

大規模言語モデルを用いたメンタリングシステムの開発に向けて

大美浪 海晟¹ 佐藤 匠¹ 奥岡 耕平¹ 佐々木 康輔²

森口 昌和² 野田 尚志² 大森 隆司¹ 大澤 正彦^{1*}

Kaisei Ominami¹, Takumi Sato¹, Kohei Okuoka¹, Kosuke Sasaki²,
Masakazu Moriguchi², Hisashi Noda², Takashi Omori¹, Masahiko Osawa¹

¹ 日本大学

¹ Nihon University

² NEC ソリューションイノベータ株式会社

² NEC Solution Innovators, Ltd.

Abstract: 探究学習において、学習者に対するメンタリングは学習者の探究活動を支援する上で重要である。しかし、教員の負荷が大きい現在の教育現場において1対1のメンタリングを導入することは困難である。そこで本研究では大規模言語モデルを用いてチャットによるメンタリングを行うメンタリングシステムを提案する。提案システムでは、メンタリングにおいて効果的に相談者の問いを深めるための質問を設計する指針である「問いのデザイン」に基づいて対話を行う。また、提案するメンタリングシステムを用いて大学生を対象に実験を行い、実現に向けた課題や有効性を検証した。

1 はじめに

探究学習とは、自らが定義した問題に対して解決に向けて情報収集や他者と連携して解決に取り組むことを通して問題解決能力や情報処理能力を獲得する教育手法である。探究学習は近年教育分野で注目されており、カリキュラムとしても導入されている。探究学習を進めるうえでは教師には従来の教える役割とは別に、ファシリテーションなどの学習者の探究を促進する役割が求められる [1]。探究学習におけるファシリテーションでは、学習者の状態に応じて探究プロセスを進めるために対話を中心とした個別支援が必要とされている [2]。一方で、教員の負荷の多さが問題視されている現状の教育体制では、学習者一人一人に対する支援を提供することは難しい。そこで、著者らは探究学習における学習者一人一人に向けた教育者の支援行動を代替するシステムの開発を目的として研究を行っている。

本研究では、1対1の対話を通して学習者の思考を支援する行動をメンタリングと呼称し、探究学習のファシリテーションを目的としたメンタリングを自動化するメンタリングシステムの開発を目指す。先行研究では「問いのデザイン」と呼ばれる、安齋らが提案するメンタリングにおいて学習者が問いを深めることを効果的に促進する質問を設計するための指針を用いたメンタリング支援システムを提案した [3, 4]。提案システムで

は、問いのデザインに基づく質問パターンの種別をメンターが選択し、選択したパターンに基づいて相談テーマからルールベースで質問文を生成しメンターに提案することで、メンターの質問生成を支援する。提案システムを用いて実験を行い、提案システムを用いることで相談者の問いを深めることに寄与する可能性を示唆したことに加え、メンター間で共通したパターンの選択傾向が存在し得ることが示された。

しかし、先行研究ではメンターの支援を目的としていたため、教育者を代替することはできていなかった。また、先行研究で提案したシステムは相談テーマからルールベースで質問文を生成するため、相談者の回答を考慮した質問文を生成することはできず、自律システムとして教育者を完全に代替することは困難である。そこで本研究では、汎用的な対話に対応できる大規模言語モデルを組み合わせることで相談者の回答を考慮して問いのデザインに基づく質問を生成するメンタリング AI の開発に取り組んだ。本研究では初期検討として、問いのデザインを用いたプロンプトを設計し、問いのデザインを導入しないプロンプトを用いた大規模言語モデルと比較して効果検証を行った。

2 背景

2.1 メンタリング

メンタリングとは、教育者や年長者といった経験や知見を有する「メンター」が学生や若者といった「メン

*連絡先：日本大学文理学部
〒156-8550 東京都世田谷区桜上水 3-25-40
E-mail: osawa.masahiko@nihon-u.ac.jp

ティ」に対して行う支援活動である。メンタリングは探究学習の分野だけではなく、職場における教育やキャリアデザインといった幅広い分野においても行われており、メンタリングの定義や目的は多岐にわたる [5, 6, 7]。例えば、Kram らはメンタリングをキャリア的機能と心理・社会的機能の 2 つの機能に分類して説明している [7]。キャリア的機能とは、「仕事のコツや組織の内部事情を学び、組織における昇進に備える」ための支援であり、心理・社会的機能とは「専門家としてのコンピテンスやアイデンティティの明確さ、有効性を高める」ような支援を指す。

