSOCIAL_p$  の交互作用項においても回帰係数 0.1748 ( $p < 0.05$ ) と有意な効果が見られた。

一方で、損失局面で分類した社会選好タイプ ( $PROSOCIAL_n$ ) は、より広範に抑制効果を示した。テーマが野球で誹謗中傷を含む投稿に対して、「いいね」では回帰係数 0.1161 ( $p < 0.01$ )、「リポスト」では 0.0693 ( $p < 0.05$ )、「引用ポスト」では 0.1056 ( $p < 0.01$ ) と複数の行動意向において有意な抑制効果が確認された。また、テーマがサッカーの場合でも、「いいね」(0.0721,  $p < 0.05$ ) や「リポスト」(0.0849,  $p < 0.01$ ) において抑制効果が見られた。

このように、利得局面よりも損失局面で測定した社会選好の方が誹謗中傷拡散抑制に寄与していたことから、本研究で提示した介入メッセージが損失回避を強調する内容であった点が影響している可能性がある。損失局面の SVO による分類は、誹謗中傷拡散抑制においてより適切な個別化指標となり得ることが示唆された。

しかしながら、本研究で導入した社会選好タイプと介入メッセージの交互作用項については、誹謗中傷投稿に対する有意な抑制効果がほとんど確認されなかった。交互作用項が有意となったのは誹謗中傷を含まない投稿に対する副作用としての結果であり、期待された「社会選好に応じたメッセージの出し分けによる誹謗中傷抑制」は十分に実証されなかった。

抑制効果が見られなかった要因として、第一にサンプルサイズの制約による統計的検出力不足が考えられる。本研究では分析対象者が 203 名に絞られており、社会選好タイプと介入群の組み合わせごとの人数が少なくなったことで、交互作用効果を検出することが困難であった可能性がある。第二に、社会規範メッセージのように短い文面で受け手に規範意識を十分伝えることが難しく、個別化しても効果が現れにくかったことが考えられる。以上より、社会選好は個別化介入の手がかりとなり得る一方で、メッセージ設計との適合性を慎重に検討する必要がある。

### 5.2 誹謗中傷を含まない投稿の拡散を抑制する副作用

本研究では、誹謗中傷投稿への拡散抑制を目的とした介入メッセージが、誹謗中傷を含まない通常投稿に対しても拡散意向を低下させる副作用を引き起こす可能性についても検討した。先行研究 [8] でも指摘されていたように、介入メッセージが誹謗中傷投稿のみを選択的に抑制することは難しく、通常投稿への拡散行動まで抑制してしまうことが課題となる。

分析結果では、誹謗中傷を含まない投稿に対しても有意な効果が複数確認された。例えば、テーマがサッカーの場合、「引用ポスト」において利他損失強調メッセージが回帰係数 0.1015 ( $p < 0.01$ ) となり、誹謗中傷を含まない投稿の拡散意向を抑制する副作用が見られた。また、損失局面の社会選好に基づく交

相互作用項でも、誹謗中傷を含まない投稿の「いいね」において利他損失強調  $\times$  PROSOCIAL<sub>n</sub> が回帰係数 0.2101 ( $p < 0.01$ ) となり、有意な副作用が確認された。さらに、テーマがサッカーで誹謗中傷を含まない投稿に対しても、「リポスト」で PROSOCIAL<sub>n</sub> が 0.0791 ( $p < 0.01$ ) と有意であり、利他的傾向を持つ参加者ほど通常投稿への拡散も抑制される可能性が示された。

これらの結果から、介入メッセージは誹謗中傷拡散を抑制するだけでなく、SNS 上の一般的な情報共有行動まで萎縮させるリスクを持つことが明らかになった。特に利他性や損失回避を強調するメッセージは、受け手に対して「拡散行動そのものが望ましくない」という印象を与える可能性がある。

