

マルチエージェントシステムにおける対話状態推定手法

小林優佳¹ 永江尚義¹ 白木佑弥¹

西田隼輔¹ 岩田憲治¹

株式会社東芝 総合研究所

{yuka.kobayashi.f34, hisayoshi.nagae.t34, yuya.shiraki.k51,
shunsuke.nishida.n18, kenji.iwata.h13}@mail.toshiba

概要

本研究は、複数の自律エージェントが協調・競合しながら議論を行うマルチエージェントシステムにおいて、対話履歴の長大化や複雑化による大規模言語モデルの性能低下を防ぐための対話状態推定手法を提案する。各エージェントはペルソナ情報や議題情報を保持し、発話ごとに個別および全体の対話状態を更新することで、履歴を簡潔化しつつ議論を進捗する。本手法を用いた実験では、ペルソナらしさや話者間の発話の違い、新規性の高いアイデアの創出において有効性が確認された。

1 はじめに

近年、生成 AI・大規模言語モデル (Large Language Model:LLM) の発展に伴い、複数の自律エージェントが協調または競合しながら意思決定を行うマルチエージェントシステムが注目を集めている。その中でも、Multi-Agent Debate(MAD) は、複数のエージェントが議論を通じてより高品質な結論や意思決定を導く手法として重要性を増している。従来の単一エージェントによる推論や生成モデルは、知識の偏りや論理的誤りを含む可能性がある。一方、複数エージェントによる討論は、より信頼性の高い出力を得ることができると期待されており、高度な推論が必要なタスク [1], 文章の評価 [2],[3], 社会シミュレーション [4],[5],[6] などに用いられている。

また、多様な視点の導入、反証のプロセスを通じて、より精練されたアイデアが提案できると期待されており、アイデア提案にも用いられている [7] [8] [9]。本研究では MAD によるアイデア提案を対象とする。各エージェントに異なる役割や視点を持たせることで、タスクに対する多角的なアプローチや集合知の活用が可能となり、単一エージェントでは得られなかった創造的な解決策の導出が期待される。

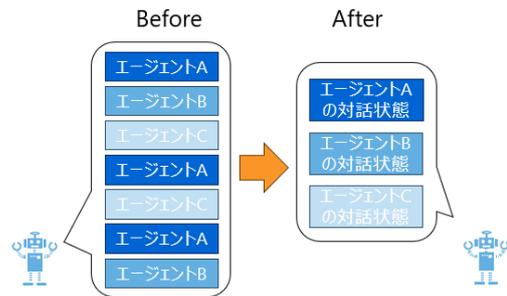


図1 エージェントの発話生成

本研究で用いる MAD システムでは、各エージェントに年齢、職業、価値観、性格などを表すペルソナ情報を与え、異なるアイデアや意見を出すようにする。

各エージェントは、全エージェントの発話を対話履歴として蓄積し、指示と対話履歴をもとに発話を生成する。図1のように、対話が続くと対話履歴が長くなり、いずれ LLM の系列長を超え、全対話履歴を扱うことができなくなる。また、他のエージェントの発話を用いるため、他のエージェントのペルソナが混同し、各エージェントの発話の差異が小さくなるという問題がある。その結果、各エージェントが提案するアイデアが類似したものになり、議論の効果が出しにくいという課題がある。

そこで、本研究ではエージェントの対話履歴から各エージェントの対話状態を推定する対話状態推定 (Dialogue State Tracking:DST) を実施し、対話履歴を簡潔な情報に集約しながら議論を進める。DST を実施することで、系列長の超過や、複雑な文章の扱いを回避することができる。本手法ではエージェントごとの対話状態と全体の対話状態を保持する。エージェントが発話する度に各エージェントの対話状態と全体の対話状態を更新し、次のエージェントの応答生成に使用する。本研究ではアイデア提案の MAD を実施し、DST の効果を確認した。

2 先行研究

従来の DST は、レストラン検索やフライト予約をするために 1 人のユーザと単一のエージェントが会話する際に、ユーザの要求や意図を追跡する技術である。近年は LLM を用いた手法が提案されている [10, 11].

