

生成系 AI による文脈情報を 反映した CAT 教材の問題生成に関する研究

A Study on Problem Generation for CAT Teaching Materials Reflecting Contextual Information Using Generative AI

青野 圭汰^{*1}, 荒関 虹希^{*2}, 上野 春毅^{*1}, 小松川 浩^{*1}

Keita AONO^{*1}, Koki ARASEKI^{*2}, Haruki UENO^{*1}, Hiroshi KOMATUGAWA^{*1}

^{*1}公立千歳科学技術大学 理工学部 情報システム工学科

^{*1}Faculty of Science and Technology, Chitose Institute of Science and Technology

^{*2}公立千歳科学技術大学大学院 理工学研究科

^{*2}Graduate School of Science and Technology, Chitose Institute of Science and Technology

Email:b2210010@photon.chitose.ac.jp

あらまし：本研究では、生成系 AI を用いた問題生成において、問題から抽出できることが期待される文脈情報に着目した。生成系 AI によって問題から解法やキーワードといった文脈を抽出し、明示的な情報に加えて生成系 AI に与えることで、文脈が問題生成に与える影響を分析した。

キーワード：生成系 AI, 文脈情報, CAT

1. はじめに

近年では、生成系 AI を用いた教材の自動作成に関する研究が進められており、問題生成の手法に関する研究が行われている⁽¹⁾⁽²⁾。実際の教育現場において、教員が問題を作成する際には、学習者の理解度に応じた適切な難易度を設定することが重要である。問題には、教員が「この問題について何をどのようにどこまでできるようにしたいか」といった出題意図が問題の文脈情報（以下文脈という）として含まれており、ルーブリックで抽象化される。CAT（Computerized Adaptive Testing）教材の設計では、通常ルーブリックが問題の難易度に読み替えて整備されており、難易度やルーブリックは開示される情報として取得可能であるが、問題の文脈は内包的な情報であり、利活用は難しい。

一方で、生成系 AI の進展により、問題文から文脈を抽出できる可能性が近年にかけて生じてきた。この技術を応用すれば、埋もれた問題の文脈を抽出し、教員の意図を反映した問題生成が可能になると考えられる。本研究では、生成系 AI を用いて問題の文脈の抽出を試みる。また、この情報に基づいて問題の生成を行い、生成系 AI による問題生成へ反映することを目指す。

2. 生成系 AI を用いた文脈情報の抽出

生成系 AI によって、問題の単元や問題文といった明示的な情報から、解法の作成や、問題を解くために必要となるような知識をキーワードとして抽出することを試みた。これらの情報は明示的に存在せず、問題の文脈情報として存在している。ここで、生成系 AI を使い、問題の単元の情報と問題文の情報をプロンプトに与え、問題の解法と問題を解くために必要となるような知識をキーワードとして作成した。中学数学/中学数学 1 年/文字と式/文字式の関係と計算の単元の問題 70 問に対し、すべての問題の解説、キーワードを作成させたところ、もとの問題とあわせてみた際に、解法に間違いがないと判定した

ものが全体で 90%、作成したキーワードに不適なものがないと判断したものが全体で 98.5%であった。このことから、今回の手法によって解法とキーワードの抽出が行えたと判断した。

3. 文脈情報を用いた問題の生成

本学で使用される e-ラーニングシステムの問題から、中学数学/中学数学 1 年/文字と式/文字式の関係と計算の問題を例とし、生成系 AI を用いた類題の生成を行った。まず、本研究において各難易度で分けた例文の中には、その難易度で共通する文脈が含まれ、例示によってその文脈を示すことができるという前提を置いた。

本研究では、生成系 AI として ChatGPT を使用した。はじめに、本研究で扱う文脈が問題文の中から取り出せると仮定し、2. で述べた生成系 AI による文脈情報の抽出を行い、問題の解法、問題を解く際に必要となるような知識をキーワードとして抽出した。次に、類似問題を作成するために ChatGPT に渡すためのプロンプトを作成した。プロンプト中に含めた情報は以下の図に示す通りである。

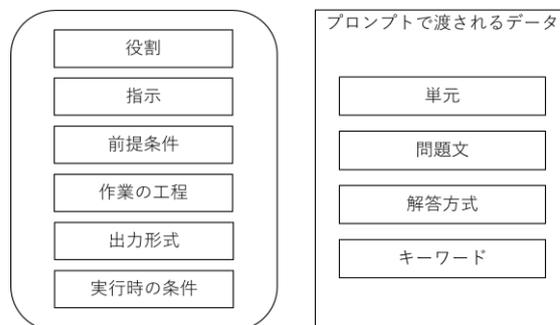


図 1：プロンプトの構造

まず、役割の部分で、GPT に対してさせたい役割を設定する。今回は、中学数学の教師として役割を与えた。次に、指示において GPT にさせたいタスク

を明確化した。今回は、問題生成を行うことを明示した。前提条件では、扱う問題がどのような単元に存在するかと、ループリックがどう定義されるかについて明示した。ループリックは、「基礎的知識の用語・概念の理解」「知識の活用」「知識の応用」の三段階で定義される。作業の工程では、GPTにどのような手順でタスクを行ってほしいかを明示した。最後に、出力形式と実行時の条件として、解答形式に関する制約やしてほしい出力の型を規定している。

