一般発表 | Track 1: 自然言語処理・機械学習基礎

2026年3月1日(日) 9:30 ~ 11:40 | C会場

[4C] 実世界応用

座長:軽部 幸起(電気通信大学)　コメンテータ:熊本 忠彦(千葉工業大学)　ジュニアコメンテータ:吉丸 直希(同志社大学)

9:30 ~ 9:55

[4C-01] ECサイトに潜むソーシャルエンジニアリング：ローカルSLMを用いたダークパターン検知

*橋本 拓也[1]、服部 峻[2]、宮城 茂幸[2] (1. 滋賀県立大学大学院、2. 滋賀県立大学先端工学研究院)

9:55 ~ 10:15

[4C-02] スマホ撮影画像への応用を見据えたナンバープレート検出モデル性能に関する予備調査

*宮木 笙伍[1]、河合 由起子[2]、栗 達[1] (1. 京都産業大学、2. 関西大学)

10:15 ~ 10:35

[4C-03] 古文単語学習の効率化のための単語色彩表現システム

*酒井 楓佳[1]、小宮 和真[2]、福原 義久[1] (1. 武蔵野大学データサイエンス学部データサイエンス学科、2. 武蔵野大学データサイエンス研究科データサイエンス専攻)

10:35 ~ 10:55

[4C-04] 自由記述嗜好を考慮した多人数レシピ推薦における嗜好統合手法の比較分析

*東 颯人[1]、宮森 恒[1] (1. 京都産業大学)

10:55 ~ 11:20

[4C-05] 強化学習による差動二輪車制御における未知実証環境での安全推論

*門垣 幸樹[1]、高井 勇志[2]、宮口 幹太[2]、北野 信吾[2]、大島 裕明[1] (1. 兵庫県立大学、2. 株式会社竹中工務店)

ECサイトに潜むソーシャルエンジニアリング：ローカルSLMを用いたダークパターン検知

橋本　拓也†　　服部　　峻††　　宮城　茂幸††

† 滋賀県立大学大学院　〒 522–8533 滋賀県彦根市八坂町 2500
†† 滋賀県立大学先端工学研究院　〒 522–8533 滋賀県彦根市八坂町 2500
E-mail: †to23thashimoto@ec.usp.ac.jp, ††{hattori.s,miyagi.s}@e.usp.ac.jp

あらまし　近年，デジタルサービスにおいてユーザの意思決定を不当に操作し不利益をもたらす UI デザインである「ダークパターン」が社会問題化している．これに対し，大規模言語モデル (LLM) を用いた自動検知手法の研究が進められているが，既存手法の多くは API 経由で利用する SaaS 型 LLM に依存しており，コストや再現性の観点から大規模な実態調査への適用が困難である．そこで本研究では，ローカル環境で動作する小規模言語モデル (SLM) を活用した，低コストかつ自律的なダークパターン検知手法を提案する．提案手法では，レンダリング結果から視覚情報を抽出し HTML 属性として埋め込む「視覚情報埋め込みを伴う構造保持型 DOM 分割」アルゴリズムを導入することで，軽量なモデルでも文脈を考慮した解析を目指す．本稿では，日本の EC サイトを対象に収集したデータセットに対し，4bit 量子化された Qwen2.5-Coder-7B を用いた評価実験を行い，VRAM が 16GB 程度の環境における提案手法の有効性と実用性を検証する．

キーワード　ダークパターン，EC サイト，UI，SLM

1　はじめに

現代のデジタル社会において，Web サービスやアプリケーションにおける UI (User Interface) および UX (User Experience) 設計の重要性が高まっている．洗練された UI/UX はユーザビリティを向上させ，ユーザに利益をもたらすが，その一方で事業者の利益を優先し，ユーザに不利益な意思決定を意図的に誘導する設計手法への懸念も広がっている．このような設計は「ダークパターン (Dark Patterns)」と呼ばれ，2010 年に Harry Brignull によって提唱された概念である．Brignull はこれを「ユーザーを騙して，たとえば，品物の購入時に保険に入らせたり，定期購入を契約させたりなど，特定の行動に誘導するため慎重に設計されたユーザーインターフェース [1]」と定義している．なお，近年では「ディセプティブパターン (Deceptive Patterns)」という呼称も提唱されているが，本稿では広く定着している「ダークパターン」という用語を用いて議論を進める．

ダークパターンの具体例としては，視覚的階層操作によって特定の選択肢を誤認させる「Interface Interference」，虚偽のカウントダウンタイマー等を用いてユーザの焦燥感を煽る「Social Engineering」，サービスの解約プロセスを意図的に複雑化する「Obstruction」などが挙げられる．これらの設計は，金銭的な損失やプライバシーの侵害といった消費者トラブルに直結することから，重大な社会問題として認識されつつある．

現在，欧米諸国を中心にダークパターンに対する法規制の議論が進展しており [2]，日本国内においても対策の必要性が叫ばれている．実効性のある規制や対策を講じるためには，Web 上におけるダークパターンの蔓延状況を把握する大規模な実態調査が不可欠である．しかし，人手による調査には限界がある上，現時点で実用的なダークパターン自動検知手法は確立しておらず，大規模調査を実施した事例は限られている．この問題を解消するためには，高精度かつ低コストな自動検知技術の確立が急務である．

近年の LLM (Large Language Model) 技術の飛躍的な進歩に伴い，LLM を用いたダークパターン検知の研究が注目を集めている．これらの研究では，従来の古典的な手法と比較して高い検出精度が報告されている．しかし，先行研究の多くは，API 経由で利用する SaaS (Software as a Service) 型の LLM に依存しており，実運用におけるいくつかの課題が浮き彫りになっている．第一にコストの問題である．SaaS 型 LLM の多くは従量課金制を採用しており，大規模なクローリングデータに対して全件検査を行う場合，API 利用料が莫大となり予算的な障壁となる．第二に再現性と持続可能性の問題である．SaaS 型モデルは頻繁にアップデートや仕様変更が行われるため，同一の入力に対しても時期によって出力が異なる可能性があり，学術的な再現性の担保が困難である．また，サービス終了やポリシー変更による利用停止のリスクも無視できない．

そこで本研究では，SaaS 型 LLM に依存しない，ローカル環境で動作する SLM (Small Language Model) に着目する．本研究の目的は，ローカル SLM を用いた低コストかつ自律的なダークパターン検知手法を確立し，大規模な実態調査を技術的に支援することである．具体的には，収集した Web ページのソースコードに対し，パラメータ数を抑えた軽量なモデルを用いて解析を行う．ローカル SLM を採用することによる主な利点は以下の通りである．

- **コスト効率**: API 利用料が発生しないため，トークン数を気にすることなく大規模なデータを解析可能である．
- **再現性と透明性**: モデルをローカル環境で管理するため，

バージョンの固定が可能であり，外部要因による実験結果の変動を防ぐことができる．

本稿では，実際に日本語 EC サイトから収集したデータセットを用いて，提案手法であるローカル SLM によるソースコード解析の有効性を検証する．

本稿の構成は以下の通りである．第 2 章では，関連研究について述べ，第 3 章では，視覚情報の埋め込みを伴う前処理とローカル SLM による推論を組み合わせた提案手法について詳述する．第 4 章では，構築したシステムを用いた評価実験の結果と考察を述べ，第 5 章で本稿の結論と今後の展望をまとめる．

2 関 連 研 究

ダークパターンに関する研究は，主に分類と検知の 2 つの領域に焦点が当てられている．本章では，これら 2 つの領域の関連研究について述べる．

2.1 ダークパターン分類法

2024 年，Gray ら [3] は，既存の 10 種類のダークパターンに関する規制上および学術上の分類を調和させ，高レベル，中レベル，低レベルの 3 階層に分類される 64 種類の統合ダークパターンタイプについて，標準化された定義を備えたオントロジーを提案した．

10 種類のダークパターンの分類の詳細は以下の通りである．

- 2010 年から Harry Brignull 自身の WEB サイト [1,2]で共有されている分類
- 学術分野におけるダークパターン分類 4 件
- EU，英国，米国の利害関係者や規制当局のダークパターン分類を含む公開文書 5 件

これらの分類法に含まれるダークパターンタイプを分析し，最終的に高レベルの分類 5 種類，中レベルの分類 25 種類，低レベルの分類 35 種類，計 64 種類のオントロジーとされている．

Gray らによるオントロジーのうち，高レベルのダークパターンタイプとその定義を以下に示す．なお，日本語訳は著者による．

- **Sneaking** is a strategy which hides, disguises, or delays the disclosure of important information that, if made available to users, would cause a user to unintentionally take an action they would likely object to.
  （**スニーキング**とは，重要な情報を隠す，偽装する，開示を遅らせる戦略で，もしその情報がユーザに提示されていれば拒否したであろう行動を，ユーザに意図せずとらせるもの.）
- **Obstruction** is a strategy which impedes a user's task flow, making an interaction more difficult than it inherently needs to be, dissuading a user from taking an action.
  （**妨害**とは，ユーザのタスク進行を阻害し，本来必要とされる以上にインタラクションを困難にすることで，ユーザーが特定のアクションをとることを思いとどまらせる戦略.）
- **Interface Interference** is a strategy which privileges specific actions over others through manipulation of the user interface, thereby confusing the user or limiting discoverability of relevant action possibilities.
  （**インターフェース干渉**とは，UI の操作を通じて特定のアクションを他のアクションよりも優遇し，それによってユーザを混乱させたり，関連するアクションの選択肢の発見可能性を制限したりする戦略.）
- **Forced Action** is a strategy which requires users to perform an additional and/or tangential action or information to access (or continue to access) specific functionality, preventing them from continuing their interaction with a system without performing that action.
  （**強制されたアクション**とは，特定の機能にアクセス（またはアクセスを継続）するために，追加的または付随的アクションの実行や情報の提供をユーザに要求し，そのアクションを実行しない限り，システムとの対話を継続できないようにする戦略.）
- **Social Engineering** is a strategy which presents options or information that causes a user to be more likely to perform a specific action based on their individual and/or social cognitive biases, thereby leveraging a user's desire to follow expected or imposed social norms.
  （**ソーシャルエンジニアリング**とは，個人的または社会的な認知バイアスに基づき，ユーザが特定のアクションを実行する可能性を高めるような選択肢や情報を提示する戦略. これは，期待される，あるいは課された社会規範に従いたいというユーザの欲求を利用するものである.）

2.2 ダークパターン検知

ダークパターン検知の研究は，近年の情報処理技術の急速な発展により，その検知手法が大きく変化してきている．

2022 年，矢田ら [4] は EC サイトを対象に，正解データが付与された自然言語テキストのデータセットを用いて，機械学習によるダークパターン自動検出のベースライン評価を行った．このベースライン評価では，ダークパターンの有無を 2 値分類で検出を行い，評価が行われた．

2023 年，Mansur ら [5] は，スクリーンショットを入力とし，画像処理技術および自然言語処理技術を用いてダークパターンを検知するシステムである AidUI を提案した．AidUI は，パターンマッチングを用いたテキスト分析，カラーヒストグラム分析技術を用いた色分析，隣接するセグメントのサイズを計算することによる空間分析によって，ダークパターンの検出を行っている．

2025 年，Zewei Shi ら [6] はマルチモーダル LLM を用いてダークパターンを自動検知する DPGuard を提案した．DPGuard は，ダークパターンの有無を判別するバイナリ分類器と，マルチモーダル LLM を用いたダークパターンのカテゴリを出

[1] 旧：https://darkpatterns.org
[2] 現：https://www.deceptive.design

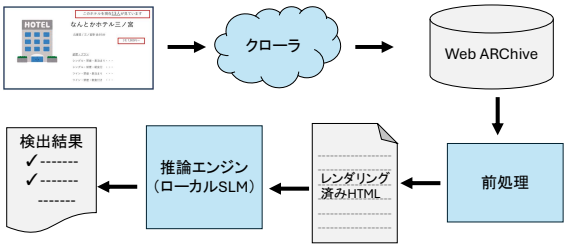

図 1　提案手法 構成図

力するダークパターン検出器からなる．これは，バイナリ分類器でダークパターンを含むと判断されたもののみマルチモーダル LLM を用いた検出器で検知を行うことにより計算コストの削減を図っている．

3　提 案 手 法

本稿では，日本語 EC サイトを対象にスクレイピングを行い，収集した WEB サイトデータを対象にローカル SLM を用いてソースコード解析を行うことにより，ダークパターンの検知を試みる．本提案手法は，スクレイピング，前処理，検知の 3 フェーズに分けられる．

本稿では，2024 年に Gray ら [3] が提案したダークパターンの分類法を採用する．しかし，すべてのダークパターンを対象に実験を行うことは難しい．例えば，カートの画面や購入手続きを進めた際に出現するダークパターンは，そのページの取得が困難である．そこで，高レベルの 5 分類のうち，Social Engineering を対象に実験を行う．これに含まれるダークパターンタイプを表 1 に示す．本稿では，Social Engineering に含まれる 8 タイプについて，検知を試みる．

また，本稿では VRAM が 16GB 程度の環境を想定しており，システムの検証環境として Google Colab(GPU:T4, VRAM:15GB)，本番環境としてデスクトップ PC(GPU:RTX 5060 ti, VRAM:16GB) を使用する．

3.1　スクレイピング

近年の WEB サイトは，JavaScript などを用いて動的に構築されるケースが多い．このため，動的なスクレイピングが可能な BrowsertrixCrawler[3]を採用する．Browsertrix Crawler とは，ブラウザベースのクロールシステムであり，Docker コンテナで実行できるよう設計されている．これにより取得された WARC ファイルを入力として用いる．

実際の EC サイトに対してスクレイピングを行うにあたり，EC サイトの URL のリストが必要である．そこで，日本語 EC サイト 200 件の URL を手動で記録した．なお，これらの EC サイトは，

- 2025 ネット通販売上高ランキング TOP100（日本ネット経済新聞）
- Shopify 導入事例集
- BASE 導入事例

より取得した．この 200 件の EC サイトに対し，Browsertrix Crawler を用いて 1 サイトあたり 15 ページのスクレイピングを行った．保存されたページは，200 サイト計 2666 ページである．

3.2　前　処　理

スクレイピングにより取得された WARC (Web ARChive) ファイルは，その状態では SLM の入力に適さない．よって，SLM への入力に適した形式に変換する必要があり，このための前処理機構を作成する．

Browsertrix Crawler を用いて収集した WARC ファイルには，HTML のほかに CSS や JavaScript，画像などが含まれる．本稿では，取得した WARC ファイルをヘッドレスブラウザでレンダリングし，レンダリング済み HTML として保存する．これにより，スクリーンショットを用いずに視覚的情報の取得を試みる．

具体的には，ヘッドレスブラウザ上で JavaScript を実行し，動的なコンテンツが展開された状態の DOM (Document Object Model) ツリーを構築する．この際，各 DOM ノードに対して window.getComputedStyle を適用し，CSS 適用後の計算済みスタイルを HTML 属性として明示的に埋め込む処理を行う．例えば，display: none や opacity: 0 といった非表示スタイルを持つ要素には不可視属性を，ボタンやリンク等の重要要素には背景色やフォントサイズを属性値として付与する．これにより，テキストベースのモデルであっても，「視覚的な隠蔽」や「配色の強調による誘導」といった視覚的特徴を DOM 構造の一部として認識可能にする．

続いて，SLM の限られたコンテキスト長を有効活用するため，意味論的浄化を行う．具体的には，ダークパターンの判定に寄与しない <script>, <style>, <svg> タグや HTML コメント等を削除し，DOM 構造を軽量化する

最後に，軽量化された DOM ツリーを SLM の入力制限（本研究では 2048 トークンと設定）に合わせて分割する．単純な文字列分割ではタグの整合性が崩れ，階層構造の欠落によりモデルが文脈を見失うリスクがある．そのため，本手法では分割された各チャンクの先頭に，ルート要素から当該ノードに至るまでの祖先タグ列（パンくずリスト）を自動的に挿入する．これにより，局所的なチャンクであってもページの全体構造における位置と文脈を保持したまま，推論を行うことが可能となる．

3.3　検　　知

3.4　検知モデルの選定と最適化

本稿では，検知に用いるローカル SLM として，Alibaba Cloud によって開発されたコーディング特化型モデルである

[3] https://crawler.docs.browsertrix.com/

表 1　対象とするダークパターンタイプ

Type	Definition
High Demand	Indicates that a product is in high-demand or likely to sell out soon, even though that claim is misleading or false.
Low Stock	Indicates that a product is limited in quantity, even though that claim is misleading or false.
Endorsements and Testimonials	Indicates that a product or service has been endorsed by another consumer, even though the source of that endorsement or testimonial is biased, misleading, incomplete, or false.
Parasocial Pressure	Indicates that a product or service has been endorsed by a celebrity, influencer, or other entity that the user trusts, even though the source of that endorsement is biased, misleading, incomplete, or false.
Activity Messages	Describes other user activity on the site or service, even though the data presented about other users' purchases, views, visits, or contributions are misleading or false.
Countdown Timer	Indicates that a deal or discount will expire by displaying a countdown clock or timer, even though the clock or timer is completely fake, disappears, or resets automatically.
Limited Time Message	Indicates that a deal or discount will expire soon or be available only for a limited time, but without specifying a specific deadline.
Confirmshaming	Frames a choice to opt-in or opt-out of a decision through emotional language or imagery that relies upon shame or guilt.

