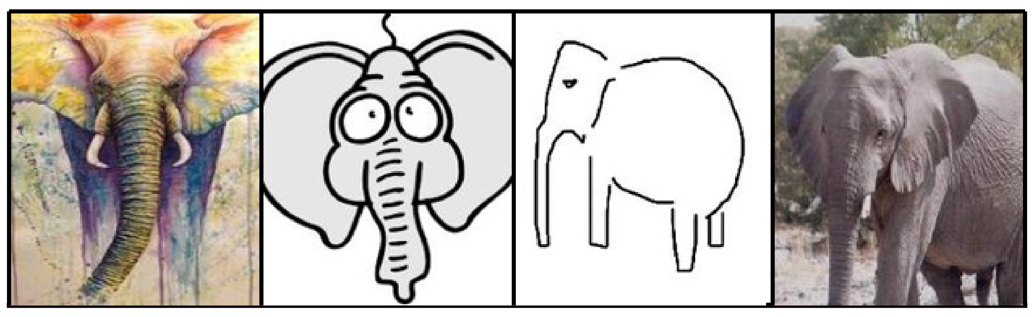
ADAPTIVE LORA MERGING FOR DOMAIN INCREMENTAL LEARNING

Eric Nuertey Coleman, Luigi Quarantiello, Julio Hurtado*, Vincenzo Lomonaco
University of Pisa , University of Warwick*

1 Introduction

Foundation models fine-tuned with PEFT methods perform well under strong iid assumptions but lose performance in dynamic contexts like DIL scenarios.



Our contributions

- Analysis of the performance of state-of-the-art LoRA merging algorithms on DIL tasks
- Novel merging method that dynamically computes the importance of each task

2 Background

- We focus on merging **LoRA** adapters in DIL scenarios
- LoRA is a PEFT technique, where the model updates are computed as:

$$W = W_0 + BA$$

- We compare **linear merging**, **TIES** and **DARE+TIES** with different weights value
- They heavily rely on manual coefficients selection, which makes it hard to pick the optimal configuration

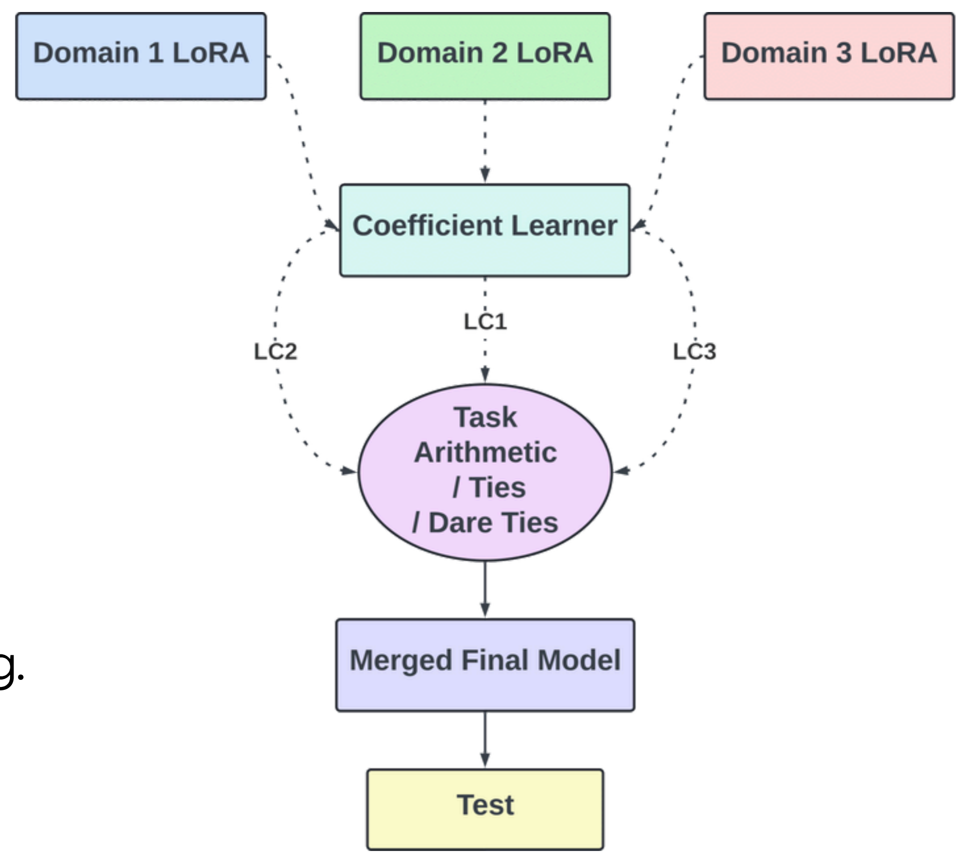
3 Adaptive LoRA Merging

Our method learns dynamically the relevance of each domain in the merging process:

$$\alpha_i = \sigma(c_i) \cdot b$$

$$b = 1 + (b_{\max} - 1) \cdot \sigma(b_{\text{raw}})$$

where ***bmax*** is a hyperparameter, allowing for adaptive coefficients scaling. We use a small memory buffer to learn the coefficients, with as low as one sample per class.



4 Results

SPC = 1	*Fixed Coefficients*			*Adaptive Coefficients*		
Merging Algorithm	**W = 1**	**W = 3**	**W = 5**	$b_{\max} = 1.0$	$b_{\max} = 3.0$	$b_{\max} = 5.0$
Task Arithmetic	29.19 ± 9.99	11.38 ± 4.76	13.34 ± 6.57	76.48 ± 14.65	29.72 ± 10.15	17.84 ± 4.44
TIES	82.42 ± 11.76	25.19 ± 9.86	16.14 ± 4.90	62.62 ± 10.10	82.62 ± 11.39	80.44 ± 12.38
DARE TIES	20.37 ± 5.02	42.58 ± 8.88	63.25 ± 14.10	16.42 ± 4.93	20.41 ± 5.27	25.02 ± 5.69
Task Arithmetic	1.83 ± 0.63	1.50 ± 0.29	1.77 ± 0.34	42.09 ± 5.95	2.17 ± 0.58	1.58 ± 0.53
TIES	66.83 ± 6.92	2.17 ± 0.42	2.16 ± 0.65	24.88 ± 3.51	62.95 ± 5.17	62.74 ± 5.83
DARE TIES	2.65 ± 0.30	13.50 ± 2.32	34.12 ± 4.85	1.68 ± 0.20	2.72 ± 0.21	3.18 ± 0.21

Experience	Adapter Weights ($b_{\max} = 3.0$)				Domain Accuracy (%)			
	Photo	Cartoon	Sketch	Art	Photo	Cartoon	Sketch	Art
1	1.09	–	–	–	100.00	–	–	–
2	0.94	1.08	–	–	98.98	90.36	–	–
3	0.88	1.05	1.01	–	96.05	86.60	85.19	–
4	0.85	1.07	1.02	0.86	97.72	80.72	76.23	87.40

- For *fixed-coefficients* method, picking the merging weights wrongly leads to huge performance loss
- Our ***adaptive-coefficients*** approach learns the importance of each adapter in the merging, assigning different weights according to the complexity of each task
- By introducing ***bmax***, we improve model robustness, with small performance variations

Computational Intelligence and Machine Learning Group
LEONARDO
Future Artificial Intelligence Research