

数学における手書き解答データを利用した 対話型教訓帰納支援システムの開発と評価

Development and Evaluation of an Interactive Lesson Induction Support System Using Handwritten Answer Data in Mathematics

山口 夢叶, 高木 正則
Yuna YAMAGUCHI, Masanori TAKAGI
電気通信大学 情報理工学域

School of Informatics and Engineering, The University of Electro-Communications
Email: y2210662@gl.cc.uec.ac.jp

あらまし：誤答から教訓を導き出す「教訓帰納」は効果的な学習方略であるが、適切な教訓を言語化することは容易ではない。本研究では、LLM との段階的な対話を通じて、質の高い教訓導出を支援するシステムを開発した。中学生を対象とした実証実験の結果、対話完遂群において教訓の質向上が示唆された。
キーワード：教訓帰納, LLM, 振り返り支援システム

1. はじめに

問題解決において、間違いからの学習は有効である。「教訓帰納」とは間違いの振り返りから得られた学びを「教訓」として言語化することで、その後の学習に活かす学習方略である⁽¹⁾。柴⁽²⁾は教訓の質を5段階に分類しており、教訓の質と数学の学業成績との間に正の相関があることを示した。一方で、LLMを数学学習の支援に活用する研究も進んでおり、対話による問題解決支援の有効性が示されている⁽³⁾。しかし、これらの研究は問題解決過程の支援が中心であり、間違いからの学習を支援する研究は少ない。

本研究では、数学学習における学習者の導出する教訓の質向上を目的として、手書き解答データを利用した対話型教訓帰納支援システムを開発し、実証実験を行う。

2. 関連研究

柴⁽⁴⁾は、学習者が問題解決後に有効な教訓を引き出すまでの思考過程を整理した。この過程では、まずどこで間違えたのかを特定し、自分の思考過程を振り返りながら正解についての理解を行う。そして、間違いが生じた原因を分析し、不足していた知識や方略を確認する。最後に、次の問題解決に活用できる教訓を導出する。しかし、学習者がこれらの思考過程を自発的に進めることは困難であることが示唆されている。一方、数学における学習者の手書き解答画像を読み取って学習支援を行う研究も進められている。Nguyenら⁽⁵⁾は、手書きの数学解答を評価する Vision-Language Model である VEHME を提案した。しかし、これらの研究は解答の評価・採点に焦点を当てており、間違いからの学習を対話的に支援するシステムは提案されていない。

これらの課題に対し本提案システムは、LLM を用いて学習者の手書き解答画像を認識し、教訓導出に至る思考過程を段階的な対話によって支援する点に特徴がある。以上を踏まえ、本研究では以下のリサ

ーチクエスチョンを設定した。

RQ1：LLM は学習者の手書き解答を正確に認識できるか。

RQ2：提案システムの利用によって、振り返り記述の質は向上するか。

3. システムの開発

本システムは、学習者が問題演習で間違えた際に、LLM (Gemini 2.5 Flash) との対話を通じて段階的に振り返りを行い、教訓の導出を支援する Web アプリケーションである。実装には TypeScript を用い、デ

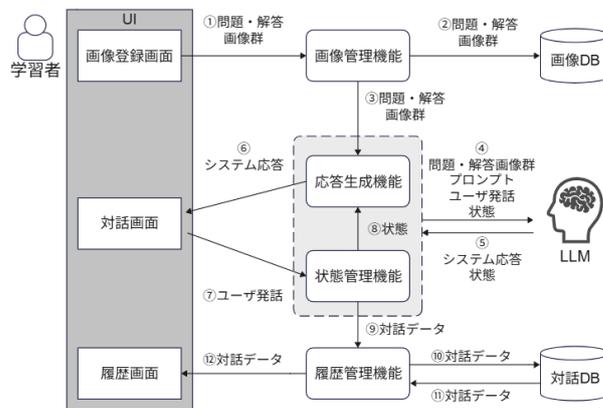


図1 システム構成図



図2 システム対話画面

表 1 振り返りスコアの群内変化

	早期終了群				完遂群			
	N	1日目 M(SD)	2日目 M(SD)	p	N	1日目 M(SD)	2日目 M(SD)	p
間違い箇所の抽出	11	1.45(0.78)	1.73(0.45)	0.250	11	1.82(0.57)	1.91(0.29)	0.500
間違いの原因分析	13	0.85(0.77)	0.54(0.63)	0.984	13	0.85(0.86)	1.08(0.73)	0.309
教訓の質	13	1.23(0.89)	1.23(0.80)	0.625	13	1.46(1.08)	2.38(1.55)	0.031

表 2 対話状態と遷移条件

状態	遷移条件
1) 表層分析	間違い箇所の特定
2) 考え方の比較	考え方の違いの説明
3) 間違いの原因分析	不足知識・方略の特定
4) 教訓導出	教訓の言語化