Algorithm 1 視覚情報埋め込みを伴う構造保持型 DOM 分割

Require: WARC records $\mathcal{W}$, Target URLs $\mathcal{U}$, Max tokens L_{max}
Ensure: Set of structured chunks $\mathcal{D}$

$\mathcal{D} \leftarrow \emptyset$
for each record $r \in \mathcal{W}$ **do**
  if $r.url \in \mathcal{U}$ **and** $r.status = 200$ **then**
    $h \leftarrow \text{HeadlessRender}(r.content)$
    $h' \leftarrow \text{InjectVisualFeatures}(h)$ ▷ Inject computed styles as attributes
    $t \leftarrow \text{CleanDOM}(h')$ ▷ Remove scripts, styles, comments
    $C \leftarrow \text{RecursiveChunking}(t.body, \emptyset, L_{max})$
    $\mathcal{D} \leftarrow \mathcal{D} \cup C$
  end if
end for
return $\mathcal{D}$
procedure RecursiveChunking($node, ancestors, L_{max}$)
  $S_{start} \leftarrow \text{StartTag}(node)$
  if $\text{Length}(buffer) + \text{Length}(S_{start}) > L_{max}$ **then**
    $\text{Flush}(buffer)$
    $buffer \leftarrow \text{Breadcrumbs}(ancestors)$ ▷ Initialize with ancestor tags
  end if
  Append S_{start} to $buffer$
  for each $child \in node.children$ **do**
    $\text{RecursiveChunking}(child, ancestors \cup \{node\}, L_{max})$
  end for
  Append $\text{EndTag}(node)$ to $buffer$
end procedure

Qwen2.5-Coder [7] を採用する．本モデルは，自然言語に加え HTML や CSS などのマークアップ言語の理解に優れており，DOM 構造の解析に適している．モデルのバリエーションには，パラメータ数が 0.5B, 1.5B, 3B, 7B, 14B, 32B のものが存在する．

モデルの選定にあたり，本研究の検証環境である Google Colab（NVIDIA T4 GPU, VRAM 15GB）において，動作確認と VRAM 使用量の検証を行った．検証の結果，14B モデルでは各パラメータを 4bit に丸める量子化（4-bit Quantization [8]）を適用しても，長い HTML コンテキストを入力した際にメモリ不足（OOM）が発生するリスクが高いことが確認された．一方，7B モデルに対し 4bit 量子化（NF4 形式）を適用した場合，モデル重みのメモリ消費を約 5.5GB に抑制できることが判明した．これにより，残りの VRAM 領域を KV キャッシュやアクティベーション計算に割り当てることが可能となり，最大シーケンス長 4096 トークンかつバッチサイズ 2 での並列推論が安定して動作することを確認した．

以上の検証より，本研究では計算資源の制約下で処理速度とコンテキスト容量のバランスが最適である “Qwen2.5-Coder-7B-Instruct” の 4bit 量子化モデルを採用することとした．また，推論エンジンの実装には ‘Unsloth’ ライブラリを用い，メモリ効率化と推論速度の向上を図っている．

4 評 価 実 験

提案手法の有効性を検証するために実施する評価実験について述べる．本実験の目的は，計算資源に制約のある環境において，4bit 量子化されたファインチューニング無しの SLM を用いたソースコード解析が，Web 上のダークパターン検知において実用的な精度と網羅性を持つかを定量的に明らかにすることである．

実験は，データ収集，前処理，モデルによる推論，および人間の評価者による正解の付与というプロセスを経て行われる．

4.1 実験用データセットの構築と前処理

評価実験の基礎となるデータセットには，独自にクローリングを行い収集した EC サイトの WARC ファイルのうち，2025 ネット通販売上高ランキング TOP100 の 100 サイトより収集

Algorithm 2 SLM を用いたダークパターン検知

Require: Structured chunks $\mathcal{D}$, Taxonomy definitions $\mathcal{T}$, Target category C_{tgt}, Batch size B, Model $\mathcal{M}$ (4-bit quantized)
Ensure: Set of detected patterns $\mathcal{P}$

$\mathcal{P} \leftarrow \emptyset$
$\delta \leftarrow \mathcal{T}[C_{tgt}]$ ▷ Retrieve specific definition to minimize context
Filter $\mathcal{D} \leftarrow \{d \in \mathcal{D} \mid \text{Length}(d.html) \geqq L_{min}\}$
Partition $\mathcal{D}$ into batches $\mathcal{B} = \{b_1, b_2, \ldots, b_m\}$ where $|b_i| \leqq B$
for each batch $b \in \mathcal{B}$ **do**
  $inputs \leftarrow \emptyset$
  for each chunk $d \in b$ **do**
    $p \leftarrow \text{ConstructPrompt}(\delta, d.context, d.html)$
    Append p to $inputs$
  end for
  $X \leftarrow \text{Tokenize}(inputs, \text{padding} = \text{True})$
  $Y \leftarrow \mathcal{M}.\text{Generate}(X)$ ▷ Batch inference on GPU
  for each output $y \in Y$ **do**
    $J \leftarrow \text{ExtractJSON}(y)$
    if $J \neq \text{null}$ **and** $J.\text{patterns} \neq \emptyset$ **then**
      $\mathcal{P} \leftarrow \mathcal{P} \cup J.\text{patterns}$
    end if
  end for
  ClearGPUMemory() ▷ Free VRAM to prevent fragmentation
end for
return $\mathcal{P}$
procedure ConstructPrompt($\delta, context, html$)
  $S_{sys} \leftarrow \text{FormatSystemPrompt}(\delta)$
  $S_{user} \leftarrow$ "Context: " + $context$ + \n + "HTML: " + $html$
  return $\text{ApplyChatTemplate}(S_{sys}, S_{user})$
end procedure

したデータからランダム抽出した 200 ページを実験用データセットとし，これを対象とした．収集したデータを SLM で解析を行うため，前章で提案した「視覚情報埋め込みを伴う構造保持型 DOM 分割」アルゴリズムを適用する．

具体的には，まずヘッドレスブラウザを用いて各ページをレンダリングし，ユーザーが視認するスタイル情報を HTML 属性として明示的に埋め込む．次に，スクリプトや広告などのノイズを除去した後，トークン制限（L_{max}）に基づき，DOM の階層構造とパンくずリスト（Context Path）を保持したまま，HTML を意味のある単位（チャンク）へと分割する．これらの処理済みデータは，メタデータと共に JSON 形式で保存し，推論用データセットとした．

4.2 推論設定とプロンプト設計

推論モデルには Qwen2.5-Coder-7B-Instruct を採用した．コンシューマー向けの限られた GPU メモリ（VRAM 16GB）環境下での推論を可能にするため，Unsloth ライブラリを用いてモデルの重みに 4bit 量子化（NF4 形式）を適用した．推論パイプラインは，バッチサイズを 2，チャンクあたりの入力最大トークン長を 4096 に設定し，並列処理によるスループットの最大化を図った．

表 2 推論パフォーマンスの比較

バッチサイズ	総処理時間	処理速度	VRAM 使用量 (GB)	
	(s)	(sec/file)	平均	最大
2	15166	75.83	6.9	7.1
4	11837	59.18	8.3	9.1
8	10003	50.01	11.3	12.7

プロンプトを記述する言語は英語とし，システムプロンプトではモデルの役割およびタスクを指示した．また，出力フォーマットは抽出要素・パターン名・推論理由を含む JSON 形式で出力するように指示した．なお，ダークパターンが存在しない場合はからのリストを出力するよう指示した．

ダークパターンの定義プロンプトは，パターンの定義文のみのプロンプト A，定義文に加え例及および除外条件を加筆したプロンプト B を用意し，それぞれで実験を行った．定義文の一例を以下に示す．

- **プロンプト A**
  Low Stock: Indicates that a product is limited in quantity, even though that claim is misleading or false.
- **プロンプト B**
  Low Stock: Indicates that a product is limited in quantity to induce urgency.
  - **Target**: "Only 3 left", "Low stock", "Almost gone", "2 people have this in their cart".
  - **EXCLUDE**: "Out of stock", "Sold out", "Not available", "Backorder", "In Stock". (Do NOT detect items that cannot be purchased immediately).

4.3 正解の付与

本実験では，提案手法の評価指標として，適合率（Precision）を評価指標として採用した．現実の Web サイトにおけるダークパターンの網羅的な目視確認（再現率の算出）は非現実的であるため，本研究では「SLM がダークパターンであると警告したものが，実際に真のダークパターンであったか」という，誤検知の少なさに焦点を当てて評価を行う．

正解の付与は以下の手順で実施した．まず，推論パイプラインを通じて出力された JSON 結果のうち，モデルが 1 つ以上のダークパターンを検出したチャンク（陽性予測データ）をすべて抽出する．次に，抽出された要素（HTML タグ）とモデルが提示した推論理由（Reasoning）を，対象ページの実際のレンダリング結果および前後のコンテキストパス（DOM の階層情報）と照らし合わせる．人間の評価者が Gray らの定義に基づいて目視判定を行い，ダークパターンであると確認できた場合を真陽性（True Positive: TP），通常の UI 要素の誤検知や，モデルによる存在しない要素の捏造（ハルシネーション）を偽陽性（False Positive: FP）として分類した．

4.4 実 験 結 果

4.4.1 推論パフォーマンス

構築した推論パイプラインを用いて，200 ページの処理を

表 3　プロンプト設計（A：定義のみ，B：定義＋除外条件）によるカテゴリ別検知結果の比較

	プロンプト A			プロンプト B（除外条件付）		
DP タイプ	TP 数	FP 数	適合率	TP 数	FP 数	適合率
Low Stock	27	59	0.3140	26	18	0.5090
Countdown Timer	1	24	0.0400	1	17	0.0556
Limited Time Messages	2	12	0.1429	2	30	0.0625
High Demand	0	5	0.0000	0	1	0.0000
Parasocial Pressure	0	5	0.0000	0	0	–
Confirmshaming	0	1	0.0000	0	0	–
Endorsements and Testimonials	0	1	0.0000	0	0	–
Activity Messages	0	0	–	0	0	–
架空のタイプ	0	4	0.0000	0	16	0.0000
全体	**30**	**111**	**0.2128**	**29**	**82**	**0.2613**

行った．この際，バッチ数を 2, 4, 8 に変化させて，それぞれの合計処理時間，平均 VRAM 使用量，最大 VRAM 使用量を記録した．この結果を表 2 に示す．この結果より，VRAM16GB 環境において Qwen2.5-coder-7B-Instruct を用いて安定した動作が可能であることが確認できる．

4. 4. 2　プロンプト設計による検知精度の比較

プロンプトの記述粒度が SLM の推論精度に与える影響を評価するため，パターンの定義文のみを与えた「プロンプト A」と，具体例および厳密な除外条件を付与した「プロンプト B」による推論結果の比較を行った．各プロンプトにおける DP タイプ別の真陽性（TP）数，偽陽性（FP）数，および適合率を表 3 に示す．

全体の結果として，プロンプト A を用いた場合の適合率は 0.2128（21.28%）であったのに対し，プロンプト B を用いた場合は 0.2613（26.13%）となり，約 5%の精度向上が確認された．しかし，カテゴリ別の検出内訳を分析すると，プロンプトの詳細化がもたらした効果は一様ではなく，特定のカテゴリにおける著しい精度改善と，別のカテゴリにおける深刻な推論の崩壊という両極端な結果が混在していることが判明した．

a）　除外条件の成功例

プロンプト B における明確な改善効果は，「Low Stock」カテゴリにおいて確認された．プロンプト A では FP が 59 件発生し適合率が 0.3140 であったが，プロンプト B では TP 数を概ね維持したまま FP を 18 件へと大幅に抑制し，適合率を 0.5090 に向上させることに成功した．これは，プロンプト B に記述した「"Out of stock" や "Sold out" は今すぐ購入できないため検知してはならない」という否定形の除外条件をモデルが正しく解釈し，単純なキーワードマッチングによる誤検知を回避できたためであると言える．

b）　複雑なルール付与による推論崩壊（ハルシネーションの増加）

一方で，プロンプトの詳細化は予期せぬ重大な副作用をもたらした．「Limited Time Messages」カテゴリにおいては，プロンプト B を適用したことで FP が 12 件から 30 件へと激増し，適合率が 0.1429 から 0.0625 へと悪化した．さらに致命的な結果として，プロンプトに定義されていない「架空の DP タイプ」を出力するハルシネーション（形式的崩壊）の件数が，プロンプト A の 4 件からプロンプト B では 16 件へと 4 倍に増加した．

これらの結果は，7B パラメータクラスの小規模言語モデルにおいて，Zero-Shot のプロンプトエンジニアリングのみで複雑なタスクを制御しようとするアプローチの限界を定量的に示している．モデルに対して「例示」や「〜してはいけない」という多数の制約を同時に与えた結果，プロンプトのコンテキストが複雑化・長大化し，モデルが情報過多（オーバーフィット）に陥ったと推察される．一部のルール（Low Stock の除外など）には適合できたものの，ルールの全体像を矛盾なく保持し続ける処理能力が不足しており，結果として定義の拡大解釈（Limited Time Messages の誤検知増）や，指示されていない架空のルールを捏造するハルシネーションを引き起こしたと考えられる．

以上の結果から，In-the-wild の無秩序な Web データに対し，ローカル SLM を用いて実用的な精度の静的解析を行うためには，プロンプトの工夫のみに依存することには限界があることが明らかとなった．

5　おわりに

本稿では，日本語 EC サイトにおけるダークパターンの実態解明に向けた第一歩として，SaaS 型 LLM に依存しないローカル SLM を用いたダークパターン検知手法を提案した．具体的には，レンダリングを伴う「視覚情報埋め込みを伴う構造保持型 DOM 分割」アルゴリズムを導入することで，LLM が解釈可能な形式で視覚情報をテキスト化し，これを 4bit 量子化された Qwen2.5-Coder-7B に入力することで，低コストかつ自律的なソースコード解析を実現するフレームワークを構築した．本手法は，API コストや外部サービスの仕様変更といった制約に左右されず，比較的小規模な計算資源（VRAM16GB 程度）で大規模な調査を可能にする点において，学術的および実用的な意義を持つ．

評価実験の結果，詳細な除外条件を付与した Zero-Shot プロンプトにより，特定のダークパターン（希少性の煽り等）にお

いては誤検知の抑制が可能であることが示された．しかし同時に，本研究のアプローチが抱える根本的な課題が浮き彫りとなった．

一つ目の課題は， 単体テストを行わずに検知パイプライン全体でのみ評価したことである．時間の都合上このような評価方法になってしまったが，単体テストを行っていないことにより，どこに問題があったのかなど，詳細な考察が行えなかった．

二つ目の課題は，前処理の方法である．本稿では，WARC ファイルをレンダリング済み HTML に変換し，これを入力としたものの，ダークパターンの判定に必須なものは，WEB ページ上にあるテキスト情報，文字やボタンなどの色や大きさなどといいた視覚的情報である．それに対して，レンダリング済み HTML はその大半が不要なデータであり，前処理方法の再検討が必要である．

三つ目の課題は，評価方法である．本稿では，陽性データに正解を付与することで適合率を算出し評価を行ったものの，ダークパターンの検知においては，再現率を用いた取りこぼしの評価が重要であると考える．しかし，実際の WEB サイトデータに対してダークパターンの取りこぼしなく正解データを付与することは難しい．このことから，検知手法の性能を評価するためには，正解ラベル付きのデータセットを用いることが必要であると考える．

今後の方針として，まずは正解ラベル付きの合成データセットを行う．架空の WEB サイトを作成し，これをデータセットとして保存することにより，適合率，再現率，F 値などの算出を可能にする．

次に，前処理をふくむ検知パイプラインの改良を行う．その上で，ダークパターンの検出実験を行い，課題を探る．

文　献

[1] ハリー・ブリヌル（著）, 高瀬みどり（訳）, “ダークパターン 人を欺くデザインの手口と対策,” 株式会社 BNN, p.12 (2024).

[2] OECD, “Dark commercial patterns,” OECD Digital Economy Papers, No.336, pp.30–45, October 2022.

[3] Colin M. Gray, Cristiana Teixeira Santos, Nataliia Bielova and Thomas Mildner, “An Ontology of Dark Patterns Knowledge: Foundations, Definitions, and a Pathway for Shared Knowledge-Building,” CHI’24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, No.289, pp.1–22 (2024).

[4] Yuki Yada, Jiaying Feng, Tsuneo Matsumoto, Nao Fukushima, Fuyuko Kido, and Hayato Yamana, “Dark patterns in e-commerce: a dataset and its baseline evaluations,” 2022 IEEE International Conference on Big Data, pp.3015–3022 (2022).

[5] S M Hasan Mansur, Sabiha Salma, Damilola Awofisayo and Kevin Moran, “AidUI: Toward Automated Recognition of Dark Patterns in User Interfaces,” ICSE ’23: Proceedings of the 45th International Conference on Software Engineering, pp.1958-1970 (2023).

[6] Zewei Shi, Ruoxi Sun, Jieshan Chen, Jiamou Sun, Minhui Xue, Yansong Gao, Feng Liu and Xingliang Yuan, “50 Shades of Deceptive Patterns: A Unified Taxonomy, Multimodal Detection, and Security Implications,” WWW ’25: Proceedings of the ACM on Web Conference 2025, pp.978–989 (2025).

[7] Binyuan Hui, Jian Yang, et al. "Qwen2.5-coder technical report," arXiv preprint arXiv:2409.12186 (2024).

[8] 久富 望, “ローカル PC での LLM（大規模言語モデル）について, ” 知能と情報，36 巻 3 号，pp.70–76 (2024).

スマホ撮影画像への応用を見据えたナンバープレート検出モデル性能に関する予備調査

宮木　笙伍† 　河合 由起子†† 　栗　達†

† 京都産業大学情報理工学部　〒 603-8555 京都府京都市北区上賀茂本山
†† 関西大学　〒 565-8585 大阪府吹田市山田南 50-2
E-mail: †{g2354558, lida}@cc.kyoto-su.ac.jp, ††ykawa@kansai-u.ac.jp

あらまし　近年, スマホによる SNS などへの風景画像投稿の増加に伴い, 人物や居住場所, ナンバープレートなど個人を特定し得る情報は, データ収集・分析・可視化において重要なリスク要因となる. 一方, 交通管理や防犯分野において, 定点カメラによる車両情報の自動取得を目的としたナンバープレート検出システムの重要性が高まっている. 本研究では, スマホによる撮影画像に対するナンバープレート検出システム構築に向けた予備検討として, 定点カメラによる公開データセットとスマートフォン撮影画像データセットを用い, 物体検出モデルによるナンバープレート検出性能の比較を行った. なお, 都市回遊促進のためのデータ収集を目的とした研究の一部として, 実環境で取得される画像データに含まれるナンバープレートなどのプライバシー情報を自動的に検出する手法の予備調査に位置づけられる. 比較対象として,YOLO シリーズに属する YOLOv11 および YOLO-World, ならびに Transformer 系モデルである RT-DETR の 3 手法を採用した. 各モデルは,Epoch 数 100, 入力画像サイズ 640 などの統一した学習条件下で学習および評価を行い,Precision（適合率）, 再現率（Recall）,F1 スコア,mAP および推論速度を指標として性能を比較した. さらに, 視覚-言語モデルである YOLO-World に対しては, テキストプロンプトの変更が検出精度に与える影響を検証する追加実験も実施した. その結果,YOLO-World は Precision 0.8155，mAP@50-95 で 0.5197 を記録し, 高精度な検出性能を示した.RT-DETR は F1 スコアおよび Recall で最高値を示した. しかし YOLO 系と比較しリアルタイム性に課題が残る結果となった.YOLOv11 は推論速度において最も優れていたが,mAP および F1 スコアでは他のモデルに及ばなかった. また,YOLO-World におけるプロンプトの変更による顕著な精度向上は確認されなかった. 本予備実験では, 学習データが少ない場合には YOLO-World が, 見逃しを減らすことを重視する場合には RT-DETR が有効であることが示された.

