

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

A Mentoring System for Inquiry-Based Learning that Selects Dialogue Strategies According to Learners

古賀 日南乃^{1*} 奥岡 耕平¹ 佐藤 匠¹ 大美浪 海晟¹

佐々木 康輔² 森口 昌和² 野田 尚志² 大澤 正彦¹

Hinano Koga¹ Kohei Okuoka¹ Takumi Sato¹ Kaisei Ominami¹

Kosuke Sasaki¹ Masakazu Moriguchi¹ Hisashi Noda¹ Masahiko Osawa¹

¹ 日本大学文理学部, ² NEC ソリューションイノベータ株式会社

Abstract: 探究学習は学習者自身が問題を設定し解決する過程を通して能力を獲得する学習法であり、教育者によるメンタリングが重要である。しかし、学習者一人一人とのメンタリングは教育者の負担増加に繋がる。そこで本研究では、学習者に応じて対話戦略を選択するメンタリングシステムを提案した。システムは、発言が積極的な場合には問いを深める質問を、そうでない場合には発言を促す傾聴質問を行う。探究学習を実施する授業にて実験を行い、提案システムの有用性を検証した。

1 はじめに

近年日本の教育において、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ために情報収集や他者と連携するプロセスを経験することで、課題解決の能力など様々な能力を習得させる学習法である。従来の教育法と異なり、探究学習では知識や問題の解決方法を直接的に教示するのではなく、学習者自らが問題を深め、探究プロセスを進めるように支援することが教育者の役割となる [10]。

支援方法の中で重要なものとして、一対一のコミュニケーションがあげられる。教育者は対話を通して学習者の関心や問題に対する理解を引き出すとともに、質問を投げかけることで思考を促すことが求められる。本研究では、このように探究学習における教育者と学習者の一対一のコミュニケーションをメンタリングと呼称する。

しかし、教育者の負担が大きいことが社会問題の1つとなっている現在の日本の教育現場では [11]、学習者一人一人に対して行わなければならない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、メンタリングを自動化するシステムが必要である。

探究学習において情報技術を導入する既存研究では、

タブレットを介した相談者のデータ収集や教育者との連携を支援するような手法の研究が多い。また、対話を主眼に置いたシステムでは、設定した問題に対してアプローチ方法やヒントを提案するといった、問題解決を補助するような研究が主であり、相談者が自身で設定した問題自体に対して内省を促すという点については十分に考慮されていなかった。この問題に対して著者らは、「問いのデザイン」と呼ばれる相談者の思考を促す対話戦略を大規模言語モデルのプロンプトに組み込むことで内省を促すメンタリングシステムが提案している [9]。しかし、自らの考えを十分に言語化できていない相談者に対しては、「問いのデザイン」による対話戦略の効果があまり発揮されない可能性が示されていた。

そこで本研究では、大規模言語モデルを用いて発話履歴から相談者の発話状態を推定し、状態に応じて対話戦略を切り替えるメンタリングシステムを提案した。具体的には、相談者の発話履歴から言語化が十分に行えていないと判断した場合に、相談者の意見に共感し、傾聴する対話戦略へと切り替える。このように対話戦略を選択することにより、内省の前段階として相談者の発話を引き出すことを目指す。システムの有用性を検証するため、実際に探究学習を実施している大学生を対象に、システムによるメンタリングを体験してもらい、評価を行った。

*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: chhi22083@g.nihon-u.ac.jp

2 関連研究

2.1 探究学習

探究学習とは学習者自らが問題を設定し、設定した問題に対して情報収集、意見交換といった問題解決のためのプロセスを進める学習方法である。これからの社会で必要な能力の向上に繋がるとして、日本では国が提示する高等学校の教育カリキュラムの指針に探究学習の導入が記載されている [12]。また、Mackenzie は探究学習の分類として、学習対象の自由度に基づいた、Structured Inquiry, Controlled Inquiry, Guided Inquiry, Free Inquiry の 4 つの分類を示している [5]。Structured Inquiry は最も自由度が低く、問題設定も解決策を探るリソースも教育者が提示するものである。対して、Free Inquiry は問題設定も学習者が行う、最も自由度が高い探究学習である。本研究で扱う探究学習は Free Inquiry に該当する。