メンタリングはその重要性が語られる一方で、メンタリングの自動化や情報技術による支援を試みた研究は多くない。脇本らは、初任教師の育成を目的としたメンタリングにおいて、授業風景の撮影や映像視聴機能を搭載したメンタリング支援システムを開発し、初任教師が感じる課題に焦点化したメンタリング支援システムを提案している [8]。チャットシステムを用いた研究例には、自殺対策といったところのケアに関する対話にチャットシステムを用いたものがある [9]。しかし、本研究で対象とする探究学習におけるメンタリングを対象とした研究は十分に行われていない。

2.2 問いのデザイン

安齋らはメンタリングにおいて、メンティの問いを効果的に深めるための「問いのデザイン」を提案しており、問いのデザインとして (I) 言葉の定義を探る問い (II) 根源的な理由を探る問い (III) 時代の変化を探る問い (IV) 真善美を探る問いの 4 点を示している。4 点について以下に仔細を述べる。

I. 言葉の定義を探る問い

「人の価値観や解釈が異なるような言葉や、時代の変化によって意味づけが異なるような言葉について、明確な定義をあえて検討するための問い」と述べられている。これは、学習者が立てた問いのなかで、何気なく使われる言葉の本質を探ることが目的と考えられる。

II. 根源的な理由を探る問い

「人間の行為の根源的な理由を探ることは、人間や社会の本質に迫る上で重要」と述べられている。これは、学習者が立てた問いにおいて、暗黙のうちに前提とされている行為の意識化が目的と考えられる。

III. 時代の変化を探る問い

「時代の変化の中で、意味合いが変わったり、関係性が変わったり、処方箋が変わったりする事象について、変化の輪郭を追いかけるための問い」

と述べられている。根源的な理由を探る問いが繰り返される行為に焦点を当てているとすると、時代の変化を探る問いはその反対に位置するもので、学習者が立てた問いのなかで、変わりゆく事象やその変遷を意識化させることが目的と考えられる。

IV. 真善美を探る問い

「人間と社会の本質に迫る問い」と述べられている。学習者が立てた問いのなかで、当人が理想とする普遍的な価値基準がどこにあるのかを探ることが目的と考えられる。

上述の問いのデザインに基づいて自らが探究する問いを深め、「明らかにする価値があるか」、「ステークホルダーの視点に立っているか」、「探究的衝動がかき立てられているか」の 3 つの観点から問いを自己評価するというプロセスを踏むことで、メンティは問いを深めることができると安齋らは述べている。本研究では、問いのデザインに基づいて大規模言語モデルに与えるプロンプトを設計することで、メンティの問いを深めるメンタリングを可能にするメンタリングシステムの実現を目指す。

3 問いのデザインに基づくメンタリングシステム

本研究では、探究学習における教育者によるメンタリングを代替するメンタリングシステムの実現を目的に、大規模言語モデルを用いてメンティの発話を考慮して問いのデザインに基づく質問文を生成するメンタリングシステムを提案する。安齋らの提案する 4 つの問いのデザインをプロンプトに組み込むことで、メンティの問いを効果的に深めることのできるメンタリングを実現する。

3.1 問いのデザインに基づくプロンプトの設計

提案システムでは、プロンプトに問いのデザインの情報を含めることで問いのデザインに基づく返答の生成を行う。具体的には、2.2 章にて述べた 4 つの定義を説明する文章とそれぞれに対応する具体例の文章を含めた。加えて、現在の発話に対してどの問いのデザインを用いて回答するかを指示も含めた。どのデザインを選択するかについては、先行研究 [4] で行なった実験におけるメンターの質問パターンの選択ログを基に、4 つのデザインの遷移確率マップを導出し、遷移確率マップに基づいて決定した。遷移確率マップは、初回に選択するデザインの選択確率と、 n 回目にあるデザインを選んだ