キーワード　物体検出, 機械学習, ナンバープレート検出

1 はじめに

スマートフォンの普及により, ほとんどすべての国民が SNS を利用できるようになった. 総務省の調査などをもとにした 2025 年のデータによれば, 日本国内の SNS 利用者は総人口の約 78.1%に当たる約 9,600 万人に達しており, インターネットユーザー全体で見てもその普及率は極めて高い [1]. 特に, 画像や動画を主体とする「ビジュアルコミュニケーション」の拡大が著しい. 月間アクティブユーザー数 (MAU) 7,370 万人を擁する YouTube や, 若年層を中心に利用率が高い Instagram(MAU 5,545 万人以上),TikTok(MAU 2,600 万人以上) など, 視覚的情報を共有するプラットフォームが日常的に利用されている. また, スマートフォンのカメラ性能は, 近年劇的な進化を遂げている. 最新のハイエンドモデルでは,2 億画素を超える超高解像度センサーや, 多くの光を取り込める 1 インチ大型センサーの搭載が進んでおり,「一眼レフ並み」と評される水準にある [2]. これらの進化は, 撮影者が意図した被写体だけでなく, 撮影に写り込んだ情報までも極めて鮮明に記録されることを意味する.SNS 上の膨大な画像データにおいて, こうした高精細な映り込み情報はプライバシー保護の観点から看過できないリスク要因となっており, 自動的かつ高精度にこれらを検出し, 適切に処理する技術の重要性が高まっている. 特に, 車両のナンバープレートは, 所有者の特定や移動履歴の把握に直結するため, プライバシー保護において極めて重要な情報である. 交通管理や犯罪捜査の現場では, 定点カメラを用いた自動ナンバープレート認識 (LPR) システムの導入が進んでいるが, これらの既存技術が, 撮影条件の多様なスマートフォン画像に対しても同様に機能するかは自明ではない. 手振れ, 角度, 照明条件が一致しないスマートフォン画像における検出性能を明らかにすることは, プライバシー保護技術の向上に寄与するものである.

そこで本研究では, スマートフォンによる撮影画像に対する検出システム構築に向けた予備検討として, 定点カメラによる公開データセットを用いた物体検出モデルの比較評価を行う. 具体的には, 最新の YOLO シリーズである YOLOv11 及び YOLO-World, さらに Transformer ベースのアーキテクチャを持つ RT-DETR の 3 手法を対象とする. これらを統一された条件下で学習・評価し,F1 スコアや mAP などの客観的指標に基づいて各モデルの検出特性を分析する. 本検証により, 各アーキテクチャの長所と短所を明らかにし, 実環境におけるナンバー

プレート検出に最適なモデルの選定の指標を示す.

本論文は以下の構成とする. 2 章では関連研究を紹介する. 3 章では物体検出モデルを比較するための実験設定を述べる. 4 章では実験結果を示し, 定量的および定性的な評価を行う. 最後に 5 章で本論文のまとめと今後の課題について述べる.

2 関連研究

本章では, 関連研究として物体検出モデルに関する研究とナンバープレート検出に関する研究を紹介する.

2.1 リアルタイム物体検出モデルの進化

Sapkota ら [3] は,Ultralytics YOLO シリーズ（v5,v8,11,26）のアーキテクチャの進化と性能評価に関する包括的なレビューを行っている. 彼らは,YOLOv8 におけるアンカーフリー予測と分離型ヘッドの導入,YOLO11 における C3k2 モジュールによる特徴抽出の効率化, そして最新の YOLO26 における NMS（Non-Maximum Suppression）および DFL（Distribution Focal Loss）の削除による推論の高速化とエッジデバイスへの適合性向上について詳述している

一方,Zhao ら [4] は,YOLO シリーズが NMS による後処理に依存している点が速度と精度のトレードオフに悪影響を与えていると指摘し, リアルタイム・エンドツーエンド物体検出器である RT-DETR を提案している. この研究では, マルチスケール特徴を効率的に処理するハイブリッドエンコーダと, 不確実性を最小化するクエリ選択手法を導入することで,RT-DETR が同規模の YOLO モデル（YOLOv5,v8 など）を速度と精度の両面で上回ることを実験により示している

2.2 オープン語彙物体検出への拡張

従来の物体検出器が固定されたカテゴリに限定されるという課題に対し,Cheng ら [5] は,YOLO-World と名付けたリアルタイム・オープン語彙物体検出器を提案している. 彼らは, 視覚と言語の情報を融合するための RepVL-PAN（Re-parameterizable Vision-Language Path Aggregation Network）と, 領域テキスト対照損失を用いた大規模な事前学習スキームを導入した. これにより, 推論時にはテキストエンコーダを再パラメータ化することで計算コストを削減しつつ, ゼロショットで多様な物体を検出できることを確認し,LVIS データセットにおいて高い精度と FPS を達成している

2.3 特定タスクへの応用：ナンバープレート検出

Fu [6] は, ナンバープレート認識（LPR）における深層学習技術の適用について調査を行っている.CNN（特に YOLO シリーズ）を用いた検出が主流である一方で, 文字認識における RNN（LSTM や GRU）の活用, 低解像度画像の超解像やデータ拡張における GAN や Diffusion Model の利用. そして大域的な特徴抽出に優れた Transformer の導入など, 各モデルのアーキテクチャごとの利点と課題を比較・整理している

2.4 関連研究のまとめと本調査の位置づけ

これらの知見を踏まえると, 物体検出技術は, 高速・軽量な CNN ベースのモデル,Transformer を用いたエンドツーエンド型モデルおよび視と言語を統合した事前学習モデルへと多様化していることが分かる.

一方で, これらのモデルが定点カメラで学習された条件からスマートフォンによる撮影画像のような撮影条件が大きく異なる環境へ適用された場合の検出性能の違いについては, 十分に検証されていない.

そこで本調査では, 事前学習の有無やモデルアーキテクチャの違いに着目し, 定点カメラ画像で学習したモデルを用いて, スマートフォン撮影画像に対するナンバープレート検出性能を比較評価する. 本検証により, ドメイン乖離環境下における各モデルの特性を明らかにすることを目的とする.

3 実験設定

本研究では, 各モデルの性能を公平に比較するため, 統一されたデータセット・学習条件下で実験を行った.

3.1 使用データセット

本実験で使用するデータは, ネガティブ画像（ナンバープレートなし）, ポジティブ画像（ナンバープレートあり）, および評価用のスマートフォン撮影画像の 3 種とする. 定点カメラで撮影された車両画像・ナンバープレート画像と, スマートフォンで撮影された日常的な風景画像の異なるドメイン間におけるモデルの検出性能を評価するため, 学習用と評価用で異なるデータセットを使用した.

3.1.1 学習用データセット

学習用データセットの構築にあたり, ポジティブ画像として Roboflow Universe で公開されている **License-plate-japan dataset**[1] および **Number Plate in Japan dataset**[2] を使用した（元データでポジティブ画像 1000 枚）. ネガティブ画像には, 背景画像として一般的によく用いられる **Microsoft COCO**[3] を採用した. ポジティブ画像に対して回転（90°, 180°, 270°）および反転（上下・左右）のデータ拡張を行い, 拡張後のポジティブ画像 6,000 枚とネガティブ画像 600 枚を用いて, 総枚数 6,600 枚のデータセットを作成した. ネガティブ画像と拡張済みポジティブ画像により,YOLOv11,YOLO-World,RT-DETR の 3 モデルに対してファインチューニングを行った.

3.1.2 評価用データセット

実環境での適用を想定し, 独自に収集したスマートフォンを用いて撮影されたデータセットを評価用として採用した. これには手ブレや多様な照明条件が含まれており,Positive(ナンバー

1 : [5] Roboflow Universe – license-plate-japan_1. https://universe.roboflow.com/new-workspace-vijtn/license-plate-japan_1 (accessed 2026-01-25).

2 : [6] Roboflow Universe – Number Plate in Japan. https://universe.roboflow.com/moriken/number-plate-in-japan (accessed 2026-01-25).

3 : [7] Microsoft COCO: Common Objects in Context. https://cocodataset.org/ (accessed 2026-01-25).

プレートあり)52 枚,Negative(なし)52 枚の計 104 枚で構成される.

3.2 比較モデルと学習条件

比較対象として, 実験当時の物体検出モデルである YOLOv11,YOLO-World, および Transformer ベースの RT-DETR の 3 種類を選定した. 学習は Google Colab 上の L4 GPU 環境で実施し, ネガティブ画像と拡張済みポジティブ画像を用いたファインチューニングを行った. ハイパーパラメータは Epoch を 100,Image Size を 640 に統一し, それ以外のハイパーパラメータは各モデルのデフォルト設定を採用した. 実験環境のハードウェア構成を表 1 および表 2 に, 各モデル共通の学習ハイパーパラメータを表 3 に示す.

表 1 学習環境

Item	Specification
GPU	NVIDIA Tesla L4
Platform	Google Colab
Framework	Ultralytics 8.3, PyTorch 2.x

表 2 推論環境 (M2 MacBook Air)

Item	Specification
Chip	Apple M2 (8-core GPU)
Memory	8 GB Unified Memory
Platform	macOS
Framework	Ultralytics 8.3, PyTorch 2.x (MPS)

表 3 学習ハイパーパラメータ

Parameter	Value
Input Image Size	640 × 640
Batch Size	16 (Default)
Epochs	100
Optimizer	Auto (SGD)
Comparison Models	
YOLOv11	`yolo11s.pt` (Small)
YOLO-World	`yolov8s-world.pt` (Small)
RT-DETR	`rtdetr-l.pt` (Large)

3.3 評価指標

性能評価には, 適合率 (Precision), 再現率 (Recall),F1 スコア, および mAP(mean Average Precision) を用いた. また, 実用性の観点から推論時間も合わせて計測した.

4 実験結果

表 4 に, 各モデルの評価結果を示す.

4.1 定量評価

実験の結果, **YOLO-World** が mAP@50 において 0.6980,Precision において 0.8155 と最も高い値を示し, 最も高精度な位置検出が可能であることが確認された. 一方, **RT-DETR** は F1 スコア (0.7271) および Recall(0.6731) で最高値を記録し見逃しが最も少なかったが, 推論時間が 523.4 ms と他の YOLO 系モデルの約 6–8 倍となり, リアルタイム性には課題が残る結果となった. **YOLOv11** は推論速度が最速（65.4 ms）であったものの,mAP および F1 スコアでは他の 2 モデルに及ばなかった.

4.2 定性評価

各モデルの実際の検出結果を図 1 に示す. 定量評価では YOLO-World が高い適合率を示したが, 実際の検出画像を確認すると, 全てのモデルにおいて共通の傾向が見られた. 正面で大きく写る対象では全モデルが検出に成功した一方, 看板などでは誤検出が, 小さい対象では見逃しが確認された. また, 明るさや大きさが十分でブレの少ない画像では全モデルが正確に検出できる一方,「強い手ブレ」などが含まれる画像では, どのモデルも検出に失敗するか, あるいは看板の文字を誤検出するケースが確認された. これは, 学習データと評価データの間で画像のドメインが乖離しており, モデルのアーキテクチャの差ではなく, 特徴抽出そのものが困難であったためと考えられる. ただしその中でも,YOLO-World はバウンディングボックスの追従性が比較的良好であり, 他のモデルよりもロバストな挙動を示した.

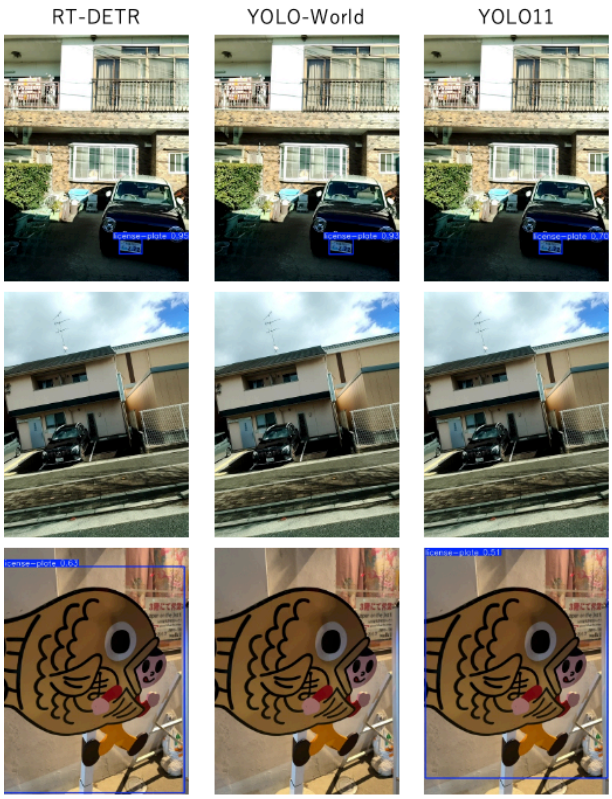

図 1 スマートフォン撮影画像に対する各モデルの検出結果比較

表 4　各モデルの性能比較（最高値を太字で示す）

Model	Precision	Recall	F1 Score	mAP@50	mAP@50-95	Inference (ms)
YOLOv11s	0.7980	0.5000	0.6148	0.5994	0.4343	**65.4**
YOLO-World	**0.8155**	0.5962	0.6888	**0.6980**	**0.5197**	81.0
RT-DETR	0.7905	**0.6731**	**0.7271**	0.6668	0.4653	523.4

4.3　追加実験（YOLO-World のプロンプト比較）

YOLO-World は学習データに含まれないクラスをテキスト（プロンプト）で指定して検出できる. そこで, プロンプトの違いが検出精度に与える影響を比較する追加実験を行った. 使用したプロンプトとそれぞれの意図を表 5 に示す. ファインチューニング済み YOLO-World に対し, スマートフォン撮影画像を入力として, 各プロンプトで推論した.

表 5　追加実験で使用したプロンプトとその意図

プロンプト	意図・特徴
`license plate`	最もシンプル
`vehicle license plate`	vehicle 追加により誤検出低減の可能性
`license plate on a vehicle`	空間制約による高精度化の可能性

表 6 に, プロンプト別の評価結果を示す. プロンプトを変更しても,Precision,Recall,F1,mAP@50,mAP@50-95 のいずれも大きな差はなく, プロンプト変更による精度の差は小さいと結論づけられる.

表 6　プロンプト別の検出性能（追加実験）

Prompt	Precision	Recall	F1	mAP@50	mAP@50-95
license plate	0.8141	0.5962	0.6888	0.6995	0.5229
vehicle license plate	0.8340	0.5798	0.6840	0.7031	0.5261
license plate on a vehicle	0.8342	0.5808	0.6848	0.6968	0.5192

4.4　考　　察

4.4.1　視覚-言語モデルによる汎化性能

実験結果より,YOLO-World が mAP および Precision において最も優れた性能を示した. この要因として学習手法の違いが挙げられる可能性がある.YOLOv11 や RT-DETR が画像のみから特徴を学習するのに対し,YOLO-World は画像とテキストのペアを用いた事前学習を行なっている. 一般に, 言語情報を補助タスクとして学習されたモデルは, 物体の意味的な概念をより深く獲得し, 未知のドメインや小規模なデータセットに対する**汎化性能（Generalization capability）**の向上に寄与すると考えられている. 本実験のような「定点カメラで学習し, スマートフォン画像で推論する」というドメインの乖離がある環境において,YOLO-World が持つ意味的な理解が, 未知の特徴に対する安定性として寄与したことが示唆される. また, 車両検出の後にナンバープレート検出を行う二段階構成とすれば, 改善の余地があると考えられる. 追加実験（4.3 節）では,`license plate` への適合が強く,`vehicle` を追加しても精度の改善は見られなかった.

4.4.2　データ量とモデル構造

一方,RT-DETR は YOLO 系モデルと比較して推論速度が遅く, 精度面でも YOLO-World に及ばなかった. これは RT-DETR の性能限界というよりも, 学習データ不足に起因する可能性が高い.Transformer ベースのモデルは,CNN ベースのモデルと比較して帰納的バイアスが弱い傾向にあり, 高い性能を発揮するにはより大規模なデータを必要とすることが多くの研究で指摘されている. 本実験のデータ規模ではドメインの乖離を解決できなかったと考えられる. YOLOv11 はシンプルなアーキテクチャにより推論速度が最速であり, したがって小規模データセットでの運用やリアルタイム性が求められる環境では YOLO シリーズ（特に YOLO-World）が適しているが, データセットの拡充が可能であれば,RT-DETR の採用も再考の余地がある.

5　ま　と　め

本研究では, スマートフォンによる撮影画像を対象としてナンバープレート検出システムの構築に向けた予備実験として, 最新の物体検出モデル 3 種の比較評価を行った. 6,000 枚の中規模データセットを用いた実験の結果, 視覚-言語事前学習を取り入れた **YOLO-World** が,Precision で 0.8155 および mAP@50-95 で 0.5197 を記録し, 最も高い検出性能と汎化能力を示した. 結論として, **学習データが少ない場合には YOLO-World が, 見逃しを減らすことを重視する場合には RT-DETR** が有効であることが示された.

今後の課題として, 以下の 5 点が挙げられる. 第一に, **車両検出の後にナンバープレート検出を行う二段階構成**の検討である. 第二に, **モバイル端末での性能評価**である. 実用化に向けては精度だけでなく, スマートフォン（エッジデバイス）上でのリアルタイム動作が求められる. 本研究では実装の安定性を重視し YOLOv11 を採用したが, 今後のモバイル実装フェーズにおいては, 最新の YOLO26 を含む複数の軽量モデルを対象に, 処理速度とバッテリー消費の観点からも比較検討を進める予定である. これに併せて, モデルの軽量化（量子化や枝刈り）と実機での遅延評価を行う. 第三に, **スマートフォン撮影画像を用いたファインチューニング**の検討である. 第四に, **エッジデバイス向け検出パイプライン**の検討である.

