

大規模言語モデルとルールベース処理を用いた 4 択問題集作成支援ツール

An Authoring Tool for 4-Option Multiple Choice Question Sets Using Large Language Models and Rule-Based Processing

小島 一秀[†] 安田 光輝[‡]
Kazuhide Kojima Koki Yasuda

1. はじめに

現在では、e ラーニングが急速に普及しているが、そこで使用される問題集は、多くが手作業で作成されており、作成の効率化の余地が残されている。このような状況を改善するため、これまでの本研究では、指導者が出題する内容を用語、その上位用語、用語説明文からなる表形式データを入力するだけで、e ラーニングでしばしば用いられる 4 択問題の問題集を自動生成するルールベース処理による作問支援ツール（既存ツール）の開発を行ってきた^[1]。これを改善するために、近年、発展の著しい大規模言語モデル（LLM）を組み合わせた作問支援ツールを提案する。

2. 先行研究

自動作問の先行研究としては、与えられたテキストデータから重要度が高いと推定された文において、その一部を空欄とするツール^[2]や、手作業でタグ付けを行ったテキストデータから穴埋め問題作るツール^[3]などが挙げられる。

これらのツールの課題としては、プログラム処理可能な電子化データなど必須であることや、問題形式が穴埋め問題に限られることなど、あまり実用的ではない。

3. 既存ツールの機能と問題

これまで、本研究では、指導者が出題したい用語やその説明などを Microsoft Excel 上に用意された表形式データを入力すると、用語に対する説明を選ぶ（図 1 内「問題集データ」と同形式）またはその逆の用語問題、用語が示すものの特徴を選ぶ特徴問題、用語に含まれる用語を選ぶグループ問題、正しい計算結果を選ぶ計算問題^[4]など 6 種類の問題集を生成する作問支援ツールを開発してきた^[1]。

ここでは、既存ツールで最も多く生成される問題種類である用語問題、特に用語から説明を選ぶ用語問題の生成を用いて既存ツールの動作について述べる。

既存ツールは、手入力された表形式データを用いて、問いを生成する。用語問題の生成で使われる表形式データは、用語、その用語が所属する上位用語、その用語を簡潔に説明する用語説明文の 3 項目からなり、形式的には図 1 中央の「表形式データ」と同じである。

まず、既存ツールは、出題対象の用語（出題用語）を 1 つ選び、用語にテンプレートに当てはめて問題文を生成し、その説明文を正解選択肢とする。次に、表形式データ内の上位用語、さらに用語自体の末尾を用いて、用語同士の階層構造を示す用語階層木を生成する。最後に、用語階層木

を用いて、出題用語と意味的に近く、抽象度合いも近い他の用語の用語説明文を、不正解の選択肢（偽選択肢）の候補として採用する。用語説明文は、1 用語に対して複数持つこともできるが、自然な選択肢となるように、これらの複数の用語説明文を接続することができる。また、偽選択肢の候補が不足している場合には、異なる用語の用語説明文を接続して、新しく偽選択肢を作成することもできる。

既存ツールは、手作業と比べて、より短時間で多くの問いを生成可能である。しかしながら、指導者は教科書の内容を要約しながら手作業で説明文を入力する必要があり、この手間が課題となっている^[1]。

4. LLM による問題集生成の問題

LLM の一つである Gemini (gemini-1.5-flash-001)^[5]を用いて前実験を行った結果、以下のようなことが判明している。単純に「<単語>について 4 択問題を作成せよ」のように、出題範囲を単語のみで記述したプロンプトを与えると、出題範囲を正確に指定できないだけでなく、誤った問題が生成される場合が少なくない。これを改善しようと、出題範囲の文書を加えたプロンプトを与えても、十分な問題数は生成されず、もっと出力するようにプロンプトを与えても、関係ない問いや重複した問いが出力された。

