



Impact of LLM on Urdu Text Summarization

Zainab Bibi ¹ Ayaz Mehmood ² Dr. Seemab Latif ²

Nust Islamabad



Overview

To evaluate the effectiveness of large language models (LLMs) on low-resource languages, this project utilizes quantized versions of LLaMA2, LLaMA3, and Mistral7B to study the effects of large language models on Urdu text summarization. This project uses the Parameter-Efficient Fine-Tuning (PEFT) technique to fine-tune these models on a variety of online sourced Urdu abstraction summaries data. To determine the quality of the summaries, the models are evaluated using ROUGE and BLUE scores. These results demonstrate how well-quantized LLMs and PEFT handle Urdu summarization, providing information and suggestions for further study and natural language processing applications.

Summary

Predicted Summary

اسرائیلی ریڈیو کے مطابق فوج کو غزہ کی پٹی میں رفاہ میں واقع سینکڑوں گھروں کو مسمار کرنے کے اختیارات دے دیئے گئے ہیں

رفاہ میں جھڑپوں سے 12 فلسطینی اور 5 اسرائیلی فوجی ہلاک ہوئے، بفر زون کی توسیع کے لیے گھروں کی مسماری جاری ہے

Summary

Predicted Summary

پاکستان کے زرمبادلہ کے بارہ ارب کے ذخائر کئی سالوں بعد پہلی دفعہ دباؤ کا شکار ہیں۔ پیسہ تیزی کے ساتھ ملک سے باہر جا رہا ہے جس کی وجہ سے روپے کی قیمت میں کمی جبکہ افراط زر کی شرح میں مسلسل اضافہ ہو رہا ہے۔ کینیڈل فلائیٹ کا سب سے بڑا سبب خلیجی ممالک میں سرمایہ کاری ہے۔ متعدد کمپنیاں جائیداد میں سرمایہ کاری کی سکیموں کے ذریعے خلیجی ممالک میں پیسہ کھینچ رہی ہیں

پاکستان میں محدود سرمایہ کاری مواقع اور کم منافع کی وجہ سے غیر ملکی کرنسی کی غیر قانونی منتقلی خلیجی ریاستوں کو ہو رہی ہے، جس سے روپے کی قیمت میں کمی اور تجارتی خسارہ بڑھ رہا ہے۔

Problem Statement

The summarization of Urdu, a low-resource language, is currently inadequate due to the limited availability of annotated data and the linguistic complexities. So, there is a significant need for more robust models capable of generating high-quality abstract summaries in Urdu. This study aims to evaluate the effectiveness of large language models such as LLaMA2, LLaMA3, and Mistral7B in enhancing the quality of summarization for Urdu texts.

Objectives

Some of the objectives of this paper are,

- Evaluate the effectiveness of advanced large language models (LLMs) such as LLaMA2, LLaMA3, and Mistral7B focusing on zero-shot and fine-tuned performance in summarizing Urdu text
- Improve the accuracy of Urdu text summarization utilizing parameter efficient fine-tuning (peft) which is a more effective model adaption as compared to traditional fine-tuning
- Implement optimized summarizing approaches into practice and validate them to achieve coherence in summaries.

Dataset

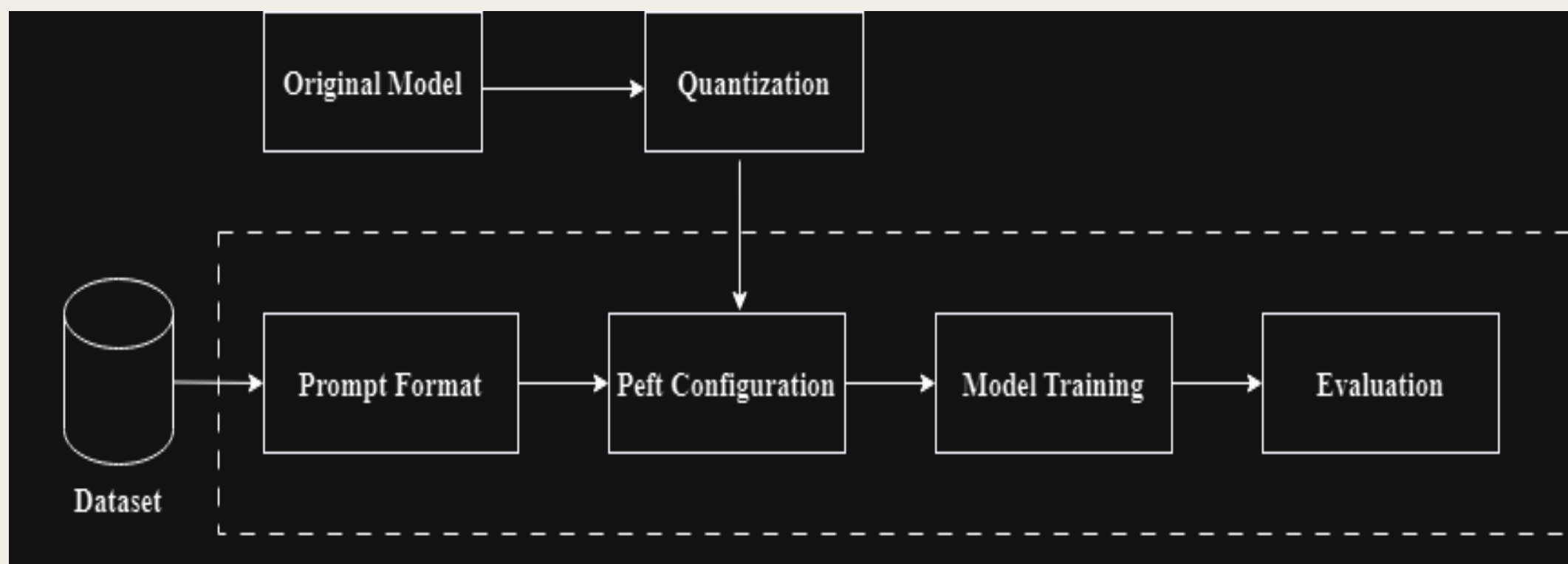
The dataset is composed of open-source Urdu abstraction summaries that were obtained from Kaggle. To help in evaluating summarization models, the data comprises question pairs along with matching summaries. The summaries offer concise abstractions of the material, while the questions represent complete text sections. Due to the lack of resources for this language, this dataset covers the last seven months and may be used to train and evaluate abstractive text summarization models in Urdu.

