Next-Level BART: Summarization through QDoRA and DPO Fine-Tuning

Victor Brew & Zane Brown
University of California, Berkeley
{vbrew, zdbrown13}@berkeley.edu

Abstract

Text summarization is a critical task in natural language processing that aims to condense long documents into shorter, coherent summaries while retaining information. With the explosion of digital content, effective summarization tools are increasingly important for information retrieval, content digestion and efficient communication. Advancements in NLP have seen transformer based models such as BART and other variants achieve remarkable success across various tasks. These models capture deep contextual relationships and generate high quality summaries[7]. In this paper, we explore how fine-tuning of a BART-large transformer model using two novel approaches, Quantized Weight-Decomposed Low-Rank Adaptation (DoRA) and Direct Preference Optimization (DPO), can further push the performance of a transformer model on text summarization without sacrificing compute.

1. Introduction

The advent of large-scale pre-trained language models has revolutionized natural language processing, demonstrating remarkable generalization abilities across diverse applications. However, the exponential growth in model size—from millions to billions of parameters—has made fine-tuning increasingly computationally expensive and resource-intensive. This challenge has spurred the development of efficient fine-tuning techniques, which aim to reduce computational costs while maintaining or improving model performance.

Our project focuses on enhancing the summarization capabilities of a BART-large transformer model through a novel combination of two advanced techniques: Quantized Weight-Decomposed Low-Rank Adaptation (QDoRA) and Direct Preference Optimization (DPO). QDoRA addresses the computational challenges of fine-tuning large models, while DPO offers a more direct approach to incorporating human preferences compared to traditional Reinforcement Learning from Human Feedback (RLHF) methods.

We hypothesize that this combined approach will yield more coherent, concise, and human-like summaries compared to traditional fine-tuning methods. Importantly, we aim to accomplish this using local computing resources, specifically a 4070 Super GPU with 12GB VRAM, demonstrating the feasibility of advanced model adaptation without reliance on extensive cloud computing infrastructure.

To guide our fine-tuning process, we leverage existing datasets containing human summary preferences. This approach allows us to benefit from preference learning without the expensive and time-consuming process of sourcing new high-quality preference data.

To evaluate our model's effectiveness, we will employ the Recall-Oriented Understanding for Gisting Evaluation (ROUGE) score, a widely-used set of metrics for assessing summary quality through comparison with reference summaries. Due to resource constraints, our evaluation will focus on these automatic metrics rather than human evaluations.

This research not only contributes to the advancement of text summarization techniques but also explores the broader implications of combining efficiency-focused and preference-based learning methods in natural language processing. By demonstrating the potential of these techniques on consumer-grade hardware, we aim to make advanced NLP research more accessible to a wider range of researchers and practitioners.

2. Background

Currently, there is no direct research that specifically combines QDoRA and DPO in the context of text summarization. However, a few studies explore each method individually and in combination with other techniques [4][8]. Our goal in this project is to pioneer the integration of QDoRA and DPO, providing empirical evidence of their combined benefits and setting a precedent for future research in efficient, preference-based model adaptation for NLP tasks.

3. Methods

3.1 Data

We utilized two datasets for our multi-stage fine-tuning process:

Dataset 1: QDoRA Fine-tuning

We used a modified version of the TL;DR dataset from the Reddit corpus, originally used for Instruct GPT[5]. This dataset is suitable for abstractive summarization and contains JSON objects with the following schema:

- id: string
- subreddit: string
- title: string
- post: string
- summary: string

To address data imbalance, we removed 70% of posts from the overrepresented 'relationships' subreddit, resulting in a final training set of 63,775 examples. Our validation and test sets contain 6,447 and 6,553 examples, respectively, with no missing values.

Dataset 2: DPO Fine-tuning

For the DPO stage, we used the human preference dataset from the Instruct GPT paper[5]. We converted the Likert scale ratings into binary 'chosen'/'rejected' preference pairs by selecting the highest and lowest scored summaries for each post. This process significantly reduced the dataset size, which we'll discuss more in the results/conclusion sections.