謝　　辞

本研究の一部は, 京都産業大学先端科学技術研究所（人間情報学研究センター）共同研究プロジェクト（M2301）の助成を受けたものである. ここに記して謝意を表す.

文　　献

[1]　いいね AI. 日本の主要 sns プラットフォーム徹底分析: 2025 年

最新データから見る利用動向, 2025. Accessed: 2026-01-04.
[2] 株式会社ノジマ 家電小ネタ帳編集部. 【一眼レフ並み！】カメラ性能が高いスマホをランキング形式でご紹介. Accessed: 2026-01-04.
[3] Ranjan Sapkota and Manoj Karkee. Ultralytics YOLO evolution: An overview of YOLO26, YOLO11, YOLOv8, and YOLOv5 object detectors for computer vision and pattern recognition. *arXiv preprint*, October 2025.
[4] Yian Zhao, Wenyu Lv, Shangliang Xu, Jinman Wei, Guanzhong Wang, Qingqing Dang, Yi Liu, and Jie Chen. DETRs beat YOLOs on real-time object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–10, 2024.
[5] Tianheng Cheng, Lin Song, Yixiao Ge, Wenyu Liu, Xinggang Wang, and Ying Shan. YOLO-World: Real-time open-vocabulary object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–15, 2024.
[6] Zhuoqun Fu. Deep learning for license plate number recognition: A survey. In *2024 IEEE International Conference*, pages 1–6. IEEE, 2024.

古文単語学習の効率化のための単語色彩表現システム

酒井 楓佳† 小宮 和真‡ 福原 義久§

†§武蔵野大学 データサイエンス学部 データイエンス学科 〒135-8181 東京都江東区有明 3-3-3

‡武蔵野大学 データサイエンス研究科 〒135-8181 東京都江東区有明 3-3-3

E-mail: †s2322028@stu.musashino-u.ac.jp, ‡g2550004@stu.musashino-u.ac.jp, §h053622@ptf.musashino-u.ac.jp

あらまし 語学学習において文章内容を的確に読み取る際，語彙の習得は必要不可欠である．特に古文の学習において，単語の意味が現在のそれとは大きく異なることが多く，古文そのものの学習の障壁となっている．そこで本稿では，古文学習における古文単語の記憶を目的とした，「古文を対象とした埋め込み表現による色彩表現システム」の実現方式を示す．具体的には，高等学校教育の対象とする古語 283 語に対して，埋め込み表現を用いて類似度の高い色彩表現を求めた．各古文単語を求められた色彩で表現することで，単語記憶にどのような変化が現れるかをペーパーテストを用いて検証した．検証の結果，古文単語に色彩表現を加味することで学習効率に一定の効果があることが認められた．

キーワード 古文学習，単語色彩表現，自然言語処理，埋め込み表現

1. はじめに

平成 27 年度学習指導要領実施状況調査では高等学校を対象としたペーパーテストでは古文や漢文の学習が大切だという質問に対して肯定的な回答をしている生徒の割合は 38.4%であり，古文の内容を的確に読み取り要約する能力を見る問題の通過率は 20.0%であることが報告されている[1]．古文を理解する上で古語を覚えるということはとても重要だが，この調査からはそれができずに古文の学習を難しく感じている生徒が多いことが推察される．そこで本研究では，“色字共感覚”から着想を得て，古語に色イメージを付与することで，意味的な印象を視覚的に捉える方法を提案する．これにより，古語の学習の難度が下がることが期待される．

共感覚とは，一つの感覚刺激から通常の感覚に加えて別の感覚が無意識に引き起こされる現象である．その中でも特に，文字に色がついているように感じるものを色字共感覚という．色字共感覚の特徴としては以下のような特徴が知られている[2][3][4]

- 感じる色を自分ではコントロールができない
- 個人内では一貫性がある
- 記憶を助けたり好きや嫌いといった情動を伴うことがある
- 感じる色は高い時間的安定性を持つ
- 文字の意味や読みが共感覚を左右するケースがあり，すでによく知っている字についても未知の読みや意味を学習させた際に，共感覚がわずかに変化することがある

これらのことから，古語の記憶に対しても，語と色を関連付けることで学習に対して補助的な役割を果たすのではないかと我々は考えた．

次に，古文を対象とした埋め込み表現による色彩表現システムの実現方式を示す．

2. 提案手法

2.1 古語単語の色彩表現を求める

本研究では，語に色彩を関連付けることで学習効果の増大を狙うものであるが，古語と現代語では，同じような語であっても意味が異なる場合があり，色のイメージと古語からの印象が合致するかは難しいと考えた．そこで，古語の主な意味をまず導き出し，それに対する色彩表現を求めることとした．

古語単語に対して色彩を割り当てる手順を図 1 および次ページに示す．なお，本研究では対象として，高校の古文教材[5]から選択した 283 語を用いた．

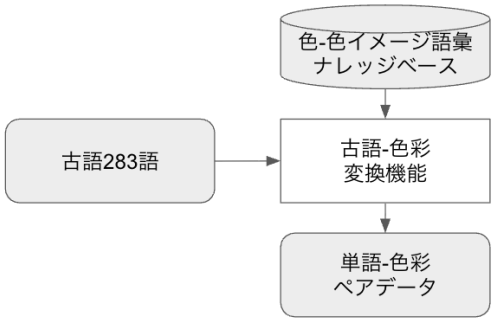

図 1 提案手法の概要図

提案手法の手順：

1. 各単語の意味を求める[5].
2. 決定版色彩心理図鑑[6]より，色彩とそれに対するイメージ語彙を求める（表 1）.
3. 色と色のイメージ語彙の関係を示すナレッジベースを構築した上で，ナレッジベース上の各イメージ語彙を OpenAI の埋め込みモデルである text-embedding-ada-002 を用いて埋め込み表現を求める.
4. 古語 283 語とその主な意味をそれぞれ同様に埋め込み表現に変換する.
5. 古語の意味の埋め込み表現とナレッジベース上の埋め込み表現のコサイン類似度を計量し，最も類似したものの色彩情報を求める.

表1　色彩とそれに該当するイメージ語彙

色彩	イメージ語彙
赤	情熱，興奮，怒り
青	落ち着き，知的，悲哀
黄	愉快，軽快，注意
橙	喜び，明るい，陽気
ピンク	可愛い，愛情，幸福
紫	上品，優雅，不安
緑	安らぎ，癒し，調和

2.2 提案手法の評価

色彩変換したものをもとに 20 人を対象としたペーパーテストを実施した．このペーパーテストでは単語に着色ありのグループと着色なしのグループの 2 つのグループを作り，10 人ずつグループに分けて実施した．着色ありのグループには色付けした単語とその意味を書いたものを，着色なしのグループには何も着色していない単語とその意味を書いたものを渡す．この時，着色ありのグループに渡す紙には色と色彩イメージ語彙も記述してある．

被験者には 10 分間で単語を覚えてもらい，その後 5 分間でテストを実施した．テストは単語の意味を書くもの，意味に当てはまる単語を書くもの，当てはまる意味を選択する問題の計 14 問を用意した．

ただし，本テストでは暗記直後にテストを行うため，短期記憶が検証対象となる．

3. 実験結果

3.1 提案手法の出力結果

表 2 は，古語をそのまま埋め込み表現に変換し，色彩を求めたものであり，表 3 は古語の意味を色彩に変換した提案システムの出力結果である．

例えば，「ののしる」をそのまま色彩表現に表すと緑色になるが，「大声で騒ぐ」を色彩表現に表すと赤色になることがわかる．

一方，「行ふ」の意味は「仏道修行を行う」であり，これはどちらも緑色になることが見て取れる．

図 2，図 3 は，古語をそのまま色彩変換した場合の割り当てられた単語の数と，古語の意味を色彩変換した場合の数を比較したものである．古語をそのまま色彩変換すると約 6 割が緑色に変換されていることがわかる．一方，古語の主な意味を色彩変換すると偏りが低減することがわかる．

表 2　古語と変換先の色彩

古語	色彩
ののしる	緑
行ふ	緑
時めく	青
なまめかし	緑

表 3　古語の意味と変換先の色彩

古語	主な意味	色彩
ののしる	大声で騒ぐ	赤
行ふ	仏道修行をする	緑
時めく	寵愛を受ける	ピンク
なまめかし	上品だ	紫

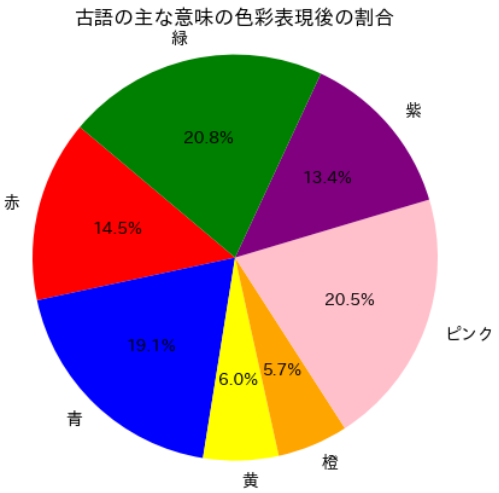

図 2 古語の意味の色彩表現後の割合

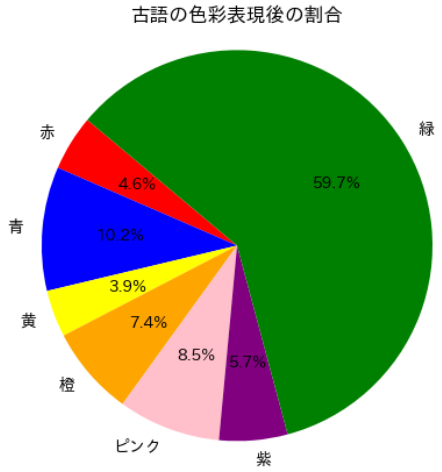

図 3　古語の色彩表現後の割合

3.2 評価実験の結果

評価実験の結果を図 4 から図 6 に示す．着色ありのグループは，最高得点と最低得点の点差が大きく標準偏差は 2.8 であった．一方，着色なしのグループは点数が偏る傾向があり，標準偏差は 2.1 であった．

なお，着色有りのグループの平均点は 8.7 点，着色無しのグループは 8.3 点であった．

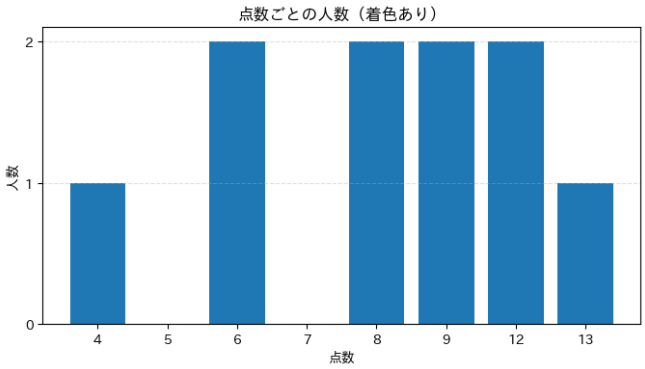

図 4　着色ありグループのテスト結果の棒グラフ

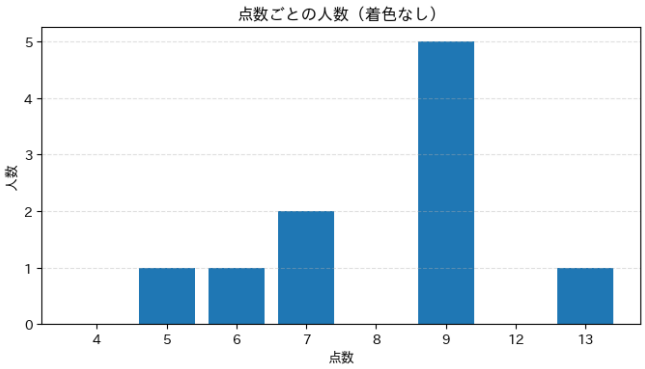

図 5　着色なしグループのテスト結果の棒グラフ

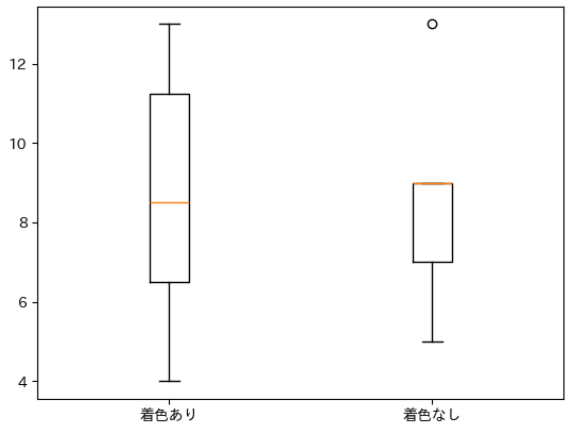

図 6　テスト結果の分布

4. 考察

4.1 古語の色彩表現について

提案手法により古語の意味的な印象を色で表現することができた．特に古語をそのまま変換した場合よりもはるかにバランスよく各色に配色されることがわかり，このことは学習する上でも有用な効果を持つと考えられる．一方で，設定された色と被検者がイメージする色がどの程度一致しているかについては今後の検証が必要である．

また，古語は同じ語でも使われる文脈の中で意味が変わる場合があるため，古語単体でのベクトル化だけでは妥当な結果が得られない可能性が考えられる．

4.2 評価実験について

実験結果より，着色ありのグループのテスト結果では，高得点を出す被験者が多い一方で，低い得点にも分布することがわかった．このことは，提案手法が用い効果をもたらす場合と逆の効果をもたらす場合がある可能性を示唆している．

また，今回の実験では，全体的にはスコアが向上することが確認されたが，被験者に対する事前のヒアリングやテストの観察結果から，古文単語を覚えることに対してポジティブな被験者や，逆に古文に対してネガティブな感情や苦手意識がある被験者がいることがわかった．また，暗記の得意不得意なども被験者ごとにあると考えられ，このような被験者固有の差異を考慮すると，より多くの被験者による実験やバイアスを取り除く仕組みが必要であるといえる．

5. 結論

本研究では，古語の主な意味を色彩表現に変換することで，意味的な印象を視覚化することができた．

また，出力結果を用いて評価実験を行った結果，古語を着色したグループでは，記憶テストで高い得点を

出す被験者が比較して多いという結果が得られた．このことから提案手法は，古語学習に対してポジティブな役割を果たしている可能性が考えられる一方，得点のばらつきが多いという結果も確認され，被験者によってはネガティブな影響を受ける可能性も示唆された．

6. 今後の展望

今後の研究では，提案手法の有効性をより明確にするため，評価実験の方法をさらに検討していく必要がある．今回の実験では，色彩表現を付与した古文単語の提示が学習効率に一定の効果をもたらす可能性が示されたが，被験者数や実験回数が十分とは言えない．今後は複数回の評価実験を実施し，実験結果に含まれるバイアスを可能な限り排除したうえで，被験者の学習効果の伸びをより詳細に分析することが求められる．また，学習時とテスト時の双方において色彩表現の有無を組み合わせた複数の実験条件を設定することで，色彩表現が学習過程および記憶想起のどの段階に影響を与えるのかを検証することも今後の課題である．

さらに，本研究では単語の意味に基づいて色彩表現を付与したが，色の定義や種類についても改善の余地がある．現在は複数の色を用いて表現しているが，色数を三原色などのより基本的な色体系に整理することで，色の意味付けをより明確にできる可能性がある．また,古文は成立した時代背景を持つ言語であるため，当時の文化や色彩に対する認識を考慮した表現方法についても検討する必要がある．

加えて，本研究では現代語を中心とした埋め込みモデルを利用しているが，古文の特徴をより適切に反映させるためには，古文コーパスを用いた埋め込み表現の構築についても検討する必要がある．古文特有の語義や文脈を考慮したモデルを利用することで，より適切な色彩表現の生成が可能になると考えられる．

最後に，本研究で提案した色彩表現を用いた語彙学習支援の手法は，古文単語に限らず，英単語などの外国語学習にも応用できる可能性がある．例えば，意味だけでなく発音や文法的特徴などに色彩表現を対応させることで，視覚的な手がかりを用いた新たな語学学習支援手法として発展させることが期待される．

参　考　文　献

[1] 平成 27 年度高等学校学習指導要領実施状況調査 教科・科目等別分析と改善点　国語総合
[2] 長田典子，“色と共感覚”，日本色彩学会誌,Vol.43,No.2,pp.111-114,2019
[3] Michiko Asano, So-ichiro Takahashi, Takuya Tsushiro, Kazuhiko Yokosawa; Synaesthetic colour associations for Japanese Kanji characters: from the perspective of grapheme learning. *Philos Trans R Soc Lond B Biol Sci* 9 December 2019; 374 (1787): 20180349. https://doi.org/10.1098/rstb.2018.0349
[4] リチャード・E・サイトウィック,デイヴィッド・M・イーグルマン，脳の中の万華鏡-「共感覚」の目くるめく世界,河出書房新社,2010
[5] 武田博幸,鞆森祥悟,読んで見て覚える重要古文単語 315[三訂版],桐原書店,2020
[6] ポーポー・ポロダクション,決定版　色彩心理図鑑,日本文芸社,2020
[7] 小川隼斗,堀尾海斗,河原大輔,和歌の埋め込みに基づく本歌取りの推定,言語処理学会,第 31 回年次大会発表論文集,2025
[8] 近藤泰弘，“和歌集の歌風の言語的差異の記述 －大規模言語モデルによる分析－“,日本語の研究,Vol.34,No.4,pp.105-117,2023

自由記述嗜好を考慮した多人数レシピ推薦における嗜好統合手法の比較分析

東　颯人†　宮森　恒†

† KyotoSangyoUniversity　〒 603–8555 京都府京都市北区上賀茂本山
E-mail: †{g2354323,miya}@cc.kyoto-su.ac.jp

