ChainThought: Enhancing Reasoning Capabilities in Language Models Through Step-by-Step Problem Solving

Brandon Wong, Derek Laboy, Eric Gulotty

March 2, 2025

Abstract

This project aims to develop a fine-tuned language model that demonstrates explicit chain-of-thought reasoning capabilities for solving mathematical problems. By training a Deepseek-7B model on the GSM8K dataset with step-by-step reasoning annotations, we seek to create an AI system that not only produces correct answers but also transparently explains its problem-solving process, similar to how students show their work in educational settings.

Introduction

Recent advancements in large language models have led to significant improvements in reasoning capabilities across various model families. Models such as GPT-o1 and Deepseek R1 now demonstrate impressive abilities to transparently explain their reasoning processes in various domains.

In this project, we seek to explore and understand these reasoning capabilities by fine-tuning a model that doesn't already have explicit chain-of-thought reasoning built in. We've selected Deepseek-7B as our starting point to explore how fine-tuning can enhance a model's ability to articulate its problem-solving steps.

Chain-of-thought reasoning—where models articulate intermediate steps leading to their conclusions—has been shown to improve performance on complex reasoning tasks. By working with grade school mathematics problems, we can learn in a domain with well-defined correct answers and structured, logical reasoning that can be explicitly articulated.

The choice of mathematics as our domain is deliberate: mathematical problems provide clear evaluation criteria and allow us to directly compare the reasoning process with the solution. This makes it an ideal domain for an educational project focused on understanding reasoning enhancement through fine-tuning.

This project will provide us with hands-on experience working with language models while exploring the practical aspects of explicitly training for reasoning transparency in a controlled educational context.

Related Work

Chain-of-thought reasoning in language models has recently emerged as an active area of research. Wei et al. [1] demonstrated that prompting large language models to generate intermediate reasoning steps before providing a final answer significantly improves performance on complex reasoning tasks, including arithmetic, commonsense, and symbolic reasoning. They showed that this approach is emergent in sufficiently large models.

Building on this work, Kojima et al. [2] found that simply prompting a model with "Let's think step by step" can elicit chain-of-thought reasoning without exemplars, demonstrating that explicit reasoning instructions can be effective even without specific training.

In terms of specific mathematical reasoning, the GSM8K dataset introduced by Cobbe et al. [3] has become a standard benchmark. It contains grade school math word problems with step-by-step solutions, making it ideal for our purposes.

For our fine-tuning approach, we will use Low-rank adaptation (LoRA) [4], which has emerged as a computationally efficient approach to fine-tuning that modifies only a small subset of model parameters, making it suitable for our computational constraints.

Deepseek-7B, our chosen base model, was released by Deepseek AI in 2023 [5] and has shown strong performance across various benchmarks including reasoning tasks, while being considerably smaller than many other state-of-the-art models. This makes it an ideal candidate for our project given computational constraints.

Proposed Work

Our project will focus on fine-tuning the Deepseek-7B model to enhance its chain-of-thought reasoning capabilities specifically for grade school mathematics problems. We propose a comprehensive approach consisting of the following components:

Data Preparation

We will use the GSM8K dataset, which contains 8.5K high-quality grade school math word problems with step-by-step solutions. We will format these examples to clearly delineate:

- Problem statement

- Step-by-step reasoning process
- Final answer

This formatting will help the model learn to generate explicit reasoning steps between receiving a problem and producing an answer.

Fine-tuning Methodology

We will employ parameter-efficient fine-tuning using LoRA (Low-Rank Adaptation) [4], which allows us to fine-tune the model efficiently by adding trainable low-rank matrices to existing weights rather than modifying all parameters. This approach has several advantages:

- Reduced computational requirements compared to full fine-tuning
- Lower memory footprint
- Faster training times
- Easier deployment through small adapter modules

Our training objective will be standard next-token prediction on the entire sequence containing the problem, reasoning steps, and answer. This encourages the model to learn the relationship between problems and their solution processes.