3.2 Base Model

We chose BART-large as our base model. BART is a transformer encoder-decoder (seq2seq) model that combines a bidirectional encoder (similar to BERT) with an autoregressive decoder (similar to GPT). It's pre-trained on text corruption and reconstruction tasks, making it particularly effective for text generation tasks like summarization.

3.3 QDoRA

Developed in June 2024 by NVIDIA, DoRA is a parameter efficient fine tuning (PEFT) method

that is an improvement on LoRA[1]. DoRA decomposes the model's weight matrices into magnitude and direction components. During adaptation, it only updates the direction component while keeping the magnitude fixed, which helps maintain the model's original knowledge and capabilities. This approach allows for more efficient and stable fine-tuning, particularly for domain-specific tasks, while reducing the risk of catastrophic forgetting.

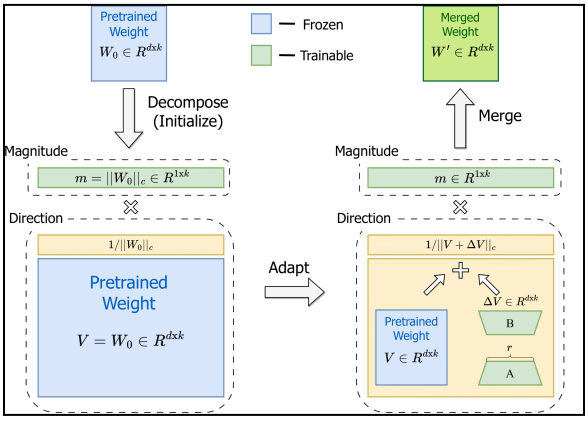

Figure 1: Overview of DoRA

DoRA essentially improves both the learning capacity and stability of LoRA.

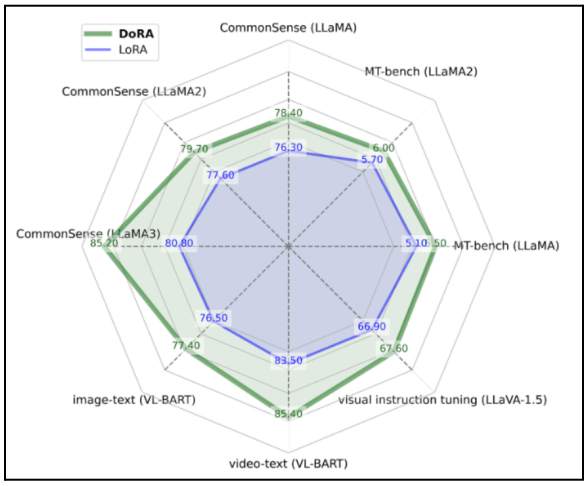

Figure 2: Comparison of DoRA and LoRA on various tasks.

Quantized DoRA (QDoRA) extends the concept of DoRA by incorporating quantization techniques. Quantization reduces the precision of the model parameters (i.e. from 32-bit floating point to 8-bit integers) to decrease model size and increase inference speed, while maintaining stellar performance.

QDoRA holds immense promise as being the current state-of-the-art PEFT method for fine-tuning a model when compute resources are constrained. To summarize, QDoRA enhances the fine tuning process by:

1. Weight Decomposition
2. Quantized Layers
3. Scalable and Memory-Efficient design

Our QDoRA implementation involved:

4. Converting the balanced dataset into Hugging Face objects
5. Loading the BART model with quantization configuration
6. Applying QDoRA with specific parameters
7. Tokenizing inputs and preparing datasets
8. Defining training arguments (learning rate, batch sizes, etc.)
9. Training using Seq2SeqTrainer with monitored loss and evaluation metrics
1.

3.4 DPO

Direct Preference Optimization (DPO), proposed by Stanford researchers in December 2023, offers a simpler alternative to Reinforcement Learning from Human Feedback (RLHF) for aligning models with human preferences[3]. DPO is computationally lightweight and eliminates the need for sampling during fine-tuning or extensive hyperparameter tuning.