また、本研究の実験デザインは先行研究と異なり、介入前後の行動意向を測定し差分を目的変数とする形式を採用した。この設計の利点として、介入メッセージが参加者の意向にどのような変化を与えたかを直接的に捉えられる点が挙げられる。介入前の意向を基準とすることで、個人差を考慮したより詳細な分析が可能となった。

一方で欠点として、意向変化を  $\{-1, 0, 1\}$  に離散化したことで情報が粗くなり、微細な変化が捨棄される可能性がある。また、介入前の意向自体が低い参加者が多く、0%回答者を除外する必要が生じたため、サンプルサイズが縮小した。さらに、「意向の変化」を測定することは実際の行動とは必ずしも一致しないため、現実の SNS 環境での拡散抑制効果を推定するには限界が残る。

以上より、本研究は介入メッセージによる誹謗中傷拡散抑制の可能性を示すと同時に、通常投稿への副作用や実験デザイン上の課題を明らかにした。今後は誹謗中傷投稿のみを選択的に抑制できる介入方法の設計や、実際の行動データを用いた検証が求められる。

### 5.3 本研究の限界と課題

**サンプルサイズの制約：**本研究ではスクリーニングで条件を満たした実験参加希望者に本実験に参加してもらった。しかし、参加希望者 15,000 名に対してスクリーニング条件を満たした実験参加者が 400 名で、その中で適切な回答ができた参加者が 203 名とさらにサンプルサイズが小さくなったため、統計的検出力が不足している可能性がある。スポーツという興味に限定的であるテーマを扱ったのは、興味の有無がはっきりしているテーマの方がより適切な参加者を募集できると考えたためであるが、ここまで絞られてしまった原因は、参加者を募集したクラウドソーシングサービスでのスポーツへの興味関心が低かったからだと思われる。考察にも挙げたが、今回の実験設計でより効果的な分析結果を得るためには、より多くの参加者を集め、説明変数を増やすことで様々な要因を考慮できる必要がある。

**若年層のサンプル不足：**本研究の実験参加者は若年層が少なく、特に 18 歳から 20 歳の参加者が非常に少なかった。若年層は SNS 利用率が高く、誹謗中傷投稿の拡散行動にも大きく関連があると考えられる。Ungaretti らによると [20]、若年層ほど加害経験が多いという研究結果もあるため、今後同様の実験を行う際は若年層の参加者をより多く集めることが重要であると

考えられる。

**誹謗中傷の定義：**先行研究では、誹謗中傷を特定の単語を含むかどうかで定義していたが、本研究ではより多様な誹謗中傷を実験に使用するため、誹謗中傷の 9 つの定義 [19] を用いて投稿を選択・設計し、反応測定を行うという実験を行った。しかし、今回の実験での誹謗中傷の捉え方が参加者にとって一様だったかは不明である。誹謗中傷の認識は個人差が大きく、文化的背景や価値観によっても異なるため、今後の研究では、より明確な定義や具体的な例を提示することで、参加者の理解を統一する必要がある。例えば、誹謗中傷の定義として、本研究でも投稿を設定することに使用した 9 つの定義を参加者に事前に説明するなどが考えられるが、詳細に説明しすぎると実験の自然性の欠如や本来のプラットフォームでの行動との乖離が顕著になる可能性もあるため、そのバランスを考慮する必要がある。

### 5.4 今後の展望

**異なる目的変数の検証：**介入前後の反応の差分を離散化したもののみを目的変数とするのではなく、介入前の反応を考慮しない最終的に投稿を拡散したかどうかの結果だけに着目した別の目的変数の設定などを試し、今回の実験手法で得られた結果と比較することなどが考えられる。特に、結果だけに着目した場合は誹謗中傷投稿の拡散を有意に抑制できる組み合わせが存在するかを再度検証する必要があると考えている。

