Adaptation-Agnostic Meta-Training

Jiaxin Chen, Li-Ming Zhan, Xiao-Ming Wu, Fu-lai Chung
{jiax.chen, lmzhan.zhang}@connect.polyu.hk,
xiao-ming.wu@polyu.edu.hk, cskchung@comp.polyu.edu.hk
Department of Computing, The Hong Kong Polytechnic University

8th ICML Workshop on Automated Machine Learning

Introduction

Many meta-learning algorithms can be formulated into an interleaved process, in the sense that task-specific predictors are learned during inner-task adaptation and meta-parameters are updated during meta-update. The normal meta-training strategy needs to differentiate through the inner-task adaptation procedure to optimize the meta-parameters. This leads to a constraint that the inner-task algorithms should be solved analytically. Under this constraint, only simple algorithms with analytical solutions can be applied as the inner-task algorithms, limiting the model expressiveness. To lift the limitation, we propose an adaptation-agnostic meta-training strategy. Following our proposed strategy, we can apply stronger algorithms (e.g., an ensemble of different types of algorithms) as the inner-task algorithm to achieve superior performance comparing with popular baselines.

Preliminary: A Unified View of Existing Meta-Learning Methods

We characterize that the existing meta-algorithms leverage a meta-training procedure as the gradient of the meta-parameters $\boldsymbol{\theta}$ is computed through the inner-task adaptation. As the meta-algorithm updates the meta-parameters $\boldsymbol{\theta}$ by minimizing the loss of the task-specific predictor $\boldsymbol{g}_{\varphi_{T_i}}$ over the query set of each task. The update rule of $\boldsymbol{\theta}$ is:

$$\theta = \theta - \nabla_\theta \mathcal{L}(\{\mathcal{D}^{ts}_{T_i}\}_{i=1}^n; \theta, \{\phi_{T_i}\}_{i=1}^n) = \theta - \nabla_{\phi_{T_i}} \sum_{i=1}^{n} \sum_{z_j \in \mathcal{D}^{ts}_{T_i}} l\left(g_{\phi_{T_i}}(x_j), y_j\right) \times \nabla_\theta g_{\phi_{T_i}},$$

This unified view is depicted as follow:

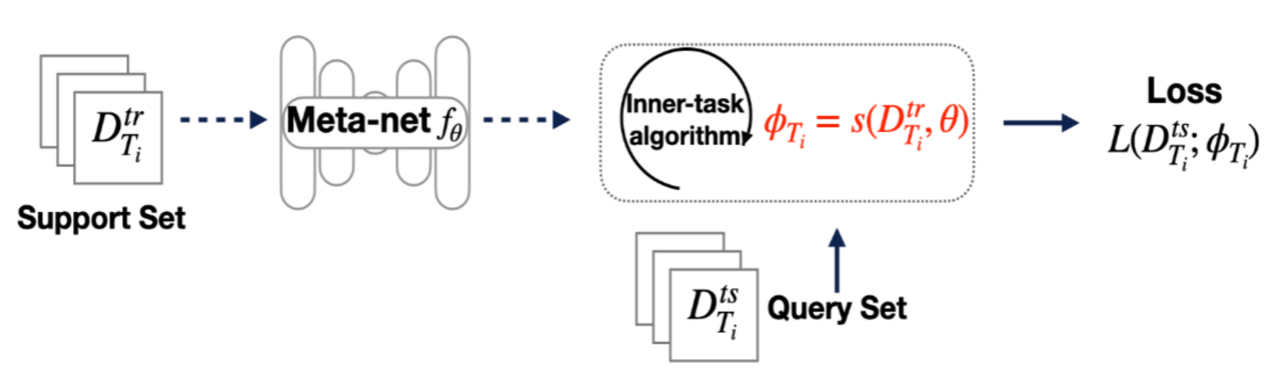

Figure 1: A common meta-training procedure of existing meta-algorithms. Note that $s(\cdot, \cdot)$ is an analytical expression.

Algorithm: Adaptation-Agnostic Meta-Training

From the unified perspective, the key constraint in designing a meta-algorithm is to find an inner-task algorithm which has an explicit analytical solutions which significantly limits the expressiveness of the inner-task algorithms. To relax this constraint, we propose an adaptation-agnostic meta-training strategy which makes no assumption on such dependency.

- In inner-task adaptation, the meta-parameters $\boldsymbol{\theta}$ is fixed, and the support set is fed to the shared embedding network and used to train the task-specific predictor.
- In meta-update, the task-specific parameters $\boldsymbol{\varphi}_{T_i}$ are fixed and the query set is used to optimize the meta-parameters $\boldsymbol{\theta}$. The iteration scheme is formulated as follows:

$$\text{Inner-task adaptation: Fix } \theta, \phi_{T_i} = \arg\min_{x_{\phi_{T_i}}} \mathcal{L}(\mathcal{D}^{tr}_{T_i}; \theta, x_{\phi_{T_i}}),$$

$$\text{Meta-update: Fix } \phi_{T_i}, \theta = \theta - l_\theta \nabla_\theta \mathcal{L}(\mathcal{D}^{ts}_{T_i}; \theta, \phi_{T_i}),$$

Algorithm: Inner-Task Algorithm

The generality and flexibility of the proposed adaptation-agnostic meta-training strategy enables us to apply a powerful algorithm as the inner-task algorithm. As shown in the flowing figure, we e combine the mean-centroid classification algorithm [1], initialization-based inner-task algorithm [2] and MLP proposed by us as the inner-task algorithm.

