

7 × Machine Learning Conference Competitions

MLRC-Bench

Correlation?

Objective

Effectiveness

Efficiency

Simplicity

Scorer

Subjective

Validity

Clarity

Rigorousness

Generalizability

Innovativeness

LLM Judge

Underlying Idea

It identifies key parameters via dynamic importance, adaptively reduces them ...

LLM Explainer

Implementation

environment/

├ data

├ test.jsonl

├ ...

└ methods

└ MyAwesomeMethod.py

edited

edited

NeurIPS 2024 LLM Merging Competition

Task Description

Develop a method for merging several pretrained language models into a generalist model ...

Human Idea

TIES Merging (Yadav et al., 2023) involves three steps: ... The merged task vector is then scaled and added to the base model: $\theta_m^p = \theta_{\text{base}} + \lambda \frac{1}{|A^p|} \sum_{i \in A^p} \hat{\tau}_i^p$

Starter Code

read-only no-read

environment/

├ data

├ dev.jsonl

├ test.jsonl

├ evaluation.py

├ main.py

└ methods

└ BaseMethod.py

└ MyMethod.py

Language Agent

Ideation Coding Execution Debugging ...

Figure 1. Overview of our MLRC-BENCH and the evaluation pipeline.