探究学習を支援するために情報技術を活用した研究も盛んに行われている [4]。例えば、Farhan ら [2] は生徒の探究学習のような生徒主導型の学習を支援するためのチャットボット、TeacherGAIA を提案している。TeacherGAIA はプロンプトエンジニアリングを用いて、大規模言語モデルによる探究学習や自己評価の支援を目的とした対話機能を実現している。例えば TeacherGAIA では、「なぜ太陽は東から昇るのか」という問題に対し、「地球の自転の向きについて考えたことはありますか」といった調査の気づきを支援する質問を行う。

ここで、本研究で扱う Free Inquiry では学習者自らが問題を設定するため、設定した問題の自らの動機との関係性や、問題の重要な点を言語化する等、問題設定に対しても思考を促すことが重要となる。しかし、既存研究の多くは設定した問題に対する解決策やアプローチ方法の探究への支援に焦点を当てており、問題設定について相談者の思考を促すという点については十分に考慮されていない。

2.2 問いのデザイン

安斎ら [7] は、「問いのデザイン」と呼ばれる、問いを深ぼるために効果的な質問の設計指針を提案している。探究学習におけるメンタリングにおいても、このデザインに基づいて質問を設計することで、学習者の問いに対する深ぼりを促すことができる。以下に「問いのデザイン」において示されている 4 つの質問のデザインパターンについて述べる。

言葉の定義 「“幸せ”とは何か？」というように、人によって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変わ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、学習者が設定した問題に関連する領域の中で、暗黙のうち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、相談者が設定した問題のなかで、時代の変化の中で、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、相談者が理想とする普遍的な価値基準がどこにあるのかを明確にすることを目的とした質問。

大美浪らの研究では、上述の「問いのデザイン」に基づいた質問を行うことで、学習者に自らの問題に対する内省を促す可能性があることが示されている [9]。しかし一方で、学習者がメンターに対して自らの考えや意図を十分に発話できていない状態では、「問いのデザイン」に基づく質問は有効でない可能性が示唆されていた。そのような状態においては、学習者の発話に対して共感した上で学習者の発話を促すような、傾聴を目的とした対話戦略が必要になると考えられる。

3 対話戦略を選択するメンタリングシステム

本研究では学習者のメンタリング時の応答内容に応じ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する。具体的には、学習者の応答から自らの問題に対する意欲や十分な言語化を行えているかを推定し、学習者の積極的な発話を促すモードと「問いのデザイン」に基づいて学習者の問いを深める質問を行うモードを切り替える。学習者の状態に適したモードで質問することで、探究学習において効果的なメンタリングを実現することができる。以下に提案システムの詳細を述べる。

図 1 に提案システムの概略図を示す。提案システムは、Iida ら [3] の研究において提案されている、Belief-desire-intention (BDI) モデルベースの認知アーキテクチャと大規模言語モデルを統合する手法をベースに構築した。以降の節で、各モジュールの動作について説明する。なお、本研究では大規模言語モデルには OpenAI の GPT-4o (バージョン:gpt-4o-2024-05-13) [6] を用いた。また、各モジュールではプロンプトと入力に加えて、直前 2 往復のシステムとユーザの発話情報がコンテキストとして与えられた。

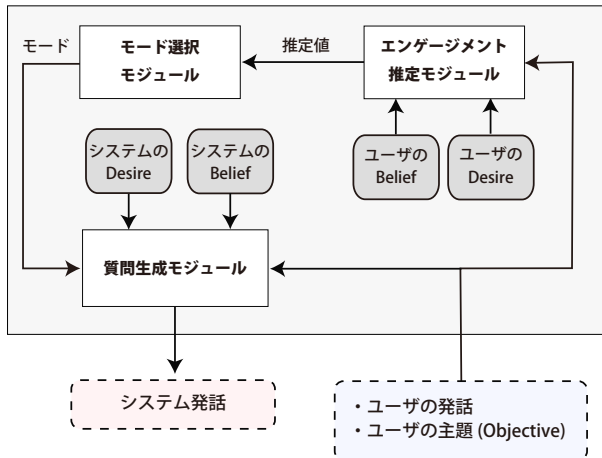


図 1: 提案システムの概略図

3.1 エンゲージメント推定モジュール

エンゲージメント推定モジュールでは、学習者の質問に対する応答から、大規模言語モデルを用いて学習者の問題に対する発話意欲や言語化の状態を推定する。入力として、ユーザの発話とユーザのメンタリングの主題、ユーザの Belief と Desire が与えられる。なお本研究では、ユーザの Belief にはユーザがメンタリングにおいて相談者であるといった役割情報が、Desire には自分の問題を深掘りしたいといった情報を固定で設定した。