Hyperparameter Optimization

We will explore various hyperparameter settings, including:

- LoRA rank (typically between 8 and 64)
- LoRA alpha (scaling factor)
- Learning rate and schedule
- Batch size
- Number of training epochs

We will use a small validation set to identify the optimal configuration.

Prompting Strategy Development

Beyond model fine-tuning, we will develop and test different prompting strategies to elicit effective chain-of-thought reasoning. This may include:

- Explicit instructions to show work step-by-step

- Formatting guides (e.g., numbered steps)
- Few-shot examples embedded in prompts

Inference-Time Techniques

We will explore basic inference techniques such as:

- Temperature sampling for reasoning diversity
- Simple verification steps where the model checks its own work
- Optimizing the prompt format to elicit better reasoning

Datasets

Our primary dataset will be GSM8K (Grade School Math 8K) [3], which consists of 8,500 high-quality grade school math word problems. These problems cover various elementary mathematical concepts including arithmetic, basic algebra, and word problems involving rates, ratios, and percentages. Each problem in the dataset comes with a step-by-step solution, making it ideal for training a model to generate chain-of-thought reasoning.

The dataset is already structured in a way that aligns with our goals, but we will need to perform the following preparation steps:

- Standardize the format of problems and solutions for consistent training
- Split the dataset into training (80%), validation (10%), and test (10%) sets
- Ensure that the test set is completely held out during training and hyperparameter tuning
- Analyze solution patterns to identify different reasoning strategies that may be present

We will carefully curate a small holdout test set from GSM8K with problems of varying difficulty to thoroughly evaluate our model's performance and generalization capabilities within the scope of grade school mathematics.

Evaluation

We will employ a multi-faceted evaluation approach to thoroughly assess our model's reasoning capabilities:

Quantitative Metrics

- **Answer Accuracy**: Percentage of problems for which the model produces the correct final answer.
- **Reasoning Completeness**: Assessment of whether all necessary intermediate steps are present.
- **Step Accuracy**: Evaluation of whether each intermediate reasoning step is mathematically correct.
- **Efficiency**: Number of steps used compared to actual solutions.
- **Partial Credit Scoring**: A rubric-based approach to evaluate solutions with correct reasoning but slightly incorrect answers:
  - 5 points: Perfect solution with correct reasoning and answer
  - 4 points: Correct reasoning with minor calculation error in final answer
  - 3 points: Mostly correct reasoning with conceptual understanding, but calculation errors
  - 2 points: Partial understanding with some correct steps
  - 1 point: Minimal correct reasoning
  - 0 points: Completely incorrect approach

Comparative Evaluation

- Compare against the base Deepseek-7B model without fine-tuning
- Compare against the base model with chain-of-thought prompting but no fine-tuning
- Where possible, benchmark against published results for larger models on the same datasets

Qualitative Assessment

- Human evaluation of reasoning clarity and naturalness
- Error analysis to identify patterns in problems where the model struggles
- Assessment of different reasoning strategies the model adopts

Generalization Tests

- Performance on held-out problems with varying complexity within GSM8K

- Ability to adapt reasoning depth to problem complexity
- Testing on a few manually created problems slightly outside the GSM8K distribution

We will also analyze how various model configurations and training strategies affect different aspects of performance, seeking to understand the trade-offs between accuracy, reasoning quality, and computational efficiency.

Timeline

We propose the following condensed 8-week timeline for project execution:

- **Week 1** (March 3-9):
  - Finalize project design and literature review
  - Set up development environment and infrastructure
- **Week 2** (March 10-16):
  - Prepare GSM8K dataset in appropriate format for fine-tuning
  - Begin implementing LoRA fine-tuning pipeline
- **Week 3** (March 17-23):
  - Complete fine-tuning pipeline implementation
  - Run initial experiments with different hyperparameter configurations
- **Week 4** (March 24-30):
  - Complete main model training
  - Develop evaluation metrics and scripts
- **Week 5** (March 31-April 6):
  - Preliminary evaluation on validation set
  - Iterate on prompting strategies
- **Week 6** (April 7-13):
  - Comprehensive evaluation on test set
  - Error analysis and model refinement
- **Week 7** (April 14-20):
  - Test generalization capabilities