表 3 手書き文字認識精度の評価結果

指標	値 (%)
数式文字一致率	650 / 680 (95.6%)
非数式文字一致率	231 / 252 (91.7%)

データベースには Firebase を採用した。図 1 にシステム構成図を示す。学習者は演習問題に解答後、間違えた問題の問題文・自身の解答・模範解答、の 3 つの画像（以後、問題・解答画像群）を登録する。登録した問題・解答画像群は画像管理機能を通じてデータベースに保存され、対話開始時に LLM へ送信される。対話中は応答生成機能が LLM を呼び出してシステム応答を生成し、状態管理機能が 4 つのステップ間の遷移を制御する。全ての対話データは履歴管理機能によって保存され、学習者は履歴画面から過去の振り返りを参照できる。実際の対話画面を図 2 に示す。対話設計は柴が提案する有効な教訓を引き出すまでの思考過程⁽⁴⁾を参考にしつつ、システムによる対話が冗長になることを防ぐために 4 つの対話状態に再構成した。表 2 に対話状態と遷移条件を示す。LLM には各状態の目標と遷移条件等を記述したプロンプトを与え、学習者の発話に応じた応答生成と状態遷移の判定を行わせる。

4. 実験と評価

4.1 実験概要

都内私立中学校の 3 年生 63 名を対象に実験を行った。科目は数学、単元は二次不等式の文章題(全 7 問)で行った。1 日目の初めに問題演習を実施した。授業時間の都合ですべての問題を解くことができなかつたため、生徒自身に正解することが容易ではないと思う問題を選択させ、時間内で演習を行わせた。次に模範解答を見ながら、間違えた問題 1,2 問について振り返りを記入させた。そして、振り返りを行った問題について事後テストを実施した。2 日目は、前日に振り返りを行った問題について提案システムと対話を行わせた。対話後に、1 日目と同じ形式で振り返りを記入させた。そして再度、事後テストを実施した。振り返りでは表 2 の 1),3),4)にそれぞれ対応する、間違い箇所の抽出、間違いの原因分析、教訓導出を記入させた。間違い箇所の抽出および間違いの原因分析については、誤りや不明瞭な表現であれば 0 点、部分的に抽出・分析ができていれば 1 点、適切に抽出・分析ができていれば 2 点で評価した。教訓の質は柴⁽⁴⁾の 5 段階尺度で評価した。さらに LLM による、手書き解答の認識精度を検証するため、

生徒の解答用紙から 20 件を無作為に抽出し、評価実験を別途行った。評価では、著者が各解答用紙を目視で書き起こした結果を正解とし、LLM の認識結果と比較した。評価指標として、数式文字と非数式文字それぞれの文字の一致率を算出した。

4.2 結果

LLM の手書き認識精度の評価結果を表 3 に示す。これにより、LLM は高い精度で手書き解答を認識できることが示唆された(RQ1)。また、参加者を対話の達成度により、完遂群（表 2 における教訓導出まで到達）と早期終了群（表 2 における考え方の比較以前で終了）に分類し、各群における振り返り結果を表 1 に示す。間違い箇所の抽出については、スコアの変化は少なかった。間違いの原因分析については、完遂群でわずかに上昇した。教訓の質については、完遂群で有意に向上した (RQ2)。

5. まとめ

本研究では、手書き解答データを用いた対話型教訓帰納支援システムを開発し、中学 3 年生を対象とした実験により評価を行った。実験の結果、対話を完遂した群では、教訓の質の向上傾向が確認された。今後は対話ログを分析し、対話設計の改善を行う予定である。

参考文献

- (1) 市川伸一. 実践的認知研究としての「認知カウンセリング」. 箱田裕司(編), 認知科学のフロンティア I, pp. 134-163. サイエンス社, 1991
- (2) 柴里実. 数学的問題解決後の振り返り記述をいかに評価するか—失敗から引き出された教訓の質を捉える新たな基準の提案—. 教育心理学研究, Vol. 70, No. 3, pp. 231-245, 9 2022.
- (3) Yuyang Ding, et al. Socraticllm: Towards socratic teaching with large language models. In Proceedings of the AAAI Conference on Artificial Intelligence, 2024.
- (4) 柴里実, 植阪友理. 問題解決の間違いからの学習を促す「教訓帰納」の利用における思考のつまずきの解明と支援. 日本テスト学会誌, Vol. 20, No. 1, pp. 91-109, 6 2024.
- (5) Thu Phuong Nguyen, Duc M. Nguyen, Hyotaek Jeon, Hyunwook Lee, Hyunmin Song, Sungahn Ko, and Tachwan Kim. Vehme: A vision-language model for evaluating handwritten mathematics expressions, 2025.