近年はマルチユーザ対話も注目されており、ユーザ間の会話による意図の変化、発話者の識別など、より複雑な手法が求められる。Jo らは複数ユーザの会話を要約し、単一ユーザ発話に変換する手法を提案した [12]. マルチユーザ対話では、全ユーザの総意が取得できれば、個々のユーザの意見を考慮する必要はないため、対話履歴の要約が有効である。MAD では各エージェントの意見に対してコメントや質問を実施するため、全体を要約するのではなく、各エージェントの対話状態を推定する必要がある。

マルチエージェントでの DST を実施している研究としては Cao の手法 [13] が挙げられる。Cao の研究ではユーザの質問に回答するチャットボットシステムの中に複数のエージェントを配置し、DST を専門に実施するエージェントを配置している。しかし、推定しているのはユーザの要求で、エージェントの状態推定を実施しているわけではない。

3 提案手法

3.1 MAD システム

本システムでは AutoGen TM¹⁾ を使用する。最初に、システムに議題を含む全体への指示を与える。その後発話順に従ってエージェントが順に発話を行う。本システムでは、対話履歴をもとに LLM が発話順を決定する AutoGen TM の機能を用いる。あらかじめ発話回数を決めておき、発話回数に到達したところで議論を終了する。また、議論中に提案されるアイデアがより多様なものになるために、各エージェントにペルソナを与え、異なるアイデアや意見を出すようにする。各エージェントにはペルソナ情報をプロンプトに追加する。ペルソナ情報の作成方法は後述する。

最後に議論中の各エージェントのアイデアを要約する。議論中の同一エージェントの発話だけを抽出し、LLM で要約を実施する。

1) <https://microsoft.github.io/autogen/stable/index.html>

3.2 MAD 用 DST

各エージェントは内部に対話状態を保持する。対話状態とは、現在の対話の内容を要約・構造化した情報である。各エージェントは同じ会話に参加していても、個別に対話状態を保持・更新する。エージェント固有の情報を入れて更新することで、各エージェントは同じ会話でも異なる対話状態を保持する。対話状態には各エージェントの対話状態 $S_b^a[i]$ (エージェント b が保持する、ステップ i のエージェント a 対話状態)、全体の対話状態 $S_b[i]$ (エージェント b が保持する、ステップ i の全体の対話状態) がある。ステップとは全エージェントの発話数である。

対話状態の更新方法について説明する。エージェント a が新規発話 $u^a[i]$ をすると、各エージェントは対話状態を更新する。ここではエージェント b について説明する。まず、エージェント a の対話状態 $S_b^a[i]$ を更新する。

$$S_b^a[i] = \text{UpdateAgentState}(S_b^a[i-1], u^a[i]) \quad (1)$$

次に、全体の対話状態 $S_b[i]$ を更新する。

$$S_b[i] = \text{UpdateState}(S_b^a[i] \oplus S_b^b[i] \oplus S_b^c[i], u^a[i]) \quad (2)$$

更新は LLM で実施する。各更新時のプロンプトは付録 (表 1.2) に載せる。上記の処理によって各エージェントの対話状態を更新し、自分の発話順が来た際は対話状態を用いて次の発話を生成する。

$$u_b[i+1] = \text{Utterance}(S_b[i] \oplus S_b^a[i] \oplus S_b^b[i] \oplus S_b^c[i]) \quad (3)$$

上記の処理では、各エージェントが保持する対話状態は本質的には同じものになる。本システムでは、対話状態の更新時に各エージェントのペルソナ情報を利用する。更新時にペルソナ情報をプロンプトに追加することで、各エージェントが保持する対話状態はペルソナ情報が反映され、固有のものになる。

本システムでは最初に議題を全体に周知するが、対話状態は議題をもとに更新されるべきである。更新時に議題をプロンプトに追加することで、各エージェントが保持する対話状態は議題に基づいた内容となる。

4 実験

4.1 マルチエージェントシステム

本実験ではビッグファイブ理論 [14] の性格特性に基づいたペルソナを作成する。ビッグファイブ理論では開放性、誠実性、外向性、協調性、神経症傾向の5つの軸で性格を表現する。この5つの軸の値に対して「高い」「低い」のどちらかを指定し、5つの値に基づいた性格を LLM に生成させる。ビッグファイブ理論の各軸の説明、各軸の値をプロンプトに入れ、性格の説明文を生成させる。また、上記のプロンプトと性格説明文を用いて、このペルソナが言いそうな発話文を10個生成させる。説明文と発話文10個を合わせてペルソナ情報とする。32個のペルソナから5つの値の異なりが大きいペルソナを3つ選択して1つのグループとし、グループを3種類作成し、実験に使用する。

議題は下記のような商品企画のアイデア検討の議題を用意する。

- 高齢者向けスマートフォンアプリの新商品
- 自動車保険の新商品
- コンビニスイーツの新商品

議題を与えるプロンプトは付録(表3)に載せる。『topic』の部分に上記の議題が入る。できるだけ異なるアイデアを提案し、議論を経てアイデアを膨らませ、多様なアイデアを出すことを目的とした指示になっている。