The dataset was obtained from the following Kaggle project, Text Summarization - Abstractive Urdu Text.

| Question | Answer |
|--|---|
| 8458 unique values | 8452 unique values |
| سرکاری عمارتوں پر قبضے کا واقعہ اس وقت پیش آیا ہے جب ایک روز قبل ہی روس کی حمایت کرنے والوں اور . . . یوکر | یوکرین کے حکام کا کہنا ہے کہ روسی اکثریت والے کرالیمیا میں مسلح افراد نے سرکاری عمارتوں پر قبضہ کر لی . . . |
| پوئل کی عمارت میں آگ لگ گئی ہے اور دھواں کے کالے پائل دکھائی دے رہے ہیں پوئل کی عمارت میں آگ لگ گئی . . . | ہمن سے اطلاعات کے مطابق منگل کو جنوبی شہر عدن میں تین مختلف دھماکوں کے واقعات پیش آئے ہیں جن میں سے . . . |
| حکام کا کہنا ہے کہ ظم ریلیز کرنے ملک بھر میں ایک بار مظاہرے شروع ہو سکتے ہیں یہ . . . دستاویزی ظم ہی ہی | یہاں کی حکومت نے سنہ 2012 میں دارالحکومت دہلی میں ہونے والے ریپہ پر پابندی . . . گئی دستاویزی ظم پر پابندی |