あらまし　本稿では，自由記述で与えられる多様で曖昧な嗜好を前提とした多人数レシピ推薦の問題に取り組む．従来のレシピ推薦研究は単一ユーザを主対象としており，多人数の嗜好衝突を体系的に扱う研究は限られている．また，グループ推薦に関する既存手法の多くは，定型的に表現された嗜好入力を前提としており，自由記述嗜好を考慮した場合の有効性は十分に検証されていない．そこで，本稿では，自由記述で与えられる嗜好に対する多人数レシピ推薦において，嗜好統合手法が推薦品質および満足度の公平性に与える影響を定量的に分析する．具体的には，自由記述嗜好を正規化する前処理として，埋め込み表現，辞書ベース手法，大規模言語モデルを用いた補完手法を比較し，それらが多人数嗜好統合に与える影響を調査する．さらに，平均満足度最大化，最小満足度最大化，および満足度分散を考慮した統合手法を対象に，グループ人数や嗜好衝突度を変化させた設定のもとで性能評価を行う．

キーワード　多人数レシピ推薦, 自由記述, 嗜好統合, 嗜好正規化, 嗜好衝突度, 公平性

1　はじめに

近年，家族や友人，同僚など，複数人で食事をする機会は日常的に存在している．一方で，レシピサイトや SNS の普及により利用可能なレシピの選択肢は爆発的に増加しており,「何を食べるか」という意思決定は必ずしも容易ではない．特に複数人で食事を決める場合，各人の嗜好や制約が異なるため，意思決定の負担はさらに大きくなる．このような背景から，レシピ推薦システムに対する需要は高まっている．

従来のレシピ推薦研究の多くは，単一ユーザを対象とした推薦を主眼としており，個人の嗜好に基づいて最適なレシピを提示する手法が数多く提案されてきた [1], [2]．一方で，複数人の嗜好を同時に考慮する多人数レシピ推薦に関する研究は限定的であり，特に嗜好が衝突する状況を体系的に扱った研究は十分とは言えない．また，グループ推薦に関する既存研究の多くは，あらかじめ定型的に整理された嗜好入力を前提としており，自然言語による自由記述嗜好を入力とした場合の有効性については十分に検証されていない．

しかし，実際の利用場面では,「さっぱりしたものが食べたい」「辛い料理は苦手」「子供向けがよい」といったように，嗜好は自由記述で曖昧かつ主観的に表現されることが多い．このような自由記述嗜好には，表記揺れ，未知語，文脈依存の意味，曖昧な評価表現などが含まれ，嗜好を数値化・ベクトル化する際の大きな課題となる [3]．さらに，多人数の嗜好を単純に平均化するような統合手法では，一部のユーザの満足度が著しく低下し，推薦結果の公平性が損なわれる可能性がある．

そこで本研究では，自由記述で与えられる嗜好を前提とした多人数レシピ推薦に着目し，嗜好の正規化手法および嗜好統合手法の違いが，推薦品質および満足度の公平性に与える影響を定量的に分析することを目的とする．具体的には，自由記述嗜好を数値表現へ変換する前処理として，埋め込み表現のみを用いる手法，辞書ベースによる正規化手法，大規模言語モデルを用いた文脈補完手法を比較する．さらに，平均満足度最大化，最小満足度最大化，満足度分散を考慮した嗜好統合手法を対象に，グループ人数や嗜好衝突度を変化させた条件下で評価実験を行う．

本研究の貢献は以下の三点にまとめられる．第一に，自由記述嗜好を対象とした多人数レシピ推薦において，異なる嗜好正規化手法が推薦結果に与える影響を体系的に比較・分析する点である．第二に，複数の嗜好統合手法を公平性の観点から評価し，満足度最大化と個人間の納得感のトレードオフを明らかにする点である．第三に，嗜好衝突度やグループ人数といった実利用を想定した条件変化が推薦性能に及ぼす影響を実験的に示す点である．

本論文の構成は以下の通りである．第 2 章では関連研究について述べる．第 3 章では，自由記述嗜好を考慮した多人数レシピ推薦の提案手法について説明する．第 4 章では，実験設定および評価指標について述べる．第 5 章では，評価実験の結果を示し，正規化手法および嗜好統合手法の違いが推薦性能に与える影響について考察する．最後に，第 6 章で本研究をまとめ，今後の課題について述べる．

2　関連研究

2.1　単語・文書のベクトル表現

自然言語処理分野では,Word2Vec や fastText などの分散表現を用いて単語の意味を捉える手法が提案されてきた [1]．レシピ推薦においても, レシピ内の単語をベクトル化することで, 従来のキーワード検索では不可能だった「材料名は異なるが意味的に近いレシピ」の抽出が可能となっている（矢野ら, 2013）．近年では,OpenAI の text-embedding-3-small に代表される高

次元の埋め込みモデルが登場し, 自由記述文全体のコンテキストを高精度に捉えることが可能となった（川嵜ら, 2024）. 一方で、これらの埋め込み表現は意味的類似性の判定には優れるものの,「特定食材の除外」といった明示的な制約条件や, ユーザーごとの嗜好の強弱を直接的に制御しにくいという課題がある.

2.2 食材嗜好に関する研究

ユーザーの食嗜好を扱う研究は, アンケートや過去の調理履歴に基づく定量化が主流であった. 例えば、食材単位の「好き・嫌い」を数値化し, それに基づいてレシピをフィルタリングする手法がある（高畑ら, 2011）. しかし、実際の利用シーンでは「こってりしたもの」や「子供が喜ぶもの」といった抽象的かつ自由な形式で嗜好が表現されることが多い. これに対し、大規模言語モデル（LLM）を用いて自由回答を特定のレシピカテゴリや類似単語へ変換・正規化する試みが行われているが, 多種多様な表現をどの程度の解像度で集計すべきかについては議論の余地がある.

2.3 多人数レシピ推薦・グループ推薦

複数ユーザーを対象とするグループ推薦では、個人の満足度の平均値を最大化する手法や, 不満を持つユーザーを最小限に抑える手法などの嗜好統合アルゴリズムが提案されてきた [4], [5]. しかし, 先行研究の多くは「ユーザーの嗜好がすでに数値化されている」ことを前提としており, 自由記述から得られた曖昧な嗜好情報をどのように統合し, 推薦結果へ反映させるべきかの検討は不十分である. 特に人数が増えるほど嗜好の把握と「すり合わせ」の難易度は増大するため（川嵜ら, 2024）, 自由記述の正規化プロセスと嗜好統合アルゴリズムの組み合わせが推薦精度およびユーザーの納得感に与える影響を分析することは、実用上極めて重要である.

3 提 案 手 法

本研究では，自由記述で与えられる多人数の嗜好を対象としたレシピ推薦において，嗜好の正規化・嗜好統合・推薦・評価を一貫したパイプラインとして設計し，正規化手法と統合手法の組み合わせが推薦品質および公平性に与える影響を分析する手法を提案する.

3.1 手法全体の流れ

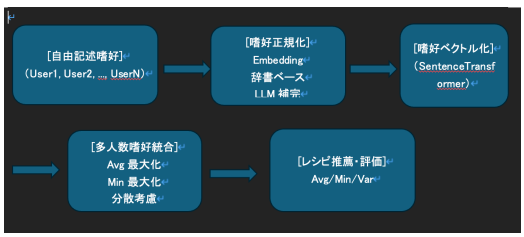

図 1 自由記述嗜好を考慮した多人数レシピ推薦の全体構成

本研究では，自由記述で与えられる多人数の嗜好に対して，嗜好正規化手法および多人数嗜好統合手法を組み合わせ，推薦品質および満足度の公平性への影響を比較分析する.

提案手法は以下の 5 段階から構成される.

1. 自由記述嗜好の入力
2. 自由記述嗜好の正規化
3. ユーザ嗜好のベクトル化
4. 多人数嗜好の統合
5. レシピ推薦および評価

各処理段階を明確に分離することで，どの要素が推薦結果に影響を与えているかを定量的に比較可能とする.

3.2 自由記述嗜好の正規化

自由記述嗜好は，表記揺れ，曖昧表現，暗黙的な制約を含むため，そのままでは推薦に利用しにくい. そこで本研究では，以下の 3 手法を用いて正規化を行い，比較する.

3.2.1 埋め込み単独手法

各ユーザの自由記述嗜好文全体を文埋め込みモデルに直接入力し，意味空間上のベクトルとして表現する. 文脈的意味を包括的に捉えられる一方で，調理制約や忌避食材といったハード制約を明示的に扱いにくいという課題を持つ. 本研究では本手法をベースラインとする.

3.2.2 辞書ベース手法

本研究では，辞書ベース手法として，形態素解析に基づくルールベースの嗜好正規化を行った. 具体的には，MeCab を用いて自由記述嗜好文を形態素解析し，品詞情報に基づいて名詞，形容詞，および動詞のみを抽出する. これにより，料理嗜好に関係しやすい語を選択的に残し，助詞や助動詞などの文法的要素を除去した簡潔な表現へと正規化する.

本手法では，あらかじめ定義した品詞選択ルールを簡易的な辞書として用いることで，自由記述文を構造化された単語列へ変換する. 正規化後のテキストは，多言語文埋め込みモデル（SentenceTransformer）を用いてベクトル表現へ変換し，推薦計算に利用する.

一方で，本手法は単語単位の抽出に基づくため，文脈的な意味関係や暗黙的な嗜好・制約を捉えることが困難であり，表現の柔軟性に限界があるという課題を有する.

3.2.3 LLM による文脈補完手法

本研究では，自由記述嗜好に含まれる曖昧表現や暗黙的制約を明示化するため，大規模言語モデル（LLM）を用いた文脈補完手法を導入する. 具体的には，ユーザの自由記述嗜好文に対して，料理推薦に適した「嗜好の正規化文」へ書き換える処理を行う.

LLM 補完は以下の手順で実施した. まず，LLM に対して,「暗黙の嗜好や制約を明示すること」,「好みの味や調理法，制約を具体化すること」,「一文から二文の簡潔な説明文とすること」,「新たな嗜好を追加しないこと」といった条件を明示的に指示する. その上で，自由記述嗜好文を入力し，文脈を保ったまま構造化された嗜好表現へと変換する.

例えば,「出汁 旨味 和食 揚げ物 NG」という自由記述に対しては，出汁の効いた和食を好み，揚げ物を避けたいという嗜

好・制約が明示的に表現された文章へと補完される．このように，LLM を用いることで，辞書ベース手法では捉えにくい文脈的意図や暗黙的制約を正規化嗜好として反映することが可能となる．

3.3 ユーザ嗜好のベクトル化

正規化後の嗜好情報は，文埋め込みモデル（multilingual-e5-base）を用いてベクトル化する．埋め込み単独手法では原文を入力とし，辞書ベースおよび LLM 手法では正規化後のテキストを入力とすることで，正規化方法以外の条件を統一した比較を行う．

3.4 多人数嗜好の統合

複数ユーザの嗜好ベクトルを統合し，グループ全体を表す嗜好ベクトルを生成する．このような嗜好統合は，グループ推薦における基本的課題として広く研究されてきた [4], [5]．本研究では以下の統合手法を比較対象とする．

3.4.1 平均満足度最大化

ユーザ嗜好ベクトルの平均をグループ嗜好として用いる手法であり，全体として高い平均満足度を得やすい一方，一部ユーザの不満が無視される可能性がある．

3.4.2 最小満足度最大化

各ユーザの満足度のうち最小値を重視する手法であり，強い不満を回避できるが，推薦結果が無難なものに偏りやすい．

3.4.3 満足度分散を考慮した手法

平均満足度に加えてユーザ間の満足度分散を考慮し，満足度の偏りが大きい推薦を抑制する．これにより，公平性と推薦品質のバランスを図る．

3.5 レシピ推薦と評価

各レシピをテキスト表現として埋め込みベクトル化し，グループ嗜好ベクトルとのコサイン類似度に基づいてランキングを生成する．評価指標として，平均満足度，最小満足度，満足度分散を用い，推薦品質と公平性を総合的に評価する．

4 実　　験

4.1 研 究 課 題

本研究では，自由記述嗜好を考慮した多人数レシピ推薦における，嗜好正規化手法および嗜好統合手法の有効性を検証するため，以下の研究課題（RQ）を設定する．

- **RQ1**：自由記述嗜好の正規化手法は，ユーザの潜在的な意図をどの程度正確にベクトル空間へ反映できるか．
- **RQ2**：異なる嗜好統合手法は，グループ全体の満足度と個人間の公平性にどのような影響を与えるか．

4.2 実 験 設 定

4.2.1 データセット

実験には，楽天レシピ API から取得した複数カテゴリ（和食・洋食・中華・その他）のレシピデータを用いた．

4.2.2 比 較 手 法

自由記述嗜好の正規化手法として，以下の 3 手法を比較対象とする．

- 埋め込み単独手法
- 辞書ベース手法
- LLM による文脈補完手法

さらに，多人数嗜好統合手法として，平均満足度最大化，最小満足度最大化，および満足度分散を考慮した手法を用いる．

4.2.3 評 価 指 標

推薦結果の評価には，以下の指標を用いる．

- **平均満足度**：グループ全体の推薦品質を表す指標．
- **最小満足度**：最も満足度の低いユーザの満足度を表し，公平性を評価する指標．
- **満足度分散**：ユーザ間の満足度のばらつきを表す指標．

4.3 実 験 手 順

まず，各ユーザの自由記述嗜好を，嗜好正規化手法ごとにベクトル化する．次に，異なる嗜好統合手法を用いてグループ嗜好ベクトルを生成し，レシピとの類似度に基づいて推薦を行う．

最後に，各評価指標を算出し，手法間で比較分析を行う．

4.4 結　　果

表 1 に，衝突度および制約強度の異なる 9 条件における，正規化手法別の top1 推薦に対する満足度指標（Avg,Min,Var）および推薦可能レシピ数を示す．

4.5 正規化手法ごとの推薦性能比較

まず，全条件を通じて平均満足度（Avg）は 0.624〜0.677 の範囲に分布しており，正規化手法間で極端な差は見られなかった．同様に，最小満足度（Min）も 0.582〜0.657 の範囲に収まり，全体としていずれの手法においても一定水準以上の推薦品質が確保されていることが確認された．一方で，条件ごとに最も高い満足度を示す正規化手法には違いが見られ，衝突度および制約の強さに応じた傾向が確認された．低衝突条件（1〜3）では，Embedding，辞書ベース，LLM のいずれの手法においても Avg および Min に大きな差は見られなかった．例えば 1【低×弱】では Embedding 手法が Avg=0.645 と最も高い値を示したものの，辞書ベース（0.633），LLM（0.624）との差は小さく，手法間の優劣は限定的であった．

衝突度が中程度の条件（4〜6）では，LLM 補完手法の優位性がより明確となった．特に 4【中×弱】および 5【中×中】では，LLM 手法が Avg および Min の両面で他手法を上回り，4 では Avg=0.673，Min=0.657 と最も高い値を示した．一方で 6【中×強】では，Avg は辞書ベース手法，Min は Embedding 手法が最大となり，制約が強い条件下では LLM 手法が必ずしも最良とはならないことが示された．

さらに衝突度が高い条件（7〜9）では，7【強×弱】において辞書ベースおよび LLM 手法が同程度の高い Avg を示したのに対し，8【強×中】および 9【強×強】では LLM 補完手法が Avg および Min の両面で最も高い値を記録した．特に 9 では，LLM 手法が Avg=0.669 と他手法を上回っており，高衝突・高

表 1　条件別・正規化手法ごとの満足度指標（Top1）

条件	衝突×制約	手法	Avg	Min	Var	推薦数
1	低×弱	Embedding	0.645	0.599	0.0009	5
		辞書	0.633	0.582	0.0012	5
		LLM	0.624	0.586	0.0004	5
2	低×中	Embedding	0.645	0.603	0.0011	5
		辞書	0.632	0.597	0.0009	5
		LLM	0.651	0.603	0.0013	5
3	低×強	Embedding	0.649	0.615	0.0005	5
		辞書	0.647	0.621	0.0004	5
		LLM	0.656	0.615	0.0008	5
4	中×弱	Embedding	0.659	0.624	0.0004	5
		辞書	0.652	0.619	0.0004	5
		LLM	0.673	0.657	0.0002	5
5	中×中	Embedding	0.658	0.616	0.0006	2
		辞書	0.654	0.612	0.0007	2
		LLM	0.677	0.639	0.0007	2
6	中×強	Embedding	0.659	0.623	0.0002	5
		辞書	0.669	0.642	0.0004	5
		LLM	0.648	0.616	0.0003	5
7	強×弱	Embedding	0.642	0.624	0.0004	5
		辞書	0.667	0.654	0.0001	5
		LLM	0.667	0.647	0.0002	5
8	強×中	Embedding	0.666	0.620	0.0009	2
		辞書	0.659	0.618	0.0008	2
		LLM	0.675	0.637	0.0007	2
9	強×強	Embedding	0.661	0.618	0.0007	5
		辞書	0.656	0.619	0.0006	5
		LLM	0.669	0.624	0.0008	5

表 2　条件別に最も高い Avg を示した正規化手法

条件	衝突×制約	最良手法（Avg）
1	低×弱	Embedding
2	低×中	LLM
3	低×強	LLM
4	中×弱	LLM
5	中×中	LLM
6	中×強	辞書
7	強×弱	辞書/LLM
8	強×中	LLM
9	強×強	LLM

表 3　衝突度別における統合手法間の順位一致傾向（5 人条件）

衝突度	平均最大化	最小最大化	分散考慮
低（1-3）	上位 1-2 位一致	上位 1-2 位一致	変動大（特に 3）
中（4-6）	1 位は概ね一致	1 位は概ね一致	全順位変動多い
高（7-9）	一致限定的	一致限定的	ほぼ全条件で変動

制約条件における有効性が示唆される．

満足度分散（Var）は，全条件において 0.0001〜0.0013 と非常に小さく，いずれの正規化手法においても極端な不満が発生していないことが分かる．ただし，低衝突条件 1 では辞書ベース手法が Var=0.0012 と比較的大きな値を示した一方，LLM 補完手法は 4【中×弱】において Var=0.0002 と最も小さく，満足度のばらつきを抑える効果が確認された．

また，推薦数については，制約が中程度かつ衝突度が中〜高の条件（5，8）において，いずれの正規化手法でも推薦レシピ数が 2 件に制限される現象が確認された．これは，嗜好制約の組合せによって候補集合自体が大きく縮小される状況を反映していると考えられる．

