

## 1 はじめに

プログラミング教育において、学習者一人ひとりに適切なフィードバックを提供することは重要であるが、容易ではない。大規模言語モデル (LLM: Large Language Model) を用いた手法 [1] などにより、フィードバック案を生成することは容易になるが、複数案から適切なフィードバックを効率的に選別することは依然として課題である。本研究では、LLM を用いた擬似学習者による適切なフィードバックの選別手法を検証する。また、学習者の能力の違いに応じた選別のために、実際の学習者の能力の違いを異なるパラメータ数のモデルを用いる方法で実現可能か、また、LLM に対して異なるプロンプトを用いる方法で実現可能か検証する。

## 2 LLM 擬似学習者によるフィードバック効率化の検証

LLM とは、膨大なテキストデータを用いて事前学習されたモデルである。LLM をプログラミング初級者に相当する擬似学習者として用いた。LLM 擬似学習者に対し、課題内容と誤ったコード、およびフィードバックを与え、フィードバックをもとに修正したコードを出力させた。異なる 23 個のフィードバックで 50 回行った。各フィードバックにおける修正成功率 (正しく修正できた割合) を表 1 に示す。直接的な答えを示すフィードバックや情報量が多いフィードバックでは修正成功率が高くなり、情報量が少なく曖昧なフィードバックでは修正成功率が低くなる傾向が確認された。LLM 擬似学習者のフィードバック間の修正成功率の違いが実際の初級者のフィードバック間の修正能力の傾向と類似する可能性を示し、LLM 擬似学習者による適切なフィードバック選別の効率化の可能性が示唆された。

表 1: 修正成功率  $r$  (単位: %)

No	r	No	r	No	r	No	r
f1	0	f7	2	f13	42	f19	70
f2	2	f8	32	f14	10	f20	30
f3	46	f9	8	f15	70	f21	42
f4	4	f10	18	f16	36	f22	16
f5	100	f11	54	f17	26	f23	66
f6	22	f12	0	f18	98		

## 3 擬似学習者による異なる能力を持つ学習者の実現の検証

LLM は学習時のパラメータ数や学習手法により性能が異なる。また、プロンプトは LLM への入力文であり内容や表現により出力が変化する。まず、パラメータ数の異なるモデルの修正成功率の違いを比較し

た。パラメータ数別の各フィードバックの修正成功率の分布を図 1 に示す。図 1 で青がパラメータ数が多いモデル、オレンジが少ないモデルである。パラメータ数の多いモデルの方が全体的に修正成功率が高い傾向が確認された。一方で、コード上の誤り箇所と修正方針の把握が容易なくつかのフィードバックにおいて、パラメータ数の少ないモデルの方が修正成功率が高かった。フィードバックの解釈の違いによる可能性が考えられる。以上より、本手法では学習者のフィードバックの活用能力の違いを実現できる可能性が示唆された。また、異なるプロンプトの修正成功率の違いを比較した。プロンプト別の各フィードバックの修正成功率の分布を図 2 に示す。青が 2 節のモデルで、赤のステップバイステップでの推論の指示を追加したプロンプトの方が単純なフィードバックにおいて修正成功率が低くなり、複雑なフィードバックにおいて修正成功率が高くなる傾向があることが確認された。したがって、複雑なフィードバックに対する学習者の能力の違いを実現できる可能性が示唆された。

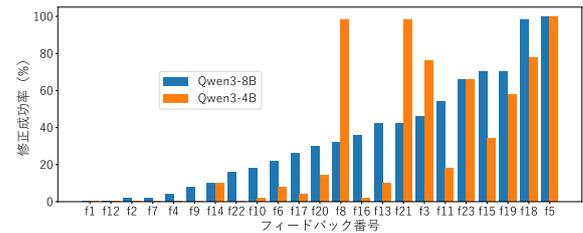


図 1: パラメータ数別の修正成功率の分布

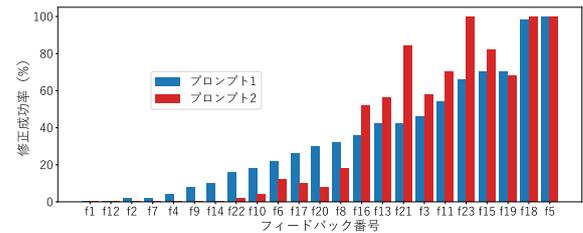


図 2: プロンプト別の修正成功率の分布

## 4 むすび

LLM 擬似学習者により適切なフィードバック選別を効率化できる可能性が示唆された。また、LLM のパラメータ数の違いや与えるプロンプトの違いにより、学習者の能力の違いを実現できる可能性が確認できた。今後の課題としては、他の課題や誤ったコードでの検証を行い、他の条件でも同様の結果となるか確認することが挙げられる。

### 参考文献

- [1] Lianne Roest et al., “Next-Step Hint Generation for Introductory Programming Using Large Language Models”, Proceedings of the 26th Australasian Computing Education Conference, pp.144-153, 2024.