
Contextual Bayesian Optimisation with In-Context-Learning via Large Language Models

Department of Computer Science

Candidate Number: 18004174 ¹

Degree: MSc Computational Statistics and Machine Learning

Supervisor: Dr Ilija Bogunovic

Submission date: 11th September 2023



¹**Disclaimer:** This report is submitted as part requirement for the MSc Computational Statistics and Machine Learning at UCL. It is substantially the result of my own work except where explicitly indicated in the text. The report will be distributed to the internal and external examiners, but thereafter may not be copied or distributed except with permission from the author.

Acknowledgements

I would like to thank my primary supervisor, Dr Ilija Bogunovic, for their guidance and deep insights during the dissertation. Additionally, I would like express my gratitude towards PhD candidate, Shyam Sundhar Ramesh, for their continual support and ideas throughout our one-to-one meetings. Without the cooperation of my peers, the dissertation would not have been successful.

Abstract

TO BE COMPLETED

Contents

1	Introduction	2
1.1	Motivations	2
1.2	Objectives	2
2	Related Work	4
2.1	Limitation Learning	4
2.1.1	Zero-Shot Learning	5
2.1.2	Few-Shot Vs In-Context Learning	5
2.2	Language Models	5
2.2.1	Recurrent-Neural Network	5
2.2.2	Long Short-Term Memory	5
2.2.3	Gated Recurrent Units	5
2.2.4	Transformers	5
2.2.5	Generative Pretrained Transformers	5
2.3	Bayesian Optimisation	5
2.3.1	Surrogate Model	5
2.3.2	Acquisition Function	5
2.4	BO-LIFT	5
2.4.1	Ask-Tell-Predict	5
2.4.2	Bayesian Optimisation Protocol	5
3	Methodology	6
3.1	BEACON	6
3.1.1	Feature Selection	6
3.1.2	Prompt Engineering	6
4	Experiments	7
4.1	Datasets	7
4.2	Hyperparameters	7
4.3	Results	7
4.3.1	Ask-Tell-Predict	7
4.3.2	Bayesian Optimisation Protocol	7

5 Conclusion	8
5.1 Discussion	8
5.2 Future Work	8
A Other appendices, e.g. code listing	9

Chapter 1

Introduction

In this chapter we provide a concise global overview of the dissertation. This covers the motivations for the dissertation and the objectives we aim to tackle and provide insight on.

1.1 Motivations

Natural Language Processing (NLP) has been extensively explored from linguistic and theoretical perspectives. Initially, NLP focused on tasks like text translation, sentiment analysis, and speech recognition. However, with advancements in hardware and software, both use cases and architectural designs have significantly improved, ushering in the era of Generative AI. The advent of Large Language Models (LLMs) has revolutionized diverse domains such as healthcare, trading, and programming, surpassing their current capabilities. These LLMs, primarily transformer architectures, undergo supervised learning on specific datasets. Notably, recent advancements have given rise to generative pre-trained transformers—frozen LLMs pre-trained on vast corpora. These task-agnostic models exhibit in-context learning (ICL), acquiring knowledge from a few examples during inference. In parallel, contextual bandit optimization has emerged as a distinct research area, addressing scenarios where experiments are framed as choices, such as ad selection based on user information or selecting drugs for optimal solubility. Bayesian Optimization (BO), an alternative to the conventional Thompson Sampling Bandit Optimization (TSBO) approach, employs a Gaussian Process Regression (GPR) surrogate model. However, the computational inefficiency of GPR becomes apparent when dealing with large sample spaces. The primary motivation of this study is to provide a comprehensive qualitative analysis of large language models, specifically examining their interplay with ICL and BO, across a range of domains.