4.6　嗜好統合手法別ランキングの安定性

平均満足度最大化，最小満足度最大化，分散考慮の 3 つの統合手法について，ランキング変動を比較した．衝突度が低い条件（1〜3）では，平均満足度最大化および最小満足度最大化において，上位 1 〜 2 位の順位が多くの手法で一致しており，ランキングは比較的安定していた．一方，分散考慮手法では，特に 3（低×強）において順位変動が大きく，公平性を重視する統合が推薦順位に強く影響することが示唆された．衝突度が中低度の条件（4〜6）では，いずれの正規化手法においても，上位 1 位は大きくの場合一致したが，2 位以降の順位は大きく変動した．特に分散考慮手法では，全順位が異なるケースが多く見られた．衝突度が高い条件（7〜9）では，平均満足度最大化および最小満足度最大化においても順位の一致は限定的であり，分散考慮手法ではほぼすべての条件で異なるランキングが生成された．

4.7　グループ人数およびユーザ評価との関係

グループ人数を変更した実験では，人数を減らすにつれて Avg および Min が上昇し，満足度分散（Var）は減少する傾向が確認された．特に 4 人グループの場合に最も高い満足度が得られた．また，5 人グループにおいて推薦数が 2 件に制限されていた条件では，人数を減らすことで推薦数が 4 件に増加することが確認された．ユーザ評価との一致については，9（衝突度：高×制約：強）においてのみ，LLM 手法の平均満足度最大化によるランキングがユーザ評価と一致した．他の条件では，一致するランキングは確認されなかった．

5　考　　察

本実験から，自由記述嗜好を前提とした多人数レシピ推薦において，正規化手法および嗜好統合手法が推薦結果に与える影響が明らかとなった．このような推薦品質と公平性のトレード

オフは，従来のグループ推薦研究においても指摘されてきた課題である [4], [5]．まず，正規化手法については，全体的な満足度指標の絶対値に大きな差は見られなかったものの，衝突度および制約が高まる条件下では，LLM 手法が平均満足度および最小満足度の両面で優位となる傾向が確認された．これは，LLM による補完が，自由記述嗜好に含まれる曖昧な表現や潜在的制約をより適切に解釈できたためと考えられる．一方で，6（中×強）のように制約が厳しい条件では，LLM 手法の性能が必ずしも最良とはならなかった．これは，制約が過度に強い場合，LLM による意味拡張がかえって制約解釈の揺らぎを生み，満足度の低下につながった可能性を示唆している．嗜好統合手法に関しては，平均満足度最大化は多くの条件で安定した数値を示した一方，ランキングの差異が小さく，衝突状況を十分に反映できていない可能性がある．最小満足度最大化および分散考慮手法は，公平性を重視する観点では有効であるが，特に衝突度が高い条件ではランキングの不安定性が顕著であった．さらに，ユーザ評価との一致が限定的であった点から，数値指標としての満足度と実際の人間の評価との間には乖離が存在することが示された．特に高衝突・高制約条件においてのみ LLM 手法が一致したことは，人間の判断が暗黙的に高度な意味統合を行っている可能性を示唆している．以上より，自由記述嗜好を扱う多人数レシピ推薦においては，LLM を用いた正規化が有効となる場面は多いものの，制約の強さや衝突度に応じて手法を使い分ける必要があることが明らかとなった．

6 まとめ

本研究では，自由記述で与えられる多様かつ曖昧な嗜好を前提とした多人数レシピ推薦問題に対して，嗜好正規化手法および多人数嗜好統合手法の違いが推薦品質および満足度の公平性に与える影響を定量的に分析した．

まず，自由記述嗜好の正規化手法として，埋め込み単独手法，辞書ベース手法，および LLM による文脈補完手法を比較した．その結果，全条件を通じて平均満足度および最小満足度の絶対値に大きな差は見られなかったものの，衝突度および制約強度が高い条件においては，LLM 補完手法が平均満足度および最小満足度の両面で他手法を上回る傾向が確認された．特に高衝突・高制約条件では，LLM 補完手法がグループ内の満足度のばらつきを抑制し，公平性の高い推薦を実現できることが示唆された．

一方で，制約が強い条件の一部では，LLM 補完手法が必ずしも最良の結果を示さないケースも確認され，意味補完による解釈の揺らぎが満足度低下につながる可能性が示された．この結果は，自由記述嗜好に対する高度な意味解釈が有効である一方で，制約条件の厳密な取り扱いには慎重な設計が必要であることを示している．

次に，多人数嗜好統合手法の比較から，平均満足度最大化は全体として安定した推薦品質を示す一方，特定ユーザの不満が反映されにくい傾向が確認された．最小満足度最大化および満足度分散を考慮した手法は，公平性の観点では有効であるものの，衝突度が高い条件では推薦ランキングの変動が大きくなることが明らかとなった．これらの結果から，多人数推薦においては，推薦品質と公平性の間に明確なトレードオフが存在することが示された．

以上より，自由記述嗜好を扱う多人数レシピ推薦においては，嗜好正規化と嗜好統合を分離して設計し，衝突度や制約強度に応じて LLM による文脈補完と公平性を考慮した嗜好統合手法を適切に組み合わせることが，推薦品質と利用者の納得感を両立する上で有効であると結論付ける．

今後の課題として，評価対象となるユーザ数およびレシピ数の拡大，実利用環境における主観評価の導入，および LLM による嗜好補完過程の説明可能性向上が挙げられる．

7 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] 矢野 達也, 林 豊洋, 大橋 健, “単語のベクトル表現を用いたレシピ推薦システム,” 人工知能学会研究資料, SIG-Challenge-053-4(3/24),2019.
[2] 高畑 麻理, 上田真由美, 中島 伸介 “食材に対する好き嫌いを考慮した料理レシピ推薦手法の提案,” DEIM Forum 2011 E3-5
[3] 川嵜翔太, 飯塚雅文, 石川玲奈, 三村明農, 森田航平, 吉村魁人, 渕田孝康 “多数の参加者の好みを考慮したレシピ推薦システムの提案,” 2024 年度電気・情報関係学会九州支部連合大会，09-1P-08, 2024．DOI: 10.11527/jceeek.2024.0_184.
[4] J. Masthoff, “Group Modeling: Selecting a Sequence of Television Items to Suit a Group of Viewers,” User Modeling and User-Adapted Interaction, 2004.
[5] S. Amer-Yahia et al., “Group Recommendation: Semantics and Efficiency,” Proc. VLDB, 2009.
[6] R. Burke, “Hybrid Recommender Systems: Survey and Experiments,” User Modeling and User-Adapted Interaction, 2002.

強化学習による差動二輪車制御における未知実証環境での安全推論

門垣　幸樹†　　高井　勇志††　　宮口　幹太††　　北野　信吾††　　大島　裕明†

† 兵庫県立大学 情報科学研究科　〒 651–2197 兵庫県神戸市西区学園西町 8-2-1
†† 株式会社竹中工務店 技術研究所　〒 270–1395 千葉県印西市大塚 1-5-1
E-mail: †mo.0709.uoh@gmail.com, ††{takai.takeshi, miyaguchi.mikita, kitano.shingo}@takenaka.co.jp, †ohshima@ai.u-hyogo.ac.jp

あらまし　本研究では，強化学習を用いて，現地学習を必要とせずに，未知実証環境で差動二輪車を安全に制御する手法を提案する．本研究における未知実証環境とは，学習環境とは異なる環境条件や予期せぬ外乱が発生する環境を指す．学習時に経験しなかった外乱が存在する環境では，ロボットの動作が不安定化し，軌道逸脱のような危険が発生する可能性がある．本研究では，タスク遂行における危険を軌道逸脱と定義し，将来の軌道逸脱を予測する危険判別器を構築する．そして，タスクの効率的達成に特化した最適方策と，安全状態への復帰に特化した安全方策を，危険判別器が推論した危険度に基づいて動的に切り替える安全推論制御を提案する．実験では，車輪の不調や路面摩擦の変化等を再現した 6 種類の未知実証環境を用いて，従来手法と提案手法の安全性と効率性を比較評価した．実験の結果，重度の車輪故障環境において，提案手法は経路誤差を抑制し，それに伴い成功率も従来手法の 79.0%から 93.2%へと向上した．本手法により，現地学習を必要とせずに，未知実証環境下でのタスク遂行の安全性を向上させる可能性が示された．

キーワード　強化学習，車体制御，未知実証環境

1　はじめに

近年，月面や災害現場といった，人間が直接調査することが困難または危険な未知実証環境において，自律型探査ロボットの活用が期待されている [1], [2], [3]．これらのロボットの制御方策は，事前に構築された学習環境で獲得されることが多い．しかし，未知実証環境について事前に情報を得ることは困難であるため，全く同じ環境条件の学習環境を設計することはできない．そのため，未知実証環境で予期せぬ外乱や環境条件に遭遇した際に，学習環境に最適化された方策は行動が不安定化し，危険な行動を引き起こす可能性がある．一度の致命的な失敗がタスクの成否を左右するような状況では，制御方策にはタスク達成能力だけでなく，高い安全性が求められる．

これまでの研究では，未知環境に遭遇した後にオンラインで追加学習を行い，方策を環境に適応させる手法が提案されている [4], [5]．この手法は環境変化に対して柔軟に適応できる利点がある一方で，現地での学習には多大な計算コストや時間が必要となるため，エネルギーや計算資源が限られる過酷な環境では適用が難しい．また，安全性を考慮した研究として，安全制約を導入することで危険な行動を抑制する安全強化学習があるが，この手法は既知の安全領域を前提としており，学習時に想定されていない未知の危険領域に対しては十分に対応できないという問題がある．

そこで本研究では，強化学習による未知実証環境での差動二輪車制御において，現地での学習を必要としない安全推論制御手法を提案する．本手法を用いた制御概要図を図 1 に示す．本手法は，タスクの効率的な達成に特化するよう事前学習した最適方策と，危険状態からの復帰に特化するように事前学習した安全方策という，役割の異なる 2 つの方策を用いる．本研究では，危険として軌道逸脱を扱い，将来の軌道逸脱を予測する危険判別器を構築する．そして，未知実証環境での制御時に危険判別器が推論した危険度に応じてこれらの方策を動的に切り替えることで，効率性と安全性を両立させた制御手法の実現に取り組む．

2　関連研究

安全性を考慮した強化学習制御や未知環境での強化学習制御に関する様々な研究が行われている．

2.1　安全強化学習

安全強化学習は，期待報酬の最大化だけでなく，事前に定義された安全制約を満たす方策を学習することを目的とする．代表的なアプローチとして，制約付きマルコフ決定過程に基づく手法が挙げられる．Achiam ら [6] が提案した Constrained Policy Optimization（CPO）は，信頼領域法を拡張し，期待報酬を向上させつつ，安全制約違反の期待値を所定の閾値以下に抑えるように方策を更新する手法である．これにより，学習プロセス全体を通じて安全性を維持することが可能となる．安全強化学習に関する包括的な調査 [7] においても，制約付き最適化や安全層の導入が主要なアプローチとして挙げられている．

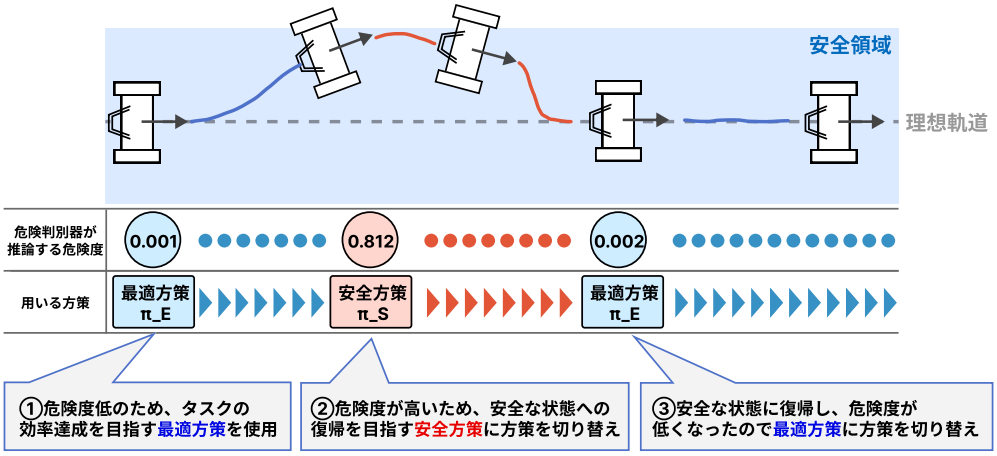

図 1　軌道逸脱を危険とした場合の本手法による制御の概要

また，タスク遂行と安全確保の役割を分離するアプローチも提案されている．Thananjeyan ら [8] による Recovery RL は，タスク方策とは別に，危険な状態からの回復に特化した回復方策を学習する．この手法では，安全判定器を用いて将来の制約違反確率を予測し，危険度が高いと判断された場合に制御権を回復方策に移すことで，探索中の安全性を高めている．

2.2　メタ学習による未知環境でのオンライン適応

未知環境への適応手法として，メタ学習が注目されている．メタ学習は学習する方法を学習する枠組みであり，新たなタスクや環境に対して，少数のデータを用いて迅速に適応することを目的とする．代表的な手法である Model-Agnostic Meta-Learning（MAML）[9] は，わずかな勾配更新で新しいタスクに適応可能な初期パラメータを学習する手法である．

ロボティクスの分野においても，シミュレーションと実環境のギャップを埋めるためにメタ学習が応用されている．Arndt ら [4] は，MAML をシミュレーション等で学習したモデルを現実世界に転移させて制御する Sim-to-Real に適用し，シミュレーションで多様な物理パラメータを経験させることで，実機上の未知の力学特性に対して数回の更新で適応可能な方策を獲得できることを示した．また，Nagabandi ら [5] は，モデルベース強化学習とメタ学習を組み合わせることで，4 脚ロボットが故障や環境変化に対してオンラインで適応する手法を提案している．Wang ら [10] は，メタ学習の過程で軌道最適化手法である iLQR [11] を用いて自己生成した教師データを利用する手法を提案している．人間のデモンストレーションを用いることなく訓練時のサンプル効率を大幅に改善し，結果として未知の操作タスクに対しても高い適応性能が得られることを示している．

しかし，これらの勾配ベースのメタ学習手法は，実環境においてエージェントがデータを収集し，その場で勾配計算とパラメータ更新を行う必要がある．Arndt ら [4] の研究でも示されているように，適応のためには実機データの収集とバックプロパゲーションが不可欠であり，計算リソースやメモリが制限された探査ロボット等においては，計算コストが大きな課題となる．また，オンライン学習中の挙動の安定性を保証することも困難である．対して本研究では，オンラインでの勾配更新を行わず，推論のみで完結する動的な方策切り替えを用いるため，計算コストを抑えつつ未知環境へ適応可能である点で優位性があると考える．

2.3　方策切り替えによる適応制御

単一の方策ではなく，状況に応じて複数の方策や制御器を動的に切り替えることで，未知環境や不確実性に対応するアプローチが提案されている．代表的なものとして，状況の不確実性に応じて行動を調整する手法や，危険な行動を事前に検知して介入する手法がある．

不確実性に応じた行動調整として，Kahn ら [12] は，衝突確率の不確実性をブートストラップ法 [13] により推定し，不確実性が高い場合にはロボットの速度を低下させる手法を提案している．同様に，Tran ら [14] は，複雑な環境における自律航行のために，探索と活用の方策を動的に切り替える協調的深層強化学習フレームワークを提案している．また，Chemingui ら [15] は，オフライン強化学習において，安全性制約に応じて複数の方策を適応的に切り替える手法の有効性を示している．

一方，危険な行動への介入による安全確保として，方策が選択しようとする行動を事前にチェックし，危険と判断された場合に別の安全な行動へ強制的に変更するシールドを用いた手法が提案されている．前述の不確実性に応じた行動調整が速度低下などの緩やかな対応を行うのに対し，シールド手法は危険な行動そのものを遮断するより直接的な介入を行う．Alshiekh ら [16] は，エージェントが将来的に危険な状態に陥る可能性のある行動を選択しようとした際に，即座に介入して安全な行動へと切り替える手法を提案している．この手法により，学習の試行錯誤の最中であっても事故の発生を防ぐことができる．Yang ら [17] は確率モデルに基づくシールド手法を提案し，安全制約を方策学習に組み込むことで安全性を向上させている．

本研究では，軌道逸脱における予測した危険度を用いて，最適方策と安全方策を推論時に切り替えることで，学習されていない未知の外乱に対しても安全な挙動を実現する．上述のシー

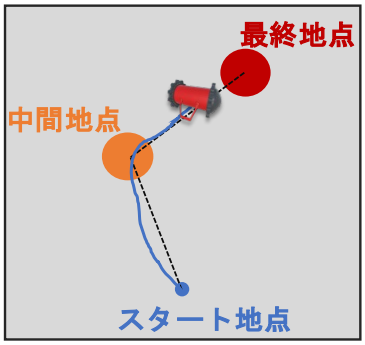

図 2　制御タスクと理想の行動

ルド手法が高度な論理推論や適応機構を用いるのに対し，本研究は計算リソースの限られたロボットにおいて，推論のみで完結する軽量な手法により，制御の効率性と安全性の両立を目指す．

3　問題定義とシミュレーション環境

本節では，本研究が対象とする制御タスクと，実験に用いるシミュレーション環境について述べる．

3.1　問題定義

本研究では，学習環境とは異なる環境条件を持つ未知実証環境での安全性を考慮した制御問題を扱う．本研究で扱う制御タスクを図 2 に示す．制御対象として差動二輪車を用いる．差動二輪車は，左右の車輪を独立に制御することで前後左右に動作可能な車両型ロボットである．制御タスクとしては，指定されたスタート地点から中間地点を経由し，最終地点に到達後，一定時間その場に留まる動作とする．このタスクにおける理想的な制御とは，各地点間を結ぶ直線経路からの逸脱を最小限に抑え，素早く効率的にタスクを達成することである．

本研究で想定する未知実証環境とは，学習時には経験しなかった環境条件や外乱が発生する環境と定義する．具体的には，車輪への異物の巻き込みやシャフトの不具合によって片輪の回転が重くなる等の車輪の出力異常や，路面摩擦の変化などが挙げられる．このような環境下で，学習環境でのみ学習を行なったモデルを用いて制御を行なった場合，直進しようとしても車体が意図せず旋回してしまうといった不安定な挙動を引き起こす．その結果，本来辿るべき軌道から大きく逸脱し，障害物への衝突や車体の転倒など，タスク継続が困難な危険な状態に陥る可能性が高い．