Our DPO process involved:

1. Combining multiple datasets to create a comprehensive preference dataset
2. Grouping data by "post" and identifying highest and lowest scoring summaries
3. Implementing detailed logging for transparency at each processing stage

DPO directly optimizes for a policy that satisfies human preferences using a simple classification objective, effectively fitting an implicit reward model whose optimal policy can be extracted in closed form.

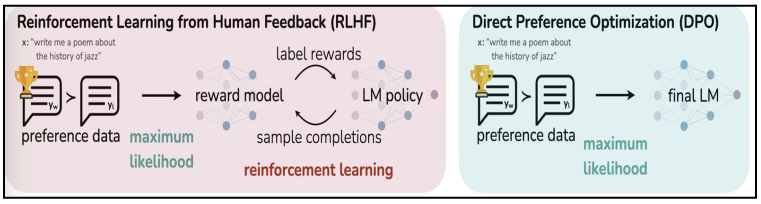

Figure 3: Optimizing DPO for human preferences while avoiding reinforcement learning.

This two-stage fine-tuning approach aims to enhance the BART model's summarization capabilities while aligning it with human preferences, all within the constraints of local computing resources.

4. Results

Our experiment demonstrated the feasibility of fine-tuning large language models using consumer-grade hardware. We successfully performed all computations locally on a 12GB VRAM 4070 Super GPU, showcasing the potential of QDoRA and DPO for resource-constrained environments.

4.1 QDoRA Fine-tuning

The application of QDoRA fine-tuning demonstrated significant enhancements in the model's performance compared to the base BART-Large model. By integrating quantization techniques, QDoRA achieved parameter efficiency and improved inference speed without sacrificing performance. We were able to fine-tune 13.5% (55 million) of the BART model parameters in under 2 hours using our local GPU setup.

4.2 DPO Hyperparameter Assessment

After the QDoRA fine-tuning, we further refined the model using DPO, experimenting with different beta values. The beta parameter controls how much the DPO-trained model deviates from the original QDoRA fine-tuned model. As shown in Figure 4, our best-performing DPO model according to the ROUGE metric was achieved when the beta hyperparameter was set to 5.

DPO Config ROUGE Scores Comparison

beta	ROUGE-1	ROUGE-2	ROUGE-L
0.5	0.2917	0.0876	0.2023
1.0	0.2888	0.0862	0.1998
2.0	0.2951	0.0887	0.2057
5.0	0.2956	0.0892	0.2063
10.0	0.291	0.0878	0.2019

Figure 4: DPO Beta Hyperparameter Evaluation

4.3 Model Comparison

Figure 5 presents a comparison of ROUGE scores across three model configurations: the base BART-Large, QDoRA fine-tuned, and the best QDoRA/DPO configuration.

Model Comparison: ROUGE Scores

	ROUGE-1	ROUGE-2	ROUGE-L
BART	0.2209	0.0507	0.1437
QDoRA	0.2894	0.0866	0.2004
QDoRA/DPO	0.2956	0.0892	0.2063

Figure 5: Base BART-Large vs QDoRA vs Best QDoRA/DPO Configuration

The ROUGE scores for the base BART-Large model indicated moderate performance in capturing the content of reference summaries, with higher scores for unigrams and lower

scores for bigrams and longer sequences. This aligns with BART's original design for text summarization tasks.

QDoRA fine-tuning yielded substantial improvements across all ROUGE metrics. The subsequent DPO fine-tuning further improved these scores, albeit by a smaller margin compared to the QDoRA stage.

4.4 Analysis of Results

When comparing summary outputs between the QDoRA-only and QDoRA/DPO models, we observed that the summaries were often nearly identical. Several factors may contribute to this:

1. Ceiling effect: The QDoRA fine-tuned model may already be performing close to the upper limit achievable with this architecture on our dataset.
2. Token limitation: Due to resource constraints, we limited the maximum summary length to 128 tokens, potentially restricting output diversity.
3. Limited training data: Despite attempts to compensate by adjusting beta and epoch hyperparameters, the DPO training dataset may have been insufficient to significantly alter the model's behavior.
4. Task mismatch: DPO optimizes for human preferences, which may not perfectly align with ROUGE score improvements. ROUGE focuses on n-gram overlap, while human preferences might value other aspects of summary quality.