5. LLM を取り入れた 4 択問題集作成支援ツール

提案ツールは、既存ツールの 6 種類の問題生成機能のうち、用語問題、特徴問題、グループ問題の 3 種類の生成機能について改良されている。ここでは用語「BMP」に対して適切な説明を選択する用語問題を生成する例を用いて説明する。

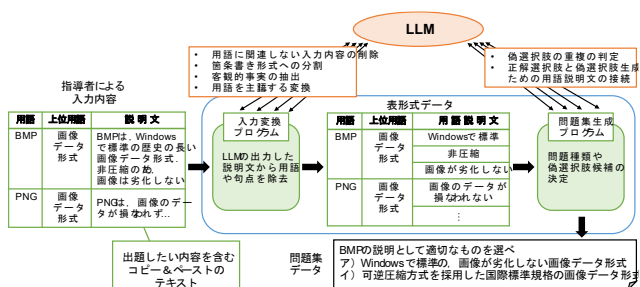


図 1: 提案ツールの構造と処理の概要

提案ツールは、入力変換プログラムと問題集生成プログラムからなる。LLM は基本的に gemini-1.0-pro-002^[5]を使用している。入力変換プログラムは、既存ツールに新たに追加されたプログラムであり、指導者がコピーした説明文を、LLM を用いて自動的に表形式データの用語説明文に変換する。用語や上位用語は手入力である。また、入力変換プ

[†] 大阪大学 D3 センター D3 Center, The University of Osaka

[‡] パナソニックオートモーティブシステムズ株式会社
Panasonic Automotive Systems Co., Ltd.

プログラムによる用語説明文に問題があれば、指導者が修正することも可能である。

入力変換プログラムは、ルールベース処理で実現しにくいために LLM を採用した処理 A1 から A5 と、処理内容が機械的であるため確実性の高いルールベース処理を用いた処理 A6 から構成される。処理 A6 で得られた結果が表形式データに格納される。図 1 の例では、指導者によってコピーされた用語「BMP」と「PNG」の説明文が、表形式データの用語説明文として変換されている。

- 処理 A1) 用語に関連する内容の選択
- 処理 A2) 説明文の分割
- 処理 A3) 主観的な内容の除去
- 処理 A4) 文が長い文の追加の分割
- 処理 A5) 各文を「<用語>は、…」形式に変形
- 処理 A6) 各文から用語や句点を除去

問題集生成プログラムの動作は以下のようにになっている。問題文生成は既存ツールと同様である。既存ツールと同様に用語階層木を用いて偽選択肢の候補を収集した後、以下の処理 B1 から B4 を用いて図 1 右下「イ）」のような自然な長さの偽選択肢が生成される。

- 処理 B1) 内容重複の削除
- 処理 B2) 固有表現を用いたランク付け
- 処理 B3) 用語説明文をルールベース処理と LLM の 2 通りで接続
- 処理 B4) 処理 B3 の 2 通りのうち自然な方を選択

処理 B3 の片方の処理以外は LLM による。処理 B1 は、単純に複数の用語説明文とプロンプトを与えるだけでは、誤りがしばしば発生するため、用語説明文のペアを全て作成し、各ペアを高度な処理に制約があるが処理速度の速い gemini-1.5-flash-001^[5]により同じかどうかを判定させている。

以上のように、提案ツールでは、問題作成処理を段階に分けることにより、LLM に対する指示を単純化し、高度な自然言語処理という LLM の特徴を、誤りを抑制しながら既存ツールに取り入れている。

