Coffee Talk #1

April 13, 2021

Ozgur Taylan Turan

Transfering Knowledge Across Learning Processes[1]

Conference paper at 2019 ICLR

[1] S. Flennerhag, P. G. Moreno, N. D. Lawrence, and A. Damianou (2018). "Transferring knowledge across learning processes". In: *arXiv*, pp. 1–23. ISSN: 23318422. arXiv: 1812.01054

Why This Paper?

- Neil Lawrence
- Transfer learning - Meta Learning

Introduction

Problem with Transfer Learning

- Structural affinity of tasks
- Information loss?

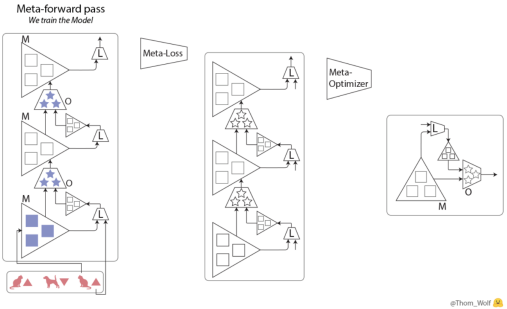
$\Longrightarrow$

Solution with Meta Learning

- Save relevant information
- Aggregate task geometry info.

Meta Learning

- Learning a new task as a learning problem
- BackProp entire learning process



A meta-training step (training the optimizer O) comprising with 3 steps of training the model M)

2

[2] **https://medium.com/huggingface/from-zero-to-research-an-introduction-to-meta-learning-8e16e677f78a**

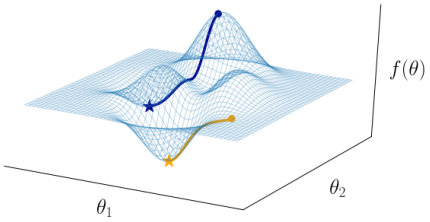
Loss Manifold

# of Updates ↑

- Loss manifold importance ↑

Ease of Learning

- Easier to navigate loss manif.



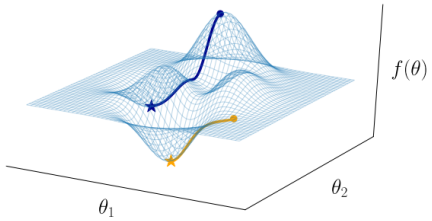
Leap

- Framework that exploits geometry
- Focus: point of initialization
- Learning and initialization relation

Learning Objective $f : \theta \in \mathbb{R}^n \rightarrow L$

$$\theta^{i+1} = \theta^i - \alpha^i S^i \nabla f(\theta^i)$$

Assume convergence after K steps

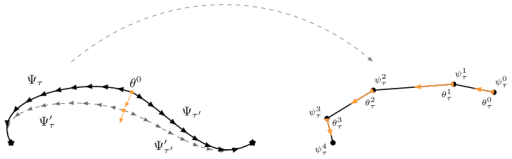


Meta objective

$$\min_{(\theta^0)} \quad \mathbb{E}_\tau[d(\theta^0; M_\tau)]$$

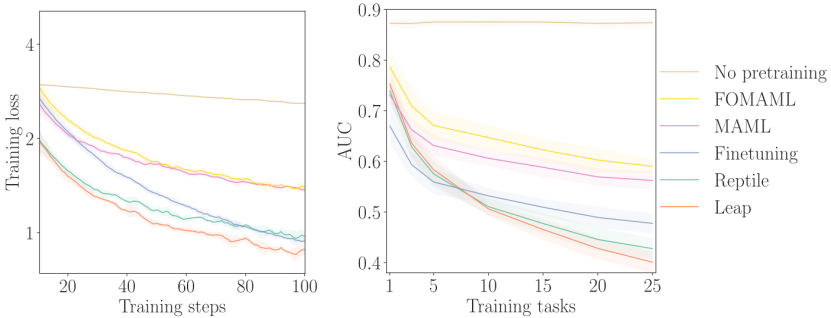
$$\text{s.t.} \quad \theta_\tau^{(i+1)} = u(\theta_\tau^i), \quad \theta_\tau^0 = \theta^0,$$

$$\theta^0 \in \mathbf{\Theta} = \cap_\tau \{\theta^0 | f_\tau(\theta_\tau^{K_\tau}) \leq f_\tau(\psi_\tau^{K_\tau})\}$$



Minimize the expected learning path for all the tasks at the same time

Results-[Omniglot]



Omniglot: dataset → 50 distinct alphabets show 1-25 and held 10 back

Results-[Multi-CV]

Train for all and held-out one

Held-out task	Method	Test (%)	Train (%)	AUC†
Facescrub	Leap	19.9	0.0	11.6
	Finetuning	32.7	0.0	13.2
	Progressive Nets‡	**18.0**	0.0	**8.9**
	HAT‡	25.6	0.1	14.6
	No pretraining	18.2	0.0	10.5
Cifar10	Leap	**21.2**	**10.8**	**17.5**
	Finetuning	27.4	13.3	20.7
	Progressive Nets‡	24.2	15.2	24.0
	HAT‡	27.7	21.2	27.3
	No pretraining	26.2	13.1	23.0
SVHN	Leap	**8.4**	**5.6**	**7.5**
	Finetuning	10.9	6.1	10.5
	Progressive Nets‡	10.1	6.3	13.8
	HAT‡	10.5	5.7	8.5
	No pretraining	10.3	6.9	11.5
Cifar100	Leap	**52.0**	**30.5**	**43.4**
	Finetuning	59.2	31.5	44.1
	Progressive Nets‡	55.7	42.1	54.6
	HAT‡	62.0	49.8	58.4
	No pretraining	54.8	33.1	50.1
Traffic Signs	Leap	**2.9**	0.0	**1.2**
	Finetuning	5.7	0.0	1.7
	Progressive Nets‡	3.6	0.0	4.0
	HAT‡	5.4	0.0	2.3
	No pretraining	3.6	0.0	2.4

Thanks for your attention!