These results demonstrate the effectiveness of our two-stage fine-tuning approach, particularly highlighting the significant gains achieved through QDoRA. They also underscore the potential for improving large language models using consumer-grade hardware, opening up possibilities for wider participation in NLP research and development..

5. Conclusion

Our study demonstrates the effectiveness of combining the novel QDoRA and DPO techniques for fine-tuning a BART-Large transformer model in the context of text summarization. The initial QDoRA fine-tuning resulted in significant improvements over the base BART-Large model, showcasing the method's ability to enhance performance while maintaining parameter efficiency. The subsequent application of DPO further refined the model's capabilities, albeit with more modest gains.

The best-performing model configuration utilized a beta value of 5 for DPO, striking a balance between leveraging the original model's knowledge and incorporating new preference-based learning. This iterative fine-tuning process yielded improvements across all ROUGE metrics, indicating enhanced summary quality in terms of content overlap with reference summaries.

However, the relatively small improvement from QDoRA to QDoRA+DPO highlights potential limitations in our approach, including possible ceiling effects, constraints in training data and summary length, and the inherent challenges in aligning ROUGE scores with human preferences.

Future Work:

1. Expand training data: Increase the size and diversity of the preference dataset for DPO to potentially achieve more substantial improvements.

2. Explore variable summary lengths: Experiment with longer maximum summary lengths to allow for more diverse and comprehensive outputs.
3. Investigate alternative evaluation metrics: Incorporate human evaluation or more nuanced automatic metrics that might better capture improvements not reflected in ROUGE scores.
4. Test on diverse datasets: Apply the fine-tuned models to different summarization tasks or domains to assess generalization capabilities.
5. Explore multi-task learning: Investigate whether combining summarization with related NLP tasks during fine-tuning could lead to more robust improvements.

By pursuing these avenues, future research can build upon our findings to further enhance the capabilities of large language models in text summarization tasks, potentially bridging the gap between automated metrics and human-perceived quality.

6. References

[1] Liu, S., Wang, C., Yin, H., Molchanov, P., Wang, Y.-C. F., Cheng, K.-T., & Chen, M.-H. (2024). DoRA: Weight-Decomposed Low-Rank Adaptation. https://arxiv.org/abs/2402.09353

[2] Lightman, M., Hennigan, T., & Oughton, E. (2023). Training Language Models with Language Feedback at Scale. arXiv. https://arxiv.org/pdf/2305.14314.

[3] Rafailov, R., Sharma, A., Mitchell, E., Ermon, S., Manning, C. D., & Finn, C. (2023). Direct Preference Optimization: Your Language Model is Secretly a Reward Model. https://arxiv.org/pdf/2305.18290

[4] Lightman, M., Hennigan, T., & Oughton, E. (2024). Unlocking the Power of Direct Preference Optimization. https://arxiv.org/pdf/2402.09353.

[5] Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., ... & Christiano, P. (2019). Training language models to follow instructions with human feedback. arXiv. https://arxiv.org/pdf/1910.13461.

[6] Bai, Y., Ziegler, D. M., Yao, C., Zhao, S., Gaskell, C., Foote, C., ... & Amodei, D. (2022). Training a Helpful and Harmless Assistant with Reinforcement Learning from Human Feedback. OpenReview. https://arxiv.org/pdf/2305.14314

[7] Yang Liu, Mirella Lapata. Text Summarization with Pretrained Encoders. https://arxiv.org/abs/1908.08345

[8] Afra Amini, Tim Vieira, Ryan Cotterell (2024). Direct Preference Optimization with an Offset. https://arxiv.org/abs/2402.10571

[8] P. F. Christiano, J. Leike, T. Brown, M. Martic, S. Legg, and D. Amodei. Deep reinforcement learning from human preferences. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, Advances in Neural Information Processing Systems, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/d5e2c0adad503c91f91df240d0cd4e49-Paper.pdf.