

議論モデルに基づく知識推論と生成 AI の統合に基づく AI 議論パートナー

森田 武史* 福澤 一吉† 山口 高平‡
青山学院大学* 明治大学† 神奈川大学‡

1. はじめに

近年、大規模意見集約・合意形成や多様な価値観の把握・可視化など、議論を支援する研究がなされている[1, 2]. 既存の研究は、統計・機械学習に基づく大局的・表層的な議論支援または形式論理学に基づく演繹的論証を含む議論支援に留まっており、論拠（明示されない隠れた根拠、発話者と議論相手の共通理解となる暗黙知）の推測を必要とする帰納的論証を含む議論の内容に立脚した意味レベルの議論支援の研究はなされていない。

一方、最近、Transformer などの極めて大規模な深層学習モデルによって作られた基盤モデルが登場し、機械翻訳や自然言語の意味理解タスクなどのベンチマークにおいて最高スコアが更新されている[3]. 基盤モデルに基づく ChatGPT などの文章生成 AI が商用サービスとして提供され、一般の人々にも普及してきている。しかし、現状の文章生成 AI は一対一の対話を基本としており、多人数による議論に応用することは困難である。文章生成 AI を議論支援に応用するためには、議論参加者の発言履歴の管理や構造化知識の利用などが必要だと考えられる。

以上の背景より、本研究では人の発言の意味を考へて議論可能な AI 議論パートナーを提案する。

2. 議論スキルを実現する AI 技術の考察

提案システムは、人と人の議論に関する哲学的研究である、イギリスの分析哲学者ステイブ・トゥールミンによる議論モデル[4]に基づいている。トゥールミン議論モデルでは、議論とは論証（根拠により主張を裏付けるプロセス）を基礎単位として話し合うことと定義され、「主張：個人の意見」「根拠：主張を支える裏付け」「論拠：主張と根拠の関係を保証する暗黙的根拠」から構成される。ここで、主張と根拠は明示化され、結論と前提の関係にあり、信頼性に依存して、論証の信頼性が変化する。

AI Discussion Partner based on Integration of Knowledge Reasoning based on Discussion Model and Generative AI

* Takeshi Morita, Aoyama Gakuin University

† Kazuyoshi Fukuzawa, Meiji University

‡ Takahira Yamaguchi, Kanagawa University

表 1 議論スキル

① 発言者の主張が明示されているか
② 発言者は自分の主張の根拠を提示しているか
③ 隠れた根拠（論拠）について言及しているか
④ 根拠から必要の飛躍が大きすぎないか
⑤ 聞かれていることに答えているか
⑥ 質疑応答はその内容が継続しているか
⑦ 根拠の信頼性は担保されているか

福澤一吉氏は、この議論モデルを具体化し、教育現場や組織内での会議などにおける議論を実現するための議論スキル（表 1）を提唱している[5].

本研究ではこの議論スキルを実現するための AI 技術群として、知識推論（オントロジー推論、知識グラフベース推論）、オントロジー統合、オントロジー学習、基盤モデル、文章生成 AI、転移学習、分散意味表現等を考察し、人の発言の意味を考へて議論可能な AI 議論パートナーを設計する。

3. AI 議論パートナーを実現するための AI 技術の詳細設計

図 1 に AI 議論パートナーのシステム構成を示す。AI 議論パートナーの各モジュールの概要を下記に示す。

(1) **議論管理**：複数の議論参加者と AI 議論パートナーの議論履歴を管理する。

(2) **議論分析**：議論履歴を知識グラフの形式に変換し、議論スキル（表 1）と領域オントロジーおよび知識グラフに基づいて、意見、議論の流れと論点、文章間の関係性などの分析を行う。

(3) **知識推論**：オントロジー推論および知識グラフベース推論により暗黙的な関係を推論し、議論知識グラフを補完する。

(4) **論証図生成**：議論分析結果から知識グラフの形式で論証図を生成する。図 2 に論証を論証図として表現した例を示す。図 2 において、①②③は根拠、④は主張、w1（感染症にかかっている人と一緒にいると感染しやすい。）と w2（発熱と倦怠感病気を診断する場合の症状である。）は論拠を表す。

