

バグモデル RAG を用いた Teachable Agent の提案

Proposal for Teachable Agent with Bug Model RAG

中谷 圭吾^{*1}, Jean Paul Pierret Robles^{*2}, 東本 崇仁^{*3}
Keigo NAKATANI^{*1}, Jean Paul PIERRET ROBLES^{*2}, Takahito TOMOTO^{*3}

^{*1} 千葉工業大学情報科学部

^{*1} Faculty of Information and Computer Science, Chiba Institute of Technology

^{*2} 千葉工業大学大学院情報科学研究科

^{*2} Graduate School of Information and Computer Science, Chiba Institute of Technology

^{*3} 千葉工業大学情報変革科学部

^{*3} Faculty of Innovative Information Science, Chiba Institute of Technology

Email: s2232108UH@s.chibakoudai.jp

あらまし：昨今 LLM が普及しており様々な分野で用いられており、教育分野でも TA として用いられた例も存在する。一定の成果をあげる一方、LLM の発言の一貫性の有無や高精度すぎる返答文による学習機会の損失など学習効果を下げる要因も散見された。本研究では LLM の改善手法の一つである RAG の技術と知識構造を構築できるオントロジーを融合させ、学習者に対し「自然言語文の対話形式による知識構造の修正活動」を提供できる TA を提案する。

キーワード：Teachable Agent, オントロジー, RAG, Learning-by-teaching, 対話, 適応的支援

1. はじめに

昨今大規模言語モデル(LLM)が普及しており、教育分野でも用いられている。筆者らは LLM を教育分野に活用した例の中でも ChatGPT ベースのシステムを Teachable Agent(TA) として活用した先行研究⁽¹⁾に着目した。TA とは学習者がエージェントの誤答を訂正をする活動を通して学習を行うシステムである。先行研究では学習者への有用性を示した一方、学習者の教授活動に見合わない高精度な返答を行い、教授活動の機会を減らす点が懸念されていた⁽¹⁾。また、先行研究では生成 AI のハルシネーションに対する対策が行われていなかった。

先行研究より、①学習者からの教授活動に見合った返答を行うための知識制限、②一貫性を有しつつ、誤りを含む知識体系の構築、が必要と考えた。筆者らは知識体系の構築手法としてオントロジーに着目した。また、LLM の解答生成の際にオントロジーを利用するため、外部情報を参照して解答生成する技術として Retrieval Augmented Generation (RAG)に着目した。本研究では、知識体系の構築をオントロジーで行い、RAG のフレームワークの外部情報リソースとして用いることで限られた範囲の知識のみを有する誤りを含む文を出力し、学習者の教授活動に見合う応答を行う TA の実現を試みる。

2. 関連研究

Angxuan らは ChatGPT をプロンプト制御し TA として振る舞わせ、学習効果を測定した⁽¹⁾。実験結果より LLM を TA として振る舞わせることの有用性を示した。一方 TA が提示する教授活動に見合わない高精度な返答文により学習者が誤りを訂正する機会が減少している可能性を指摘した。筆者らはこのよ

うな生成 AI は誤りの一貫性の有無の点で問題が考えられる。もし誤りに一貫性がない場合、会話が進行するほど誤りの内容が変化する恐れがあるためである。先行研究より、筆者らは TA に用いる LLM に対し、(1)知識を制限すること、(2)解答に含まれる誤りに一貫性を持たせること、の必要性を認識した。本研究では「一貫した誤り」を、誤った知識によって発生する一連の誤った解釈・解答、と定義する。

LLM の解答生成の改善手法としてユーザからの質問と、それを元にした外部情報への検索結果を LLM に渡すことで、専門分野や企業の内部情報などの限定的な範囲での解答の精度向上を図る RAG が注目されている。Sharma らはこの外部情報としてオントロジーを用いることでさらなる正確性、応答速度の向上を図っている⁽²⁾。オントロジーは概念同士の関係性を示し、概念の意味・使い方を示す情報構造体である。これは辞書の様に用いられるだけでなく、構築の際に暗黙的な情報、例として前提知識・背景などを明示的に記述できる。本研究では明示化された前提知識を意図的に操作し、一貫した誤りを引き起こす知識構造をオントロジーで再現する。

誤答の理由を定義・内包している TA の先行研究として古池らの研究が存在する⁽³⁾。しかしこの研究では解答形式が選択式、部品の組み合わせ形式で行われており、TA の応答もテンプレート形式のものとなっている。筆者らは、学習者が行う教授活動に合う適応的なフィードバックが学習者に対して有用であると考え、自然言語形式の TA の実装を目指す。

3. 提案手法

本研究では学習者に対し自然言語による教授ができるシステムの提供を目指す。本手法の骨子は以下

の点である。(1)一貫した誤りを引き起こす知識体系の構築をオントロジーで行い、RAGの外部情報として取り込ませる。これによって出力文が一貫した誤りを含んだ制限された知識状態をもとに出力される。(2)学習者からの教授活動に合わせてオントロジーを選択・修正し、適切な知識状態を元に返答を行う。本項目では(1)を重点的に記述する。

本研究では会話を行う上での知識状態をオントロジーで構築する。正しい知識状態を表現したオントロジーを適正オントロジー、前提知識が誤っている知識状態のオントロジーを不適正オントロジーと定義する。不適正オントロジーの誤り箇所は学習者からの教授活動に合わせて、随時修正される。