  - Begin drafting final report
- **Week 8** (April 21-27):
  - Complete final report
  - Prepare presentation materials

Midway Update

In this stage of our project, we have made significant progress in both data analysis and model development. Our problem statement has been refined to focus on enhancing chain-of-thought reasoning in large language models for solving grade school math problems. To this end, we have chosen the GSM8K dataset, which offers a diverse set of math problems along with detailed step-by-step solutions.

Baseline Model Evaluation

We implemented a baseline model using the Deepseek-7B architecture. Our baseline evaluation, as detailed in `example_output.json`, shows that the model often struggles with arithmetic computations. These initial results have provided valuable insights into the areas where the model needs improvement.

Repository and Experimental Setup

A GitHub repository has been established to collect all our work. The repository includes:

- `baseline_evaluate.py` – a script to evaluate the baseline model.
- `tune_model.py` – a script for fine-tuning the model using parameter-efficient methods (PEFT with LoRA).
- `clean_cache.py` and `testing.py` – utility scripts for cache management and preliminary testing.

These resources have enabled us to streamline our experiments and ensure reproducibility of our results.

Literature Influence and Future Work

Our work has been greatly influenced by the recent research presented in [6]. The paper highlights the benefits of explicitly training language models with chain-of-thought reasoning, and its insights are shaping our approach to fine-tuning. Moving forward, our focus

will be on tuning the baseline model to address its weaknesses, particularly in mathematical reasoning. We plan to experiment with advanced fine-tuning techniques using PEFT and LoRA to enhance the model's performance on the GSM8K dataset.

Summary

In summary, we have:

1. Trained and evaluated a baseline Deepseek-7B model on GSM8K, identifying significant shortcomings in arithmetic reasoning.
2. Established a GitHub repository to manage our code and experiments.
3. Initiated our literature survey, with the work in [6] providing key insights into chain-of-thought training.

Our next steps will involve refining our tuning process to improve the model's performance and address the identified weaknesses.

Conclusion

In this project, we propose to fine-tune the Deepseek-7B model to enhance its chain-of-thought reasoning capabilities specifically for grade school mathematical problem-solving. By leveraging the GSM8K dataset and parameter-efficient fine-tuning techniques, we aim to develop a model that not only produces correct answers but also generates clear, step-by-step explanations of its reasoning process.

This project will give us valuable hands-on experience with language model fine-tuning techniques while exploring an interesting capability - explicit reasoning. The educational context provides a well-defined problem space with clear evaluation metrics, making it suitable for a course project of this scope.

The relatively compact size of our chosen base model (Deepseek-7B) makes this project feasible with the computational resources available to us, while still allowing us to explore meaningful improvements in model capabilities.

Through this project, we expect to gain practical insights into the challenges and opportunities of enhancing reasoning transparency in language models, while developing a useful prototype that demonstrates step-by-step problem solving in an educational context.

References

[1] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*. NeurIPS, 2022.

[2] Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. *Large Language Models are Zero-Shot Reasoners*. NeurIPS, 2022.

[3] Cobbe, K., Kosaraju, V., Bavarian, M., Chen, M., Jun, H., Kaiser, L., Plappert, M., Tworek, J., Hilton, J., Nakano, R., Hesse, C., & Schulman, J. *Training Verifiers to Solve Math Word Problems*. arXiv preprint arXiv:2110.14168, 2021.

[4] Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. *LoRA: Low-Rank Adaptation of Large Language Models*. ICLR, 2022.

[5] Deepseek AI. *Deepseek-7B: An Open Source Large Language Model*. 2023.

[6] Wei, J., et al. *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*. NeurIPS, 2022. Available at: `https://arxiv.org/abs/2402.03300`.