Spreading Factor Impact on Llama3 QLoRA Fine-Tuning Performance in QuixBugs

Assignee Research

June 1, 2026

Abstract

This report synthesises findings from 7 peer-reviewed papers addressing the following research question: How does the choice of spreading factor (SF) in LoRa modulation affect the F1-score performance of Llama3 on QuixBugs when using QLoRA fine-tuning compared to full fine-tuning. Pre-training Large Language Models (LLMs) on web-scale datasets becomes fundamental for advancing general-purpose AI. In contrast, enhancing their predictive performance on downstream tasks typically involves adapting their knowledge through fine-tuning. 10 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Layer-wise LoRA fine-tuning: a similarity metric approach. Research question: How does the choice of spreading factor (SF) in LoRa modulation affect the F1-score performance of Llama3 on QuixBugs when using QLoRA fine-tuning compared to full fine-tuning?.

2 Methodology

Systematic literature search across multiple databases yielded 7 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.5/10.

3 Results

7 papers retrieved. 10 claims extracted; 1 independently verified. Quality review score: 4.5/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
The experiments were conducted on encoder-only transformer architectures with RoBERTa-base and DeBERTa-v3base, and on de	×	0.04
For RoBERTa-base and DeBERTa-v3base, the LoRA rank was set to 8 and the alpha to 16.	×	0.02
For LLaMA 2-7B, Mistral-7B-v0.1, and Gemma-7B, the LoRA rank and alpha were set to 128.	×	0.03
For the multimodal model LLaVA-1.5-7B, the LoRA rank and alpha were set to 128.	×	0.03
The method was evaluated on the GLUE benchmark for NLU tasks using RoBERTa-base and DeBERTa-v3base.	×	0.06
The method achieved a parameter reduction of 50% with only an average drop in predictive performance of 0.27 percentage	×	0.09
The method by Lodha et al. 2023 suffered from an average drop of 7.68 percentage points with the same configuration.	×	0.02
The method identifies subsets of layers to fine-tune that exhibit competitive predictive performance while significantly	✓	0.15
LoRA adapts two matrices B $\in$Rd$\times$r and A $\in$Rr$\times$k, with r min(d, k), and their product is the update ΔW $\in$Rd$\times$k, further scale	×	0.03
The method reduces the number of trainable parameters while preserving predictive performance compared to fine-tuning al	×	0.14

References

- http://arxiv.org/abs/2412.09827v1

- http://arxiv.org/abs/2411.14961v3
- http://arxiv.org/abs/2602.05988v1