Evaluating the Stability and Generalization of Parameter-Efficient Adaptation for Low-Resource Indic Languages

Submitted by: Adil Mohamed

sadilmohammed0004@gmail.com

Research Proposal

Contents

1 Abstract

Parameter-efficient fine-tuning (PEFT) methods (LoRA, BitFit, IA³) adapt pretrained language models using far fewer trainable parameters than full fine-tuning. This project systematically compares PEFT and full fine-tuning for low-resource Indic languages, focusing on (i) stability (variance across random seeds and consistency of learned predictions/representations) and (ii) generalization (held-out test performance and cross-lingual transfer) on public benchmarks such as IndicGLUE and IndicGenBench. Multi-seed experiments across representative classification and generation tasks will quantify trade-offs between parameter efficiency, reliability, and task performance.

2 Introduction / Background & Significance

2.1 Why Pretrained Models Struggle with Low-Resource Indic Languages

India's linguistic diversity is immense, yet most pretrained models are trained on high-resource language data. Indic languages suffer from scarce labeled data and limited benchmarks, making stable adaptation critical. In such regimes, full fine-tuning becomes unstable.

2.2 Why Full Fine-Tuning is Unstable in Low-Data Settings

Fine-tuning all parameters with limited data results in high variance and poor reproducibility. Prior work (e.g., Mosbach et al., 2020) shows substantial variance due to random seeds. These problems are amplified in Indic contexts due to smaller datasets (100–3k examples).

2.3 Why Stability and Generalization Are Under-Studied

Although PEFT methods are known to be computationally efficient, their effects on variance and generalization in Indic NLP are unclear. Prior work mostly evaluates high-resource benchmarks. Our study adapts metrics and evaluation from Du
Nguyen (2023), and Mai et al. (2025) to the low-resource setting.

3 Research Aim and Questions

Main Aim: Evaluate and compare the stability and generalization of PEFT methods versus full fine-tuning in low-resource Indic NLP.

Research Questions:

RQ1 Do PEFT methods exhibit lower test accuracy variance across random seeds than full fine-tuning?

RQ2 Do PEFT methods generalize comparably or better than full fine-tuning in small-data settings?

RQ3 Does PEFT improve zero-shot cross-lingual transfer from higher- to lower-resource languages?

4 Research Design and Methodology

4.1 Base Models and Adaptation Methods

XLM-RoBERTa or IndicBERT will be used as base models. Four strategies will be compared: Full fine-tuning, LoRA, BitFit, and IA^3. All experiments will be standardized by learning rates, schedulers, and model checkpoints.

4.2 Tasks and Languages

Classification (News Category, NER) and Generation (MT, QA) from IndicGLUE and IndicGenBench. Languages: Hindi, Tamil (high-resource); Assamese, Odia (medium); Konkani, Santali (low).

4.3 Evaluation

- **Stability:** Accuracy/F1 mean ± SD, disagreement across seeds, representation similarity.
- **Generalization:** Held-out test performance, overfitting gap, cross-lingual transfer accuracy.
- **Experimental scale:** 10–15 seeds × 4 methods × 6 languages × 2–3 tasks = 500–800 runs.

4.4 Feasibility

Each run costs 15–45 GPU minutes. This is tractable with 2–4 GPUs over 3–4 weeks. Prioritized focus on classification tasks ensures coverage under resource constraints.

5 Expected Outcomes and Pitfalls

- **Stability Gains:** PEFT methods expected to yield lower variance.
- **Generalization:** Minimal trade-off; PEFT may outperform via implicit regularization.
- **Transfer:** PEFT may better preserve cross-lingual features.

Potential Pitfalls:

- Underfitting $\rightarrow mitigatewithparametersweeps.$
- High seed variance $\rightarrow usebootstrapconfidenceintervals.$
- Resource limits $\rightarrow focusonkeylanguagesandtasks.$

6 Timeline

- **Weeks 1–2:** Setup
- **Weeks 3–4:** Implement LoRA, BitFit, IA^3
- **Weeks 5–8:** Main experiments
- **Weeks 9–10:** Transfer evaluation, analysis
- **Weeks 11–13:** Write-up and documentation

References

1. Mosbach, M. et al. (2020). On the Stability of Fine-tuning BERT.
2. Du, S. & Nguyen, T. (2023). Measuring the Instability of Fine-Tuning.
3. Hu, E. J. et al. (2021). LoRA: Low-Rank Adaptation.
4. Zaken, E. et al. (2022). BitFit.
5. Kakwani, D. et al. (2020). IndicBERT.
6. Mai, K. et al. (2025). Unifying Study of PEFT.