プロンプトにはまず、システムの Belief と Desire と相談者が設定した主題が記述される。次に、ユーザの発話からユーザが主題について自らの考えを話したいといった意図を持っているか数値化しなさい、といった旨の指示文が与えられる。なお、値は 0 から 100 までの自然数で、値が大きいほどユーザが自ら積極的に発話できている状態となるように指示した。最後に、出力は数値のみにするといった出力に関する制約を記載した。

3.2 モード選択モジュール

モード選択モジュールでは、エンゲージメント推定モジュールで推定した値を元に、問いのデザインに基づいて質問生成するモードと傾聴するモードのどちらのモードで質問生成を行うかを決定する。ここで、モード選択はエンゲージメント推定モジュールが逐次的に出力する推定値の時刻 t における移動平均 μ_t が閾値 T を超えた場合に問いのデザインモード、閾値以下の場合に傾聴モードとした。なお本実験では、移動平均は直近 3 回分の推定値の移動平均とし、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。



図 2: システムのインターフェース

3.3 質問生成モジュール

質問生成モジュールでは、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を生成する。入力にはシステムの Belief と Desire とユーザの発話とユーザのメンタリングの主題が与えられ、システムの発話が出力される。なお、本研究ではシステムの Belief にはシステムがメンターであるといった役割の情報を、Desire にはメンターとして相談者の問いの深掘りを促す質問をしたいという願望の情報を固定で設定した。

3.3.1 問いのデザインモード

問いのデザインモードでは、第 2.2 章で述べた 4 つの問いのデザインに基づいたプロンプトを用いて質問を生成する。プロンプトにはまず、システムの Belief と Desire と相談者が設定した主題が記述される。その後、4 つの観点の中から 1 つが選ばれ、その観点に基づいて質問を生成する指示する文章と、2.2 章で示すような説明と例文が記載される。最後に、質問は端的に短くするといった出力形式に関する制約事項が記載される。なお、本研究では問いの観点の選び方については 2.2 章で述べ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」に戻るようにした。

3.3.2 傾聴モード

傾聴モードでは、学習者の発話を促すために傾聴するような発話を生成する。プロンプトには、問いのデザインモードのプロンプトの内、問いのデザインに基づいて質問の生成を指示する箇所を傾聴する質問を生成させるような指示文に置き換えたものを用いた。なお傾聴する際には、学習者の設定した主題に対して学習者自身の経験や考えを踏まえ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ように指示した。

表 1: アンケート項目

	質問文	形式	質問タイミング
Q1	現在の自分の Objective を記述してください	自由回答	事前・事後
Q2	自分の Objective を設定した理由があれば記述してください	自由回答	事前・事後
Q3	自分の Objective について十分に考えられている	リッカート	事前・事後
Q4	自分の Objective に自信がある	リッカート	事前・事後
Q5	自分の Objective について、他人が正しく理解できるように説明することができる	リッカート	事前・事後
Q6	自分の Objective が好きである	リッカート	事前・事後
Q7	自分の Key-Results を箇条書きで記述してください	自由回答	事前・事後
Q8	自分の Objective について、メンターに理解してもらったと思う	リッカート	事後

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。画面左上のテキストフォームからメンタリングの題目として相談者の問いを入力し、画面左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、提案するメンタリングシステムの有効性を検証することである。探究学習に実際に取り組んでいる学習者を対象に検証するため、日本大学文理学部にて実施された探究学習をカリキュラムに導入した授業において実験を実施した。実験参加者に授業内で設定した目標を主題としてメンタリングを受けてもらい、アンケートによる主観評価と対話ログによる分析を行った。

4.2 比較条件

傾聴だけではなく、既存の問いの観点の対話戦略と組み合わせる用いることが学習者の問いを深めることに寄与することを明らかにするために、提案条件と傾聴のみ条件の 2 条件での比較実験を行った。傾聴のみ条件は他者意図推定を行わず、常に傾聴モードを用いて質問生成した条件である。なお、本実験ではメンタリングをした経験の影響を考慮し、被験者間実験を行った。