**参加者の増加・テーマの多様化：**より多くの参加者を集めることで、年齢や性別、年収や生活の満足度など誹謗中傷に関連性がありそうな要素を説明変数として分析に使用できるようになる。また、誹謗中傷に関するテーマをスポーツ以外にも多様化する工夫も今後行っていきたい。例えば、政治、芸能、社会問題など、スポーツ以外の誹謗中傷が発生しやすいテーマを取り入れることが考えられる。

**個人の社会選好以外の個人傾向との比較：**本研究では、個人の社会選好の傾向に基づいて介入メッセージの効果を検証したが、今後は他の個人傾向（例えば、性格特性、価値観、過去の SNS 利用経験など）と比較し、どの個人傾向が誹謗中傷拡散行動に最も影響を与えるかを明らかにすることを検討している。

## 6 おわりに

本研究は、SNS 上の誹謗中傷拡散を抑制するため、個人の社会選好 (SVO) を考慮したメッセージ介入の有効性を検証することを目的として実施したオンライン実験である。利己損失強調、利他損失強調、多数派帰属強調、少数派脱却強調の 4 種類の介入を用い、 $SVO_p$  および  $SVO_n$  の両面から効果を分析した。

主な知見は以下のとおりである。今回の介入前後の反応の差分を離散化したものを目的変数とした分析では、介入群のみ、社会選好タイプのみを考慮する場合には有意な抑制効果を持つものが一部見られた。また、社会選好タイプは損失局面で分類した方がより効果的に介入メッセージの出し分けができた。一方で誹謗中傷拡散を抑制する有意な効果を持つ介入メッセージ

と個人の社会選好傾向の組み合わせはなく、有意だった組み合わせは誹謗中傷を含まない通常投稿の拡散まで抑制してしまう副作用としての効果だった。

これらの結果から、社会選好は個別化の有力な手がかりになりうる値ではある一方で、メッセージ内容との適合度を十分に検討したうえで個別化指標であることが分かった。しかし現状の実験設計では相性の良い組み合わせの場合であっても抑制効果が副作用としてのみ確認されたため、今後はより効果的な介入メッセージの設計や異なる要素の導入を検討する必要がある。

本研究はナッジメッセージを SVO による個別化を利用して出し分けることで効果的な誹謗中傷拡散の抑制が可能かを検証しメッセージとの適合度が高い社会選好タイプは有意な抑制効果を持ち、個人の社会選好と介入メッセージの組み合わせによって有意な抑制効果を示すものが存在するということが分かったため、いくつかの課題を残してはいるものの SNS 上の有害情報拡散抑制に向けた個別化介入の可能性を示す第一歩となった。今後も多様な個人の属性や個人差指標を活用した介入手法の開発と検証を進めたい。