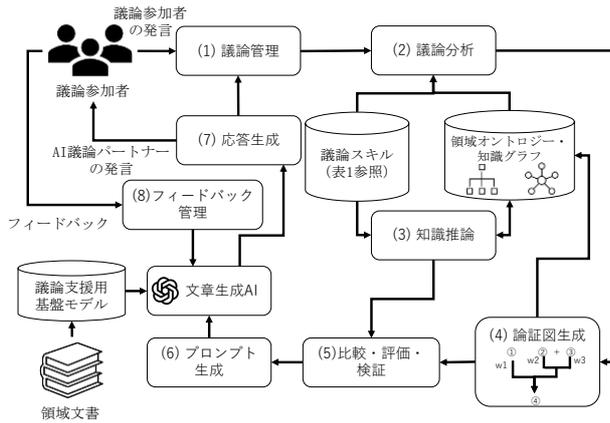


図1 AI 議論パートナーのシステム構成

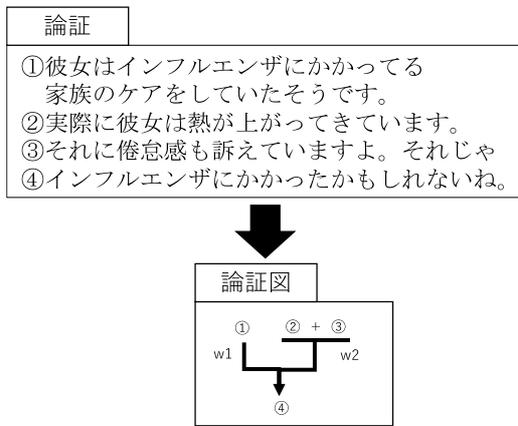


図2 論証を論証図として表現した例

(5) 比較・評価・検証：知識推論の結果と論証図から議論スキルを考慮しつつ、下記の四つの観点から議論を点検し洗練する。

(5-1) 議論モデル構成要素の検証 (表1 ①②③④ ⑦)：各参加者の発言に、a) 根拠・前提, b) 主張・結論, c) 論拠・仮定が含まれているかを確認する。どれか要素が欠けている場合にはそれを指摘して、それを補充するように指摘する。(表1 ①②③)

(5-1-1) 表1②のさらなる確認事項⑦：根拠・前提が経験的事実であるかどうかを確認する。根拠・前提の引用先などの裏が取れるかどうかの確認をする。裏が取れない根拠を使っている論証を無効にする。

(5-1-2) 表1③④及び、その評価：論拠は事実ではなく仮定になっていることを確認する。さらに、その論拠が根拠と結論を結合させるに十分な論拠かを評価する。論拠として不十分な場合は、それを指摘し、最終的な論証が成立していないことを発話者に伝える。

(5-2) 議論展開の検証 (表1 ⑤⑥)：議論が論理的に展開しているかを確認する。a) 各自の発言が、その直前または、以前に出された他者

の発言内容のどこに直接触れているのかに言及しているかを確認する。b) 自分の発言は他の者の発言に対して、どう反応するかを明示しているかを確認する。

(5-3) 論証検証：各自の論証が複雑になる場合は、論証間の関係が論理的に結合しているかをチェックする。複数の根拠が同時に使われ、それらが一つの結論を導くために使用される場合には、根拠間の関係をチェックする。性質の異なる根拠を複数使用する場合、その理由についても発話者に確認する。

(5-4) 議論対立軸明示：複数の発話者間が対立する場合、その対立軸を明確にして議論の参加者に示す。

(6) プロンプト生成：比較・評価・検証の結果から文章生成 AI が応答生成するために必要なプロンプトを生成する。議論の領域文書からファインチューニングにより構築した議論支援用基盤モデルを文章生成 AI は参照する。

(7) 応答生成：文章生成 AI の出力結果から AI 議論パートナーの発言を生成する。

(8) フィードバック管理：AI 議論パートナーの発言が不適切な場合には、フィードバック管理モジュールを介して、議論参加者は文章生成 AI にフィードバックすることにより、本システムを利用しながら性能を向上させる。

4. おわりに

本研究では人の発言の意味を考えて議論可能な AI 議論パートナーを提案した。本研究における知識推論は、オントロジーと知識グラフに基づいている。しかし、GPT-4 などの最近の生成 AI は知識推論の性能が向上し、マルチメディアの入力にも対応している。今後は、生成 AI に基づく知識推論とマルチメディアの入力を検討し、大学生を対象とした実証実験を通して、AI 議論パートナーの実現を目指す。

参考文献

- [1]. Ito, T., et al. (2020). D-Agree: Crowd Discussion Support System Based on Automated Facilitation Agent. AAI Conference, 34(09), 13614-13615.
- [2]. Slonim, N., et al. (2021). An autonomous debating system. Nature, 591(7850), 379-384.
- [3]. JST CRDS. (2023). 人工知能研究の新潮流 2 ～基盤モデル・生成 AI のインパクト～, CRDS-FY2023-RR-02.
- [4]. Toulmin, S. E. (2003). The Uses of Argument. Cambridge University Press.
- [5]. 福澤一吉. (2018). 新版 議論のレッスン. NHK 出版.