4.3 設定

実験は、日本大学文理学部内で開講されている「学問の扉」の 2024 年 11 月 13 日に実施された授業内で行った。授業では、OKR (Objective and Key Results) と呼ばれる、個人や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が行われている。学習者はそれぞれ自らの目標である「Objective」を定めており、その Objective に対して Key Results を考えること等を通して、探究学習を行っていた。本実験ではメンタリングの題目として、実験参加者各自が持つ Objective を設定してもらい、システムとのメンタリングを通して Objective に対する問いを深めてもらった。

4.4 評価方法

システムの有用性を検証するため、アンケートによる主観評価と対話ログの分析を行った。アンケートはメンタリングの前後それぞれで実施した。メンタリング前アンケートでは参加者自身の Objective と Key Results に関する質問に答えてもらった。表 1 に質問内容と質問を行ったタイミングを示す。メンタリング後ではメンタリング前に行った 7 つの質問に加えて、メンターに自身の Objective を理解してもらった印象を評価するための 7 段階のリッカート尺度の質問 (Q8) についても回答してもらった。また、SUS(System Usability Scale)[1] と呼ばれる製品やサービスの使いやすさを図るためのアンケートにも回答してもらった。

4.5 手順

はじめに、実験内容とシステムの使い方の説明を行った後に、実験参加への同意の確認を行った。チャットシステムによるメンタリングは授業の一環としても実施したため、実験に同意しない学生もメンタリングを行っ

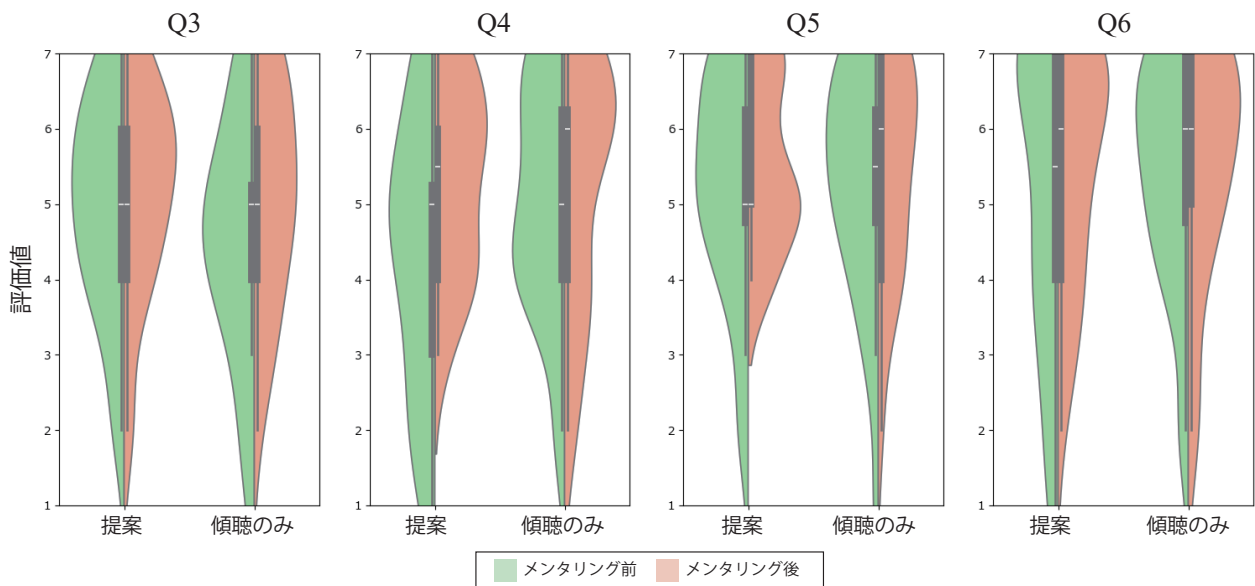


図 3: メンタリング前後で実施した質問項目の評価

でもらったが、データは実験参加に同意した学生についてのみデータを取得した。なお、本実験は日本大学文理学部研究倫理委員会の承認を受けて実施した（承認番号:06-64）。実験説明後、メンタリング前アンケートに回答してもらった。アンケートの回答後、各実験参加者の個人のPCでweb上のチャットシステムにアクセスしてもらい、メンタリングの題目として自らのObjectiveをフォームに入力してもらった。Objective入力後、通信テストとして、メンタリング開始を示すテキストを送信してもらった。なお、テキストは事前にフォームに入力されている全参加者共通のテキストを送信してもらった。通信テスト実施後、実験者の開始の合図で一斉に15分間のメンタリングを行ってもらった。15分後、途中でであってもメンタリングをその場で終了してもらい、メンタリング後アンケートに回答してもらった。なお、実験終了後、複数のシステムで実施した旨を説明した後に参加者全員が両条件のシステムでメンタリングできるように機会を提供した。