## 謝 辞

本研究の一部は、JSPS 科研費 24K03046, 24K23877 の助成を受けたものです。ここに記して謝意を示します。

## 文 献

- [1] 総務省 2025. 令和 7 年版情報通信白書 第 1 部 第 2 章 第 3 節 インターネット上の偽・誤情報等への対応, 2025. <https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r07/html/nd123100.html> (参照 2026-01-05).
- [2] 総務省 2023. 令和 5 年度インターネット上の違法・有害情報対応相談業務等請負業務報告書 (概要版). [https://www.soumu.go.jp/main\\_content/000946765.pdf](https://www.soumu.go.jp/main_content/000946765.pdf) (参照 2026-01-05).
- [3] Masanori Takano, Fumiaki Taka, Chiki Ogiue, and Natsuki Nagata. Online harassment of japanese celebrities and influencers. *Frontiers in psychology*, Vol. 15, p. 1386146, 2024.
- [4] Ruth Jeong, Megan Gilbertson, Logan N Riffle, and Michelle K Demaray. Participant role behavior in cyberbullying: An examination of moral disengagement among college students. *International journal of bullying prevention*, Vol. 6, No. 1, pp. 28–40, 2024.
- [5] 総務省. プロバイダ責任制限法の一部を改正する法律 (概要), 2022. [https://www.soumu.go.jp/main\\_content/000836903.pdf](https://www.soumu.go.jp/main_content/000836903.pdf) (参照 2026-01-06).
- [6] 西谷千乃与, 安藤一秋. アスリートに対する sns 上の誹謗中傷の分析と検出. Web インテリジェンスとインタラクション研究会 予稿集 第 20 回研究会, pp. 25–32. Web インテリジェンスとインタラクション研究会, 2024.
- [7] 正木博明, 柴田健吾, 星野秀偉, 石濱嵩博, 齋藤長行, 矢谷浩司. 若年層 sns ユーザに対するプライバシー・安全上の行動に関するナッジの大規模定量調査. 情報処理学会論文誌, Vol. 61, No. 12, pp. 1892–1902, 2020.
- [8] 村田拓介, 秋田航汰, 小野祐紀, 河合求真, 森実輝. Sns 上の誹謗中傷投稿への同調に関する分析: ナッジメッセージによる誹謗中傷抑制の試み. 大阪大学経済学, Vol. 73, No. 4, pp. 55–55, 2024.
- [9] Susan Athey, Niall Keleher, and Jann Spiess. Machine learning who to nudge: causal vs predictive targeting in a field experiment on student financial aid renewal. *Journal of Econometrics*, Vol. 249, p. 105945, 2025.
- [10] Benjamin E Hilbig, Isabel Thielmann, Ingo Zettler, and Morten Moshagen. The dispositional essence of proactive social preferences:

- The dark core of personality vis-à-vis 58 traits. *Psychological Science*, Vol. 34, No. 2, pp. 201–220, 2023.
- [11] Ryan O. Murphy, Kurt A. Ackermann, and Michel Handgraaf. Measuring social value orientation. *Judgment and Decision Making*, Vol. 6, No. 8, pp. 771–781, 2011.
  - [12] Ziqi Zhang, David Robinson, and Jonathan Tepper. Detecting hate speech on twitter using a convolution-gru based deep neural network. In *European semantic web conference*, pp. 745–760. Springer, 2018.
  - [13] Thomas Davidson, Dana Warmsley, Michael Macy, and Ingmar Weber. Automated hate speech detection and the problem of offensive language. In *Proceedings of the international AAAI conference on web and social media*, Vol. 11, pp. 512–515, 2017.
  - [14] Richard H. Thaler and Cass R. Sunstein. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press, New Haven, CT, 2008.
  - [15] Kevin Munger. Tweetment effects on the tweeted: Experimentally reducing racist harassment. *Political Behavior*, Vol. 39, No. 3, pp. 629–649, 2017.
  - [16] Jan Luca Pletzer, Daniel Balliet, Jeff Joireman, D Michael Kuhlman, Sven C Voelpel, and Paul AM Van Lange. Social value orientation, expectations, and cooperation in social dilemmas: A meta-analysis. *European Journal of Personality*, Vol. 32, No. 1, pp. 62–83, 2018.
  - [17] Alex Williams and Li Chen. Rapid diffusion of sports-related abusive posts after events. *Journal of Online Behavior*, Vol. 12, No. 1, pp. 15–32, 2025.
  - [18] 笹川スポーツ財団. スポーツ観戦率:スタジアムや体育館などでの直接スポーツ観戦率は 26.2%, 12 2025. [https://www.ssf.or.jp/thinktank/sports\\_life/data/sportsspectating.html](https://www.ssf.or.jp/thinktank/sports_life/data/sportsspectating.html) (参照 2026-01-07).
  - [19] 国際大学グローバル・コミュニケーション・センター (GLOCOM). Innovation nippon 2022: ジャーナリストへの誹謗中傷の実態 報告書〈概要版〉. Technical report, 国際大学グローバル・コミュニケーション・センター (GLOCOM), April 2023. 概要版 PDF, 調査報告書.
  - [20] Joaquín Ungaretti, Talía Gómez Yepes, María Laura Sánchez Pujalte, and Edgardo Etchezahar. Cyberbullying perpetration among spanish adults: The roles of fear of missing out and critical thinking. *Societies*, Vol. 15, No. 9, p. 249, 2025.

