

FT-LLM2026 数学タスク オープン枠 再現手順報告

稲留 隆之
Team005 Lion

概要

本稿では、FT-LLM 主催の「チューニングコンペティション 2026」における数学タスク・オープン枠への参加アプローチについて報告する。運営から提供されたベースモデルに対し、2段階の Supervised Fine-Tuning (SFT) を施すことで、日本の中学・高校レベルの数学問題に対する解答精度向上を試みた。学習データには、評価ベンチマークの傾向を模倣して生成したデータセットと、外部 LLM を用いて生成した大規模な蒸留データセットの2種類を合成して使用した。さらに、推論時には Self-Consistency を適用することで、解答の頑健性を高める工夫を施した。結果として、運営から提供された開発用評価データにおいて 95% の正解率を達成した。本稿では、ベースモデルの選定からデータセット構築、学習戦略、そして推論手法に至るまでの一連の再現手順を詳述する。

1 はじめに

近年、大規模言語モデル (LLM) の能力向上は目覚ましく、様々な分野への応用が期待されている。特に、数学のような論理的推論を必要とする領域においても、その性能が注目されている。

FT-LLM が主催する「チューニングコンペティション 2026」は、LLM のチューニング技術を探求し、その応用事例を開拓することを目的としている。本コンペティションの数学タスクでは、日本の中学・高校レベルの数学問題を対象とし、提供されたベースモデル (llm-jp-4-8b) のファインチューニングによって解答精度を競う。

今回、本コンペティションのオープン枠に参加し、提供された計算資源を活用してモデルの性能向上に取り組んだ。本稿では、最終的に提出したモデルの開発アプローチ、特にデータセット構築と2段階の学習戦略、そして推論時の工夫について、その再現手順を詳細に報告する。

2 モデル学習のアプローチ

最終的に提出したモデルの開発は、「ベースモデル選定」「学習データセット構築」「事後学習 (SFT)」「推論」の4つのフェーズで構成される。全体の戦略として、まず多様な数学問題に対応できる汎用的な能力を大規模データで学習させ、その後、コンペティションの評価形式に特化したチューニングを施すという段階的なアプローチを採用した。

2.1 ベースモデルの選定

運営からは、特性の異なる3つの llm-jp のベースモデルが提供された。

- 事前学習のみを終えたモデル
- 継続事前学習を行ったモデル
- インストラクションチューニングされたモデル

開発用評価ベンチマーク (100 問) に対する各モデルの初期性能を評価した結果を表 1 に示す。

表 1 ベースモデルの初期性能評価

モデル	正解率
事前学習のみ	7%
継続事前学習済み	48%
インストラクションチューニング済み	44%

継続事前学習済みモデルとインストラクションチューニング済みモデルの精度はほぼ同等であった。当初から継続事前学習とファインチューニングの両方を行う方針であったため、追加の継続事前学習が行いやすいと考えられる「2. 継続事前学習を行ったモデル」を以降のチューニングのベースとして選定した。

2.2 学習データセットの構築

オープン枠のライセンス要件を満たしつつ、高品質な学習データを大量に確保するため、特性の異なる2種類のデータセットを構築した。

2.2.1 評価ベンチマーク準拠データセット

コンペティションの評価形式への最適化を目的とし、運営から提供された開発用評価データを参考に、類似の傾向を持つ問題と解答のペアを生成した。ライセンス問題を回避するため、問題作成に LLM は使用していない。設問のバリエーションは手動で複数パターンを用意し、計算部分は SymPy を用いてプログラムで自動生成することで、問題と解答の正確性を担保した。このアプローチにより、コンペティションの出題カテゴリを網羅する約 3 万件のユニークな問題・解答ペアを作成した。

さらに、このデータセットに思考の連鎖 (Chain of Thought; CoT) を付与した。CoT の生成には、ライセンス上問題のない Qwen3-235B-A22B-Instruct-2507 を使用した。各問題に対して、Temperature と Top-p を調整しながら 20 個の異なる CoT を生成し、生成された CoT と正解が一致するもののみを学習データとして採用した。最終的に、3 万問 × 20 個/問 ≒ 約 60 万件の CoT 付きデータセットを構築した。

