小口研究室　研究紹介　(2025年度)
(お茶の水女子大学理学部情報科学科)

サイバー犯罪グループ判定モデル構築におけるLLMによる自動ラベリングの検証
（研究担当:伊藤　純菜）

背景・目的

背景

- クレジットカード不正利用被害は増大傾向，2024年の番号盗用被害は513.5億円
- 窃取されたカード情報はTelegramのグループ機能を通じてSNS上で売買されている
- 先行研究によりモニタリングの有効性は確認されたが，効果最大化には監視対象グループの継続的な拡大が必要
- グループ探索手法の判定部分では機械学習モデルの学習に専門家によるラベル付きデータが必要

人手によるラベル付けの課題

①膨大な時間とコストがかかる
②専門知識が必要で人材が限られる
③犯罪種ごとに同様の作業を繰り返す必要がある

→ **LLMによる自動ラベリング**で課題を解決し探索手法の汎用性向上を目指す

ラベリング手法

入力プロンプト構成

使用モデル：GPT-4o（OpenAI）

タスク1　関連性判定：カード売買・フィッシングに関連する単語・意味が含まれるか（緩い基準・文脈理解重要）

タスク2　証拠検出：カード番号やその一部が実際に含まれるか（厳しい基準・パターン検出重要）

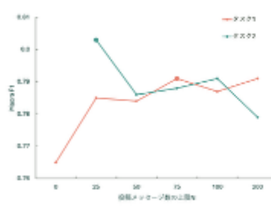
使用データ

1169	タスク1		タスク2	
総グループ数	285 関連グループ	884 非関連グループ	86 関連グループ	1083 非関連グループ

グループ名，説明文，投稿メッセージ，画像（最大10枚）多数等

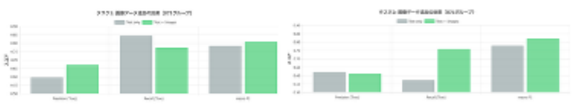
実験

結果①ー投稿メッセージ数の最適化



- タスク1：N=75が最適（macroF1 = 0.791）
- タスク2：N=25が最適（macroF1 = 0.803）
- N=75で51.9%が上限到達
- コスト約14ドル
- 誤分類改善122件（False Positive 74件減少）
- 投稿日時による期間制限の併用が有効

結果②ー画像データ追加の効果

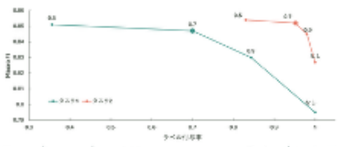


- タスク2のRecallが0.573→0.760に18.7pt改善
- OCRによる画像内カード番号検出が極めて有効
- 画像追加コスト約16ドルで，タスク1：+2.3pt，タスク2：+4.2pt

費用対効果の観点から画像情報追加は有効

※GPT-4oへの直接画像入力は安全機能により拒否されたため Tesseract OCRによるテキスト化を採用

結果③ー不確実性フィルタリング



7種類のプロンプトバリエーションで揺らぎを定量化
閾値0.7が最適バランス
（ラベル付与率70%，macroF1=0.847，5.2pt向上）
閾値0.7以降は精度向上に対するコストが著しく増大

実用運用：70%を自動ラベリング，残り30%を人手でアノテーション

生成 AI を用いたパーソナライズしたイベント推薦システムの構築
（研究担当:大本　詩織）

研究背景

- 国際観光の回復に伴う異なる背景や嗜好を持つユーザに適応したイベント推薦システムの必要性の高まり
- LLMの急速な発展
  ・様々なソースからユーザの情報を推定可能
  ・複数条件の柔軟な統合が可能

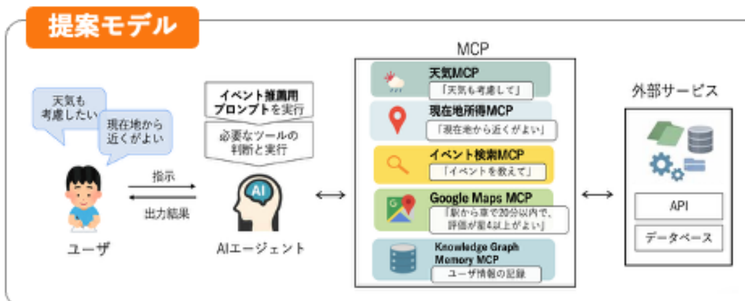
LLMを用いたユーザ適応型推薦システムの構築
ユーザのパーソナリティを用いた推薦精度の向上