表 4: テーマが野球であり, かつ  $SVO_p$  を説明変数に用いたときの重回帰分析の結果, 推定された回帰係数. 表内の値は回帰係数と括弧内に標準誤差を示す. 有意な値には,  $p < 0.05$  のとき下線,  $p < 0.01$  のとき太字を付す.

説明変数	いいね		リポスト		引用ポスト	
	誹謗中傷あり	誹謗中傷なし	誹謗中傷あり	誹謗中傷なし	誹謗中傷あり	誹謗中傷なし
切片	<b>0.1104 (0.021)</b>	<b>0.0862 (0.024)</b>	<b>0.0927 (0.021)</b>	0.0350 (0.021)	0.0289 (0.019)	0.0332 (0.020)
利己損失強調	0.0458 (0.047)	-0.0371 (0.053)	0.0176 (0.046)	0.0220 (0.045)	0.0286 (0.041)	0.0086 (0.043)
利他損失強調	0.0480 (0.044)	0.0863 (0.050)	0.0293 (0.043)	-0.0252 (0.043)	0.0174 (0.039)	-0.0100 (0.041)
多数派帰属強調	0.0115 (0.048)	0.0281 (0.055)	0.0267 (0.048)	0.0317 (0.047)	-0.0130 (0.043)	0.0227 (0.045)
少数派脱却強調	0.0050 (0.048)	0.0089 (0.055)	0.0191 (0.047)	0.0064 (0.047)	-0.0042 (0.043)	0.0119 (0.045)
PROSELF <sub>p</sub>	0.0081 (0.034)	0.0337 (0.039)	0.0480 (0.034)	0.0482 (0.033)	-0.0171 (0.030)	0.0049 (0.032)
PROSOCIAL <sub>p</sub>	<b>0.1023 (0.034)</b>	0.0525 (0.038)	0.0447 (0.033)	-0.0132 (0.033)	0.0460 (0.030)	0.0283 (0.031)
利己損失強調 × PROSELF <sub>p</sub>	0.0629 (0.076)	0.0990 (0.086)	0.0691 (0.075)	0.0312 (0.074)	-0.0404 (0.067)	0.0442 (0.070)
利己損失強調 × PROSOCIAL <sub>p</sub>	-0.0171 (0.071)	-0.1361 (0.081)	-0.0515 (0.070)	-0.0092 (0.069)	0.0690 (0.063)	-0.0356 (0.066)
利他損失強調 × PROSELF <sub>p</sub>	0.0687 (0.067)	-0.0886 (0.076)	0.0359 (0.066)	-0.0286 (0.066)	0.0296 (0.059)	0.0014 (0.062)
利他損失強調 × PROSOCIAL <sub>p</sub>	-0.0207 (0.072)	<u>0.1749 (0.082)</u>	-0.0066 (0.071)	0.0034 (0.070)	-0.0122 (0.064)	-0.0114 (0.067)
多数派帰属強調 × PROSELF <sub>p</sub>	-0.0500 (0.075)	0.0520 (0.085)	-0.0473 (0.074)	0.0851 (0.073)	0.0012 (0.066)	0.0592 (0.069)
多数派帰属強調 × PROSOCIAL <sub>p</sub>	0.0615 (0.078)	-0.0239 (0.088)	0.0740 (0.077)	-0.0534 (0.076)	-0.0142 (0.069)	-0.0365 (0.072)
少数派脱却強調 × PROSELF <sub>p</sub>	-0.0735 (0.079)	-0.0288 (0.089)	-0.0097 (0.078)	-0.0396 (0.077)	-0.0076 (0.070)	-0.1000 (0.073)
少数派脱却強調 × PROSOCIAL <sub>p</sub>	0.0786 (0.073)	0.0376 (0.083)	0.0288 (0.072)	0.0460 (0.071)	0.0034 (0.065)	0.1118 (0.068)

表 5: テーマがサッカーであり, かつ  $SVO_p$  を説明変数に用いたときの重回帰分析の結果. 表内の値は回帰係数と括弧内に標準誤差を示す. 有意な値には,  $p < 0.05$  のとき下線,  $p < 0.01$  のとき太字を付す.

