

研究テーマ

議論支援システムの研究

福富隆大

2025年1月7日

研究の動機

背景と目的

議論をもっと円滑に進めたい

よくある課題

話が噛み合わない → 議論が停滞 → なかなか決まらない

議論支援システムで改善

支援システム → 建設的な対話 → スムーズな議論 → 納得できる合意

研究の動機（続き）

目指すこと

- 非建設的な発言を減らす
- 生産的な対話を促す

関連するキーワード

- 対話調整、合意形成プロセス
- バイアス緩和、感情理解
- 談話理解、論理構造化、皮肉等についての言語学

関連研究①

Large Language Models in Argument Mining: A Survey

Hao Li, Viktor Schlegel, et al., 2025, arXiv

どんなもの？

LLM時代の議論マイニング技術をまとめたサーベイ。250本の論文と40のデータセットを分析し、主張の抽出や品質評価がどう変わったかを整理。

先行研究と比べてどこがすごい？

従来のサーベイと違い、プロンプトやRAG、few-shot学習など2021年以降の新しいやり方を体系化。教育・法律・医療への応用例も紹介。

技術や手法のキモはどこ？

Chain-of-Thought、RAG、LoRAなどの効率的な学習、合成データ生成、統合的データセット(IAM、ARIES)の活用。

関連研究①（続き）

どうやって有効だと検証した？

既存研究を分析して、GPT-4のfew-shot学習が従来のfine-tunedモデルと同等以上の性能を出すことを確認。議論品質評価では人間評価者と高い相関($\rho=.46-.93$)。

議論はある？

LLMがブラックボックスなこと、バイアス、計算コストが課題。多くの研究が特定分野だけの評価で、異なる分野や多言語での性能はまだ不十分。

次に読むべき論文は？

Lawrence & Reed (2019)（基礎）、Chen et al. (2024)（LLM応用実験）、Ivanova et al. (2024)（品質評価）

関連研究②

Towards Detecting and Mitigating Cognitive Bias in Spoken Conversational Search

Kaixin Ji, et al., 2024, arXiv

どんなもの？

音声での会話検索における認知バイアス（確証バイアスなど）の検出と軽減を目指した研究。脳波や皮膚反応などの生理データと行動データを組み合わせて使う。

先行研究と比べてどこがすごい？

テキストではなく音声に特化。脳波、皮膚電位などを統合。話速調整や質問促しなど実際に使える軽減策を提案。倫理的な問題（認知の自由の侵害）にも言及。

技術や手法のキモはどこ？

EEG（前頭葉の活動）、EDA（皮膚の電気反応）、音声の文字起こしや記憶テストで検出。話す速さの調整や質問で偏りを軽減。

関連研究②（続き）

どうやって有効だと検証した？

7人の予備実験で、簡単なトピックと難しいトピックでの脳波・皮膚反応の違いを観察。難しいトピックで前頭葉の活動が増え、ストレスレベルが上昇。明示的な質問が最も効果的。

議論はある？

生理信号は個人差があり解釈が難しい。リアルタイムで内容进行操作することの倫理的問題。短時間の実験ではセンサーの信頼性が低い。EEG分析には専門知識が必要。

次に読むべき論文は？

Args.meコーパス（議論データ）、認知負荷理論の基礎、音声でのナッジに関するHCI研究

関連研究③

MindScope: Exploring cognitive biases in LLMs through Multi-Agent Systems

謝振濤、趙加宝 et al., 2024, arXiv

どんなもの？

LLMの認知バイアスを検出するマルチエージェントシステム。72種類のバイアスを調べる質問データと対話シナリオでGPT-4など12モデルを評価。

先行研究と比べてどこがすごい？

従来研究より多い72種類を体系的に調査。複数回の対話で静的なテストでは見つからないバイアスを発見。検出精度がGPT-4単体より35%向上。

技術や手法のキモはどこ？

RuleGenフレームワーク：記憶・計画・反省・行動の4つのモジュール。調査対象、協力者、司会者の役割分担。自動シナリオ構築と二重チェック。

関連研究③（続き）

どうやって有効だと検証した？

12モデルの評価で、モデルが大きくなるとバイアスが減ることを確認（Llama2-7b～70b）。対話ではサンクコスト誤謬や計画誤謬が静的テストより目立つ。GPT-4と人間評価者で高い一致度（Kappa=0.72、精度88%）。

議論はある？

IKEA効果は全モデルで低性能。ファインチューニング（Vicunaシリーズ）でバイアスが増える可能性。多言語対応や長時間の対話での安定性が課題。

次に読むべき論文は？

認知バイアスの心理学の基礎文献、マルチエージェントの対話生成研究、LLMの安全性・公平性の研究

まとめ&研究テーマ案

関連研究からわかったこと

議論マイニング: LLMで議論の構造を自動で抽出・評価できるように

バイアス検出: 対話中の偏った考えを見つけて改善できる

評価手法: マルチエージェントで議論をシミュレーションして評価できる

研究テーマ

LLMで議論の流れを見える化して、評価・改善につなげるシステム

議論の中で「誰が何を主張しているか」を自動で抽出して、論理の流れをリアルタイムで見える化。話が脱線しそうになったら気づけるようにして、議論が終わった後に品質を評価してフィードバック。

どうやって実現するか

主な機能

1. 議論の流れを見える化

- LLMで主張と根拠を自動抽出
- 論理のつながり（主張同士の関係、論点の展開）をリアルタイム表示

2. 議論の評価とフィードバック

- 議論の質を測る（論理的か、いろんな視点があるか、偏りはないか等）
- どこを改善すればいいかを見つけてフィードバック
- マルチエージェントシステムで客観的に評価

期待される効果

議論の進め方が改善される

- 何を話しているか明確になって、焦点がブレない
- 話が逸れたらすぐ気づいて軌道修正できる

議論の質が上がる

- 論理的な議論の組み立てをサポート
- 評価とフィードバックで次回以降も改善

みんなが納得できる結論に

- 議論全体が見えるので建設的な対話になる
- 非建設的な発言が減って生産的な議論になる

参考文献

1. Hao Li, Viktor Schlegel, et al. "Large Language Models in Argument Mining: A Survey." arXiv:2506.16383, 2025.
2. Kaixin Ji, et al. "Towards Detecting and Mitigating Cognitive Bias in Spoken Conversational Search." arXiv:2405.12480, 2024.
3. 謝振濤, 趙加宝, et al. "MindScope: Exploring cognitive biases in large language models through Multi-Agent Systems." arXiv:2410.04452, 2024.