Bachelor Project



F3

Faculty of Electrical Engineering Artificial Intelligence Center

Methods of Evolutionary Optimization of Prompts for Large Language Models

Enhancing task-solving capabilities of general models by prepending an optimal prefix

Vojtěch Klouda

Supervisor: Ing. Jan Drchal PhD. Field of study: Artificial Intelligence Subfield: Natural Language Processing

May 2025

Acknowledgements

Declaration

=)))

Prohlašuji, že jsem předloženou práci vypracoval samostatně, a že jsem uvedl veškerou použitou literaturu.

 $\rm V$ Praze, 10. May 2025

Abstract

Abstrakt

TODO

Keywords: language model, evolutionary algorithm

Supervisor: Ing. Jan Drchal PhD. Resslova 307/9 Praha, E-322

TODO

Klíčová slova: jazykový model, evoluční algoritmus

Překlad názvu: Metody evoluční optimalizace vstupních řetězců pro velké jazykové modely — Vylepšení řešících schopností obecných modelů připojením optimálního prefixu

Contents

1 Introduction	1
2 Literature Review	3
2.1 Large Language Models	4
2.1.1 Word2Vec, tokenization	4
2.1.2 Transformer architecture	4
2.2 Prompts	4
2.2.1 Prompting techniques	4
2.2.2 Prompt optimization methods	5
A Bibliography	7

Figures

Tables

Chapter 1

Introduction

This is all very important

Chapter 2

Literature Review

2.1 Large Language Models

Since the Transformer architecture Cite attention is all you need brought us the first GPT and BERT models, LLMs have changed the NLP field. Now there are many powerful proprietary (OpenAI, Anthropic) and open-source (Llama, Qwen) models. cite

2.1.1 Word2Vec, tokenization

abcd

2.1.2 Transformer architecture

abcd

2.2 Prompts

Prompts are instructions we give to the LLM to get the output we want. It has been shown experimentally that small changes in prompt design, like simple paraphrasing, can influence the accuracy of the model's output. cite Naturally this inspired research into optimizing LLM performance by changing its prompts.

2.2.1 Prompting techniques

In solving challenging reasoning tasks, it is beneficial to get the model to show its reasoning steps. As LLMs just predict the next token, building on its previous thoughts improves performance on tasks like the GSM8k benchmark. This technique, called Chain-of-Thought has been built upon which resulted in methods such as Tree-of-Thought, ReAct and Reflexion. cite prompt survey

2.2.2 Prompt optimization methods

There has been work of optimizing soft and discrete prompts. As proprietary models usually do not allow access to its internal states, soft prompt optimization is not and option there. Several search and optimization methods have been employed in this task, like beam search and evolutionary algorithms. sota rewrite here

Appendix A

Bibliography