

探究学習を支援するメンタリングシステムの実践的評価: メンタリングの戦略の切り替えが学習者に与える影響

A Practical Evaluation of a Mentoring System for Inquiry-Based Learning: Effects of Switching Mentoring Strategies on Learners.

大信田 和華[†], 角森 穂佳[†], 古賀 日南乃[†], 奥岡 耕平[†],
佐々木 康輔[‡], 野田 尚志[‡], 森口 昌和[‡], 大澤 正彦[†]
Kazuha Oshida[†], Honoka Tsunomori[†], Hinano Koga[†], Kohei Okuoka[†],
Kosuke sasaki[‡], Hisasi Noda[‡], Masakazu Moriguchi[‡], Masahiko Osawa[†]

[†] 日本大学, [‡] NEC ソリューションイノベータ株式会社

[†]Nihon University, [‡]NEC Solution Innovators, Ltd.

chkz23007@g.nihon-u.ac.jp

概要

本研究の目的は探究学習において教育者が学習者の思考を促すために行うメンタリングの自動化である。著者らは先行研究として、学習者の発話に応じて対話戦略を選択するメンタリングシステムを開発したが、評価実験において比較条件と対話時間の設定に課題があった。そこで本研究では、比較条件を修正し1週間の期間を設けて実験を行った。実験の結果、条件間で差は見られなかったが、メンタリング前後で探究的衝動が強まった。

キーワード: メンタリング, 探究学習, 大規模言語モデル, 問いのデザイン, 教育支援

1. はじめに

近年、探究学習と呼ばれる学習方法への関心が高まっている。探究学習とは学習者自らが問題を設定し、設定した問題を解決するプロセスを経験することで、課題解決能力などを習得する学習法である。探究学習はデューイの理論を起源にもつもので、学習者は探究的衝動と呼ばれる、自らがその問題を探究したいという強い欲求をもって学習に取り組むことが重要とされている (Dewey, 1933)。

探究学習において教育者は一対一のコミュニケーションであるメンタリングを通して学習者の思考を促すことが求められる。しかし、教育者の負担が大きい教育現場では、学習者一人一人に対するメンタリングを導入することは困難である。そのため、教育者の負担を増加させずに質の高い探究学習を実現するためにメンタリングの自動化が求められる。

探究学習において対話支援を自動化する既存研究では、情報収集のヒントを提案するといった、問題解決を補助するような対話を主眼に置いたシステムの研究がある (Farhan 他, 2023)。しかし、相談者が自身で設

定した問題に対して内省を促すという点については取り組まれていなかった。

この問題に対して著者らは、「問いのデザイン」を大規模言語モデル (LLM) のプロンプトに組み込み、探究的衝動を持てるように支援するメンタリングシステムを提案した (古賀他, 2025)。さらに、提案システムは相談者の発話から探究的衝動を推定し、探究的衝動が十分ないと判断された場合に、相談者の意見を傾聴する対話戦略へと切り替えることができる。このように前準備として相談者の発話を引き出すことで、その後の相談者に問いのデザインに基づく質問に対して探究的衝動が強い回答を促すことを目指している。

しかし先行研究では、傾聴のみを用いた条件とのみ比較して問いのデザインのみ条件と比較しておらず、対話戦略の選択が与えた影響について十分に評価できていなかった。また、メンタリング時間が15分であったため、システムの影響を評価するための十分な時間が確保できていなかった。

そこで本研究では、先行研究で提案したメンタリングシステムの対話戦略の選択による影響を評価するために、提案システムと問いのデザインのみ条件を比較する実験を行った。実験は実際に探究学習を実施している授業内で大学生を対象に行い、メンタリングの期間を1週間設けて実施した。

2. 関連研究

2.1 探究学習

MacKenzie は探究学習の形式を学習者の自由度に基づいて、自由度の低いものから順に Structured Inquiry, Controlled Inquiry, Guided Inquiry, Free Inquiry の4つに分類している (MacKenzie, 2016)。本研究で扱う探究学習は、問題の設定と解決策の探索を学習者自らが考え実施する、Free Inquiry に分類される。Free Inquiry

では学習者自らが問題を設定するため、問題設定に対する思考を促すことが重要となる。

2.2 探究学習における対話支援の自動化

探究学習における対話支援を自動化する既存研究として、Farhan らは探究学習のような生徒主導型の学習を支援するチャットボット、TeacherGAIA を提案している (Farhan 他, 2023)。

しかし、既存研究の多くは設定した問題に対する解決策の探究を支援することに焦点を当てており、問題設定について学習者の思考を促すという点については取り組まれていない。