説明変数	いいね		リポスト		引用ポスト	
	誹謗中傷あり	誹謗中傷なし	誹謗中傷あり	誹謗中傷なし	誹謗中傷あり	誹謗中傷なし
切片	<b>0.0789 (0.022)</b>	0.0259 (0.022)	<u>0.0444 (0.020)</u>	0.0217 (0.019)	0.0126 (0.017)	-0.0055 (0.017)
利己損失強調	0.0809 (0.047)	-0.0374 (0.048)	0.0319 (0.043)	-0.0102 (0.041)	-0.0393 (0.036)	-0.0254 (0.037)
利他損失強調	0.0136 (0.045)	0.0435 (0.046)	-0.0311 (0.041)	0.0548 (0.038)	<u>0.0729 (0.034)</u>	<b>0.1015 (0.035)</b>
多数派帰属強調	0.0220 (0.049)	0.0458 (0.050)	0.0591 (0.045)	0.0075 (0.042)	0.0083 (0.038)	-0.0371 (0.039)
少数派脱却強調	-0.0376 (0.049)	-0.0259 (0.050)	-0.0154 (0.045)	-0.0303 (0.042)	-0.0293 (0.038)	-0.0445 (0.039)
PROSELF <sub>p</sub>	0.0411 (0.035)	0.0157 (0.035)	0.0078 (0.032)	-0.0278 (0.030)	-0.0124 (0.027)	-0.0138 (0.027)
PROSOCIAL <sub>p</sub>	0.0378 (0.034)	0.0102 (0.035)	0.0367 (0.031)	0.0495 (0.029)	0.0250 (0.026)	0.0083 (0.027)
利己損失強調 × PROSELF <sub>p</sub>	0.0717 (0.077)	-0.0042 (0.079)	0.0069 (0.071)	0.0163 (0.066)	-0.0064 (0.059)	0.0901 (0.061)
利己損失強調 × PROSOCIAL <sub>p</sub>	0.0092 (0.072)	-0.0332 (0.074)	0.0250 (0.066)	-0.0265 (0.062)	-0.0329 (0.056)	<u>-0.1155 (0.057)</u>
利他損失強調 × PROSELF <sub>p</sub>	-0.0160 (0.069)	0.0031 (0.070)	-0.0211 (0.063)	-0.0192 (0.059)	0.1034 (0.053)	0.0060 (0.054)
利他損失強調 × PROSOCIAL <sub>p</sub>	0.0296 (0.073)	0.0404 (0.074)	-0.0100 (0.067)	0.0740 (0.063)	-0.0305 (0.056)	0.0956 (0.058)
多数派帰属強調 × PROSELF <sub>p</sub>	0.0179 (0.076)	0.0325 (0.077)	0.0087 (0.070)	0.0386 (0.065)	-0.0885 (0.058)	-0.0237 (0.060)
多数派帰属強調 × PROSOCIAL <sub>p</sub>	0.0041 (0.079)	0.0133 (0.081)	0.0503 (0.072)	-0.0311 (0.068)	0.0969 (0.061)	-0.0134 (0.063)
少数派脱却強調 × PROSELF <sub>p</sub>	-0.0325 (0.080)	-0.0157 (0.082)	0.0132 (0.073)	-0.0635 (0.069)	-0.0209 (0.062)	-0.0862 (0.064)
少数派脱却強調 × PROSOCIAL <sub>p</sub>	-0.0051 (0.075)	-0.0102 (0.076)	-0.0286 (0.068)	0.0332 (0.064)	-0.0084 (0.057)	0.0417 (0.059)

表6: テーマが野球であり, かつ  $SVO_n$  を説明変数に用いたときの重回帰分析の結果. 表内の値は回帰係数と括弧内に標準誤差を示す. 有意な値には,  $p < 0.05$  のとき下線,  $p < 0.01$  のとき太字を付す.

