

生成 AI と解答確信度を用いた適応型問題生成システムの構築

Development of an Adaptive Question Generation System Using Generative AI and Answer Confidence

中澤 佑午^{*1}, 小島 篤博^{*2}
Yugo NAKAZAWA^{*1}, Atsuhiko KOJIMA^{*2}

^{*1}大阪公立大学 現代システム科学域

^{*1} College of Sustainable System Sciences, Osaka Metropolitan University

^{*2}大阪公立大学大学院 情報学研究科

^{*2}Graduate School of Informatics, Osaka Metropolitan University

Email: so22384i@st.omu.ac.jp

あらまし：従来の e ラーニングにおける小テストは正誤のみで評価するため、当て推量を見抜けない課題がある。そこで本研究では、生成 AI と解答確信度を用いた適応型問題生成システムを提案する。本システムは教材から問題を自動生成し、学習者の正誤と確信度に基づいて弱点スコアを更新、重点的な復習を促す。実験の結果、確信度を考慮した群は自信なしの正解から自信ありの正解への移行が有意に多く、知識定着における有効性が示唆された。

キーワード：生成 AI, 確信度, 適応型学習, 問題生成, 学習支援システム

1. はじめに

学習者の弱点に応じて問題を調整する適応型学習支援が重要である。しかしながら、従来の e ラーニングでは解答の正誤のみで理解度を評価するため、当て推量による正解や思い違いによる誤答を見抜くのが難しい⁽¹⁾。一方、近年の大規模言語モデル（生成 AI）の発展により、教材から問題を自動生成する学習支援の可能性が広がっている⁽²⁾。そこで本研究では、生成 AI と解答確信度を用いて学習者ごとに弱点を補強する適応型問題生成システムを構築する。

2. システム概要

2.1 システム構成

本研究で提案するシステムは、図 1 に示すように、教材資料（PDF 等）を根拠として生成 AI が多肢選択式問題を自動生成し、学習者の解答履歴に基づいて次に出題する問題内容を更新する適応型学習支援システムである。実装には Google スプレッドシートおよび Google Apps Script（GAS）を用い、システム全体を Web アプリとして提供する。

教材を PDF 形式で Google Drive 上に保存し、GAS から生成 AI（Gemini API）へ教材ファイルの URI と出題指示を送信することで、教材内容に沿った四択の多肢選択式問題（問題文、選択肢、正答、解説）を生成する。生成された各問題には、対応する学習分野を表すタグ（例えば“演算子”、“if 文”など）を 1 つ付与し、学習者の理解状況をタグ単位で管理する。

学習者は Web 画面上で問題に解答するとともに、各解答に対して解答確信度（自信あり／なしの二段階）を入力する。解答結果および確信度は Google スプレッドシートに記録され、後述する出題内容の更新に用いられる。

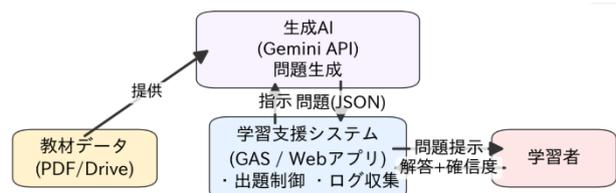


図 1 システム概要図

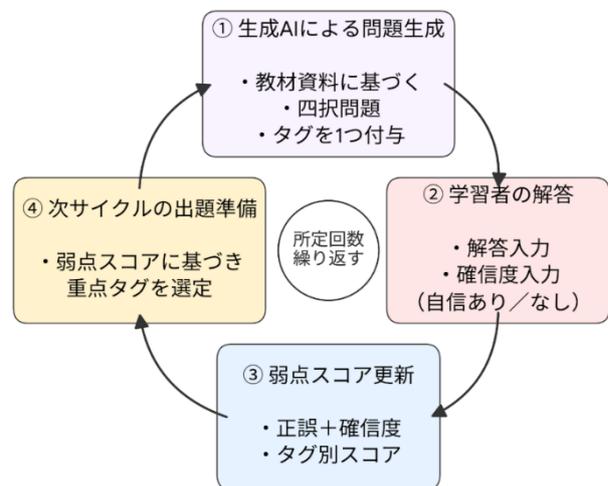


図 2 出題とスコアの更新

2.2 システムの動作

本システムでは、複数の問題からなる 1 回分の出題を 1 セットとし、事前に指定した回数だけ繰り返す。初回の出題セットでは、全学習者に共通の固定問題を提示する。

各セットにおける処理の流れを図 2 に示す。まず、生成 AI が教材資料に基づいて四択問題を生成し、各問題にはタグを 1 つ付与する。次に、学習者が問題に解答し、あわせて解答確信度を入力する。シス