2.3 問いのデザイン

安齋ら (安齋他, 2021) は、自らの問題設定において探究的衝動が重要だと述べており、自身で深めるための観点として「問いのデザイン」を提案している。以下に問いのデザインで示されている4つの観点を示す。

言葉の定義 「“幸せ”とは何か？」というように、時代の変化によって意味づけが変わるような言葉や人によって解釈が異なるような言葉について、改めて定義を学習者に問う質問。

根源的な理由 「なぜ人は嘘をつくのか？」というように、設定した問題の中で、暗黙的に前提となっていた人の行為の根源的な理由を問う質問。

時代の変化 「これからの子育てとは？」といったように、設定した問題において、時代の中で変わりゆく事象やその変遷を意識化させるための質問。

真善美 「正しい～とは何か？」のように、学習者が理想とする普遍的な価値基準を明確にする質問。

3. 対話戦略を選択するメンタリングシステム

先行研究では、メンタリング時の学習者の回答によって対話戦略を切り替えることで、効果的なメンタリングを可能にするメンタリングシステムを提案した。提案システムは lida ら (2024) の研究において提案された LLM と Belief desire-intention(BDI) モデルベースの認知アーキテクチャを統合する手法を基に、3つのモジュールで構成されている (図1)。以下に各モジュールの動作を説明する。なお、本研究で用いた LLM は OpenAI の GPT-4o である。また、各モジュールではプロンプトとユーザの回答に加えて直前2往復の入出力の履歴が与えられた。

3.1 探究的衝動推定モジュール

学習者の発話文から LLM を用いて学習者の問題に対する探究的衝動の強さを推定する。プロンプトにはユーザの Belief と Desire とメンタリングの題目が

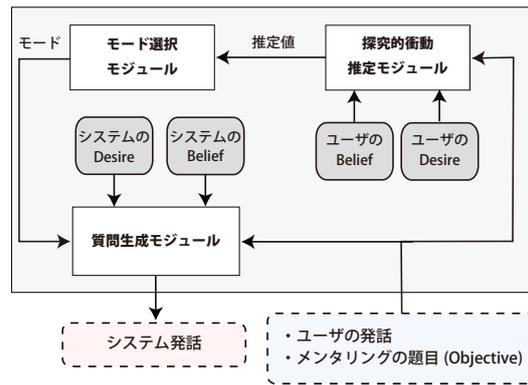


図1 提案システムの概略図

記述された後、ユーザの発話から題目に対する探究的衝動の強さを推定して、100までの自然数に数値化する旨の指示文が記述されている。また、ユーザの Belief にはユーザが相談者であるといった役割情報を、Desire には自分の問題を深堀りたいといった情報をそれぞれ設定した。

3.2 モード選択モジュール

探究的衝動推定モジュールで推定した値から、質問生成を行う対話戦略を決定する。時刻 t における推定値の移動平均 μ_t が閾値 T を超えた場合に、問いのデザインに基づいて質問を生成する問いのデザインモード、閾値以下の場合に傾聴モードになるようにした。

3.3 質問生成モジュール

モード選択モジュールで選択したモードでシステムの発話の生成を行う。システムの Belief にシステムがメンターであるといった役割の情報を設定し、Desire にはメンターとして相談者の問いの深堀りを促す質問をしたいという情報を固定で設定した。

3.3.1 問いのデザインモード

問いのデザインに基づいた4つのプロンプトを用いて質問の生成を行う。各プロンプトにはシステムの Belief と Desire、相談者の題目が記述された後、その観点に関する説明が記述され、その観点に基づいて質問を生成するように指示する文章が記載される。最後に、出力に関する制約事項に記載される。なお、4つのプロンプトの選択方法は2.3章で記載した順序で選ばれるようにし、最後の「真善美」の次は再び「言葉の定義」が選ばれるようにした。

3.3.2 傾聴モード

学習者の発話を促すために傾聴するような発話の生成を行う。プロンプトは、問いのデザインモードのプロンプトの内、問いのデザインに関する箇所を傾聴する質問の生成を指示する文に置き替えた。指示文としては、題目に対して学習者の経験などに基づいた主観的な想いについての発話を促し、その想いに同調する



図2 システムのインターフェース

ように指示した。

3.4 インターフェース

提案システムのチャット画面を図2に示す。まず、画面左上のテキストフォームからメンタリングの題目を入力する。次に画面左下のテキストフォームにシステムの発話に対する回答を入力する。送信ボタンを押すとシステムに回答が送信され、画面右上部にシステムからの返答が表示される。