1.2 Objectives

This thesis aims to reproduce, adapt, and extend the works of LIFT and BO-LIFT, with three clear objectives:

1. Analyse the in-context sample complexity across different Large Language Models (LLMs).
This objective entails two aspects:

- a) Examining the number of examples used during the training and inference phases.
 - b) Investigating the impact of the amount of context provided in the examples/templates i.e. token lengths.
2. Reproduce the findings, obtained in BO-LIFT, for sequential-model based-optimisation i.e. Bayesian Optimisation via LLMs, in alternative datasets and assess its viability as an alternative approach.
 3. Apply a refined prompt-engineered interface on a distinct yet related dataset. This dataset includes various associated features for each compound name, such as temperature and SMILES representations. Our goal is to assess whether the LLM predicts consistent outputs for the same compounds, considering the varying feature values.

Chapter 2

Related Work

In this chapter, we delve into the captivating realm of Limitation Learning i.e. learning with limitations, which in this case is sparse data. Our exploration begins with a thorough review of the existing literature, where we critically examine the concepts of Zero-Shot Learning (ZSL), Few-Shot Learning (FSL), and In-Context Learning (ICL). Through a comparative analysis of these approaches, we gain valuable insights into their respective limitations and strengths. Then we provide an in-depth analysis of the development of Language Models, from the foundational Recurrent Neural Networks (RNNs) to the remarkable advancements achieved with Generative Pretrained Transformers (GPT). Next, we shift our focus to statistical optimisation algorithms, specifically Bayesian Optimisation (BO), and explore the intricate workings of surrogate models and acquisition functions that facilitate this process. Finally, we unveil a recent intriguing framework called BO-LIFT, which seeks to connect these seemingly unrelated concepts, and demonstrate the sheer power LLMs possess.

2.1 Limitation Learning

The term *limitation learning* is an alternative short-form to describe the limitations that come when algorithms try to *learn*. In our dissertation, we focus on the tradeoff between performance against access to minimal data.

- 2.1.1 Zero-Shot Learning
- 2.1.2 Few-Shot Vs In-Context Learning
- 2.2 Language Models
 - 2.2.1 Recurrent-Neural Network
 - 2.2.2 Long Short-Term Memory
 - 2.2.3 Gated Recurrent Units
 - 2.2.4 Transformers
 - 2.2.5 Generative Pretrained Transformers
- 2.3 Bayesian Optimisation
 - 2.3.1 Surrogate Model
 - 2.3.2 Acquisition Function
- 2.4 BO-LIFT
 - 2.4.1 Ask-Tell-Predict
 - 2.4.2 Bayesian Optimisation Protocol

Chapter 3

Methodology

In this chapter, we introduce Bayesian Enhancement via Adaptive Context Optimisation (BEACON), a modification of the BO-LIFT framework. This chapter provides a detailed explanation of the Context Selection process and the Prompt Engineering techniques used in BEACON.

3.1 BEACON

3.1.1 Feature Selection

3.1.2 Prompt Engineering

Chapter 4

Experiments

In this chapter, we describe the experimental design of our study. We introduce real scientific datasets used in our experiments, AqSolDB and BigSolDB, and discuss the selection and tuning of various hyper-parameters of the LLMs.

4.1 Datasets

4.2 Hyperparameters

- Due to financial reasons, we only tested the datasets on two OpenAI models: text-curie-001 & text-davinci-003.

4.3 Results

In this chapter, we present the results of our experiments. We provide a detailed analysis of the performance of the Ask-Tell-Predict method and the Bayesian Optimisation Protocol on the different datasets, as well as against the BO-LIFT framework.

4.3.1 Ask-Tell-Predict

4.3.2 Bayesian Optimisation Protocol

Chapter 5

Conclusion

In this chapter, we conclude the dissertation by summarising our findings, discussing their implications, identifying the limitations of our study, and suggesting potential directions for future research.

5.1 Discussion

5.2 Future Work

- Try to implement feedback in the BO-protocol.
- Instead of feature selection to obtain top m contextual inputs, try combinatorial context addition (this however will require large compute power and cost so may not be feasible).
- Can we use this directly now for bandit problem? Or even for hyper-parameter tuning (which BO can do)?
- Attempt transfer learning (but note that fine-tuning will require large compute power and cost so may not be feasible).

Appendix A

Other appendices, e.g. code listing