The Role of Deconfounding in Meta-Learning

Reporter: Yinjie Jiang

5/27/2022

Contents

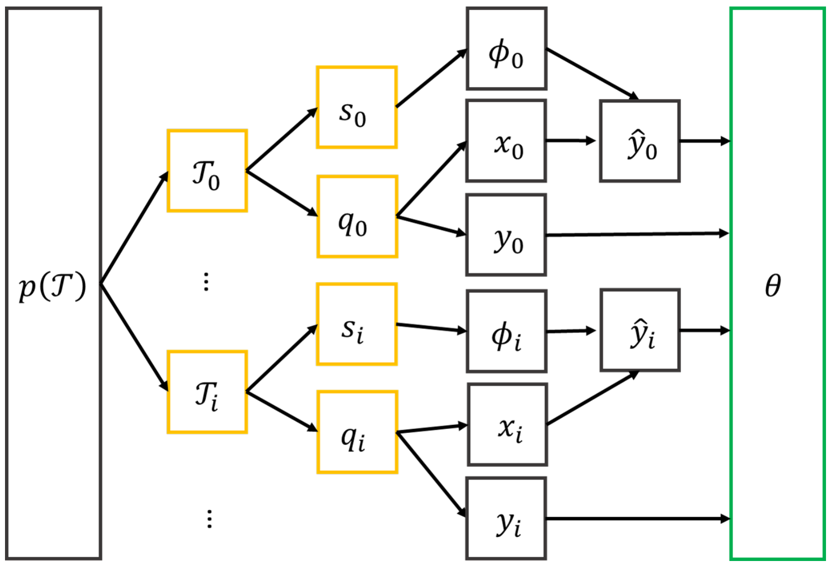
- Meta-learning and Memorization Overfitting
- A Causal View of Meta-Learning
- Deconfounded MAML
- Experimental Results

Meta-learning and Memorization Overfitting

Formulate meta-learning and memorization overfitting

Formulation of Meta-learning

- Meta-learning learns the model initialization $\boldsymbol{\theta}$ from a series of tasks $\mathcal{T}_i$ sampled from a task distribution $p(\mathcal{T})$.
- Gradient-based meta-learning formulate learning such a initialization $\boldsymbol{\theta}$ as a bi-level optimization problem.



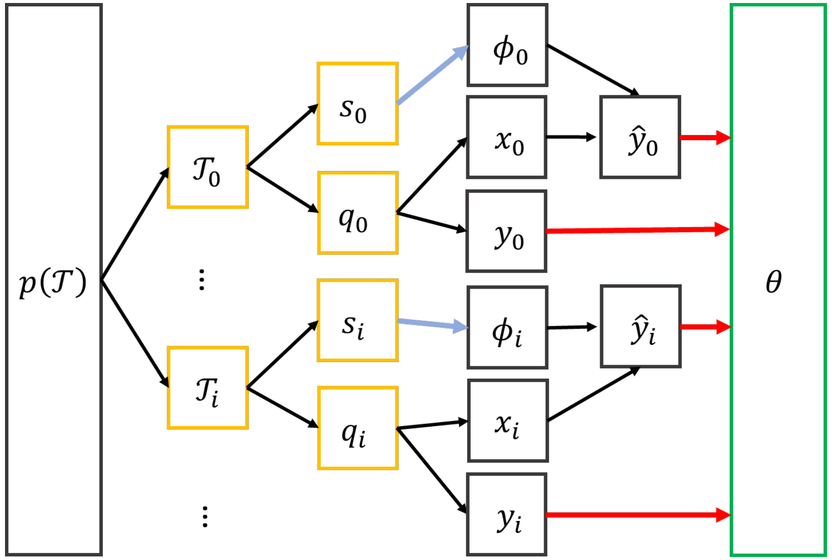
Formulation of Meta-learning

- The inner-loop optimizes the task objective:

$$\mathcal{L}(\phi_i) = \frac{1}{K^s}\sum_{j=1}^{K^s} \mathcal{L}(f_{\phi_i,\theta}(x_{i,j}^s), y_{i,j}^s)$$

