Boosting of Thoughts: Trial-and-Error Problem Solving with Large Language Models

Sijia Chen, Baochun Li
Department of Electrical and Computer Engineering
University of Toronto
Toronto, Ontario, Canada
sjia.chen@mail.utoronto.ca, bli@ece.toronto.edu

Di Niu
Department of Electrical and Computer
Engineering, University of Alberta
Edmonton, Alberta, Canada
dniu@ualberta.ca

Abstract

The reasoning performance of Large Language Models (LLMs) on a wide range of problems critically relies on chain-of-thought prompting, which involves providing a few chain of thought demonstrations as exemplars in prompts. Recent work, e.g., Tree of Thoughts, has pointed out the importance of exploration and self-evaluation in reasoning step selection for complex problem solving. In this paper, we present Boosting of Thoughts (BoT), an automated prompting framework for problem solving with LLMs by iteratively exploring and self-evaluating many trees of thoughts in order to acquire an ensemble of trial-and-error reasoning experiences, which will serve as a new form of prompting to solve the complex problem. Starting from a simple prompt without requiring examples, BoT iteratively explores and evaluates a large collection of reasoning steps, and more importantly, uses error analysis obtained from the LLM on them to explicitly revise prompting, which in turn enhances reasoning step generation, until a final answer is attained. Our experiments with GPT-4 and Llama2 across extensive complex mathematical problems demonstrate that BoT consistently achieves higher or comparable problem-solving rates than other advanced prompting approaches. The source code is available under the folder *examples/BoTReasoning* of https://github.com/iQua/llmpebase.

1 Introduction

Large language models (LLMs) with the autoregressive paradigm have gained remarkable performance across various tasks due to their potential reasoning ability Brown et al. (2020); Lewkowycz et al. (2022). The guarantee of such ability in complex tasks heavily relies on chain-of-thought (CoT) Wei et al. (2022) prompting, which provides step-by-step reasoning examples. This approach suggests that the reasoning ability can be elicited through a chain of thoughts, where a thought serves as an intermediate step toward problem solving.

Thus, subsequent studies Fu et al. (2022); Wang et al. (2022); Yao et al. (2024); Besta et al. (2023), especially Tree-of-Thought (ToT) Yao et al. (2024), have been proposed to improve CoT. To guarantee effectiveness, the prompt of these approaches generally includes human annotations on one specific task. Such a reliance limits their scalability. Recent work that either employs a double-check with LLMs to improve answers Paul et al. (2023); Weng et al. (2023); Madaan et al. (2023) or boosts prompts based on feedback Zheng et al. (2023); Zhang et al. (2023a); Hou et al. (2023); Pitis et al. (2023) has demonstrated significant promise. The existing literature generally tends to discard ineffective thoughts from the prompt. However, humans typically can continuously learn from errors by carefully analyzing them to gain experience, thereby gradually improving performance. We therefore ask: whether the thought generation of LLMs can dispense with human annotation and mimic such a problem-solving way of humans to achieve effective reasoning across various tasks?

This paper proposes a novel framework, shown in Fig. 1, referred to as the Boosting of Thoughts (BoT), which achieves the *boosting* mechanism that embraces *aggregation* and *experience*, thereby enabling the progressive refinement of unreliable reasoning steps (weak thoughts) by learning from errors to eventually solve various problems. Starting with a simple prompt without human annotations