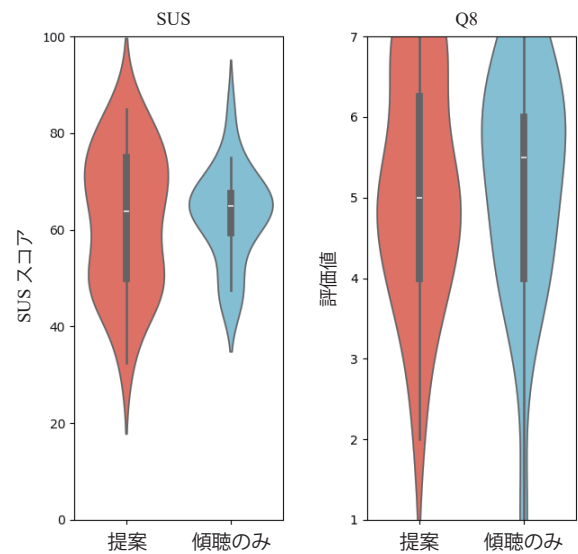


図 4: SUS スコアと Q8 の評価

4.6 結果

実験に参加した78名のうち、システムの動作不備等の理由で無効となった26名のデータを除いた、52名のデータに対して分析を行った。アンケートの質問項目の内、メンタリングの前後で共通する質問項目について結果を図3に示す。メンタリング経験要因とシステム要因が与える影響を分析するために、整列ランク変換を施した二元配置分散分析 (Aligned Rank Transformed Two-Way ANOVA) を行った。分析の結果、Q4にのみ交互

作用が見られた ($F(1, 50) = 7.46, p < .01, \eta_p^2 = 0.13$)。そこで事後検定として、ウィルコクソンの符号順位検定を用いて単純主効果の検定を行ったところ、提案条件において、使用前後で有意に値が増加していることが示された ($p < .01, r = 0.53$)。

交互作用が示されなかった項目の内、Q3については、システム要因間では有意差は見られなかったものの ($F(1, 50) = 0.39, p = 0.54, \eta_p^2 = 0.01$)、メンタリング経験要因で有意差が得られた ($F(1, 50) = 9.74, p < .01, \eta_p^2 = 0.16$)。このことから、両条件ともにメンタリ

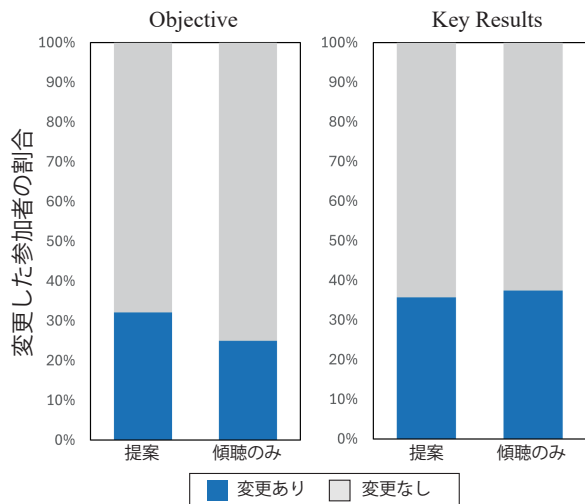


図 5: Objective と Key Results を変更した参加者の割合

ングを経験することで Objective に対する思考が深まったという印象があったことが分かる。なお、Q5 と Q6 については両要因ともに有意差は見られなかった。次に、メンタリング後のみ質問した項目 (Q8) と SUS の結果を図 4 に示す。各項目について、ウィルコクソンの符号順位検定を用いて検定を行った所、両項目共に有意差は見られなかった (Q8: $p = 0.75, r = 0.05$, SUS: $p = 0.99, r < .01$)。

