Final Project Proposal

Zhou Shouchen, Wei Zhengxuan

Introduction

The development of large language models (LLMs) has sparked interest in their ability to solve optimization problems. One key area of exploration is the transformation of natural language descriptions of optimization problems into standard mathematical formulations [2, 3, 1]. The NL4Opt dataset [2] provides a benchmark for this task, facilitating research in this domain. In this project, we aim to evaluate the performance of existing LLMs on this task and identify their challenges. Furthermore, for open-source models, we will fine-tune them to better address the task. We hope our work will provide valuable insights and encourage further research in this field.

Problem Description

Given an input text $T_{\text{in}} \in \mathbb{R}^{L_1}$, where L_1 is the input length, an LLM maps it to an output text $T_{\text{out}} \in \mathbb{R}^{L_2}$, where L_2 is the output length. This can be expressed as:

$$T_{\text{out}} = \text{LLM}(T_{\text{in}})$$

where $\text{LLM}(\cdot)$ represents the model's transformation function. The goal is for T_{out} to represent a standard optimization problem of the form:

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g(x) \leq 0, \\ & h(x) = 0, \end{aligned}$$

where f, g, and h are functions, and x represents the decision variables.

Motivation

1. Many real-world optimization tasks are described in natural language, making manual conversion into standard forms time-consuming. LLMs can automate this process.
2. Current LLMs face issues like lack of interpretability and hallucinations. Rigorous evaluation is needed to understand their practical applicability.
3. Despite extensive benchmarking of LLMs on various tasks, their performance on optimization problem transformation remains underexplored. This project aims to fill this gap.
4. Open-source LLMs are advantageous for deployment and debugging. Comparing their performance with proprietary models can aid practical use.
5. By analyzing model performance, we aim to identify strengths, weaknesses, and potential improvements for future research.
6. Most LLMs are designed for general tasks. Fine-tuning open-source models for this specific task can significantly improve their performance.

Plan

1. Preprocess the NL4Opt dataset to eliminate redundancy and ensure alignment with standard optimization formats.
2. Evaluate the performance of various LLMs, including GPT-3.5, GPT-4, Gemini, Claude 3, ERNIE-4.0, GLM-4, Moonshot-v1, Baichuan 3, Qwen1.5 (open-source), and LLaMA 3 (open-source).
3. Fine-tune open-source models to develop task-specific solutions for optimization problem conversion.

References

[1] Yanghua Peng, Yixin Bao, Yangrui Chen, Chuan Wu, and Chuanxiong Guo. Optimus: an efficient dynamic resource scheduler for deep learning clusters. In *Proceedings of the Thirteenth EuroSys Conference*, pages 1–14, 2018.

[2] Rindranirina Ramamonjison, Timothy Yu, Raymond Li, Haley Li, Giuseppe Carenini, Bissan Ghaddar, Shiqi He, Mahdi Mostajabdaveh, Amin Banitalebi-Dehkordi, Zirui Zhou, and Yong Zhang. Nl4opt competition: Formulating optimization problems based on their natural language descriptions. In Marco Ciccone, Gustavo Stolovitzky, and Jacob Albrecht, editors, *Proceedings of the NeurIPS 2022 Competitions Track*, volume 220 of *Proceedings of Machine Learning Research*, pages 189–203. PMLR, 28 Nov–09 Dec 2022. URL `https://proceedings.mlr.press/v220/ramamonjison23a.html`.

[3] Ziyang Xiao, Dongxiang Zhang, Yangjun Wu, Lilin Xu, Yuan Jessica Wang, Xiongwei Han, Xiaojin Fu, Tao Zhong, Jia Zeng, Mingli Song, et al. Chain-of-experts: When llms meet complex operations research problems. In *The Twelfth International Conference on Learning Representations*, 2023.