

DONOD Threshold Variation and Its Impact on Training Efficiency and Cross-Domain Generalization in LLaMA-2 Models

Assignee Research

May 30, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: How does varying the DON threshold in DONOD affect the trade-off between training efficiency (measured in training tokens reduced) and cross-domain generalization (measured by performance on MBPP and Ad-hoc instruction fine-tuning of large language models (LLMs) is widely adopted for domain-specific adaptation. While domain-specific supervised fine-tuning (SFT) is effective and efficient, it often weakens cross-domain generalization and struggles with noisy training data. 6 claims were extracted from source literature; 3 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 6.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: DONOD: Efficient and Generalizable Instruction Fine-Tuning for LLMs via Model-Intrinsic Dataset Pruning. Research question: How does varying the DON threshold in DONOD affect the trade-off between training efficiency (measured in training tokens reduced) and cross-domain generalization (measured by performance on MBPP and HumanEval) when applied to LLaMA-2-13B compared to LLaMA-2-7B?.

2 Methodology

Systematic literature search across multiple databases yielded 15 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 6.5/10.

3 Results

15 papers retrieved. 6 claims extracted; 3 independently verified. Quality review score: 6.5/10.

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
DONOD improves target-domain accuracy by 14.90% and cross-domain accuracy by 5.67% by filtering out 70% of the dataset.	✓	0.24
Data pruned by smaller models using DONOD generalizes effectively on larger models.	✓	0.16
DONOD demonstrates comparable or superior performance to existing methodologies while being dataset-agnostic.	✓	0.18
The evaluation benchmark incorporates AGIEval and IFEval to assess logical reasoning, mathematics, reading comprehension	×	0.06
DONOD achieves state-of-the-art performance in core reasoning tasks such as math and logic, outperforming full-data base	×	0.06
DONOD enables lossless training acceleration with only 20% of the data.	×	0.03

References

- <http://arxiv.org/abs/2504.14810v2>
- <http://arxiv.org/abs/2505.18227v3>
- <http://arxiv.org/abs/2308.13515v5>