また、Objective と Key Results の変更の有無についても分析を行った。図 5 に各条件において Objective と Key Results を変更した参加者の割合を示す。Objective の変更と Key Results の変更についてそれぞれカイ 2 乗検定を行った結果、Objective の変更 ($\chi^2(1) = 0.32; p = 0.57; \phi = 0.07$) と Key Results の変更 ($\chi^2(1) = 0.01; p = 0.89; \phi = 0.01$) の両方で有意差は見られなかった。

5 考察

アンケートの分析の結果、提案システムとのメンタリングによって傾聴のみ条件に比べて Objective に対する自信を向上させていることが分かった。ここで、実験参加者のコメントを見ると「自分の目標が明確になった気がした」「抽象的な言葉の具体化ができたと感じました。」というような自らの問題設定に対する理解や言語化が明確になったことを言及するコメントが見られた。このことから、問題設定に対する理解が深まったことが、自らの問題設定に対する自信の向上に繋がった可能性がある。

しかし、他のアンケート項目や Objective と Key Re-

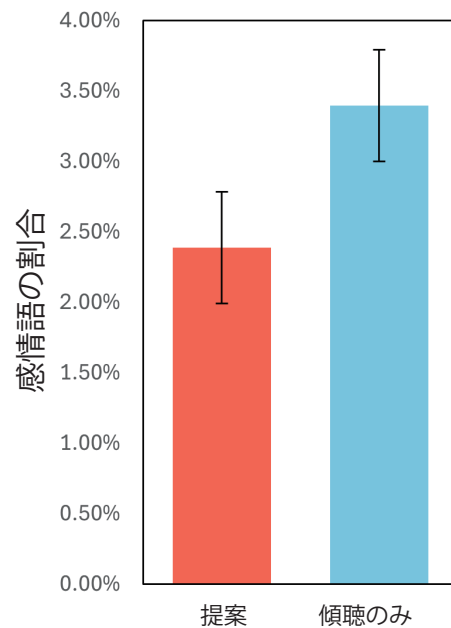


図 6: 発話における感情語が占める割合

sults の変更の有無については両条件間で有意差は見られなかった。一方で、Q3 において両条件共にメンタリング後に有意に値が向上していることから、両条件ともにメンタリングによって学習者の思考を促した印象を与えていた。このことから、両条件共に学習者の思考を促すことにはある程度寄与していたと考えられる。

ここで、傾聴のみ条件のコメントを分析すると、「自分がその objective を立てた理由に気がつけた気がした」といったように問題設定の動機について言及するコメントが見られた。これは、学習者の問題に対する主観的な発話を促すことで、自らの問題に対する動機を明らかにすることができたためと考えられる。また、追加分析として参加者の対話文に含まれる感情語の割合を調査した。割合は発話文を分かち書きした単語の内、感情語に分類される単語の割合を示す。なお、感情語の分類に長岡技術科学大学自然言語処理研究室の公開する日本語感情表現辞書 [8] を用いた。分析の結果、参加者の発話文に感情語が含まれている割合について条件間で有意傾向 ($(t(50) = 1.79; p = 0.08; d = -0.50)$) が見られ、傾聴のみ条件の方が感情語が含まれる割合が多かった (図 6)。このことから、傾聴のみ条件の方が参加者自らの感情を吐露するような主観的な発話が多かった可能性が示唆された。

つまり、提案条件では問いのデザインに基づいた質問によって問題設定に対する理解や言語化が促されていたのに対して、傾聴のみ条件では設定した問題に対する自身の想いに関する発話を促されることで、自らの動機という観点で思考が促されていた可能性がある。

