

ChatGPT と数学の解答の変化

節賀日々輝 鈴木寿弥 池吉陸斗 豊福和奏

⁽¹⁾延岡高等学校 Nobeoka High School

Abstract

ChatGPTは数学が苦手という事実からこの着想を思いつき、この研究では、私たちが普段ChatGPTを使う時により使いやすくすることを目的にして、ChatGPTへの問題の与え方を変化させることによって数学の問題の正答率がどのように変化するかを調べた普通に問題を与える場合と計算方法にpythonを指定する場合とChatGptに数学教師を演じてもらう場合の3つを検証した。全体で見ると、与え方によって正答率に大きな変化があるとは言えなかったが、分野別では与え方による正答率の違いが見られた。今後は試行回数を多くしてより精度を高めていきたい。

Keyword

ChatGPT

1. 序論

(1)研究背景

一般に、ChatGPTは計算問題をことが苦手だと言われていることを知った。そこで、問題の与え方の変化の工夫によって正しい答えを導きやすくなるのではないかと思った。

(2)研究の目的(or動機or意義)

精度に変化がある与え方を知れば、普段の検索などでChatGPTを使うときも、命令の方法を工夫することで高精度な出力を得ることができるのではないだろうか。

(3)先行研究

ChatGPTは計算問題を解けるの？正しく解かせるコツやプラグインの利用方法をご紹介！

<https://bocek.co.jp/media/exercise/chatgpt/8209/>

(4)研究仮説

ChatGPTに数学の教師を演じてもらうことで、回答の過程を詳しく丁寧になるよう促したり、計算方法を数値計算を得意とするPythonに指定することで計算ミスを減らそうとしたりなど、このように問題の与え方を工夫することで、出力の精度を高められ、問題の完答率を高くすることができるのではないか。

2. 調査方法

①学校で全員に配られてる問題集(NEWACTION LEGEND)の問題をChatGPTにそのまま与えて解かせる。

②ChatGPTに特定のプロンプトを用いて数学教師を演じてもらって解答を得る[1]、計算方法をPythonで指定したり[2]などの工夫をし、①との出力の変化を調べる。

③①、②の出力結果のデータを比較し、どの与え方が一番ChatGPTの完答率が高かったかをまとめて、そのような結果となった原因を考察する。

上記の①、②は以下の条件で行う。

◇代数、解析、確率・統計のそれぞれの分野に関する問題を15問ずつ与える。

◇問題の難易度レベルは☆1から☆4の中の☆2(普通レベル)である。

◇有料版ChatGPT4oを使用する。

◇検証方法を変化させるごとに別個体のChatGPTを使用する。

◇単純に表しにくい数式はLaTeX文法で与える。

◇小問同士が関連していない場合は小問番号(2)まで与える。

3. 結果

①そのまま与えた場合の完答率:

代数分野	60%
解析分野	46%
確率分野	66%
合計	57%

②工夫をした場合の完答率:

[1] 数学教師を演じてもらう場合

代数分野	66%
解析分野	53%
確率分野	53%
合計	57%

[2] 計算方法にPythonを指定した場合

代数分野	80%
解析分野	40%
確率分野	46%
合計	55%

4. 考察

代数分野で計算方法にPythonを指定した場合の完答率が最も高かったのは、計算が得意なPythonを用いたことでChatGPTの計算ミスが減ったことが原因と考える。

```
python
# 変数の定義
x = sp.Symbol('x')

# (1) 2(x^4) + x^2 - 3 の因数分解
expr1 = 2*x**4 + x**2 - 3
factored_expr1 = sp.factor(expr1)

# (2) x^4 - 6x^2 + 1 の因数分解
expr2 = x**4 - 6*x**2 + 1
factored_expr2 = sp.factor(expr2)

factored_expr1, factored_expr2

結果
((x - 1)*(x + 1)*(2*x**2 + 3), (x**2 - 2*x - 1)*(x**2 + 2*x - 1))
```

確率分野でそのまま与えた場合の完答率が最も高かった原因は、数学教師を演じてもらう場合では、計算ミスが多く発生し、計算方法にPythonを指定した場合では、間違えた解法のまま、Pythonを用いたことが原因と考えられる。また、与えた問題数が45問と不足していたのが原因だとも考える。数学教師を演じてもらう場合は、数学教師を演じてもらうことで、計算過程がより丁寧になっていた。

$$(a + b + c)(ab + bc + ca) - abc$$

分配すると、

$$a \cdot (ab + bc + ca) + b \cdot (ab + bc + ca) + c \cdot (ab + bc + ca) - abc$$
$$a^2b + abc + aca + ab^2 + b^2c + abc + abc + bc^2 + c^2a - abc$$
$$a^2b + ab^2 + abc + abc + abc - abc + b^2c + bc^2 + c^2a + aca$$

同類項を整理すると、

$$a^2b + ab^2 + b^2c + bc^2 + c^2a + aca + abc$$

ここで、共通因数を見つけると、この式は $(a + b)(b + c)(c + a)$ に因数分解できます。

$$(a + b)(b + c)(c + a)$$

そのまま与えた場合

$$(a + b + c)(ab + bc + ca) - abc$$

展開すると、

$$a \cdot ab + a \cdot bc + a \cdot ca + b \cdot ab + b \cdot bc + b \cdot ca + c \cdot ab + c \cdot bc + c \cdot ca - abc$$
$$a^2b + abc + a^2c + b^2a + b^2c + bc^2 + c^2a + c^2b + abc - abc$$

+abcと-abcが消えるので、整理すると、

$$a^2b + a^2c + b^2a + b^2c + c^2a + c^2b$$

ここで、共通因数を見つけると、

$$ab(a + b) + bc(b + c) + ca(c + a)$$

これが 因数分解の形 です。

答え

$$ab(a + b) + bc(b + c) + ca(c + a)$$

数学教師を演じてもらう場合

5. 結論

合計の完答率からは与え方の違いによって完答率が上がったとは言えない。代数分野に関しては、計算方法にPythonを指定した場合の完答率が最も高かった。完答率が最も高かった原因は、Pythonを用いたことでChatGPTの計算ミスが減ったことが、計算が多い代数分野で効果が大きかったからだと考える。解析分野に関しては、数学教師を演じてもらう場合の完答率が最も高かったが、大きな差はみられなかった。確率分野に関しては、そのまま与えた場合の完答率が最も高かった。

6. 謝辞

本研究の遂行にあたり、指導教官として終始多大なご指導を賜った早田先生と寺崎先生にこの場で感謝申し上げます。

7. 参考文献

前提知識を考慮した数学の確率問題自動解答の精度向上: <https://x.gd/Ec9dJ>