6. 評価

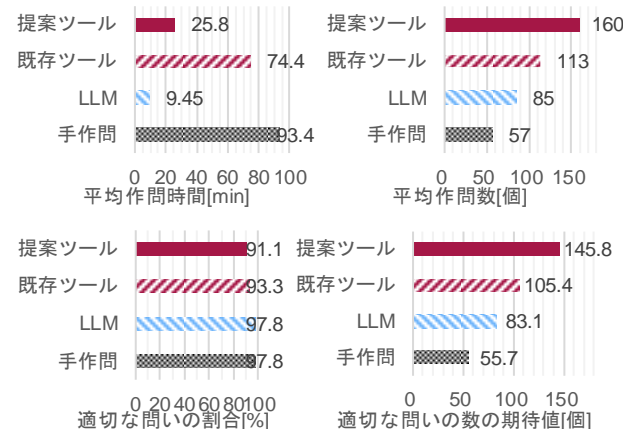


図 2: 評価結果

提案ツールの評価のために、基本情報技術者試験の教科書^[6]から 4 ページの出題範囲を用いて、提案ツール、既存ツール、LLM、手作問により問題集を作成した。被験者はそれぞれの方式で 2 名であるが、作成者による違いがでにくい LLM では 1 名のみとしている。

問題作成時間は、被験者が入力などの操作を行うのにかかった時間であり、提案ツール、既存ツール、LLM が動作している時間は含まれない。LLM は問いの重複が無視できないためそれらを除く作業が含まれている。各手法の問題作成時間は、図 2 左上の通りで、LLM が 93.4[min]と手作問の 10 分の 1 と最も短く、提案ツールが 25.8[min]と手作問の 4 分の 1 と 2 番目に短い。また、提案ツールは、既存ツールの 74.4[min]の 3 分の 1 と入力変換プログラムの効果が現れている。

平均作問数は、提案ツールが 160 個と最も多く既存ツールからは 1.5 倍と増加しており、LLM の 2 倍、手作問の 3 倍となっている (図 2 右上)。

問いの品質を評価するために、各種法で生成された問いから 45 問をサンプリングし、問題文、正解選択肢、偽選択肢を 3 名の評価者により適切、不適切の二段階の評価を行い、以下のような問いを適切な問いとした。

- 問題文と正解選択肢、少なくとも 1 つの偽選択肢が全評価者から適切と判定
- 偽選択肢に正解がないと全評価者が判定
- 出題範囲内であると全評価者が判定

各手法の適切な問いの割合は、LLM と手作問が 98.7% と高く、提案ツールと既存ツールがそれぞれ 91.1% と 93.1% となり少し低い値となった (図 2 左下)。ただし、これらから、適切な問いの数の期待値を求めても、図 2 右下のように、提案ツールは 145.8 個と、LLM の 1.8 倍、既存ツールの 1.4 倍と最も多いことに変化はなかった。

7. おわりに

4 択問題集を効率的に作成するための作問支援ツールに、高度な自然言語処理が実現可能な LLM を導入し、問題作成時間を短縮しながら、より多くの問題作成を実現した。今後の課題としては、LLM に若干劣る問いの品質の改善や、情報処理技術者試験以外の分野への応用が挙げられる。

参考文献

- [1] 安田 光輝, 神田 将吾, 龍宮寺 嵩士, 小島 一秀, “機械的な処理や用語の階層構造を用いた 4 択問題生成ツールの開発と評価”, 研究報告コンピュータと教育(CE), 2025-CE-178, No. 2, pp. 1-7 (2025).
- [2] N.Afzal, R.Mitkov, “Automatic Generation of Multiple Choice Questions using Dependency-based Semantic Relations”, *Soft Computing*, vol.18, pp.1269-1281 (2014).
- [3] J.Araki, D.Rajagopal, S.Sankaranarayanan, S.Hoim, Y.Yamakawa, T.Mitamura, “Generating Questions and Multiple-Choice Answers using Semantic Analysis of Texts”, *Proceedings of the 2016 the 26th International Conference on Computational Linguistics: Technical Papers*, pp.1125-1136 (2016).
- [4] 小島 一秀, 神田 将吾, “計算問題のための偽選択肢生成方式”, 情報処理学会第 85 回全国大会, 4H-04, pp.435-436 (2023).
- [5] G. Team, et al., “Gemini: A Family of Highly Capable Multimodal Models”, arXiv:2312.11805 (2024).
- [6] 月江伸弘, “徹底攻略基本情報技術者教科書令和 3 年度”, インプレス (2020).