MLO Research Notes
Tuesday 8th December, 2020

MLO Master Thesis Proposal – Spring 2021

Inspecting different learning schemes for fast adaptation/transfer on tasks

Tao LIN*

1 Introduction

Meta Learning methods like MAML [FAL17] aim to learn good meta-initialization such that the model can quickly adapt to a newly sampled test task. Recent work [RRBV20] questions the effectiveness of MAML-like meta learning methods, and debates between fast adaptation and feature reuse: the main body of the network (a.k.a. feature extractor) does not incur significant parameters change during the inner-loop adaptation phase (either early phase of the meta-training, or test time adaptation phase), and thus the feature reuse is the dominant factor.

Transfer learning, a different learning paradigm compared to MAML, also studies the importance level of feature reuse. One very recent work [NSZ20] "What is being transferred in Transfer Learning?" states the importance of feature reuse for transfer learning (especially for layers close to the inputs); they also point out that the additional benefits of pre-trained weights that are not directly coming from feature reuse.

Contrastive Learning (CL), as a new unsupervised learning paradigm [CKNH20, HFW+20, GSA+20, CKS+20] emerging in computer vision this year (2020), aims to learn a good feature extractor from unlabeled data thus it can be further used (fine-tuned/transferred) on other tasks.

Meta Learning, Transfer Learning, and Contrastive Learning, are similar to some extent: they rely on a specifically learned model initialization for other (downstream) tasks, i.e., meta-learned model initialization v.s. parameters trained from scratch under the label supervision v.s. parameters trained from scratch through unsupervised contrastive learning.

However, it is yet unclear

- the performance difference of the learning algorithms (i.e., standard supervised learning, MAML, and CL schemes) on different downstream tasks.
- can we use some analytical tools to better probe the representation difference introduced by different learning algorithms?
- can we better understand the training/learned feature representation, through the aspect of (1) training with random labels [MAT+20], (2) memorization [FZ20].
- can be observe the phenomenon of double descent [NKB+20] (the impact of model size/data size) under the scope of this study?

*Machine Learning and Optimization Laboratory (MLO), EPFL

- (optional) can we utilize the insights on representation learning (i.e., standard supervised learning, meta learning, unsupervised contrastive learning) to design a better federated learning algorithm?

References

[CKNH20] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. *ICML*, 2020.

[CKS+20] Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey E Hinton. Big self-supervised models are strong semi-supervised learners. *Advances in Neural Information Processing Systems*, 33, 2020.

[FAL17] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. *ICML*, 2017.

[FZ20] Vitaly Feldman and Chiyuan Zhang. What neural networks memorize and why: Discovering the long tail via influence estimation. *Advances in Neural Information Processing Systems*, 33, 2020.

[GSA+20] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in Neural Information Processing Systems*, 33, 2020.

[HFW+20] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9729–9738, 2020.

[MAT+20] Hartmut Maennel, Ibrahim M Alabdulmohsin, Ilya O Tolstikhin, Robert Baldock, Olivier Bousquet, Sylvain Gelly, and Daniel Keysers. What do neural networks learn when trained with random labels? *Advances in Neural Information Processing Systems*, 33, 2020.

[NKB+20] Preetum Nakkiran, Gal Kaplun, Yamini Bansal, Tristan Yang, Boaz Barak, and Ilya Sutskever. Deep double descent: Where bigger models and more data hurt. In *International Conference on Learning Representations*, 2020.

[NSZ20] Behnam Neyshabur, Hanie Sedghi, and Chiyuan Zhang. What is being transferred in transfer learning? *Advances in Neural Information Processing Systems*, 33, 2020.

[RRBV20] Aniruddh Raghu, Maithra Raghu, Samy Bengio, and Oriol Vinyals. Rapid learning or feature reuse? towards understanding the effectiveness of maml. In *International Conference on Learning Representations*, 2020.