本研究における危険な状態とは，タスクの失敗に直結，あるいはその可能性を高める状態と定義し，タスク遂行における危険として軌道逸脱を用いる．

本研究の目的は，上述した未知実証環境において，エージェントが危険な状態に陥ることを防ぎ，あるいは陥った際に速やかに安全な状態へ復帰させ，制御タスクを効率よく安全に達成する制御方法を実現することである．

3.2　MDP の定式化

本研究では，上述の制御問題を観測空間 $\mathcal{O}$，行動空間 $\mathcal{A}$，遷移確率 $\mathcal{P}$，報酬関数 $\mathcal{R}$ からなるマルコフ決定過程（MDP）として定式化する．エージェントは各ステップ t において，環境からの観測 $o_t \in \mathcal{O}$ に基づき行動 $a_t \in \mathcal{A}$ を決定し，その結果として次の観測 o_{t+1} と報酬 r_t を受け取る．

本研究で対象とする差動二輪車の行動空間 $\mathcal{A}$ は，左右の車輪への出力値からなる 2 次元の連続空間として定義される．

$$\mathcal{A} = \{(a^{(left)}, a^{(right)}) \mid -1.0 \leqq a^{(left)}, a^{(right)} \leqq 1.0\} \tag{1}$$

ここで，1.0 は最大速度での前進，-1.0 は最大速度での後退を意味する．

観測空間 $\mathcal{O}$ および報酬関数 $\mathcal{R}$ の具体的な設計については，最適方策と安全方策の必要な情報が異なるため，詳細な定義は第 4 節にてそれぞれ述べる．

3.3　シミュレーション環境

本研究では，方策の学習と評価を行うためのシミュレーション環境として Unity ML-Agents [1]を使用する．制御対象は，本体と左右の車輪から構成される差動二輪車である．車体には，外部環境および自身の状態を認識するためのセンサとして，3 つの機能が搭載されている．1 つ目は IMU（慣性計測装置）である．IMU は車体の 3 軸加速度および角速度を計測し，自身の姿勢の不安定化の予兆を検知することができる．2 つ目は，IR センサである．これは，車体前方の左右に配置され，障害物や壁までの距離を計測することができる．3 つ目は，自己位置推定機能である．これは理想軌道に対する相対位置や，目標地点までの距離を正確に取得できる機能である．これらのセンサ情報に基づき，エージェントは環境の状態を観測する．

本研究における学習環境と評価環境の設計を図 3 に示す．学習環境では，Domain Randomization（DR）を用いて路面の凹凸，車輪の不調，および地面摩擦をエピソードごとにランダムに変動させる．これらのパラメータはエピソード開始時に決定され，エピソード中は一定である．路面の凹凸は，地形の高さをフラクタルノイズで生成し，振幅とスケールを変動させることで多様な起伏を表現する．車輪の不調は，エージェントが出力しようとしたトルクに対して実際に出力されるトルクが減少する状態を表す．学習環境では出力しようとしたトルクに 0.7 ～ 1.0 の範囲で変動する出力係数を乗じることで，車輪の不調を模擬する．地面摩擦係数は 0.3 ～ 1.0 の範囲で変動させる．これにより，エージェントは多様な環境条件を経験し，未知環境に対するロバスト性を獲得する．

評価環境である未知実証環境には，学習環境とは地形設計そのものが異なる富士麓の洞窟環境を使用する．この環境上で，以下の 2 種類の条件を設定し評価を行う．

OOD 条件　学習時に経験した外乱や環境条件の変動範囲から逸脱した条件である．車輪の不調では出力係数を 0.3 ～ 0.7

1：https://github.com/Unity-Technologies/ml-agents

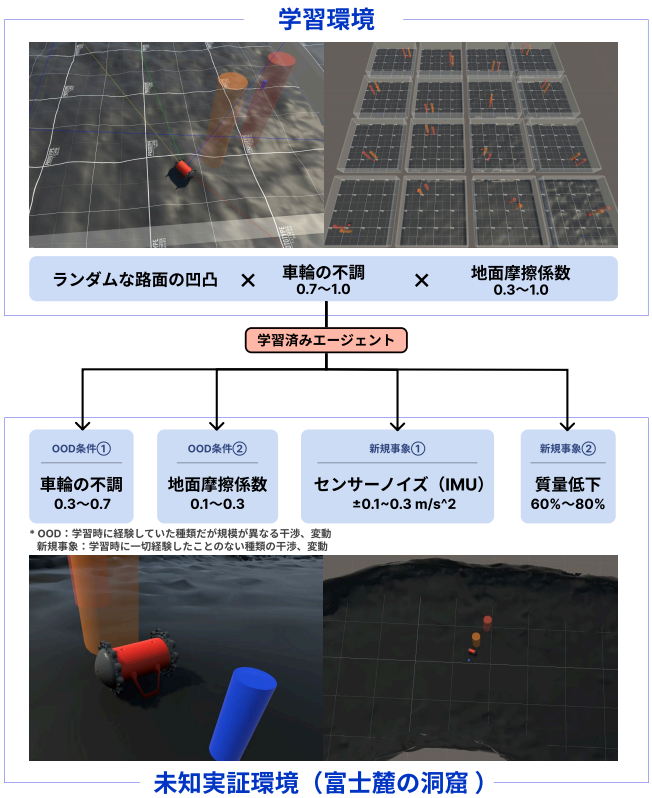

図 3　学習環境と未知実証環境の設計概要.

とし，学習時より深刻な故障を模擬する．地面摩擦係数では 0.1 ～ 0.3 とし，学習時より滑りやすい路面を設定する．

新規事象　学習時には一切経験していない種類の干渉である．センサノイズでは，IMU の加速度に $\pm 0.1 \sim 0.3$ m/s^2 の定常バイアスを付加し，センサ異常を模擬する．車体の欠損では，部品欠損等による質量低下を想定して，車体質量を 60% ～ 80% に低下させる．

4　方策切り替えを用いた安全推論制御

本節では，未知実証環境において現地学習を必要とせずに安全性とタスク遂行効率を両立する安全推論制御手法について述べる．本手法は，第 4.1 節と第 4.2 節で述べる役割の異なる 2 つの事前学習済み方策を，第 4.3 節で述べる危険判別器が予測する危険度に応じて動的に切り替える．これにより，安全性の高い状況では効率的な制御を，危険性の高い状況では安全性を優先した制御を実現する．

4.1　最適方策

最適方策は，タスクを素早く効率的に達成することを目指して，制御を行うように学習される方策である．

4.1.1　学習環境とカリキュラム学習

最適方策の学習アルゴリズムには，Proximal Policy Optimization（PPO）[18] を採用した．未知実証環境に対するロバスト性を獲得するため，Domain Randomization（DR）を用いて，物理パラメータがエピソードごとにランダムに変動する環境で学習を行なった．DR はシミュレーション環境において摩擦係数や車輪の不調，路面の凹凸などの環境パラメータに意図的なばらつきを導入することで，多様な環境条件に対するロバストな制御方策の獲得を目指す手法である．近年の研究では，DR が Sim-to-Real 問題の解決に有効であることが示されている [19], [20], [21]．また，著者らの先行研究 [22] において，本研究と同様の差動二輪車制御タスクにおいて DR が未知実証環境への適応に有効であることを確認している．本研究では，この結果を踏まえ，DR で学習した方策をベースラインとして採用し，その上で危険判別器による動的な方策切り替えを導入することで，さらなる安全性の向上を目指す．なお，多様な環境条件を一度に学習することは困難であるため，3 段階のカリキュラム学習を導入し，段階的に環境の変動範囲を拡大することで安定した学習と高い汎化性能の獲得を図った．

表 1　最適方策と安全方策の観測空間一覧

方策	観測項目	次元	内容概要
最適	中間目標情報	4	相対位置，ヨー角誤差，距離
	最終目標情報	4	相対位置，ヨー角誤差，距離
	到達フラグ	1	中間地点への到達状態
	姿勢情報	2	車体のピッチ角およびロール角
	IMU 情報	3	車体座標系における 3 軸線形加速度
	IR センサ	2	左右のセンサによる障害物までの距離
	角速度	1	Y 軸周りの角速度
安全	直近目標情報	4	直近目標への相対位置，ヨー角誤差，距離
	復帰基準情報	2	復帰軌道始点からの変位
	制御誤差情報	3	クロストラックエラー，方位角誤差など
	姿勢情報	2	車体のピッチ角およびロール角
	IMU 情報	3	車体座標系における 3 軸線形加速度
	IR センサ	2	左右のセンサによる障害物までの距離

4.1.2　観測空間

最適方策の観測空間は，タスクの進行状況と自己の状態を把握するための計 17 次元のベクトルで構成される．観測空間の詳細を表 1 に示す．タスク達成に直接関わる中間地点や最終地点に関する情報のほか，姿勢安定性の監視のための IMU 情報や，障害物回避のための IR センサ情報が含まれている．

強化学習モデルへ入力される情報としては，現在の観測情報 o_t と 1 ステップ前の観測情報 o_{t-1}，および 1 ステップ前の行動情報 a_{t-1} をスタックした，計 36 次元のベクトル o_{t-1}, o_t, a_{t-1} を用いる．

4.1.3　報酬設計

最適方策の報酬関数は，効率的なゴール到達を促しつつ，無駄な時間経過や危険な挙動を抑制するように設計した．各ステップの報酬 r_t は，以下の式で計算される．

$$r_t = r_{progress} + r_{time} + r_{posture} + r_{event} \tag{2}$$

進捗報酬 $r_{progress}$ は，車体が目標方向を向いている場合に，目標への接近距離 Δd に係数 200 を掛けた値を付与する．これは「目標に 1m 近づくと報酬 200 を得る」ことを意味し，後退による接近は評価しない．時間ペナルティ $r_{time} = -0.1$ は

表 2　最適方策と安全方策の報酬設計一覧

方策	報酬項	値/係数	説明
最適	進捗報酬	200.0	目標接近距離に乗算
	中間地点到達	+50	中間地点到達時
	最終地点到達	+100	停止条件達成時
	時間ペナルティ	−0.1	毎ステップ
	姿勢ペナルティ	−0.1	車体の傾きに応じて
	逸脱ペナルティ	−1.0	到達時に蓄積値を減算
安全	復帰進捗	1000.0	軌道接近距離に乗算
	復帰成功	+5.0	安全領域復帰時
	方向ペナルティ	−0.1	軌道付近での角度誤差
	姿勢ペナルティ	−0.1	車体の傾きに応じて
	時間ペナルティ	−0.01	毎ステップ

毎ステップ付与され，無駄な停滞を抑制する．姿勢ペナルティ $r_{posture}$ は，車体の傾きに応じて最大 -0.1 を付与し，安定走行を促す．

イベント報酬 r_{event} として，中間地点到達時に+50，最終地点での停止成功時に+100 を付与する．また，理想軌道からの逸脱については，走行中の逸脱距離を蓄積しておき，目標地点に到達した時点で係数 -1.0 を掛けてペナルティとして減算する．これにより，学習の安定性を保ちつつ直線的な走行を促進する．各報酬項の具体的な値を表 2 に示す．

4.2 安全方策

安全方策は，理想軌道から逸脱した危険な状態から，軌道上へ復帰することに特化した方策である．本研究では，タスク遂行における危険として軌道逸脱を対象とし，安全方策は軌道復帰に特化した設計とする．

4.2.1 学習環境と目的

安全方策の学習には，カリキュラム学習的なアプローチを用いる．エピソード開始時に，エージェントを意図的に軌道から大きく外れた位置や不安定な姿勢に配置し，そこから短時間で安全領域へ復帰することをタスクとする．また，最適方策と同様に DR を適用し，摩擦係数や車輪出力などの物理パラメータをランダムに変動させた環境で学習を行う．これにより，未知実証環境においても安定した復帰制御が可能な方策の獲得を目指す．

4.2.2 観測空間

安全方策の観測空間は，復帰制御に必要な情報に特化した 16 次元のベクトルと，過去 1 フレーム分の履歴で構成される．観測情報の詳細を表 1 に示す．最適方策と比較して，クロストラックエラーや復帰方向誤差など，現在の軌道からの逸脱状況を示す情報が含まれている．

強化学習モデルへ入力される情報としては，最適方策と同様に，現在の観測情報 o_t と 1 ステップ前の観測情報 o_{t-1}，および 1 ステップ前の行動情報 a_{t-1} をスタックした，計 36 次元のベクトルを用いる．

4.2.3 報酬設計

安全方策の目的は，速やかに軌道上かつ目標方向を向いた安全な状態へ復帰することである．そのため，復帰行動の進捗を評価する報酬設計を行なった．各ステップの報酬 r_t は，エージェントの状態に応じて以下のように計算される．

$$r_t = \begin{cases} w_{rec} \cdot \Delta d_{err} & \text{(復帰中)} \\ -w_{align} \cdot |\theta_{err}| & \text{(軌道付近)} \end{cases} + r_{common} \tag{3}$$

復帰中は，軌道への復帰方向を向いている場合に限り，クロストラックエラーの減少量 Δd_{err} に比例した正の報酬を与える．これにより最短経路での復帰を促す．軌道付近に到達した後は，方位角誤差 θ_{err} に対するペナルティを与え，軌道方向への整列を促す．また，共通項 r_{common} として，姿勢安定化や時間経過に対するペナルティ，および復帰成功時の報酬を含む．具体的なパラメータを表 2 に示す．

4.3 方策切り替えのための危険判別器

本節では，最適方策から安全方策への切り替え判断を担う危険判別器について述べる．未知実証環境における予期せぬ悪路や外乱に対して，効率性を重視する最適方策は行動が不安定化する可能性がある．そこで，観測情報から車体の状態を監視し，危険な状態に陥る予兆を検知するモデルを構築する．本研究では，タスク遂行における危険として軌道逸脱を対象とし，将来の逸脱発生度を予測する．

危険判別器の設計において決定すべきパラメータは，逸脱判定閾値 D_{th}，予測時間 T_{pred}，入力系列長 T_{ctx}，および分類閾値 θ_{cls} の 4 つである．これらのパラメータは，以下に述べる学習データセットを用いて決定した．

4.3.1 学習データセットの構築

判別器の学習には，最適方策を用いて多様な DR 環境を走行させることで収集した約 1,070 万ステップのログデータを使用した．学習の安定化のため，チャタリング除去を行なった．危険フラグは一度発生すると短期間に 0 と 1 を繰り返す傾向があるため，各エピソードにおいて最初にフラグが立った時点のみを正例とし，それ以降のデータは学習から除外した．データを学習用とテスト用に 8 : 2 で分割し，約 420 万ステップの学習データは正例と負例の割合をアンダーサンプリングで 1 : 1 に揃えた．約 105 万ステップのテストデータは実運用時の分布での性能評価を行うため，サンプリングを行わずに元の分布のまま使用した．テストデータにおける正例の割合は約 1.7% である．以下では，このデータセットを用いた各パラメータの導出方法について述べる．

4.3.2 逸脱判定閾値の決定

逸脱判定閾値 D_{th} は，最適方策が軌道からのずれを自力で修正できる限界の距離として定義する．この閾値を超えると，最適方策だけでは軌道に戻ることが難しくなり，安全方策への切り替えが必要となる．

閾値を決定するため，収集したログデータを用いて，軌道からの距離が時間とともにどのように変化するかを分析した．具体的には，ある時点で軌道から距離 D だけ離れている場合に，次の時点でその距離が縮まる傾向にあるか，それとも広がる傾向にあるかを調べた．

表 3　分類閾値と各評価指標

分類閾値 θ_{cls}	再現率	誤警報率	適合率	F1 値
0.5	0.937	0.115	0.122	0.215
0.6	0.917	0.096	0.141	0.244
0.7	0.883	0.078	0.162	0.273
0.8	0.818	0.056	0.199	0.320

分析の結果，軌道からの距離が 0.15m 以下の場合は，最適方策が誤差を減少させる方向に制御できていた．しかし，距離が 0.16m を超えると，誤差が増大する傾向に転じることが確認された．この結果から，最適方策が自力で回復できる限界は約 0.16m であると判断し，逸脱判定閾値を $D_{th} = 0.16$m と決定した．

4.3.3　予測時間と入力系列長の決定

逸脱判定閾値 $D_{th} = 0.16$m を固定した上で，危険予測に最適な予測時間 T_{pred} と入力系列長 T_{ctx} をグリッドサーチにより決定した．

危険判別器には，時系列データからの特徴抽出に優れた Long Short-Term Memory（LSTM）[23] を採用した．1 ステップ分の入力特徴量は，第 4.1.2 節で述べた最適方策の 17 次元の観測空間にクロストラックエラーと角度誤差，左右の車輪出力を加えた，21 次元のベクトルである．モデルは隠れ層サイズ 64 の LSTM 層と，それに続く全結合層によって構成され，シグモイド関数を通して危険度 $\hat{y} \in [0, 1]$ を出力する．

評価指標には，正例が少ない不均衡データにおいて正例の検出性能を重視する PR-AUC（Precision-Recall AUC）を採用した．入力系列長 T_{ctx} を 0.0 秒から 1.8 秒，予測時間 T_{pred} を-0.4 秒から 10.0 秒の範囲で変化させ，グリッドサーチを実施した結果を図 4 に示す．

結果より，$T_{pred} = 1.2$ 秒において PR-AUC $= 0.30$ を達成し，これはランダム分類器の約 0.017 と比較して約 18 倍の性能である．$T_{pred} > 2.0$ 秒では性能が急落し，長期予測の限界が確認された．入力系列長に関しては，増加させても PR-AUC の向上は見られなかった．以上の結果から，必要な情報量を最小限に抑えつつ，将来の危険を取りこぼしなく予測できる値として $T_{pred} = 1.2$ 秒，$T_{ctx} = 0.6$ 秒を採用した．

4.3.4　分類閾値の決定

モデル出力を「危険」と「安全」に二値化する分類閾値 θ_{cls} は，再現率で評価される安全性と，誤警報率の抑制で評価される効率性のバランスを考慮して決定した．閾値を変化させたときの各指標を表 3 に示す．