4.2 実験条件

ペルソナ情報作成、対話状態推定、応答生成には GPT™-4.1 を用いる。temperature は 0.7 とする。発話回数は 30 回とする。実験パターンは下記とする。

- baseline: 各エージェントの発話をそのまま蓄積して応答生成に用いる。
- struct: DST で対話状態に変換されることを想定して、要約しやすい構造の発話を生成させる。応答生成時の指示に付録(表4)のプロンプトを追加する。
- dst: DST を実施
- persona: ペルソナ情報を利用して DST を実施
- topic: 議題を利用して DST を実施
- persona+topic: ペルソナ情報と議題を利用して DST を実施

DST 実施時の発話はすべて上記の構造化された発話を生成させる。

4.3 評価手法

本稿では各エージェントの発言の一貫性、エージェント間の発言の多様性、アイデアの新規性の3つの観点で評価する。これらの3つの観点を以下の指標によって算出した。

- ペルソナ正解率: 対話履歴とペルソナ情報をもとに、どのペルソナの発話かを LLM に当てさせ、正解率で評価。
- 話者間の発話の相違: 各発話を埋め込み表現に変換し、発話間のコサイン類似度を算出し、(同一話者の発話間の類似度の平均値)-(異なる話者間の発話間の類似度平均値)を算出する。埋め込み表現の算出には BAAI/bge-m3²⁾を用いる。
- アイデアの新規性: 最終提案の TFIDF で評価する。エージェント a の最終提案 I^a の TFIDF は下記で算出される。

$$TFIDF(I^a) = \frac{1}{N_{I^a}} \sum_{w \in I^a} N(w) \times IDF(w) \quad (4)$$

$$IDF(w) = \log \frac{N_D}{N(w \in d)} \quad (5)$$

(N_{I^a} : I^a の単語数, $N(w)$: 単語 w の頻度, N_D : 全文書数, $N(w \in d)$: 単語 w を含む文書数)

IDF() は Wikipedia のダンプ記事³⁾をもとに算出する。TFIDF は文書中に珍しい単語が入っているかどうかを算出する指標である。最終提案中の珍しい単語の分布に基づいてアイデアの新規性を評価する。

4.4 実験結果

評価手法に基づいて評価結果を算出する。ペルソナのグループが3種類、議題が3種類あるので、ひとつのパターンについて9回の試行がなされているので、9回の平均値、標準偏差を算出した。

まず、ペルソナ正解率について評価を実施した(図2)。ペルソナ情報を用いて DST を実施することで、0.71(baseline) から 0.84(persona) に向上したことを確認できた。他の手法では効果が確認できなかったため、ペルソナ情報を用いた DST が重要であることがわかる。