説明変数	いいね		リポスト		引用ポスト	
	誹謗中傷あり	誹謗中傷なし	誹謗中傷あり	誹謗中傷なし	誹謗中傷あり	誹謗中傷なし
切片	<b>0.1149 (0.021)</b>	<b>0.0790 (0.024)</b>	<b>0.0913 (0.021)</b>	0.0340 (0.021)	0.0299 (0.018)	0.0356 (0.020)
利己損失強調	0.0461 (0.046)	-0.0355 (0.052)	0.0195 (0.045)	0.0136 (0.045)	0.0270 (0.040)	0.0001 (0.043)
利他損失強調	0.0387 (0.043)	0.0707 (0.049)	0.0252 (0.043)	-0.0197 (0.043)	0.0059 (0.038)	-0.0109 (0.041)
多数派帰属強調	0.0151 (0.048)	0.0250 (0.055)	0.0231 (0.047)	0.0329 (0.047)	-0.0020 (0.042)	0.0241 (0.045)
少数派脱却強調	0.0150 (0.047)	0.0188 (0.054)	0.0235 (0.046)	0.0071 (0.046)	-0.0010 (0.041)	0.0223 (0.044)
PROSELF <sub>n</sub>	-0.0012 (0.033)	0.0638 (0.038)	0.0220 (0.033)	0.0438 (0.033)	<u>-0.0757 (0.029)</u>	-0.0060 (0.031)
PROSOCIAL <sub>n</sub>	<b>0.1161 (0.034)</b>	0.0152 (0.038)	<u>0.0693 (0.033)</u>	-0.0098 (0.033)	<b>0.1056 (0.029)</b>	0.0416 (0.031)
利己損失強調 × PROSELF <sub>n</sub>	0.1446 (0.074)	0.0232 (0.085)	0.1281 (0.073)	-0.0914 (0.073)	-0.0247 (0.064)	-0.0297 (0.069)
利己損失強調 × PROSOCIAL <sub>n</sub>	-0.0985 (0.071)	-0.0587 (0.081)	-0.1086 (0.070)	0.1051 (0.070)	0.0517 (0.061)	0.0298 (0.066)
利他損失強調 × PROSELF <sub>n</sub>	-0.0042 (0.070)	-0.1394 (0.080)	-0.1015 (0.069)	0.0160 (0.068)	-0.0713 (0.061)	0.0554 (0.065)
利他損失強調 × PROSOCIAL <sub>n</sub>	0.0429 (0.067)	<b>0.2101 (0.077)</b>	0.1267 (0.066)	-0.0357 (0.066)	0.0772 (0.058)	-0.0663 (0.063)
多数派帰属強調 × PROSELF <sub>n</sub>	-0.0547 (0.073)	0.0915 (0.083)	0.0488 (0.072)	0.0374 (0.072)	-0.0263 (0.063)	0.0203 (0.068)
多数派帰属強調 × PROSOCIAL <sub>n</sub>	0.0698 (0.080)	-0.0666 (0.091)	-0.0257 (0.078)	-0.0045 (0.078)	0.0243 (0.069)	0.0039 (0.074)
少数派脱却強調 × PROSELF <sub>n</sub>	-0.0870 (0.074)	0.0884 (0.085)	-0.0534 (0.073)	0.0818 (0.073)	0.0467 (0.064)	-0.0520 (0.069)
少数派脱却強調 × PROSOCIAL <sub>n</sub>	0.1019 (0.075)	-0.0696 (0.086)	0.0769 (0.074)	-0.0747 (0.074)	-0.0477 (0.065)	0.0743 (0.070)

表7: テーマがサッカーであり, かつ  $SVO_n$  を説明変数に用いたときの重回帰分析の結果, 推定された回帰係数. 表内の値は回帰係数と括弧内に標準誤差を示す. 有意な値には,  $p < 0.05$  のとき下線,  $p < 0.01$  のとき太字を付す.