課題

- システム構築に多くの労力・学習コストを必要とし、様々な要因を柔軟に組み合わせたイベント推薦が困難
  → 一連の流れをLLMによって行うことで、拡張性の高い柔軟なイベント推薦システムの構築を目指す
- 推薦要素が固定的であり、ユーザの状況や要求に応じた変更が困難
  → MCPを用いることで、ユーザ入力に応じ推薦要素を動的に変更する柔軟なイベント推薦システムを構築

MCP (Model Context Protocol)
- LLMが応答生成時に必要な外部情報を取得するためのクライアント・サーバ型プロトコル
- LLMと外部システム/アプリケーション間の連携方式を標準化

提案モデル



AI エージェントが、**マルチMCPサーバ**を介して必要な情報を動的に取得・統合

LLMとAIエージェントの連携により、**ユーザの状況や要求に応じた外部ツールをLLMが自動判断・実行**

実験

GPT-5を用いてツール選択挙動および推薦生成の有効性を検証

1 指示内容に応じたツール選択挙動

- **ユーザの指示に応じ必要なツールが自動選択されることを確認**
  例：「天気も考慮してイベントを推薦してほしい」
  → 天気情報取得ツールが実行され、雨天時には屋内イベントが推薦されるなど、天候・気温を考慮した推薦生成を確認
- **複数ツールが依存関係を考慮して連続実行されることを確認**
  例：「今週末開催されているイベントを推薦してほしい」
  → ① 現在日時を取得するMCP
  ② 相対的な時間表現を具体的な日付に変換するMCP
  ③ イベント検索MCP
  が順次実行され、開催期間を考慮したイベント推薦が行われることを確認

MCP×AIエージェントを用いることで、**ユーザの指示に応じて外部情報を自動的に取得し、推薦要素を動的に組み替えることが可能**

ただし
LLMはツール名や説明文などの表層情報に基づいてツール選択を行うため、誤選択が生じ得ることを確認

セキュリティ上の懸念

2 性能改善に関する検証

外部ツール連携は柔軟性を高める反面、ツール呼び出しや推論過程の増加により実行時間が増大し得る

1. Memory の利用により実行時間が短縮できるか検証
Knowledge Graph Memoryを導入し保存情報の粒度による実行時間を比較

Knowledge Graph Memory MCPサーバ
ユーザについての情報をローカルの**ナレッジグラフ**として永続的に保存可能なMCPサーバ

- **全情報を保存した時のみ**実行時間の短縮を確認
  ツール呼び出しのオーバーヘッドが生じ、**保存情報の粒度によっては実行時間が増加**するため、適切な設計が必要

Memory 利用条件	平均実行時間 [秒]
Memoryなし	157.2
性格特性のみ保存	187.4
性格特性および趣味趣向を保存	182.2
全情報保存	119.2

2. 利用するLLMの違いが実行時間・推薦品質に与える影響を検証

使用モデル：
- GPT-5：高い推論能力を持つ最新モデル
- GPT-5-mini：GPT-5をベースに高速性/コスト効率を重視して設計された軽量版モデル
- GPT-4o：従来世代のモデル

指示：「品川駅から車で20分以内で行けるイベントを検索して教えて」

- **軽量モデル**：高速である反面、制約条件が満たされない可能性
- **高性能モデル**：要件を満たしやすい反面、外部ツール呼び出しの回数増加や推論過程の増大による実行時間の長期化

モデル	平均実行時間[秒]	備考
GPT-5	318.2	必要なツールは安定して実行されたが，過剰呼び出しが発生
GPT-5 mini	126.5	ツール省略が生じた
GPT-4o	15.7	精度に問題あり（比較対象外）

実行信頼性と処理時間のトレードオフを考慮し、用途に応じたLLM選択が必要

今後の課題

- 悪意のある MCP サーバ混入を想定した異常検知・冗長化を含むセキュリティ対策の検討
- MCPサーバ数増加時におけるツール選択挙動およびシステム安定性の評価

ブロックチェーンプラットフォームHyperledger Irohaの性能評価と性能向上手法の研究
（研究担当:坂本　明穂）

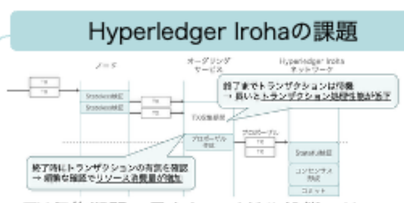
研究背景

【背景】企業間データ連携によるイノベーションの創出が期待
【課題】利用範囲が広く，データの信頼性担保が重要
→**ブロックチェーン**の利用：耐改ざん性とデータ検証