4. 実験

4.1 目的

本実験の目的は、提案システムの対話戦略の選択により、探究的衝動の弱い回答を行う相談者に探究的衝動が強い回答を促すことができるようになるか検証することである。

4.2 条件

提案システムが探究的衝動を強めることを明らかにするために、提案条件と問いのデザインのみ条件の2条件で実験を行った。問いのデザインのみ条件は探究的衝動を推定せず、常に問いのデザインモードを用いる条件である。なお、本実験ではメンタリングの経験による影響を考慮し参加者間実験とした。

4.3 設定

本実験は日本大学文理学部内で開講されている探究学習を導入した授業である「自主創造の基礎」の受講生を対象として2週間実施した。授業ではOKR(Objective and Key Results)と呼ばれる目標設定フレームワークを用いて自らが取り組む問題を記述している。OKRは目標である「Objective」とその目標の達成を測るための成果「Key Results」で構成される。本実験では、実験参加者のObjectiveをメンタリングの題目として入力してもらい、メンタリングを行ってもらった。なお、本実験ではモード選択モジュールの移動平均は窓幅3とし、閾値 T は80とした。また、最初の2回の発話は常に問いのデザインモードになるようにした。

4.4 評価方法

本実験では、探究的衝動の強さを評価するために実験参加者の回答文に対する客観評価を行った。まず、メンタリングの前後で問いのデザインに基づいた4つの質問に回答してもらい、回答文を収集した。収集後、4つの回答文を1つのセットとして、内容に対して「相談者が自分の目標について、自ら考えたい、話したい、明らかにしたいという思いを持っているかどうか」を第1から第4著者までの4名が5段階のリッカート尺度で1セット毎に評価した。なお、参加者や回答タイミングが分からないように、セットの評価順序はランダムに決定し、回答文のみを表示して評価した。

また、参加者へ質問する際は形式を統一するために、文中の一部を変数として実験参加者のObjectiveを代入して生成する、テンプレート形式を用いた。また、自らのOKRに加えてメンタリング後にはシステムを利用した印象等についても回答してもらった。

4.5 手順

まず実験内容について説明を行った後に、実験参加者への同意確認を行った。なお、メンタリングは授業の一環として実施したため同意しない学生もメンタリングを行ったが、実験に使用するデータは同意した学生からのみ取得した。本実験は日本大学文理学部研究倫理委員会の承諾を受けて実施した(承諾番号:07-21)。

次に、メンタリング前アンケートに回答してもらった。その後、実験参加者のPCでweb上のシステムにアクセスしてもらい、自らのObjectiveを入力してもらった後に、全参加者共通のテキストを送信してもらい動作テストを行った。動作テスト後、1週間メンタリングをしてもらった。なお、期間中は任意のタイミングで中断と再開ができるものとし、再開時は中断時の続きからメンタリングが再開するものとした。

1週間後システムを停止し、メンタリング後アンケートに回答してもらった。なお、アンケートはシステム停止後から1週間以内に回答してもらった。実験終了後デブリーフィングを行い、実験参加者全員に実験内で試していない条件のシステムでメンタリングできる機会を1週間提供した。

4.6 結果

実験参加者のうち、対話を1回以上を行い、メンタリング前後のアンケートに回答した53名のデータに対して分析を行った。そのうち、探究的衝動の弱い学習者に対する影響を評価するために、メンタリング前の平均評価値が3未満だった学習者25名のデータに対して分析を行った。

探究的衝動の客観評価の結果を図3に示す。メンタ

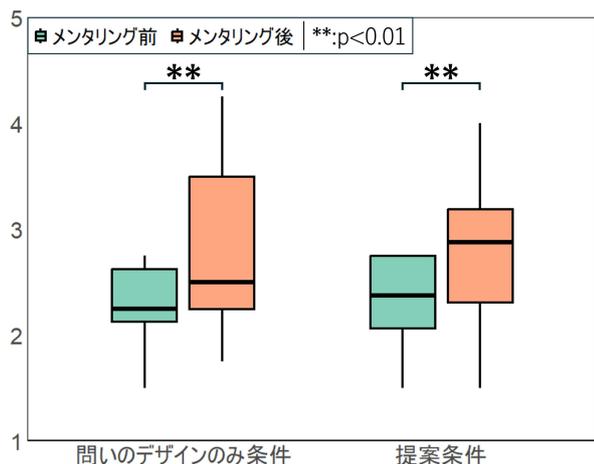


図3 探究的衝動の客観評価の結果

リング前後と条件間で比較するために、整列ランク変換を施した二元配置分散分析を行った。分析の結果、交互作用は見られなかったため各要因の主効果を見ると、条件間では有意差は見られなかった ($F(1, 23) = 0.01$, $p = 0.91$, $\eta_p^2 < 0.01$) もの、メンタリング前後では、メンタリング後に客観評価が有意に向上していた ($F(1, 23) = 8.39$, $p < 0.01$, $\eta_p^2 = 0.27$)。