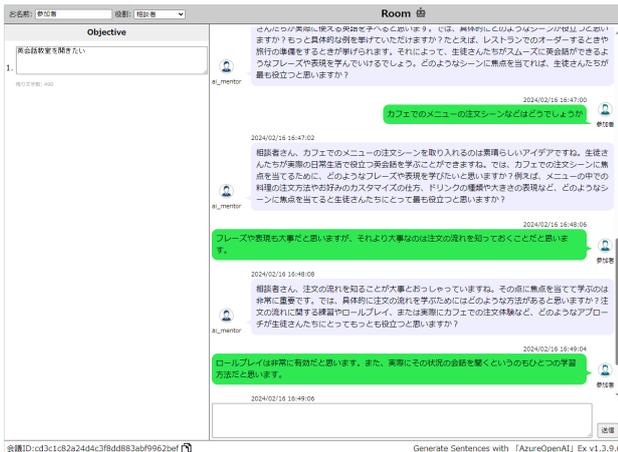


図 1: メンタリングシステムのチャット画面

時の次の $n+1$ 回目を選ぶデザインの選択確率で構成されている。

また、メンタリングを実現するための基本的な情報として以下の情報をプロンプトに含めた。

- 対話履歴
- メンタリングの目的
- メンティの問いの内容
- 出力時のフォーマットの指示

なお、大規模言語モデルには Azure OpenAI の gpt-3.5-turbo バージョンを用いた。

3.2 インターフェース

提案システムのチャット画面を図 1 に示す。画面左上のテキストフォームからメンタリングの題目の入力を行い、画面右下の入力フォームにメンタリングシステムに対する回答を入力し、送信ボタンを押すことでメンタリングシステムに回答が送信される。また、入力フォームの上部にメンタリングのシステムとのチャットの履歴が表示される。

4 実験

4.1 目的

本実験の目的は、著者らが提案する問いのデザインに基づくメンタリングシステムの有効性を検証することである。本実験では、提案システムを用いて実験参加者にメンタリングを受けてもらい、メンタリングに対する実験参加者による主観評価とメンタリング内容の対話

分析を行なった。また本実験は、「学問の扉」という授業内で実施した。この授業はグループでのプログラム開発を通して、情報技術やグループワークの技術を高めることを目指した授業である。この授業では OKR (Objective and Key Result) と呼ばれる、個人、チーム、および組織が測定可能な目標を定義し、その結果を追跡するために使用する目標設定フレームワークを用いて探究学習も行われており、一人ひとりの自己分析や目標設定を丁寧に実施されている。本実験ではメンタリングの題目を OKR の目標にあたる「Objective」を題目として設定してもらった。

4.2 条件

本実験では、問いのデザインをベースに開発した提案システムの有効性を検証するために提案システム条件とシンプル LLM 条件の 2 条件で比較を行なった。シンプル LLM 条件では、提案システムで用いるプロンプトから問いのデザインに関連する部分のみを除いたプロンプトを用いた。なお、本実験ではシステムとのメンタリングの経験が与える影響を考慮して被験者間実験を行なった。

4.3 評価

評価方法としてはアンケートによる主観評価を行なった。アンケートはメンタリングに対する満足度を評価する 1 問と、メンタリングで問いを深めることができた度合いを示すフカボリ度を評価する 1 問の計 2 問で、各質問には 5 段階のリッカート尺度を用いた。満足度に関する質問 Q1 は「今回のメンタリングについて満足していますか？」で、1. 満足できなかった 5. 満足できた、の 5 段階で回答してもらった。フカボリ度を評価する質問 Q2 は「メンタリングを通して、気づきはありましたか？」で、1. 深めることができなかった 5. 深めることができた、の 5 段階で回答してもらった。加えて、これらの 2 問に対してそれぞれ高く評価できる箇所と低く評価する箇所を自由記述で回答してもらった。最後に、実験全体に対する印象を自由記述で回答してもらった。

4.4 手順

まず初めに、実験参加者に対して実験内容とシステムの使い方について説明を行なった後に、実験参加への同意を得た。なお、本研究は日本大学文理学部研究倫理委員会の承認を受けた（承認番号:05-32）。本実験ではこちらが制作したメンタリングシステムを、教育用に学生に提供した。データ取得、分析に同意した学生についてのみチャット履歴および使用後アンケートを分析した。