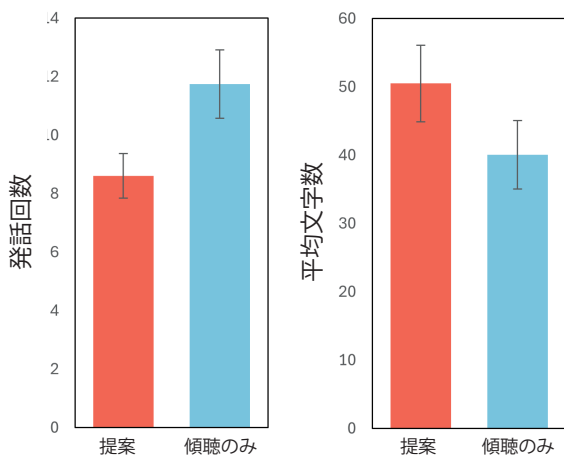


図 7: 各条件ごとの発話回数と平均文字数

そのため今後は、それぞれの対話戦略による思考の深め方の違いについてより詳細に分析していく必要がある。

また、提案条件では先述の通り自らの問題設定に対する言語化が進んだことに言及するコメントがある一方で、「質問攻めで自らの整理にはなったが不足点や改善案が分からなかった」といった質問への不満感を示すコメントが散見された。これは、問いのデザインモードでの質問の回答に対する負荷が高く、十分に回答できなかつたと感じた際に不快感を与えていた可能性を示唆している。ここで、追加分析として発話回数と平均文字数について比較を行った(図7)。すると、提案条件に比べて傾聴のみ条件の方が有意に発話回数が多かった($t(50) = -2.31; p = 0.03; d = -0.64$)。また、ユーザの平均文字数については有意差は見られなかったが($t(50) = 1.37; p = 0.18; d = -0.38$)、提案条件の方が平均文字数が多い傾向にあった。これらのことから、傾聴のみ条件の方が回答に対する負荷が低かつた可能性が考えられる。これは、問いのデザインに基づく質問の性質である可能性もあるが、エンゲージメント推定モジュールの推定精度が十分でなかつた可能性や、発話意欲以外にも学習者の状態として考慮すべき要素がある可能性もあるため、今後は推定精度や発話意欲以外の考慮すべき状態についても検討していく必要がある。

6 おわりに

本研究では、探究学習における教育者の負担を軽減するために、学習者を指導するためのチャットシステムを開発した。提案システムは、大規模言語モデルを用いて発話履歴から学習者のエンゲージメントを推定し、推定結果に基づいて対話を選択する。対話戦略には、問いのデザインに基づいて質問する戦略と、学習者の意見に

共感して積極的に傾聴する戦略の2つの戦略を用いた。提案システムの有効性を検証するために、探究学習に取り組む学生を対象に実験を行った。結果として、提案システムが学習者の問題設定に対する自信を高めるのに役立つことを示した。一方で、比較条件も問いのデザインとは異なる視点から、問題設定に対する思考を促せていた可能性が示された。したがって今後の研究では、問題設定に対して思考を深める際の観点について整理し、各対話戦略がそれぞれの観点での思考の深化にどのように貢献するかについて、より詳細に分析する必要がある。

参考文献

- [1] John Brooke. Sus: A quick and dirty usability scale. *Usability Evaluation in Industry*, Vol. 189, , 11 1995.
- [2] Shanti Divaharan Hui Yong Tay Farhan Ali, Doris Choy and Wenli Chen. Supporting self-directed learning and self-assessment using teachergaia, a generative ai chatbot application: Learning approaches and prompt engineering. *Learning: Research and Practice*, Vol. 9, No. 2, pp. 135–147, 2023.
- [3] Ayu Iida, Kohei Okuoka, Satoko Fukuda, Takashi Omori, Ryoichi Nakashima, and Masahiko Osawa. 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*, p. 260–269, New York, NY, USA, 2024. Association for Computing Machinery.
- [4] Qing Li, Lynn Moorman, and Patti Dyjur. Inquiry-based learning and e-mentoring via videoconference: a study of mathematics and science learning of canadian rural students. *Educational Technology Research and Development*, Vol. 58, No. 6, pp. 729–753, Dec 2010.
- [5] Trevor MacKenzie. *Dive into Inquiry: Amplify Learning and Empower Student Voice*. EdTechTeam Press, 2016.
- [6] OpenAI. Hello gpt-4o, 2024.
- [7] 安齋勇樹, 小田裕和. 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, 2021.

- [8] 長岡技術科学大学自然言語処理研究室. 長岡技術科学大学 自然言語処理研究室:snow d18:日本語感情表現辞書.
- [9] 大美浪海晟, 佐藤匠, 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. 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. HAI シンポジウム 2024, pp. P-52, 2024.
- [10] 金井達亮. 探究学習に初めて取り組む高校教師の経験—指導への意識の問い直しについての語りに着目して—. 教師学研究, Vol. 24, No. 2, pp. 39-48, 2021.
- [11] 日本教職員組合. 2021 学校現場の働き方改革に関する意識調査 | 日本教職員組合 (jtu) , 2022.
- [12] 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解説 (平成 30 年告示) , 2018.