【データ共有基盤に求められる条件】
- 即時的な情報交換が不可欠なシナリオにおける，高性能トランザクション処理
- 継続的に運用を続けるため，低い運用負荷

両条件を同時に満たすブロックチェーンプラットフォームを構築

Hyperledger Irohaの課題

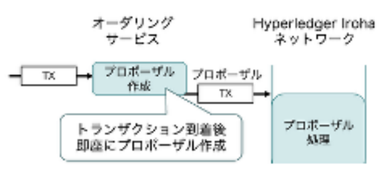


TX収集期間の長さとアイドル状態のリソース消費量はトレードオフ
- タイムアウト長→性能低，CPU使用率低
- タイムアウト短→性能高，CPU使用率高

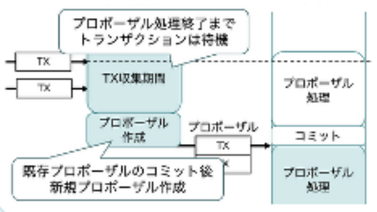
【目的】高い処理性能と低リソース消費の両立

研究背景

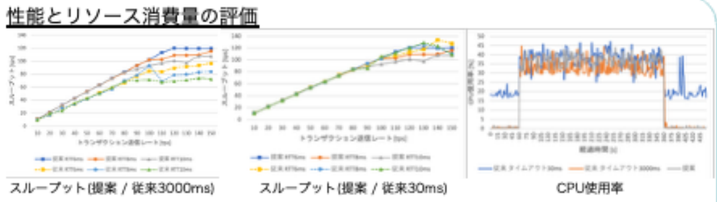
提案手法
- アイドル時



- プロポーザル処理時



性能とリソース消費量の評価



スループット(提案 / 従来3000ms)　スループット(提案 / 従来30ms)　CPU使用率

トランザクション処理性能とリソース消費の評価

	提案	従来3000ms	従来30ms
性能限界以前の処理性能	◎	△	◎
性能限界	○	△	◎
処理性能の安定性	◎	◎	△
リソース消費量(アイドル時)	◎	◎	×
リソース消費の安定性	○	◎	×

→ 提案手法は，処理性能とアイドル時のリソース消費を両立し，安定した状態で継続的に処理可能

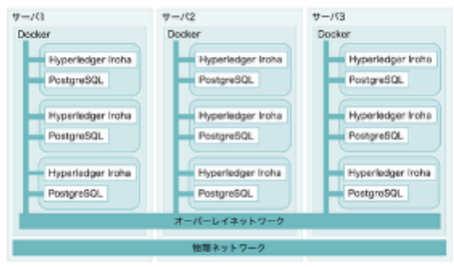
投票遅延の最適化

投票遅延
コンセンサス形成時に次の代表ノードに投票を送信する間隔
最適値はネットワーク内のノード数とネットワーク遅延で変化[1]
- ノード数による最適値の変化は調査されている[2]
- ネットワーク遅延による最適値の調査は不十分
→ ネットワーク遅延値に応じた適切な投票遅延を調査

[1] https://iroha.readthedocs.io/en/main/configure/index.html
[2] YAC: BFT Consensus Algorithm for Blockchain (Muratov et al. 2018)

【目的】
提案手法と投票遅延の最適化を組み合わせ，ネットワーク遅延環境下での処理性能の向上

実験環境



スループットの平均改善率 (従来 投票遅延5000msを基準)

	従来+最適値	提案(5000ms)	提案+最適値
RTT 0ms	1.76%	20.02%	22.00%
RTT 10ms	2.15%	41.11%	43.87%
RTT 20ms	21.21%	88.14%	93.22%

レイテンシの平均改善率 (従来 投票遅延5000msを基準)

	従来+最適値	提案(5000ms)	提案+最適値
RTT 0ms	2.96%	52.44%	59.71%
RTT 10ms	2.46%	67.90%	69.91%
RTT 20ms	29.81%	75.94%	77.45%

→ 提案手法+投票遅延最適化は特に高遅延環境で有効に機能

まとめ

まとめ
- 性能とリソース利用効率の両立に向けた手法を提案
- 提案手法と投票遅延の最適化を組み合わせることでネットワーク遅延環境下でのHyperledger Irohaのトランザクション処理性能が最大93%性能が改善

今後の課題
- 利用状況に合わせて動的に環境パラメータを変更する仕組みの導入

https://www.is.ocha.ac.jp/~oguchi_lab/　　oguchi@is.ocha.ac.jp