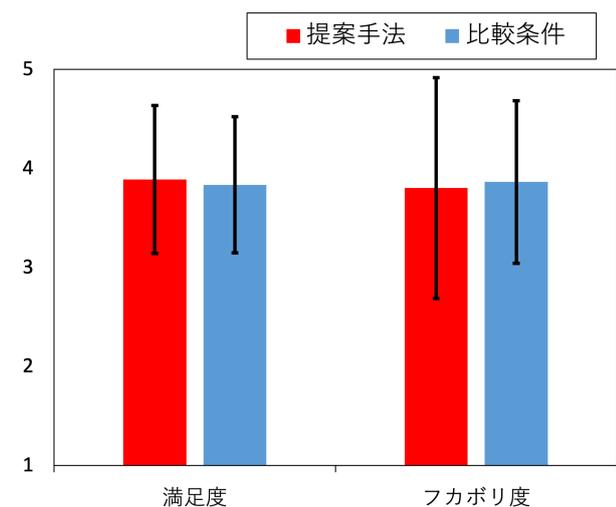


図 2: アンケートの結果

その後、チャットシステムの使い方を確認するために実験参加者が web 上のチャットシステムにアクセスしてもらい、テキストを送信するテストを行なった。なお、アクセスの際には授業を実施した教室に備え付けられている各自の PC からアクセスするようにしてもらった。テスト実施後、メンタリングの題目を入力してもらいメンタリングを 15 分間行なってもらい、終了後アンケートに回答してもらった。

4.5 実験参加者

実験参加者は日本大学文理学部内で実施された「学問の扉」という授業を受講した大学生の内、実験に同意した 85 名である。学生は授業を通して Objective に対してある程度探究のプロセスを進めている学生であった。

4.6 結果

実験に参加した 85 名のうち、実験の指示に従わなかった 14 名の無効なデータを除いた 71 名のデータに対して分析を行なった。アンケートの結果を図 2 に示す。アンケートの結果に対して 5%有意水準でマン=ホイットニーの U 検定を行なったところ、Q1, Q2 ともに条件間で有意な差は得られなかった (Q1:p=0.985, Q2:p=0.851)。そのため、実験参加者の主観的な評価においては提案手法の有効性は見られなかった。

5 考察

実験参加者による主観評価の結果、主観評価では満足度とフカボリ度の両方において提案手法とシンプル

LLM 条件の間に差は見られなかった。ここで、Q1, Q2 それぞれの項目で高く評価できる部分についての自由記述の回答を見ると、両条件ともに「自分の題目について深ぼるような質問をしてくれた」「自分の Objective の輪郭が明確になった」と好意的な評価が見られた。実際に両条件ともに、Q1 では 65%以上、Q2 では 70%以上が 4 以上の評価を回答していることから、両条件ともに実験参加者にとって問いを深めることに効果的であったが故に差が見られなかったと考えられる。

ここで、対話ログについて分析すると提案システムは「目標の中の「無関心」という言葉の本質や意味について考えてみましょう」のように、題目の意味や意図をメンティに考えさせる対話をする傾向があった。一方で、シンプル LLM 条件では「会話の中で自分の意見や感情を適切に伝える方法についてどのように考えていますか?」のように、題目において次に行動する内容や具体的なタスクについて対話する傾向が見られた。つまり、提案システムの方が題目をフカボリするような対話であったのに対して、シンプル LLM 条件では題目について具体的なタスクについて話す、行動を促すような対話であったといえる。この傾向について、著者らが対話を分類した所、題目をフカボリするような発話は提案条件では 109 回であったが、シンプル LLM 条件では 3 回であった。また、具体的なタスクに関する対話は提案条件では 17 回であったが、シンプル LLM 条件では 60 回であった。このことから、提案システム条件の方が題目についてフカボリするような対話を行なっていた可能性がある。

また、シンプル LLM 条件の方が具体的なタスクについて言及するため、メンティに対する質問よりも具体的なタスクに対する情報を提供するような対話が見られ、発話量が長くなる傾向が見られた。ここで、システム側の発話量について比較したところ、提案システム条件は平均 259 文字であったのに対してシンプル LLM 条件は 381 文字であり、t 検定の結果有意にシンプル LLM 条件の方が文字数が多かった ($p < .001$)。また、長くなる原因の 1 つとして、シンプル LLM 条件の方が 1 度の発話で複数の質問を同時に行なっている様子が見られた。これは、提案条件では一つの問いのデザインをプロンプトで指定することで質問の方向性を指示しているのに対して、シンプル LLM 条件では指示をしていなかったためと考えられる。