2) <https://huggingface.co/BAAI/bge-m3>

3) <https://dumps.wikimedia.org/jawiki/>

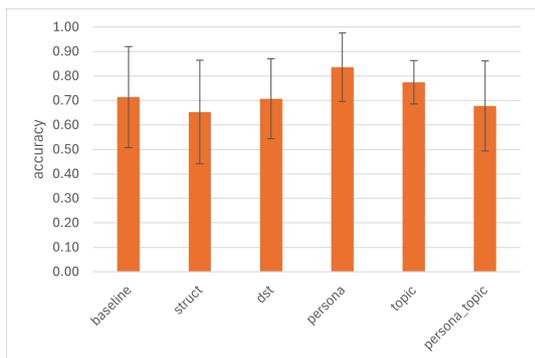


図2 パルソナ正解率

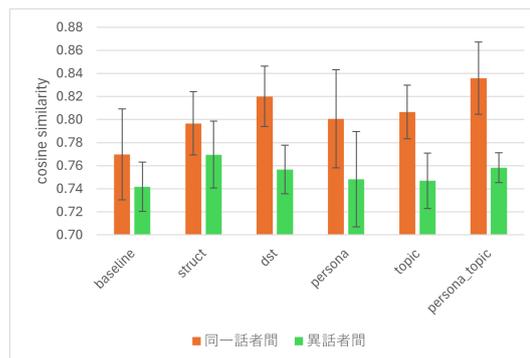


図4 同一話者内と異話者間の発話間類似度

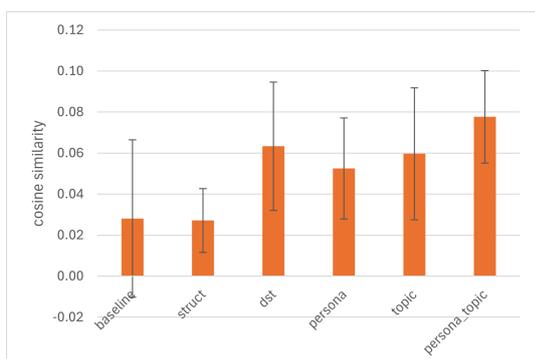


図3 話者間の発話の相違

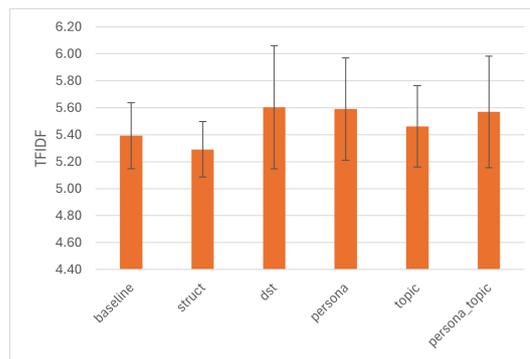


図5 アイデアのTFIDF

話者間の発話の相違については DST を実施することで、異なるエージェント間での差異が大きくなることが確認できた(図3)。図4は同一話者内の発話間類似度の平均値と異話者内の発話間類似度の平均値である。まず、発話を構造化することで、全発話同じ構造になり、同一話者内も異話者間も類似度が上昇する。DST を実施しない場合、各エージェントの発話がそのまま、次の発話に用いられるので、エージェントの話し方の癖などが他のエージェントに伝搬する可能性がある。DST を実施することで、伝搬を防ぐことができ、同一話者内の発話間類似度が向上し、異話者間の類似度が低下する。そのため、同一話者間・異話者間の類似度の差が向上していると考えられる。

最後に、各話者のアイデアの TFIDF 平均値を図5に載せる。DST を実施した実験結果はしないものよりも値が高く、DST を実施することでアイデアの新規性が向上することを確認できた。

上記3種類の評価により、DST を導入することで、各エージェントは自身のペルソナ情報に基づいた発話を生成し、新規性の高いアイデアを提案することが確認できた。従来の手法では、応答生成時に他のエージェントの発話が入力に含まれるため、ペルソナが混同し、違いが出にくいという問題が

あったが、DST によって防ぐことができた。また、応答生成時に全対話履歴を処理する必要があったが、DST を実施することで、各エージェントのアイデアの古い情報は新しい情報に上書きされ、応答生成時に古い情報を処理する必要がなくなり、議論が前に進むことを後押しできると考えられる。

5 まとめ

マルチエージェントシステムにおいて、各エージェントの発話を蓄積した文章を次のエージェントの応答生成に用いると、LLM の系列長を超える、長文を解釈する必要があり性能が落ちるなどの問題が発生する。それに対して、本手法では各エージェントの発話から対話状態を推定し、各エージェントの対話状態と全体の対話状態を用いて次のエージェントの応答生成を実施する。本手法を適用することで、各エージェントは自身のペルソナ情報に沿った発話ができるようになり、新規性の高いアイデアを提案することができるようになった。

本手法では各エージェントの発話を個別に用いて対話状態を更新するので、エージェント間の会話を表現しにくい。今後は上記のような課題に取り組み、より効率よく議論できるマルチエージェントシステムの開発を目指す。

