

CONTINUAL FINE-TUNING AND STABLE ADAPTATION OF LARGE MODELS

Project Description

Modern AI systems increasingly rely on adapting large pretrained models to new tasks through continual fine-tuning. However, sequential updates often lead to catastrophic forgetting and instability, where newly learned capabilities degrade previously acquired knowledge.

This project investigates scalable algorithms for **continual model merging** and **stability-aware fine-tuning**, focusing on projection-based parameter updates that reduce interference between tasks while preserving previously learned capabilities. The central objective is to understand how parameter updates interact across tasks and to design scalable methods for stable continual adaptation of large models.

Project Type

- Research–Engineering balance: **50% research** and **50% implementation**.
- Research components include studying recent literature on continual learning and model merging, analyzing parameter-space geometry of fine-tuned models, and evaluating stability and knowledge preservation across tasks.
- Engineering components include implementing task-vector merging algorithms, building experimental pipelines for sequential fine-tuning, and evaluating models on multi-task benchmarks.

Duties / Activities

- Implement task-vector based model merging methods.
- Reproduce baseline algorithms for continual learning and model merging.
- Develop projection-based techniques to reduce interference between tasks.
- Conduct experiments on sequential multi-task benchmarks.
- Analyze parameter updates and representation drift during sequential training.
- Evaluate catastrophic forgetting and task performance across sequential updates.
- Document experimental findings and prepare technical reports.
- Outstanding interns may contribute to a research manuscript or open-source implementation.

Required Skills

- Strong Python programming skills.
- Experience with PyTorch or similar deep learning frameworks.
- Background in machine learning.
- Solid foundations in linear algebra and optimization.
- Experience training or fine-tuning neural networks.

Preferred Academic Level

- Senior undergraduate students in Computer Science, Artificial Intelligence, or related fields.
- Master's students in Machine Learning, Data Science, or Applied Mathematics.

Learning Opportunities

- Research experience in continual learning and model merging.
- Training and evaluation of large neural networks.
- Understanding parameter-space geometry and stability of deep models.
- Exposure to reproducible machine learning experimentation.
- Participation in ongoing research directions on large generative models.

Expected Team Size

It is preferable to have a team of 2 interns.

Deliverables / Expected Outcomes

- Reproducible experimental pipeline for sequential model merging.
- Benchmark results analyzing stability and catastrophic forgetting.
- Documented code repository.
- Final technical report summarizing methods and findings.
- Potential contribution to a research manuscript.

Mentor

[Dr. Abdelkader Baggag](#)