説明変数	いいね		リポスト		引用ポスト	
	誹謗中傷あり	誹謗中傷なし	誹謗中傷あり	誹謗中傷なし	誹謗中傷あり	誹謗中傷なし
切片	<b>0.0776 (0.022)</b>	0.0232 (0.022)	<u>0.0425 (0.020)</u>	0.0184 (0.018)	0.0154 (0.017)	-0.0087 (0.017)
利己損失強調	0.0834 (0.047)	-0.0377 (0.048)	0.0290 (0.043)	-0.0116 (0.040)	-0.0418 (0.036)	-0.0270 (0.037)
利他損失強調	0.0038 (0.044)	0.0393 (0.045)	-0.0378 (0.040)	0.0403 (0.037)	<u>0.0738 (0.034)</u>	<b>0.0940 (0.035)</b>
多数派帰属強調	0.0244 (0.049)	0.0437 (0.050)	0.0648 (0.045)	0.0063 (0.042)	0.0126 (0.038)	-0.0439 (0.039)
少数派脱却強調	-0.0341 (0.048)	-0.0220 (0.049)	-0.0135 (0.044)	-0.0166 (0.041)	-0.0292 (0.037)	-0.0318 (0.038)
PROSELF <sub>n</sub>	0.0055 (0.034)	-0.0190 (0.035)	-0.0425 (0.031)	<u>-0.0608 (0.029)</u>	-0.0250 (0.026)	-0.0256 (0.027)
PROSOCIAL <sub>n</sub>	<u>0.0721 (0.034)</u>	0.0422 (0.035)	<b>0.0849 (0.031)</b>	<b>0.0791 (0.029)</b>	0.0403 (0.027)	0.0168 (0.027)
利己損失強調 × PROSELF <sub>n</sub>	0.1379 (0.075)	-0.0100 (0.077)	-0.0290 (0.069)	-0.0329 (0.064)	0.0079 (0.059)	0.0613 (0.060)
利己損失強調 × PROSOCIAL <sub>n</sub>	-0.0545 (0.072)	-0.0277 (0.074)	0.0580 (0.066)	0.0213 (0.061)	0.0268 (0.053)	-0.0883 (0.057)
利他損失強調 × PROSELF <sub>n</sub>	-0.1240 (0.071)	-0.0435 (0.072)	-0.0733 (0.065)	-0.0720 (0.060)	0.0470 (0.055)	-0.0227 (0.056)
利他損失強調 × PROSOCIAL <sub>n</sub>	0.1278 (0.068)	0.0828 (0.070)	0.0355 (0.062)	0.1122 (0.058)	0.0268 (0.053)	<u>0.1166 (0.054)</u>
多数派帰属強調 × PROSELF <sub>n</sub>	0.0406 (0.074)	0.1002 (0.076)	0.0463 (0.068)	0.1101 (0.063)	-0.0770 (0.058)	0.0782 (0.059)
多数派帰属強調 × PROSOCIAL <sub>n</sub>	-0.0162 (0.081)	-0.0565 (0.082)	0.0184 (0.074)	-0.1038 (0.068)	0.0896 (0.063)	-0.1221 (0.064)
少数派脱却強調 × PROSELF <sub>n</sub>	-0.0490 (0.075)	-0.0656 (0.077)	0.0135 (0.069)	-0.0660 (0.064)	-0.0028 (0.059)	-0.1423 (0.060)
少数派脱却強調 × PROSOCIAL <sub>n</sub>	0.0149 (0.076)	0.0436 (0.078)	-0.0269 (0.070)	0.0495 (0.065)	-0.0264 (0.059)	0.1106 (0.060)