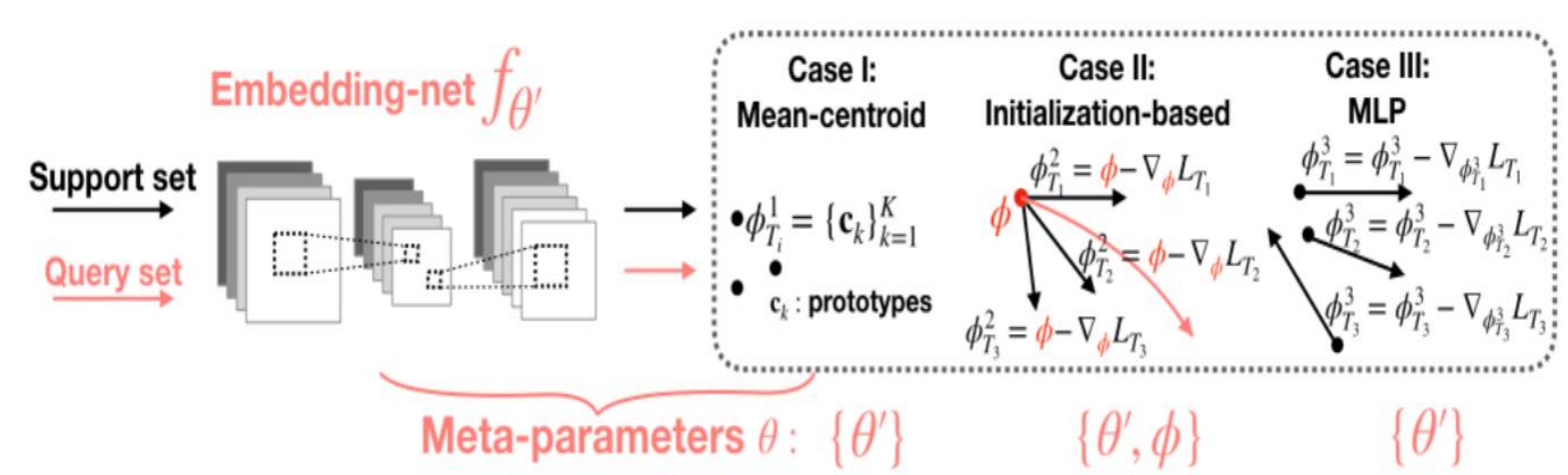

Figure 2: Inner-task adaptation of an instantiation (mean-centroid classification algorithm of (Snell et al., 2017), initialization-based inner-task algorithm in (Raghu et al., 2020) and MLP proposed by us (Eq. (11)) of A2M.

Results

- standard few-shot classification

Table 1: Results of 5-way classification tasks using Conv-4 (the above set) and ResNet-18 (the below set) respectively. See the complete table in Appendix B.

*mini*ImageNet test accuracy		
Model	5-way 1-shot	5-way 5-shot
Matching Net (Vinyals et al., 2016)	43.56 ± 0.84	55.31 ± 0.73
Relation Net (Sung et al., 2018)	49.31 ± 0.85	66.60 ± 0.69
MAML (Finn et al., 2017)	46.70 ± 1.84	63.11 ± 0.92
Protonet (Snell et al., 2017)	44.42 ± 0.84	64.24 ± 0.72
MetaOptNet (Lee et al., 2019)	49.20 ± 0.42	65.54 ± 0.38
A2M (Mean-centroid + MLP+ Init-based)	$\mathbf{50.31 \pm 0.87}$	$\mathbf{68.55 \pm 0.67}$
Matching Net (Vinyals et al., 2016)	52.91 ± 0.88	68.88 ± 0.69
Relation Net (Sung et al., 2018)	52.48 ± 0.86	69.83 ± 0.68
MAML (Finn et al., 2017)	49.61 ± 0.92	65.72 ± 0.77
Protonet (Snell et al., 2017)	54.16 ± 0.82	73.68 ± 0.65
MetaOptNet (Lee et al., 2019)	50.83 ± 0.45	71.01 ± 0.38
A2M (Mean-centroid + MLP+ Init-based)	$\mathbf{57.04 \pm 0.84}$	$\mathbf{75.65 \pm 0.71}$

- cross-domain few-shot classification

Table 2: Results for a 5-way cross-domain classification task.

*mini*ImageNet→CUB		
	5-way 1-shot	5-way 5-shot
Matching networks (Vinyals et al., 2016)	41.10 ± 0.74	53.07 ± 0.74
Prototypical networks (Snell et al., 2017)	42.71 ± 0.78	62.02 ± 0.70
Relation net (Sung et al., 2018)	40.74 ± 0.76	57.71 ± 0.73
MAML (Finn et al., 2017)	32.77 ± 0.64	51.34 ± 0.72
D-MLP	35.88 ± 0.66	57.78 ± 0.76
A2M (Mean-centroid + MLP + Init-based)	$\mathbf{43.55 \pm 0.80}$	$\mathbf{64.63 \pm 0.82}$

Conclusion

We provided a unified view on the commonly used meta-training strategy and proposed an adaptation-agnostic meta-training strategy that is more general, flexible and less prone to overfitting. In future work, we target to analyze the theoretical properties of the adaptation-agnostic meta-training strategy and explore more powerful inner-task algorithms

References

[1] Snell, J., Swersky, K., & Zemel, R. (2017). Prototypical Networks for Few-shot Learning. *Advances in Neural Information Processing Systems*, *30*, 4077-4087.

[2] Raghu, A., Raghu, M., Bengio, S., & Vinyals, O. (2019, September). Rapid Learning or Feature Reuse? Towards Understanding the Effectiveness of MAML. In *International Conference on Learning Representations*.