Program of Thoughts Prompting for Accurate Numerical Reasoning

SoonHo Kim
Team 14 - 20200703

1   Problem Definition

Numerical reasoning has been a long-standing challenge in natural language processing (NLP). Existing methods like Chain-of-Thought (CoT) prompting enable large language models (LLMs) to perform step-by-step reasoning for math word problems. However, CoT has a fundamental limitation in that LLMs must handle both reasoning and computation within natural language, which often leads to incorrect results due to arithmetic errors, especially with large numbers or iterative processes. The paper proposes a solution that separates these two responsibilities.
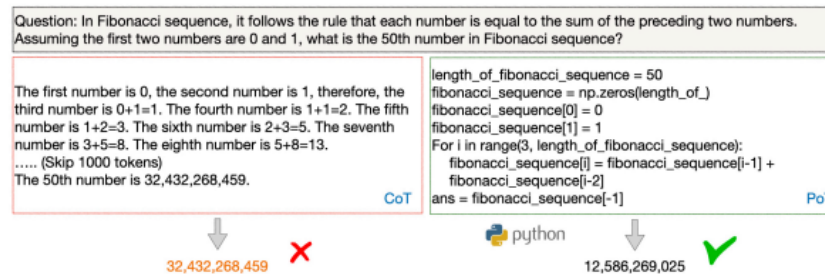
2   Proposed Method

Figure 1: Comparison between Chain of Thoughts and Program of Thoughts [1].

The paper introduces a new method called **Program of Thoughts (PoT)** prompting. Instead of solving math problems entirely through language, PoT allows LLMs to express the reasoning process using code (e.g., Python), and offloads actual computation to a program interpreter. This separation enables models to focus on logical structure while leveraging the precision and speed of traditional computing systems. PoT also uses semantically meaningful variable names and multi-step decomposition, which significantly improves performance across several math and finance benchmarks. Compared to CoT, PoT achieves an average 12% improvement across datasets like GSM8K, AQuA, and FinQA.

3   Discussion

Our team's project focuses on building a general game-playing AI using LLMs, with chess as the primary testbed. In chess, a game state leads to many deterministic possibilities that are best handled through algorithmic simulation. Inspired by PoT, we plan to offload these deterministic calculations (e.g., generating legal moves, evaluating material scores) to a Python-based backend. The LLM will then focus only on strategic reasoning, such as choosing the best move based on the situation. This division allows for more accurate move predictions and better interpretability, much like how PoT achieves more robust reasoning by separating logic and calculation.

References

[1] Wenhu Chen, Xueguang Ma, Xinyi Wang, and William W Cohen. Program of thoughts prompting: Disentangling computation from reasoning for numerical reasoning tasks. *arXiv preprint arXiv:2211.12588*, 2022.