作成したプロンプトを使い、実際に生成を行った。作成した問題数は、初級難易度が20問、中級難易度が30問、上級難易度が20問である。問題生成において、プロンプトで渡す情報の条件を変え、生成結果の比較を行った。プロンプトで渡す情報の条件を以下に示す

- Few-shot Learning と文脈情報を使わず、問題文と解答形式のみを参考問題として与えて生成する
- 参考問題以外の情報として Few-shot Learning を使い、生成したい難易度と同じ難易度の問題から問題文と解答形式を例示する
- 参考問題以外の情報として Few-shot Learning を使い、生成したい難易度と同じ難易度の問題から問題文と解答形式、キーワードを例示する

また、問題生成の結果は

- 生成された問題が解け、かつ生成された解答形式や解答に即しているか
- 作成したい問題のループリックとして適当か
- 作成したい問題の単元として適当か

の3点でチェックを行った。

以上の条件で生成を行ったところ、まず初級難易度においては条件1で65%、条件2で60%、条件3で75%であった。

条件1と条件2において、結果が大きく変化せず、条件3において正しく生成できた問題の数が増加したことから、結果に影響を与えているのは Few-shot Learning によるキーワードの例示と考えられる。

次に、中級難易度では、条件1で60%、条件2で76.6%、条件3で63%であった。

中級難易度の問題においては、キーワードの情報よりも、Few-shot でキーワード以外の明示的な情報を与えた方が精度は高くなるのが分かった。

最後に、上級問題では、問題が成立していてもループリックや単元のずれた問題が散見された。条件1では、正しく生成できていた問題が50%、そのうちループリックが合っていたものが50%であった。単元については、条件1においては問題が成立していた中でずれたものはなかった。条件2では、正しく生成できた問題が40%、そのうち25%でループリックが合っていない問題だった。単元については、条件2においても問題が成立していた中でずれたも

のはなかった。条件3では、正しく生成できていた問題が45%、そのうち22.2%でループリックが合っていない問題だった。単元のずれた問題は、正しく生成できていた問題に対して11.1%であった。

今回の検証では、問題の難易度が上がるたびに正しく生成できる問題の割合は下がった。特に上級問題において精度は条件1よりも低くなっている。これは、Few-shot によって参考にできる問題が増えた分、複雑な問題では考慮すべきポイントが増え、結果的にうまく結果に反映できなかったと考えられる。

以下に、実際の生成結果の例を一つ示す。図の表示形式は、上から

- 生成した問題文
 - 生成した問題の解答方式
 - 選択問題の選択肢
 - 解答
- となっている

文字式において、文字に具体的な数を当てはめて計算することを何とよいか。

['dropdown']
['代入', '置換', '変換']
['代入']

図 2: 実際に生成されたもの (初級難易度, 条件 3)

4. まとめ

今回の結果から、問題の文脈を用いた問題生成においては、初級問題の例ではキーワードの情報が影響を与えており、Few-shot を用いて例示を行うことで結果に反映することが可能であるとわかった。

一方、難易度が上の問題では例示の効果がうまく出ないことが分かった。

今後の展望として、文脈の抽出の際に与える情報を増やしてより多くの文脈を得られるかの検証や、問題生成時にプロンプトで渡す情報を変更して精度が向上するかの検証などがあげられる。

参考文献

- (1) 田中 英武, 前田 悠翔, 井垣 宏, Few-Shot Prompting を用いた言語系生成 AI によるプログラミング演習問題の自動生成手法の検討
<https://jssst.or.jp/files/user/taikai/2023/papers/46-R-S.pdf>
- (2) 前田 悠翔, 田中 英武, 井垣 宏, 福安 直樹, 多様な問題パターンおよび難易度を考慮した言語系生成 AI によるプログラミング演習問題自動生成手法の検討
https://www.jstage.jst.go.jp/article/repit/2024/0/2024_88/pdf-char/ja
- (3) 荒関虹希, 大規模言語モデルを用いた CAT 教材の難易度分類アルゴリズムに関する研究
公立千歳科学技術大学修士論文