参考文献

- [1] Yilun Du, Shuang Li, Antonio Torralba, Joshua B. Tenenbaum, and Igor Mordatch. Improving Factuality and Reasoning in Language Models through Multiagent Debate. In **the 41st ICML**, Vol. 235, pp. 11733–11763, 21–27 Jul 2024.
- [2] Chi-Min Chan, Weize Chen, Yusheng Su, Jianxuan Yu, Wei Xue, Shan Zhang, Jie Fu, and Zhiyuan Liu. ChatEval: Towards Better LLM-based Evaluators through Multi-Agent Debate. <https://arxiv.org/abs/2308.07201>, 2023.
- [3] Zhaopeng Feng, Jiayuan Su, Jiamei Zheng, Jiahua Ren, Yan Zhang, Jian Wu, Hongwei Wang, and Zuozhu Liu. M-MAD: Multidimensional Multi-Agent Debate for Advanced Machine Translation Evaluation. In **ACL (Volume 1: Long Papers)**, pp. 7084–7107, Vienna, Austria, July 2025.
- [4] Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative Agents: Interactive Simulacra of Human Behavior. <https://arxiv.org/abs/2304.03442>, 2023.
- [5] Chen Gao, Xiaochong Lan, Zhihong Lu, Jinzhu Mao, Jinghua Piao, Huandong Wang, Depeng Jin, and Yong Li. S³: Social-network Simulation System with Large Language Model-Empowered Agents. <https://arxiv.org/abs/2307.14984>, 2025.
- [6] Xiaoli Hu, yang shen, and Keke You. Multi-Agent Social Simulation: An Experimental Framework for Language-Native Social Experiments. In **Open Conference of AI Agents for Science 2025**, 2025.
- [7] Tian Liang, Zhiwei He, Wenxiang Jiao, Xing Wang, Yan Wang, Rui Wang, Yujia Yang, Shuming Shi, and Zhaopeng Tu. Encouraging Divergent Thinking in Large Language Models through Multi-Agent Debate. In **EMNLP**, pp. 17889–17904, Miami, Florida, USA, November 2024.
- [8] Nate Straub, Saara Khan, Katharina Jay, Brian Cabral, and Oskar Linde. Persona-based Multi-Agent Collaboration for Brainstorming. <https://arxiv.org/abs/2512.04488>, 2025.
- [9] Mizuki Hoshino, Shun Shramatsu, and Fuminori Nagasawa. A Business Idea Generation Framework Based on Creative Multi-Agent Discussions. In **Proceedings of the 2nd Workshop on Agent AI for Scenario Planning**, pp. 82–85, Montreal, Canada, 16 August 2025.
- [10] Yujie Feng, Zexin Lu, Bo Liu, Liming Zhan, and Xiaoming Wu. Towards LLM-driven Dialogue State Tracking. In **EMNLP**, 2023.
- [11] Zekun Li, Zhiyu Zoey Chen, Mike Ross, Patrick Huber, Seungwhan Moon, Zhaojiang Lin, Xin Luna Dong, Adithya Sagar, Xifeng Yan, and Paul A. Crook. Large Language Models as Zero-shot Dialogue State Tracker through Function Calling. In **ACL**, 2024.
- [12] Yohan Jo, Xinyan Zhao, Arijit Biswas, Nikoletta Basiou, Vincent Auvray, Nikolaos Malandrakis, Angeliki Metallinou, and Alexandros Potamianos. Multi-User Multi-WOZ: Task-Oriented Dialogues among Multiple Users. In **EMNLP-Findings**, 2023.
- [13] Lang Cao. DiagGPT: An LLM-based and Multi-agent Dialogue System with Automatic Topic Management for Flexible Task-Oriented Dialogue. <https://arxiv.org/abs/2308.08043>, 2024.
- [14] D. P. McAdams. The five-factor model in personality: A critical appraisal. **Journal of personality**, Vol. 60, No. 2, pp. 329–361, 1992.

商標について：

"AutoGen"および"GPT"はそれぞれ Microsoft Corporation, OpenAI 社の登録商標です。また、その他本稿に掲載の商品、機能等の名称は、それぞれ各社が商標として使用している場合があります。

A 参考情報

表1 各エージェントの対話状態の更新時のプロンプト

下記は agent A の発話内容に基づいて agent A の意見をまとめたものです。

agent A の意見：

{前回の agent A の対話状態}

agent A の新規発話：{agent A の新規発話}

新規発話をもとに agent A の意見を更新してください。

agent A の意見：

他のメンバーからの質問・コメントに対する回答

他のメンバーへの質問・コメント

提案アイデア

表2 全体の対話状態の更新時のプロンプト

agent A の意見

{agent A の対話状態}

agent B の意見

{agent B の対話状態}

agent C の意見

{agent C の対話状態}

agent A の新規発話：{agent A の新規発話}

新規発話をもとに全体意見を更新してください。

全体意見：

表3 議題を与えるプロンプト

『topic』についての企画をしてください。

自分のアイデアを出してください。必ず、他の人と異なるアイデアを提案してください。他の人のアイデアに対して、質問やコメントをしてください。自分宛の質問やコメントに対して回答してください、またそれらを踏まえてアイデアを改良してください。

提案アイデア

【商品名】

【商品内容】

【価格】

【顧客ターゲット】

【宣伝方法】

表4 発話構造化のプロンプト

下記のフォーマットで発話してください。

他のメンバーから{persona_name}への質問・コメントに対する回答 (質問・コメントがない場合は空欄)

他のメンバーへの質問・コメント

提案アイデア