

PARAMETER-SPACE MODEL MERGING FOR LARGE LANGUAGE MODELS

Project Description

Modern AI systems increasingly rely on sequential fine-tuning of large pretrained models across tasks and domains. However, this sequential adaptation often leads to catastrophic forgetting, where previously learned capabilities degrade when new tasks are introduced.

This project studies **continual model merging**, an alternative paradigm where task-specific models are combined through structured parameter updates rather than retraining from scratch. The goal is to develop scalable algorithms that integrate new task knowledge while preserving previously learned representations.

The project investigates projection-based merging strategies that operate directly in parameter space. These methods analyze the geometry of parameter updates and construct orthogonal update directions that reduce interference between tasks. By exploiting low-rank structure and task-vector representations, the project aims to design efficient merging algorithms that support stable continual learning in large neural models.

Project Type

- Research–Engineering balance: **50% research** and **50% implementation**.
- Research tasks include studying recent literature on continual learning and model merging, analyzing parameter-space geometry of task updates, and evaluating stability of sequential adaptation.
- Engineering tasks include implementing merging algorithms, building experimental pipelines, and running large-scale experiments.

Duties / Activities

- Implement baseline methods for sequential fine-tuning and model merging.
- Develop projection-based merging strategies that reduce task interference.
- Analyze task vectors and parameter update geometry.
- Run experiments on sequential multi-task learning benchmarks.
- Evaluate catastrophic forgetting and task retention.
- Document experimental results and prepare technical reports.

Required Skills

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

Preferred Academic Level

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

Expected Team Size

- 2 interns working collaboratively under the supervision of the research team.

Research Outcomes

- Implementation of scalable continual model merging algorithms.
- Experimental evaluation of stability and forgetting across sequential tasks.
- Reproducible experimental pipeline and documented codebase.
- Final technical report summarizing methods and findings.
- Potential contribution to a research manuscript or open-source release.

Mentor

[Dr. Abdelkader Baggag](#)