Landmark-Guided Policy Optimization for Multi-Objective Language Model Selection

Márcio Monteiro · Weichen Li · Puyu Wang · Marius Kloft · Sophie Fellenz
RPTU University Kaiserslautern-Landau, Germany

Motivation

Fine-tuning LLMs is the dominant paradigm for achieving SOTA performance in specialized tasks. But how to choose a base model to fine-tune?

Common but Suboptimal Approaches

Select the largest model possible

- Smaller specialized models may outperform larger general-purpose ones.
- Larger models can incur higher deployment costs and greater environmental impact.

Select high-scoring models from benchmarks

- High scores on benchmarks may not transfer / align well to the new target task (no free lunch).

Test as many base models as possible

- Search costs can be prohibitively high.

The LAMPS Framework

Open-source AutoML Framework
https://github.com/ML-LAMPS/lamps

Multi-objective: Tracks Pareto-efficient models using the hypervolume indicator over arbitrary objectives (*e.g.*, performance vs. model size).

Landmark Fine-Tuning: Fine-tunes and evaluates candidate models on incrementally larger subsets of the training data (multi-fidelity method).

Meta-Learning via Reinforcement Learning:
Learns an efficient search policy from past fine-tuning trajectories stored in a meta-dataset.

Multi-Objective Formulation

Consider a target dataset $\mathcal{D}$ and a set $\mathcal{X}$ of candidate pretrained language models to be fine-tuned. Given n metrics of interest (objectives), the problem can be formulated as a multi-objective optimization problem:

$$\min_{x \in \mathcal{X}} \quad \big(f_1(x, \mathcal{D}), \ldots, f_n(x, \mathcal{D})\big)$$
$$\text{s.t.} \quad f_i(x, \mathcal{D}) \leq f_i^{\max} \quad \text{for all } i = 1, \ldots, n,$$

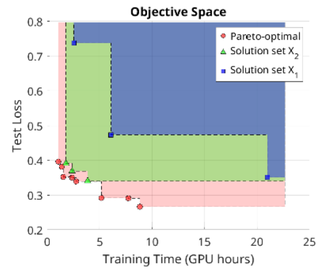
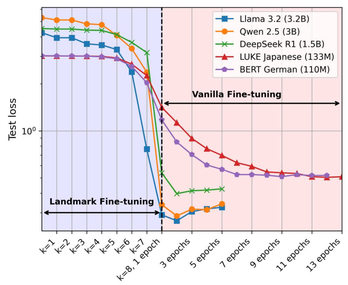


Illustration of the hypervolume indicator in a bi-objective setting, corresponding to the shaded areas. The higher the hypervolume, the closer the solution set is to the true Pareto front.

Landmark Fine-Tuning: 20 Newsgroups / 5 Base Models

Larger models:

- Start with have **higher initial loss**
- Eventually **outperform smaller models**



Early learning behavior:

- Models that improve **more rapidly in the early stages** tend to achieve a **lower final loss**.

Key takeaway:

- The **initial portion of the learning curve** can help predict the overall model **performance**.

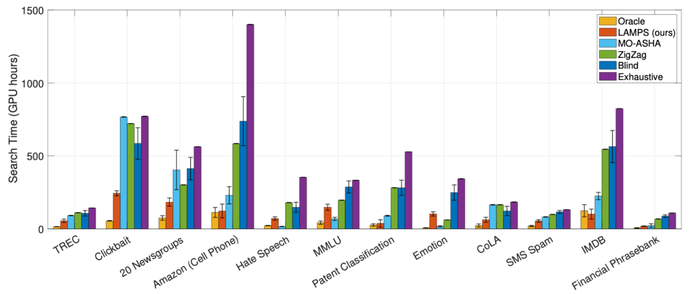
Search Algorithm

Algorithm 1 LAMPS Search Procedure

Input: training dataset $\mathcal{D}^{\text{train}}$, validation dataset $\mathcal{D}^{\text{val}}$, policy π_θ, number of landmarks K
Output: set of non-dominated fine-tuned models $\hat{X}^\star$
Initialize time step $t \leftarrow 0$
Initialize selected model set $X \leftarrow \varnothing$
Initialize landmark level for each model x_i: $k_i \leftarrow 0$
Evaluate candidate models on $\mathcal{D}^{\text{val}}$ and construct initial state s_0
repeat
  Select action $a_t \leftarrow \arg\max_a \pi_\theta(a \mid s_t)$
  Update landmark level: $k_{a_t} \leftarrow \min(k_{a_t} + 1, K)$
  Fine-tune x_{a_t} for one epoch on $\mathcal{D}^{\text{train}}_{k_{a_t}}$
  Evaluate updated performance on $\mathcal{D}^{\text{val}}$
  if stopping criterion met for model x_{a_t} **then**
    $X \leftarrow X \cup \{x_{a_t}\}$
  end if
  Update environment state s_{t+1}
  $t \leftarrow t + 1$
until search budget is exhausted
Compute the non-dominated set $\hat{X}^\star = \{x \in X \mid \nexists y \in X : y \succ x\}$
return $\hat{X}^\star$

Main Results

- Mean search time (GPU-hours) to reach **99% of the optimal hypervolume indicator** on held-out datasets. For reference, also showing the time required to complete an exhaustive search.
- **LAMPS reduces search time by 73.6%** on average, compared with exhaustive search and outperforms other feasible methods on 9 out of 12 datasets.



Case Study: Amazon (Cell Phone)

Reference GPU: 1x NVIDIA A100 40GB at $3.67 hourly

Search Method	Cost	Reduction
Exhaustive (baseline)	$ 5,141.67	-
MO-ASHA	$ 845.20	↓ 83.5 %
LAMPS (ours)	**$ 449.58**	↓ **91.2 %**

Key Takeaways

- ✓ LAMPS demonstrates an average of **73.6% reduction in search time** across 12 held-out text classification tasks.
- ✓ LAMPS outperforms other methods in 9 out of 12 held-out datasets.

References

[1] You et al. LogME: Practical assessment of pre-trained models for transfer learning, ICML 2021.
[2] Arango et al. Quick-Tune: Quickly learning which pretrained model to finetune and how. ICLR 2024.
[3] Pfahringer et al. Meta-learning by landmarking various learning algorithms. ICML 2000.

Correspondence to:
Márcio Monteiro <marcio.monteiro@cs.rptu.de>

Forty-Third International Conference on Machine Learning
(ICML 2026)

Funded by
Deutsche Forschungsgemeinschaft
German Research Foundation
Federal Ministry of Research, Technology and Space
Carl Zeiss Stiftung