2.2.2 蒸留データセット

モデルの汎用的な数学能力を底上げするため、NVIDIA の研究を参考に、LLM を用いて問題と解答を完全に生成する「蒸留データ」を構築した。

- 問題生成: Qwen3-235B-A22B-Instruct-2507 を使用し、日本の中学・高校の数学カリキュラムに準拠した問題を約 5 万件生成した。評価ベンチマークの傾向に過度に依存することを避け、多様性を確保することを目的とした。
- 回答・CoT 生成: 生成した問題に対し、Qwen3-235B-A22B-Instruct-2507 を用いて回答と CoT を生成した。解答の正当性を高めるため、1 問あたり 20 回の生成を行い、80%以上で回答が一致したものを採用する「マジョリティ・ボートティング」を適用した。これにより、約 100 万件規模のデータセットを構築した。

最終的に、これら 2 つのデータセットを全体の学習データとして使用した。

2.3 事後学習(Supervised Fine-Tuning)

事後学習は、特性の異なるデータセットを用いて 2 段階に分けて実施した。

- 第 1 段階 (フルファインチューニング): ベースモデルに対し、汎用的な数学能力を注入するため、2.2.2 で作成した 100 万件規模の「蒸

留データセット」を用いてフルファインチューニングを行った。

- 第 2 段階 (LoRA ファインチューニング): 第 1 段階で得られたモデルに対し、コンペティションへの特化を目的として、2.2.1 で作成した 60 万件規模の「評価ベンチマーク準拠データセット」を用いて LoRA (Low-Rank Adaptation)による追加学習を行った。

各段階で使用したハイパーパラメータを表 2 に示す。

表 2 SFT のハイパーパラメータ

パラメータ	第 1 段階 (Full)	第 2 段階 (LoRA)
学習率	2e-5	1e-4
LoRA Rank	-	64
LoRA Alpha	-	128
Dropout	-	0.03

2.4 推論時の工夫

最終的な提出モデルでは、推論の頑健性を向上させるため、Self-Consistency を採用した。これは、同一の問題に対して複数回 (今回は 20 回) 推論を実行し、最も多く得られた回答を最終的な出力とする多数決のアプローチである。開発用評価データでは精度向上への寄与は限定的だったものの、学習に使用していない他の開発用ベンチマークでは顕著な精度向上が見られたため、最終提出モデルに採用した。

3 結果と考察

本アプローチによって開発したモデルは、運営から提供された 100 問の開発用評価データに対し、95%という高い正解率を達成した。これは、データセット生成に使用した Qwen3 シリーズの中で、ベースモデルより少し重みの大きい

Qwen3-30B-A3B-Instruct-2507 モデルが記録した 84% を上回る結果であり、本学習戦略の有効性を示唆している。

一方で、今回のチューニングは日本の中学・高校レベルの数学ベンチマークに特化したものであり、未知の形式の問題に対する汎化性能については、最終評価の結果を待つ必要がある。特に、学習データの大部分を合成データに依存しているため、想定外のパターンに対する脆弱性が懸念される。

また、当初計画していた継続事前学習については、利用可能なデータセットの調査に時間を要し、開発

スケジュールと計算リソースの兼ね合いから最終的に断念した。初期段階で学習プロセス全体のリソース配分を見通せなかった点が反省点として挙げられる。もし継続事前学習を実施できていれば、さらなる汎化及び性能向上が見込めた可能性がある。

4 おわりに

本稿では、チューニングコンペティション 2026 の数学タスクに向けて開発したモデルのアプローチについて報告した。2種類の合成データセットを用いた2段階の SFT と、推論時のセルフコンシステンシーを組み合わせることで、開発データにおいて高い性能を達成した。今回のコンペティション参加を通じて、限られたリソースの中で効果的なチューニング戦略を立案・実行する貴重な経験を得ることができた。このような機会を提供いただいた運営の方々に深く感謝申し上げる。

参考文献

- [1] Shubham Toshniwal, Wei Du, Ivan Moshkov, Branislav Kisacanin, Alexan Ayrapetyan, Igor Gitman. **OpenMathInstruct-2: Accelerating AI for Math with Massive Open-Source Instruction Data**. 2024.
- [2] Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, Denny Zhou. **Self-Consistency Improves Chain of Thought Reasoning in Language Models**. 2022.