適正・不適正オントロジーの例を図1の命題に関するオントロジーをもとに説明する。適正オントロジーでは含意(ならば、 \rightarrow)の論理演算は通常、逆、裏、対偶という派生系があり通常と対偶、逆と裏の真理値は同値である、と定義している。一方、不適正オントロジー①では含意の通常・逆・裏・対偶の真理値が同値である、と定義している。また不適正オントロジー②では、論理演算には項ロールを付与された命題ではなく、主張ロール、を格納している。これは命題で主張する事柄を表しており、「Aである」のAが該当する。この構造の場合、論理演算の項として他の命題が格納できない構造となっている。

これらの不適正オントロジーを元に応答する場合を考察する。不適正オントロジー①を適用した場合、ある含意が真であるならば逆・裏も真であると考え応答する。また不適正オントロジー②を適用した場合、「 $(A \vee B) \Rightarrow C$ 」のような三項以上の命題は解答できない。学習者はこのような前提知識が誤っている状態の知識構造を有するTAに対して教授活動を行い、知識体系を修正する学習活動を実現する。

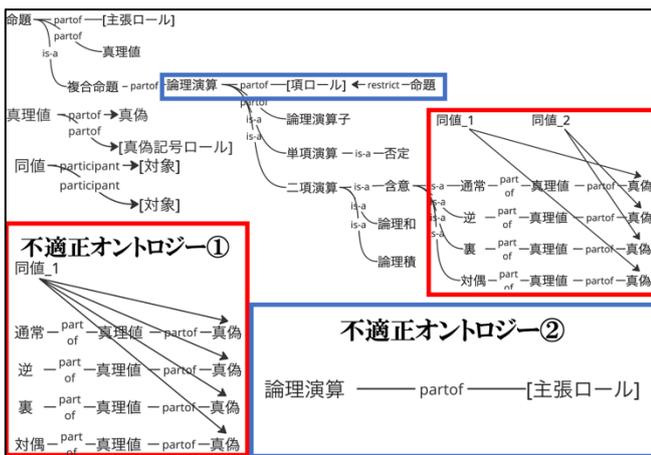


図1 適正・不適正オントロジー

4. 提案システム

本提案システムの構成図を図2に示す。本システムではRAGの技術をもとに外部情報として適正オントロジーと不適正オントロジーを用いる。使用する不適正オントロジーは、学習者の知識状態に合わ

せた選択機能、学習者との教授活動に基づき段階的に適正オントロジーに近づける修正機能で変化する。LLMは返答文とオントロジーを元にしたプロンプトを元に応答文を作成する。

学習の流れを示す。(1)事前問題を提示し学習者の解答から不適正オントロジーを選択する。(2)システムは不適正オントロジーをもとに誤答を生成する。(3)学習者は誤答の問題点・修正点を返答する。(4)返答文に合わせてオントロジーを修正し(2)に戻る、これを適正オントロジーになるまで繰り返す。図3に図2の不適正オントロジー①の会話例を提示しており、命題の逆・裏が真であると考えた会話を行っている。この会話の場合、学習者は逆・裏が真ではないことを示すだけでなく、同値関係の組み合わせが誤っていることをシステムに教授する必要がある。

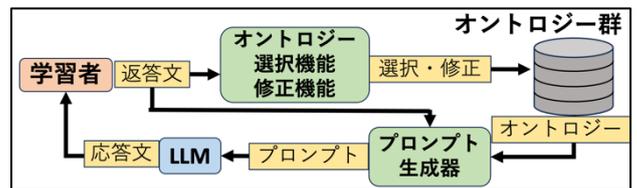


図2 システム構成図

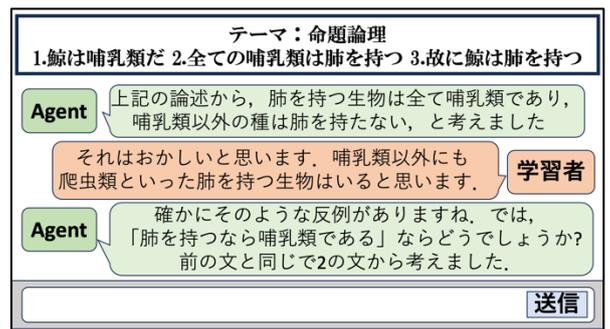


図3 システムでの会話例

5. おわりに

本研究では、誤った知識体系を構築する手段としてオントロジーを用いた。またこのオントロジーを用いて一貫した誤りを出力し、教授活動に合わせて知識状態を修正するTAを提案した。今後の課題として、対話システムの開発及び対話システムの評価実験を検討している。

参考文献

- (1) Angxuan Chen, Yuang Wei, Huixiao Le, Yan Zhang: "Learning-by-teaching with ChatGPT: The effect of teachable ChatGPT agent on programming education", <https://arxiv.org/abs/2412.15226>(参照 2025.5.21)
- (2) Kartik Sharma, Peeyush Kumar, Yunqing Li: "OG-RAG: Ontology-Grounded Retrieval-Augmented Generation For Large Language Models", <https://arxiv.org/abs/2412.15235>(参照 2025.5.21)
- (3) 古池謙人, 樋村いづみ, 東本崇仁, "TAME: 誤りの可視化を重視した Teachable Agent の挙動制御手法", 電子情報通信学会論文誌, vol.J106-D, No.2, pp.110-122(2023)