5. 考察

客観評価の分析の結果、両条件ともメンタリング前後で有意に探究的衝動を高められていたが、条件間で違いは見られなかった。ここで、対話と推定値のログを分析すると、事前評価値3未満の参加者で、提案条件下であった10名のうち、戦略の切り替えが行われた者は8名であった。しかし、最初に入力した全員共通の対話開始の依頼文に対する推定値が低かったため、3回目の発話が傾聴に切り替わりやすくなっていた。そこで、3回目以降で切り替えが生じた参加者を確認すると、1名のみであり、提案システムによる切り替えがあまり生じていないことがわかった。そのため、条件間での差が見られなかったと考えられる。

ここで、事前評価値3以上で提案条件下で参加した16名に対して、同様に3回目以降での切り替えを確認すると、事前評価値が低い参加者よりも多い5名に3回目以降での切り替えが生じていた。そこで、推定値の確からしさを見るために、全参加者の事前評価値と各対話の平均推定値のスピアマンの順位相関係数 ρ を求めた。その結果、弱い相関 ($\rho = 0.30$) が有意にあることが分かった ($p = 0.02$)。このことから、推定値は対話全体ではある程度傾向を捉えていた可能性がある。

そこで、推定値が低く評価された対話ログを分析すると、参加者がシステムに「どういうことですか？」など質問した場合や、短い文章で返した場合に著しく低

い値を返す傾向があった。そのため、探究的衝動が強い参加者であってもそのような対話があると局所的に値が低下し、傾聴モードに切り替わったと考えられる。そのため、移動平均の窓幅を大きくするなど、切り替え判定手法の改善が必要である。一方で、初回の固定文に対する推定を除く全ての推定値の内、90%以上が70以上で推定されており、値が高めに推定される傾向にあったことも判定を困難にした原因と考えられる。そのため、探究的衝動推定のプロンプトに推定値を含めた例示を含めるなど、推定方法の改善も必要である。

また、全参加者の対話回数について分析すると、事前評価値3以上の参加者は平均10.7回であったのに対して3未満の参加者は平均6.56回であり、この差は有意であった ($p = .03$, $d = 0.59$)。そのため、探究的衝動が十分でない参加者は対話の発話回数が少ないことから、切り替えが行われるほど対話していない場合や傾聴による効果が得られるほど対話が行えていなかった可能性がある。そのため、メンタリング自体の継続を促す戦略も必要と考えられる。

6. おわりに

著者らは先行研究として、学習者の発話から探究的衝動を推定して対話戦略を選択するメンタリングシステムを提案している。本研究では、提案システムの対話戦略の選択によって、探究的衝動が弱い相談者に対して探究的衝動を向上を促すことができるか評価する実験を行った。実験は、探究学習を実施する授業にて1週間期間を設けて、問いのデザインのみ条件と提案条件の比較実験を行った。実験の結果として、条件間で差は見られなかったが、メンタリング前後で探究的衝動を強める効果が見られた。

文献

- Trevor MacKenzie (2016). *Dive into Inquiry: Amplify Learning and Empower Student Voice*. EdTechTeam Press.
- 安斎勇樹, 小田裕和 (2021). リサーチ・ドリブン・イノベーション「問い」を起点にアイデアを探究する. 翔泳社.
- John Dewey (1933). *How We Think*. Heath.
- 古賀日南乃, 奥岡耕平, 佐藤匠, 大美浪海晟, 佐々木康輔, 森口昌和, 野田尚志, 大澤正彦 (2025). 学習者に応じて対話戦略を選択する探究学習のためのメンタリングシステム. *HAI シンポジウム*, pp. P-41.
- Shanti Divaharan, Hui Yong Tay, Farhan Ali, Doris Choy, and Wenli Chen (2023). Supporting self-directed learning and self-assessment using teachergaia, a generative AI chatbot application: Learning approaches and prompt engineering. *Learning: Research and Practice*, 9(2), 135-147.
- Ayu Iida, Kohei Okuoka, Satoko Fukuda, Takashi Omori, Ryoichi Nakashima, and Masahiko Osawa (2024). Integrating large language model and mental model of others: Studies on dialogue communication based on implicature. In *Proceedings of the 12th International Conference on Human-Agent Interaction (HAI '24)*, pp. 260-269.