テムは、解答の正誤と確信度の組合せに基づいてタグ別の弱点スコアを更新する。弱点スコアは、タグごとの理解度の低さを数値化した指標であり、数値が高いほどその分野を苦手としていることを示す。

更新後の弱点スコアを用いて、次の出題セットで重点的に出題すべきタグを選定し、それらに対応する類題を同一形式で生成・出題する。この手順を繰り返すことで、学習者の弱点に沿った復習を段階的に提示することを狙う。

3. 実験

提案手法の効果を検証するため、弱点スコア更新ロジックの差異によりグループ分けした実験を行った。

3.1 実験方法

実験の概要を図3に示す。実験の参加者15名を、IDの奇数をA群(n=8)、偶数をB群(n=7)として割り当てた。グループAは確信度を考慮して弱点スコアを更新する提案システムを使用し、グループBも確信度の入力を行うが、システムはそれを無視して正誤のみでスコア更新を行う。両群における実験手順は共通であり、全参加者が事前テスト、3セットの学習セッション、事後テストを実施する。

また、比較を公正にするため事前テストと事後テストの問題は同一である。

3.2 実験結果

群間比較には Welch の t 検定を用い、p 値が 0.05 未満の場合を有意と判定した。

表1に示すようにどちらのグループもテストの正答率は15ポイント以上伸びた。一方正答率の群間比較では、事前テスト・事後テストともに有意差は認められなかった。正答数の伸び(事後-事前)についても群間差は有意ではなかった。

一方、確信度を含む遷移指標として「事前で自信なしの正解だった設問が、事後で自信ありの正解へ遷移した数(同一設問対応)」を比較したところ、A群がB群より有意に大きかった。さらに自信ありの正解数の増加(事後-事前)も p が 0.0603 であり A群が大きい傾向を示した。

3.3 考察

本結果より、確信度を弱点スコア更新に取り入れる設計(A群)は、正解ではあるが自信がない状態を、同一設問で自信ありの正解へ移行させる点で効果が示唆された。

特に、同一設問対応の遷移数(自信なしの正解→自信ありの正解)が群間で有意差が見られた。これは、正答率のみでは把握しにくい理解の安定度を、確信度情報によって補足し、出題選定に反映するという本研究の狙いと整合する。

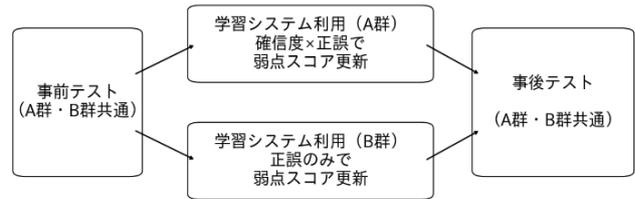


図3 実験の概要

表1 テストの平均正答率

正答率	事前テスト	事後テスト	事前と事後の差
A	0.725	0.875	0.150
B	0.600	0.857	0.257

一方で、学習中に出题されたタグ数の平均はA群2.9、B群2.3であり、A群の方が扱われたタグの幅が広がった。これは、確信度を用いたスコア更新が弱点タグの選定に影響し、出題内容の分布にも差を生じさせた可能性を示す。したがって、正解の確信が高まった改善は、スコア更新そのものに加えて、出題されるタグの多様性の違いも関与している可能性がある。

また、正答率や正答数の伸びに差が見られなかった理由として、事後テストの正答率が高水準に近づき差が出にくくなる可能性、A群では誤答だけでなく正解だが自信なしも弱点として扱うため、短い介入では得点上昇より先に正解の確信が高まる改善が現れた可能性が考えられる。

4. まとめ

本研究では、生成AIと解答確信度を用いた適応型問題生成システムを提案し、その設計と実験による評価を行った。正誤と確信度を組み合わせることで弱点をより精緻に推定し、学習者ごとに最適化された復習問題を出題することで、効率的な知識定着とメタ認知能力の向上を図る。

今後は、学習成果・自己評価精度・利用者体験の観点から有効性を検証することや、Moodleへの移植、また、スコア更新アルゴリズムのさらなる洗練や、復習の多様性確保といった設計上の課題についても検討を進める予定である。

参考文献

- (1) 鄭健俊, 泉隆: “情報技術学習支援システムによる学習者の評価—確信度を追加した学習者の評価の検討—”, 2012年度第28回ファジィシステムシンポジウム, pp1006-1007
- (2) 岩澤孝徳, 久保田真一郎, 喜多敏博, マジュンダール・リトジツ: “生成AIを活用した短答質問作成および評価のための教師向けシステムの設計”, 日本教育工学会 2025年春季全国大会講演論文集, pp149-150