MathGenie: Generating Synthetic Data with Question Back-translation for Enhancing Mathematical Reasoning of LLMs

Zimu Lu, et al

The Chinese University of Hong Kong

September 11, 2024

Abstract

- There is a performance gap between open / closed source LLMs
  - New method for generating diverse and reliable math problems

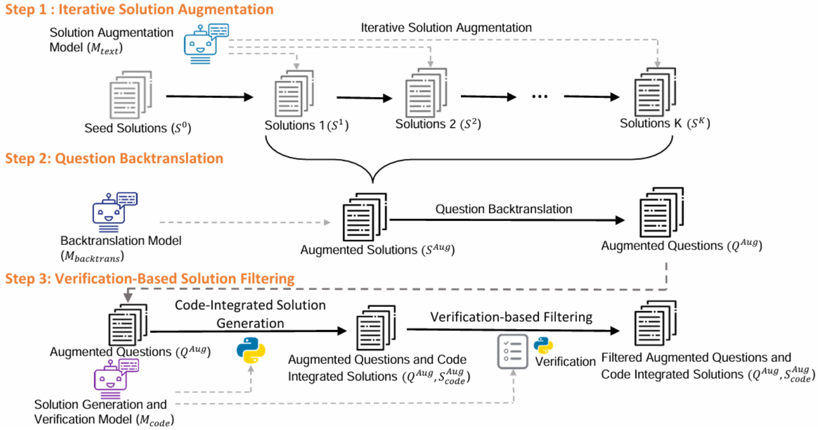
- We suggest MathGenie
  - This augmentation increased open-source models' performance

Introduction

- Three main types of solution
  - CoT, PoT, Code-Integrated solution
  - Code-Integrated solution is superior
  - MathGenie generates Code-Integrated solution

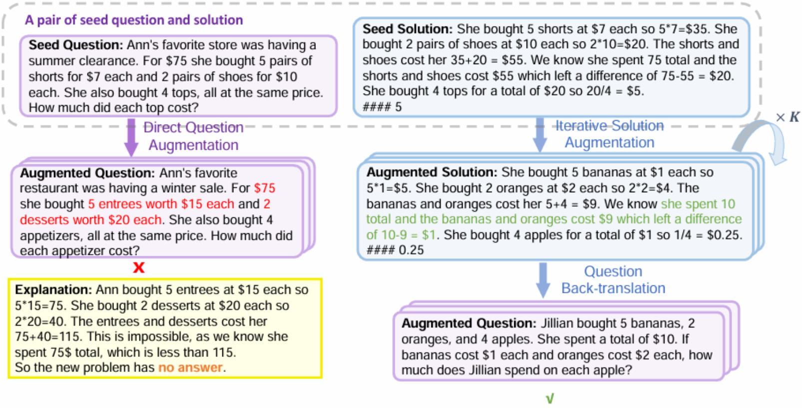
MathGenie Framework

- Iterative Solution Augmentation: Gives variation to solutions
- Question Back-Translation: Invert solutions to questions
- Verification-based Solution Filtering: Verify question-solution pair



Question Back-Translation

- Direct Question Augmentation (Left)
  - It may produce question with no answer
- Question Back-Translation (Right)
  - Correctly augments the question



Experiment

- Fine-tunes pretrained models (Llama Family) with MathGenie
  - This results in MathGenieLM
- Prepare 5 datasets
  - In-domain: GSM-8K, MATH
  - Out-domain: SVAMP, Simuleq, Mathematics
- Compare MathGenieLM with various models
  - Open source: Mammoth, MathCoder, ToRA
  - Closed source: GPT-3.5, GPT-4, PaLM-2

Result and Conclusion

- MathGenieLM achieved SOTA for open-source models
  - However, there was noticable gap compared to GPT-4

- Limitation
  - It requires significant GPU resource
  - It cannot process images as input