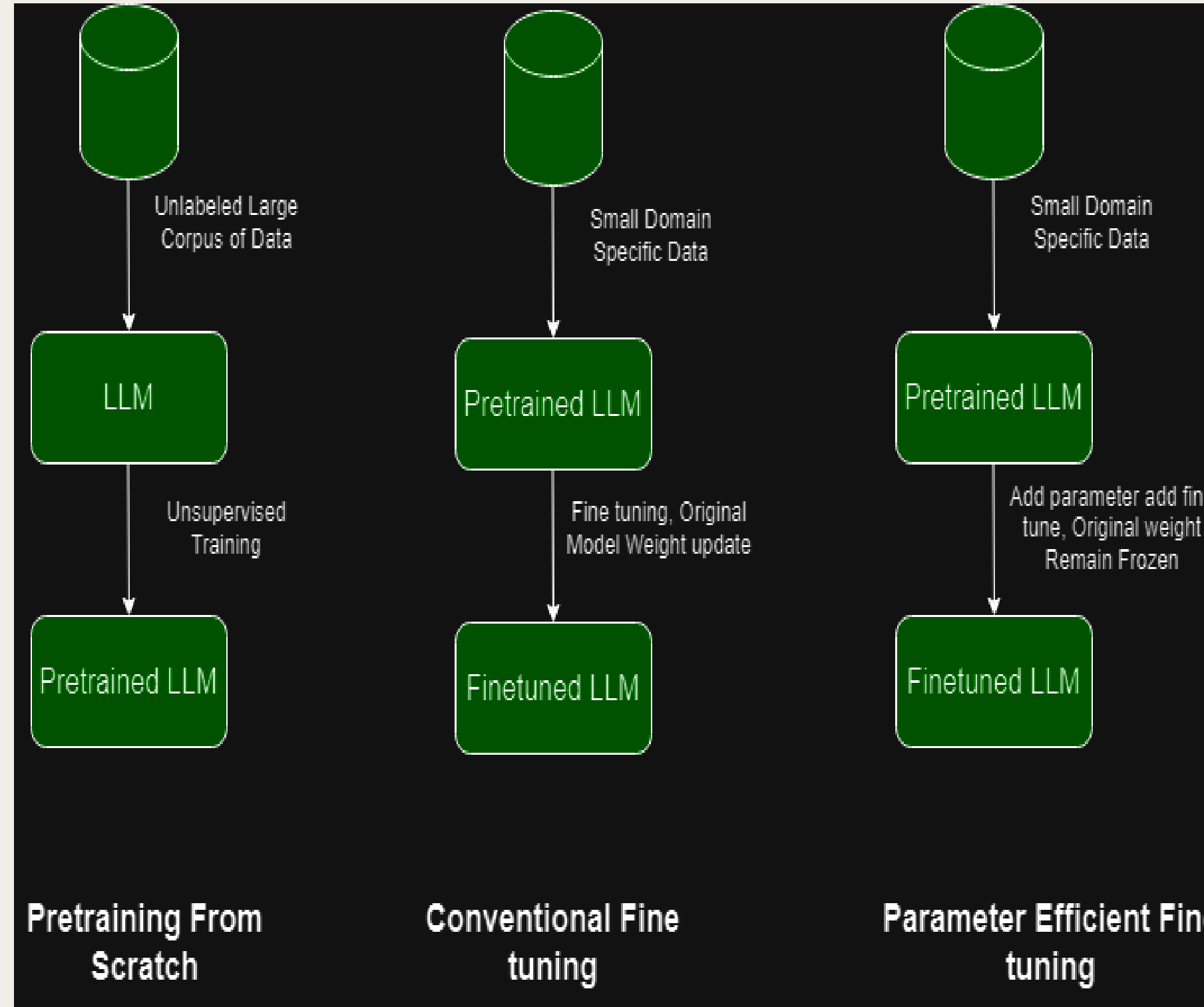
Methodology

- Dataset Preparation**, Begin with a comprehensive dataset that includes detailed text passages and corresponding abstractive summaries in Urdu.
- Quantization**, Convert the original language models (LLaMA2, LLaMA3, and Mistral7B) into their quantized versions due to limited resources.
- Prompt Format**, Design appropriate prompts to guide the model in generating summaries.
- PEFT Configuration**, Implement the Parameter-Efficient Fine-Tuning (PEFT) technique and Low-rank Adaptation instead of traditional fine-tuning to optimize the models for Urdu text summarization.
- Model Training**, Train the models using the quantized versions and low-rank adaptation on the train dataset.
- Evaluation**, Assess the performance of the trained models using standardized metrics to ensure the quality and accuracy of the summaries.



Aanalysis

In this project, we explore many methods of improving language models, such as parameter-efficient fine-tuning (PEFT), conventional fine-tuning, and pretraining from scratch, for Urdu text summarization. We use Low-Rank Adaptation (LoRA) with PEFT because of limited computing resources. This method is a good fit for our resource-constrained setting since it enables us to fine-tune by adding parameters while freezing the original model weights.



Results

The models' initial performance was quite low and required computational resources. Initially, we used only half of the data and fine-tuned a limited number of parameters due to computational resources. By increasing the number of fine-tuned parameters using LORA parameters r and alpha, increasing the training epochs, and utilizing Colab Pro+ for GPU resources, we observed significant improvements in results. Furthermore, using the full dataset further enhanced the outcomes.

| Model | ROUGE Score | BLEU Score |
|-----------|-------------|------------|
| LLaMA3 | 12.43 | 15.27 |
| LLaMA2 | 5.32 | 7.41 |
| Mistral7B | 1.63 | 4.17 |

Discussion and Future Work

Discussion

The zero-shot performance of the models was zero and even after training the model on complete data, the results were not very good. LLaMA3 achieved quite good results then LLaMA2 and Mistral7B, and the evaluation matrix ROUGE and BLEU ratings of 12.43 and 15.27, respectively. However, by utilizing PEFT with LoRA, we were able to optimize our resource utilization and attain some level of improvement via hyper parameter tuning. This suggests that significant enhancements in Urdu text summarizing might be made with more thorough training, increase in data and improved resources.

Future Work

Given the current limitation and low performance, there is a need to explore additional large language models and fine-tuning techniques to identify more effective approaches to improve the performance of LLM on Urdu summarization. To improve the quality and consistency of the generated summaries, it will also be essential to add more extensive and diverse datasets for Urdu text. Finally, further research into novel structures and language modeling methodologies could provide more effective solutions for Urdu text summarizing tasks.RAZAe 2024

References

- RAZAe, W. S. H. (2024). *End to end urdu abstractive text summarization with dataset and improvement in evaluation metric.*