$\theta_{cls} = 0.7$ において，再現率 88.3%，誤警報率 7.8%を達成した．これは危険状況の約 88%を事前検知可能であり，かつ誤警報による不要な方策切り替えを約 8%に抑制できることを意味する．$\theta_{cls} > 0.7$ では再現率が低下し，閾値 0.8 では 82%となるため，安全性を優先する観点から $\theta_{cls} = 0.7$ を採用した．

4.3.5　方策切り替え方法

推論制御時には，学習済みの危険判別器を用いて各ステップごとに危険度を算出する．判別器が出力する軌道逸脱危険度

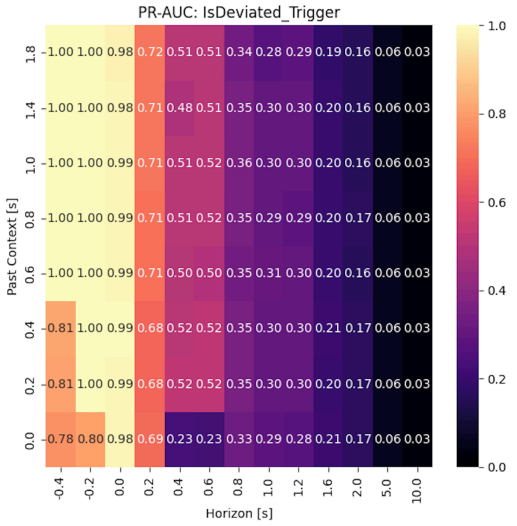

図 4　グリッドサーチによる PR-AUC ヒートマップ

P_{dev} が，事前に設定した閾値 $\theta_{danger} = 0.7$ を超えた場合，制御方策を最適方策から安全方策へと切り替える．安全方策によって車体が安定し，かつ軌道付近への復帰が完了して危険度が十分に低下した $\theta_{danger} < 0.7$ と判断された時点で，再び最適方策へと制御を戻す．これにより，危険度が低い状態では最適方策による高速な移動を行い，危険度が高い状態では，即座に安全方策による安全な状態への復帰を行う効率性と安全性を両立した制御を実現する．

5　実　　験

本節では，提案手法である安全方策と最適方策の動的切り替え制御の有効性を検証する評価実験について述べる．実験の目的は，未知の環境変動や外乱に対して，提案手法が既存の古典的制御手法や単一の強化学習方策と比較して，より高い安全性とタスク遂行能力を発揮できることを定量的に示すことである．

5.1　実 験 設 定

5.1.1　比 較 手 法

提案手法の性能を相対的に評価するため，以下の 3 つの手法を比較対象とした．

PID 制御　軌道追従のための直列 PID 制御器である．学習環境と同様の凹凸路面において，遺伝的アルゴリズム [24] を用いてゲインパラメータを事前にチューニングしたものを使用する．

最適方策　第 4.1 節で述べた DR 環境で学習された，タスクの効率達成に特化した強化学習方策である．

提案手法　本研究で提案する，将来の危険度に基づき最適方策と安全方策を動的に切り替える手法である．

5.1.2　評 価 環 境

各手法の汎化性能とロバスト性を評価するため，表 4 に示す複数の環境で実験を行なった．評価環境の基本地形には，実際の富士麓の洞窟のスキャンデータに基づいた凹凸地形を使用し，未知実証環境を再現した．Normal 環境は学習時と同様

表 4　評価環境の設定

環境名	設定内容
Normal	学習時と同様のパラメータ範囲
WheelFault	車輪出力係数 0.3 ～ 0.7
WheelFaultHard	左車輪出力係数 0.3，右車輪 1.0
LowFriction	地面摩擦係数 0.1 ～ 0.3
IMUNoise	加速度センサに ±0.1 ～ 0.3 m/s^2 の定常バイアス
MassLoss	車体の質量を 60% ～ 80% に低下

のパラメータ範囲を持つ基準環境である．WheelFault 環境と WheelFaultHard 環境は車輪の出力異常を模擬しており，学習時の出力係数範囲 0.7 ～ 1.0 より深刻な故障状態を再現している．特に WheelFaultHard 環境は左右の車輪で大きく異なる出力係数を設定し，重度の非対称故障を表現している．LowFriction 環境は，学習時の摩擦係数範囲 0.3 ～ 1.0 を下回る 0.1 ～ 0.3 を設定することで，学習時より滑りやすい路面を再現した．IMUNoise 環境と MassLoss 環境は，学習時には一切経験していない新規の外乱である．IMUNoise 環境では加速度センサに定常バイアスを付加することでセンサ異常を模擬し，MassLoss 環境では部品欠損等による質量低下を想定した設定とした．

各環境において，1 エピソードあたり最大 5,000 ステップを制限時間とし，500 エピソードの試行を行い，各手法の安全性と効率性を比較評価した．エピソードの終了条件は，タスク成功時，すなわち最終目標地点に到達し停止条件を満たした時，または制限時間到達時，すなわち 5,000 ステップ経過時のいずれかとした．

5.2　評 価 指 標

制御の安全性と効率性を多角的に評価するため，以下の 7 つの指標を用いる．

成功率　全エピソードのうち，制限時間内に最終目標地点に到達し，停止条件を満たした割合．

平均ステップ数　1 エピソードの終了までにかかった平均ステップ数．

成功時平均ステップ数　タスクに成功したエピソードのみを対象とした平均ステップ数．タスク遂行の効率性を評価する指標である．

平均総経路誤差　各エピソードにおける，理想経路からの逸脱距離の累積値の平均．この値が小さいほど，外乱に対してふらつくことなく，安全な軌道を維持できていることを示す．

成功時平均総経路誤差　タスクに成功したエピソードのみを対象とした累積経路誤差の平均．成功したエピソードにおける経路追従性を評価する．

平均最大経路誤差　各エピソードにおける，理想経路からの逸脱距離の最大値の平均．一時的に大きく逸脱したかどうかを評価する指標である．

成功時平均最大経路誤差　タスクに成功したエピソードのみを対象とした最大経路誤差の平均．成功したエピソードにおける一時的な逸脱の程度を評価する．

5.3　結果と考察

各環境における評価結果を表 5 および表 6 に示す．表 5 は全エピソードを対象とした評価結果であり，表 6 は成功エピソードのみを対象とした評価結果である．

成功率について，提案手法は 6 環境中 5 環境で最も高いスコアとなった．特に，左右の車輪出力が大きく異なる重度の非対称故障の状況を模した WheelFaultHard 環境において，提案手法の優位性が確認された．最適方策の成功率が 79.0%まで低下したのに対し，提案手法は 93.2%を達成し，14.2 ポイントの大幅な改善が確認された．これは，危険判別器が車輪故障から発生する軌道逸脱の予兆を検知し，安全方策へ切り替えることで実現されたと考えられる．

タスク達成の効率性について，成功時平均ステップ数では最適方策が多くの環境で最短を示した．一方，提案手法は WheelFaultHard 環境において成功時平均ステップ数が 2,050 と，PID の 1,559 と比較して約 31%増加した．これは安全方策への切り替えにより慎重な制御が行われるためであり，安全性と効率性の間にトレードオフが存在することを示していると考える．ただし，効率性を優先してタスク自体が失敗しては本末転倒であり，PID が 91.8%に対し提案手法は 93.2%と成功率の向上を実現している点で，この効率性の低下は許容できると考える．Normal 環境や LowFriction 環境など外乱が比較的軽微な環境では，提案手法の成功時平均ステップ数は最適方策と同等である．これは困難な環境においては安全性を考慮し，比較的容易な環境においては効率性を重視する提案手法の特性を表していると考えられる．

経路追従の安全性について，平均総経路誤差と平均最大経路誤差から考察する．WheelFault 環境において，PID 制御は平均総経路誤差 29.70 と最適方策の 6.39，提案手法の 5.475 と比較して著しく大きな値を示した．定性的な観察により，PID 制御は車輪故障という想定外の状況に適応できず，車体が転倒して大きな軌道逸脱を引き起こすケースが確認された．PID 制御では軌道追従と転倒抑制を同時に考慮した制御設計が困難であるのに対し，強化学習では報酬関数に転倒ペナルティを組み込むことで，これらを統合的に学習できる．提案手法は最適方策よりも低い経路誤差を達成しており，危険予測に基づく事前の方策切り替えが転倒を含む重大な軌道逸脱の抑制に効果的であることが確認された．平均最大経路誤差においても，WheelFaultHard 環境で提案手法が 0.116 と最適方策の 0.117，PID の 0.136 を下回り，一時的な大きな逸脱を抑制できていることが示された．成功エピソードのみの結果では，PID 制御が多くの環境で低い経路誤差を示しているが，これはタスクに失敗した困難なエピソードが除外されているためである．全エピソードを対象とした評価において，提案手法が PID 制御より優れた経路追従性を示したことは，困難な状況下でも安全な走行を維持できることを表していると考える．

表 5　各手法の評価結果（全エピソード）

環境	手法	成功率 [%] ↑	平均ステップ数↓	平均総経路誤差↓	平均最大経路誤差↓
Normal	最適方策	99.0	867.3	2.31	0.072
	PID	98.0	959.9	3.79	**0.064**
	提案手法	**99.6**	**867.0**	**2.19**	0.073
WheelFault	最適方策	96.4	**1557.3**	6.39	0.084
	PID	89.2	1836.6	29.70	0.145
	提案手法	**96.8**	1580.8	**5.47**	**0.076**
WheelFaultHard	最適方策	79.0	2645.8	**11.24**	0.117
	PID	91.8	**1841.4**	18.30	0.136
	提案手法	**93.2**	2250.5	11.58	**0.116**
LowFriction	最適方策	**99.0**	**873.6**	2.77	0.075
	PID	97.6	968.3	4.08	**0.065**
	提案手法	**99.0**	874.5	**2.37**	0.074
IMUNoise	最適方策	**99.6**	**843.1**	**2.07**	0.072
	PID	97.6	962.0	3.88	**0.068**
	提案手法	99.4	868.8	2.13	0.072
MassLoss	最適方策	98.4	901.0	2.59	0.075
	PID	96.8	1002.3	4.36	**0.056**
	提案手法	**99.2**	**888.2**	**2.26**	0.071

表 6　各手法の評価結果（成功エピソードのみ）

環境	手法	成功率 [%] ↑	成功時平均ステップ数↓	成功時平均総経路誤差↓	成功時平均最大経路誤差↓
Normal	最適方策	99.0	**825.6**	1.92	0.070
	PID	98.0	877.5	**1.88**	**0.053**
	提案手法	**99.6**	850.4	2.07	0.073
WheelFault	最適方策	96.4	**1428.7**	4.52	0.078
	PID	89.2	1453.6	**2.94**	**0.055**
	提案手法	**96.8**	1467.8	3.89	0.072
WheelFaultHard	最適方策	79.0	2019.9	8.84	0.114
	PID	91.8	**1559.2**	**3.52**	**0.081**
	提案手法	**93.2**	2049.9	9.30	0.110
LowFriction	最適方策	**99.0**	**832.0**	2.24	0.074
	PID	97.6	869.2	2.00	**0.055**
	提案手法	**99.0**	832.8	**1.96**	0.073
IMUNoise	最適方策	**99.6**	**826.4**	1.99	0.072
	PID	97.6	862.7	2.14	**0.057**
	提案手法	99.4	843.8	**1.93**	0.072
MassLoss	最適方策	98.4	**834.3**	2.03	0.070
	PID	96.8	870.1	2.43	**0.050**
	提案手法	**99.2**	855.1	**2.02**	0.071

6 ま と め

本研究では，学習環境とは異なる外乱や環境条件を持つ未知実証環境において，現地学習を必要とせずに差動二輪車のタスク遂行と安全性を両立させる安全推論制御手法を提案した．提案手法では，タスクの効率達成に特化した最適方策と，危険状態からの復帰に特化した安全方策という役割の異なる 2 つの方策を事前に学習し，LSTM を用いた危険判別器によって推定された危険度に基づき，方策を動的に切り替える．この切り替えにより，タスクの効率性と安全性を両立させることが可能になると考えた．

シミュレーション実験により，6 種類の未知実証環境において提案手法の有効性を検証した．車輪の深刻な非対称故障を模擬した WheelFaultHard 環境では，最適方策のみの成功率 79.0%に対し，提案手法は 93.2%を達成し，14.2 ポイントの改

善が確認された．この結果は，危険判別器が学習時に経験していない故障パターンに対しても，軌道逸脱の予兆を捉えて有効に機能することを示している．また，Normal，WheelFault，LowFriction，MassLoss 環境においても，提案手法は成功率と経路追従性の両面で優れた性能を示した．効率性の面では，困難な環境において成功時平均ステップ数が増加したが，これは安全方策による慎重な制御の結果であり，成功率の向上という形で補われていることが示された．

提案手法は状況に応じて効率性と安全性を適応的に切り替える制御を実現しており，未知実証環境における車体制御の安全性向上に寄与できると考える．

謝　　辞

本研究は，JSPS 科研費 JP24K03228 の助成，ならびに，JST【ムーンショット型研究開発事業】【JPMJMS2238】の支援を受けたものです．ここに記して謝意を表します．

文　　献

[1] Amr Eldemiry, Yajing Zou, Yaxin Li, Chih-Yung Wen, and Wu Chen. Autonomous exploration of unknown indoor environments for high-quality mapping using feature-based RGB-D SLAM. *Sensors*, Vol. 22, No. 14, 2022.

[2] Daniel Lawson and Ahmed Hussain Qureshi. Control transformer: Robot navigation in unknown environments through PRM-guided return-conditioned sequence modeling. In *Proceedings of the 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2023)*, pp. 9324–9331, 2023.

[3] Ravi Raj and Andrzej Kos. Intelligent mobile robot navigation in unknown and complex environment using reinforcement learning technique. *Scientific Reports*, Vol. 14, , 2024.

[4] Karol Arndt, Murtaza Hazara, Ali Ghadirzadeh, and Kyrre Glette. Meta reinforcement learning for sim-to-real domain adaptation. In *Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA 2020)*, pp. 2725–2731, 2020.

[5] Anusha Nagabandi, Ignasi Clavera, Simin Liu, Ronald S. Fearing, Pieter Abbeel, Sergey Levine, and Chelsea Finn. Learning to adapt in dynamic, real-world environments through meta-reinforcement learning. In *Proceedings of the 7th International Conference on Learning Representations (ICLR 2019)*, pp. 1–17, 2019.

[6] Joshua Achiam, David Held, Aviv Tamar, and Pieter Abbeel. Constrained policy optimization. In *Proceedings of the 34th International Conference on Machine Learning (ICML 2017)*, pp. 22–31, 2017.

[7] Shangding Gu, Long Yang, Yali Du, Guang Chen, Florian Walter, Jun Wang, Yaodong Yang, and Alois Knoll. A review of safe reinforcement learning: Methods, theory and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 46, No. 12, pp. 11216–11235, 2024.

[8] Brijen Thananjeyan, Ashwin Balakrishna, Suraj Nair, Michael Luo, Krishnan Srinivasan, Minho Hwang, Joseph E. Gonzalez, Julian Ibarz, Chelsea Finn, and Ken Goldberg. Recovery RL: Safe reinforcement learning with learned recovery zones. In *Proceedings of the 17th Robotics: Science and Systems (RSS 2021)*, pp. 1–10, 2021.

[9] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning (ICML 2017)*, pp. 1126–1135, 2017.

[10] Lin Wang, Yu Zhang, Daqi Zhu, and Sarah Coleman. Supervised meta-reinforcement learning with trajectory optimization for manipulation tasks. *IEEE Transactions on Cognitive and Developmental Systems*, Vol. 16, No. 2, pp. 681–691, 2024.

[11] Weiwei Li and Emanuel Todorov. Iterative linear quadratic regulator design for nonlinear biological movement systems. In *Proceedings of the 1st International Conference on Informatics in Control, Automation and Robotics (ICINCO 2004)*, pp. 222–229, 2004.

[12] Gregory Kahn, Adam Villaflor, Vitchyr Pong, Pieter Abbeel, and Sergey Levine. Uncertainty-aware reinforcement learning for collision avoidance. *arXiv preprint*, Vol. abs/1702.01182, , 2017.

[13] Bradley Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, Vol. 7, No. 1, pp. 1–26, 1979.

[14] Van Manh Tran and Gon-Woo Kim. Cooperative deep reinforcement learning policies for autonomous navigation in complex environments. *IEEE Access*, Vol. 12, pp. 101053–101065, 2024.

[15] Yassine Chemingui, Aryan Deshwal, Honghao Wei, Alan Fern, and Janardhan Rao Doppa. Constraint-adaptive policy switching for offline safe reinforcement learning. In *Proceedings of the 39th AAAI Conference on Artificial Intelligence (AAAI 2025)*, pp. 15722–15730, 2025.

[16] Mohammed Alshiekh, Roderick Bloem, Rüdiger Ehlers, Bettina Könighofer, Scott Niekum, and Ufuk Topcu. Safe reinforcement learning via shielding. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI 2018)*, pp. 2669–2678, 2018.

[17] Wen-Chi Yang, Giuseppe Marra, Gavin Rens, and Luc De Raedt. Safe reinforcement learning via probabilistic logic shields. In *Proceedings of the 32nd International Joint Conference on Artificial Intelligence (IJCAI 2023)*, pp. 5739–5749, 2023.

[18] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal Policy Optimization Algorithms. *arXiv preprint*, Vol. abs/1707.06347, , 2017.

[19] Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world. In *Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2017)*, pp. 23–30, 2017.

[20] Eugene Valassakis, Zihan Ding, and Edward Johns. Crossing the gap: A deep dive into zero-shot sim-to-real transfer for dynamics. In *Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2020)*, pp. 5372–5379, 2020.

[21] Yevgen Chebotar, Ankur Handa, Viktor Makoviychuk, Miles Macklin, Jan Issac, Nathan Ratliff, and Dieter Fox. Closing the sim-to-real loop: Adapting simulation randomization with real world experience. In *Proceedings of the 2019 IEEE International Conference on Robotics and Automation (ICRA 2019)*, pp. 8973–8979, 2019.

[22] 門垣幸樹, 大島裕明. 強化学習による差動二輪車制御における未知実証環境への適応. 第 17 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2025), 2025.

[23] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, Vol. 9, No. 8, pp. 1735–1780, 1997.

[24] John H. Holland. Genetic algorithms. *Scientific American*, Vol. 267, No. 1, pp. 66–73, 1992.