UNIVERSIDAD DE INGENIERÍA Y TECNOLOGÍA

CARRERA DE CIENCIA DE LA COMPUTACIÓN

Large Language Models for the Generation of reviews for products in e-commerce

AUTOR

Luis Antonio Gutiérrez Guanilo
luis.gutierrez.g@utec.edu.pe

ASESOR

Cristian López
clopezd@utec.edu.pe

Lima - Perú
2023

Contents

Chapter 1

Context and Motivation

Introduction

Large Language Models (LLMs) such as GPT-4, BERT, LLama, and LLama2 are transforming sectors like healthcare [1] [2], finance, and e-commerce by their remarkable ability to understand and generate text that closely resembles human communication. These models play a pivotal role in enhancing decision-making processes, automating customer service, and improving data analysis [3].

Although these models perform well across various applications, there are scenarios where they require specific training to handle particular tasks effectively. Fine-tuning is a strategic approach to enhance model performance by training pre-existing models with specialized datasets to better meet domain-specific needs [4]. Examples of such specialized applications include LLama2-chat [5], Mistral Instruct [6], and StructLM [7], each tailored with unique datasets. However, the lack of high-quality, focused datasets, particularly in areas like product attributes and e-commerce, remains a significant challenge, emphasizing the need for comprehensive datasets that enable models to interact effectively with detailed product information.

Creating a dataset involves a deep understanding of the data types collected. While Audio and Video are significant, Text and Tabular data are more common in real-world applications, appearing in formats such as Excel tables, Wikipedia pages, and other spreadsheets. These data can be formatted in several styles, including HTML, CSV (Comma Separated Values), TSV (Tab Separated Values), Markdown, DFLoader, Data-Matrix, and JSON. JSON, in particular, is highly valued for its readability and easy integration with contemporary web technologies [8].

Using JSON-centric methods to fine-tune models significantly enhances their capacity to process and generate structured data accurately [9]. This capability is crucial for e-commerce platforms, where product data's structure and content frequently vary. By focusing on JSON-structured data to fine-tune LLMs like LLama2-chat, Mistral Instruct, and StructLM, this

project seeks to significantly refine the extraction and normalization of product attribute values. This will lead to more accurate and contextually relevant product reviews, directly improving user interaction and satisfaction.

Problem Description

Despite the advancements of LLMs in various sectors, they often struggle with domain-specific tasks without precise and targeted training. A significant problem in e-commerce is the interaction with detailed product information due to the lack of high-quality, focused datasets. This deficiency affects the models' ability to accurately extract and normalize product attribute values, leading to suboptimal product reviews and recommendations. Additionally, the diverse structure and content of product data on e-commerce platforms pose a challenge. There is a pressing need to create and utilize datasets that cater specifically to the structure and nuances of product data, particularly in JSON format, to enhance the performance of LLMs in accurately processing and generating structured data.

Motivation

The key challenge in leveraging LLMs effectively in e-commerce and other sectors is the absence of high-quality, focused datasets, especially concerning product attributes. This gap hinders the models' ability to interact efficiently with detailed product information. Fine-tuning pre-existing models with specialized datasets is a strategic approach to enhance model performance and meet domain-specific needs. Moreover, the prevalent use of JSON in modern web technologies underscores the importance of fine-tuning models to process and generate structured data accurately.

Objectives

Genetal Objective

The primary objective of this project is to refine product recommendations based on the specifications of a product in e-commerce by enhancing the accuracy of product reviews and focused reviews.

Specific Objectives

Specifically, this project aims to fine-tune LLMs like LLama2-chat, Mistral Instruct, and StructLM with a bespoke, product-related JSON dataset. This will enable proficient analysis and processing of JSON-formatted product data, leading to more accurate and contextually relevant product reviews. The initiative also seeks to compare the improved performance of these trained

models against baseline models, demonstrating significant enhancements in handling structured product data based on the metrics of hallucination, fluency, and relevance.

Aportes

Chapter 2

Marco Teórico

Chapter 3

Revisión de la Literatura

Chapter 4

Metodología

4.1 Descripción de la Metodología

Chapter 5

Experimentaciones y Resultados

5.1 Experimentos y Resultados

Chapter 6

Conclusiones y Trabajos Futuros

6.1 Conclusiones

6.2 Trabajos Futuros

Bibliography

[1] K. He, R. Mao, Q. Lin, Y. Ruan, X. Lan, M. Feng, and E. Cambria, "A survey of large language models for healthcare: from data, technology, and applications to accountability and ethics," 2023.

[2] S. Reddy, "Evaluating large language models for use in healthcare: A framework for translational value assessment," *Informatics in Medicine Unlocked*, vol. 41, p. 101304, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352914823001508

[3] T. Varshney, "Build an llm-powered data agent for data analysis," Feb 2024. [Online]. Available: https://developer.nvidia.com/blog/build-an-llm-powered-data-agent-for-data-analysis/

[4] D. Bergmann, "Build an llm-powered data agent for data analysis," March 2024. [Online]. Available: https://www.ibm.com/topics/fine-tuning

[5] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale, D. Bikel, L. Blecher, C. C. Ferrer, M. Chen, G. Cucurull, D. Esiobu, J. Fernandes, J. Fu, W. Fu, B. Fuller, C. Gao, V. Goswami, N. Goyal, A. Hartshorn, S. Hosseini, R. Hou, H. Inan, M. Kardas, V. Kerkez, M. Khabsa, I. Kloumann, A. Korenev, P. S. Koura, M.-A. Lachaux, T. Lavril, J. Lee, D. Liskovich, Y. Lu, Y. Mao, X. Martinet, T. Mihaylov, P. Mishra, I. Molybog, Y. Nie, A. Poulton, J. Reizenstein, R. Rungta, K. Saladi, A. Schelten, R. Silva, E. M. Smith, R. Subramanian, X. E. Tan, B. Tang, R. Taylor, A. Williams, J. X. Kuan, P. Xu, Z. Yan, I. Zarov, Y. Zhang, A. Fan, M. Kambadur, S. Narang, A. Rodriguez, R. Stojnic, S. Edunov, and T. Scialom, "Llama 2: Open foundation and fine-tuned chat models," 2023.

[6] A. Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. de las Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier, L. R. Lavaud, M.-A. Lachaux, P. Stock, T. L. Scao, T. Lavril, T. Wang, T. Lacroix, and W. E. Sayed, "Mistral 7b," 2023.

[7] A. Zhuang, G. Zhang, T. Zheng, X. Du, J. Wang, W. Ren, S. W. Huang, J. Fu, X. Yue, and W. Chen, "Structlm: Towards building generalist models for structured knowledge grounding," 2024.

[8] A. Singha, J. Cambronero, S. Gulwani, V. Le, and C. Parnin, "Tabular representation, noisy operators, and impacts on table structure understanding tasks in llms," 2023.

[9] C. Gao, W. Zhang, G. Chen, and W. Lam, "Jsontuning: Towards generalizable, robust, and controllable instruction tuning," 2024.