以上のことから、提案条件、シンプル LLM 条件共に実験参加者にとって満足度の高い、題目を深めるようなメンタリングであったと評価される一方で、実際のメンタリングの内容の傾向としては、問いを深めるメンタリングと問いの次のタスクを明確にするメンタリングという違いがあったと考えられる。本研究で目的とする探究学習におけるメンタリングでは、自らの問いに対する探究のプロセスを学習者が進めることを目的として

いるため、提案システムの方が探究のプロセスを進めるメンタリングであった可能性がある。

一方で、フカボリ度に関する評価においても差が見られなかったことから、実験参加者に対してフカボリ度の基準を提示できていなかった可能性がある。また、対話ログに関する分類についても、分類を担当した著者の基準に依存しているため、十分に客観的な評価とはなっていない。そのため今後は、本研究で目的とする問いを深めるメンタリングの評価方法について適切な評価方法を検討する必要がある。具体的には、主観評価においてはフカボリ度の評価について十分に実験参加者が理解できるようなインストラクションや、質問内容を検討する必要がある。また、対話ログの分析については基準について言語化した後、複数名の評価者によって分類することが必要である。加えて、より客観的な対話ログの分析方法としてトピック分析といった自然言語処理を用いた分析も検討する必要がある。

6 おわりに

本研究では、探究学習における教育者の負荷の削減を目的に、学習者とのメンタリングを行うメンタリングシステムの開発に取り組んだ。問いを深めるために有効とされる質問設計の指針である問いのデザインと大規模言語モデルを組み合わせることで、効果的に学習者の問いを深めることのできるメンタリングシステムを提案した。提案システムを用いて大学生を対象にメンタリングを行ってもらい、有効性を検証した。実験の結果、シンプルな大規模言語モデルを用いた場合と比べて、相談者の問いについて定義や意図といった問いを深めるようなメンタリングを行なっている傾向が見られた。

参考文献

- [1] 金井達亮. 探究学習に初めて取り組む高校教師の経験 一指導への意識の問い直しについての語りに着目して一. 教師学研究, Vol. 24, No. 2, pp. 39–48, 2021.
- [2] 文部科学省. 【総合的な探究の時間編】高等学校学習指導要領（平成 30 年告示）解説, 2018.
- [3] 佐々木康輔, 田足井昇太, 森口昌和, 野田尚志, 大森隆司, 宮田章裕, 大澤正彦. 問いかけ文の半自動生成機能付チャットシステムによるメンタリングの試み. *HCS*, pp. 112–117, 2023.
- [4] 大美浪海晟, 田足井昇太, 奥岡耕平, 佐々木康輔, 森口昌和, 野田尚志, 大森隆司, 大澤正彦. メンタリングにおける相談者への問いかけ文生成システムのメ
- ンターに対する影響の評価. 電子情報通信学会教育工学研究会, pp. 5–9, 2024.
- [5] 麓仁美. 組織における協力行動のマネジメント：仕事の設計がメンタリング行動と向社会的モチベーションに与える影響. 組織科学, Vol. 53, No. 2, pp. 43–56, 2019.
- [6] 久村恵子. メンタリングの概念と効果に関する考察. 経営行動科学, Vol. 11, No. 2, pp. 81–100, 1997.
- [7] K. E Kram. *Mentoring at work: Developmental relationships in organizational life*. University Press of America, 1988.
- [8] 脇本健弘, 苅宿俊文, 八重樫文, 望月俊男, 中原淳. 初任教师が感じる課題に焦点化したメンタリングを支援するシステムの開発と評価. 教育システム情報学会誌, Vol. 30, No. 2, pp. 160–171, 2013.
- [9] 富澤浩樹, 川乗賀也. 「こころの相談窓口」への誘導を目的とした若年層向けチャットボットシステムに関する考察. 経営情報学会全国研究発表大会要旨集, Vol. 202011, pp. 285–288, 2021.