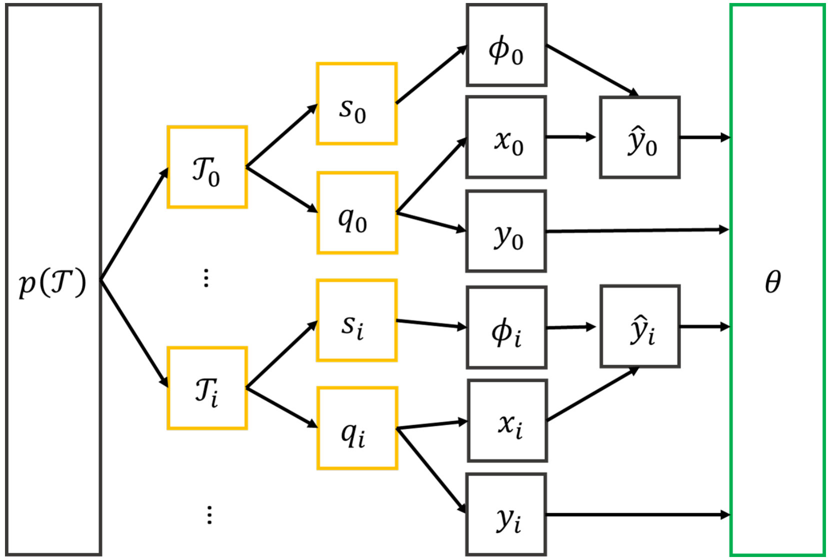
- The outer-loop optimizes the meta objective:

$$\mathcal{L}(\theta) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}_{p(\phi_i|\theta,s_i)}[\frac{1}{K^q}\sum_{j=1}^{K^q} \mathcal{L}(f_{\phi_i,\theta}\left(x_{i,j}^q\right), y_{i,j}^q)]$$



Memorization in Meta-learning

- Memorization overfitting [1] means the metaknowledge memorizes all query sets in meta-training tasks even without adapting on the support sets



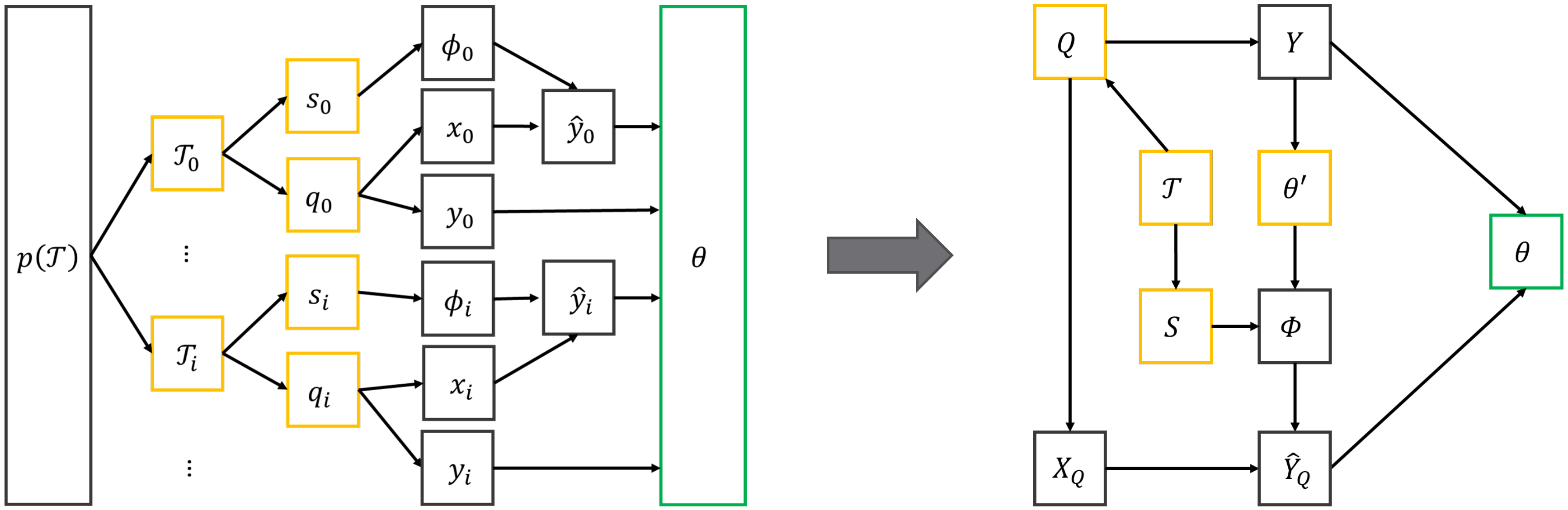
[1] Yin, M., Tucker, G., Zhou, M., Levine, S., & Finn, C. (2019, September). Meta-Learning without Memorization. In *International Conference on Learning Representations*.

A Causal View of Meta-Learning

Explain the memorization overfitting under a causal perspective

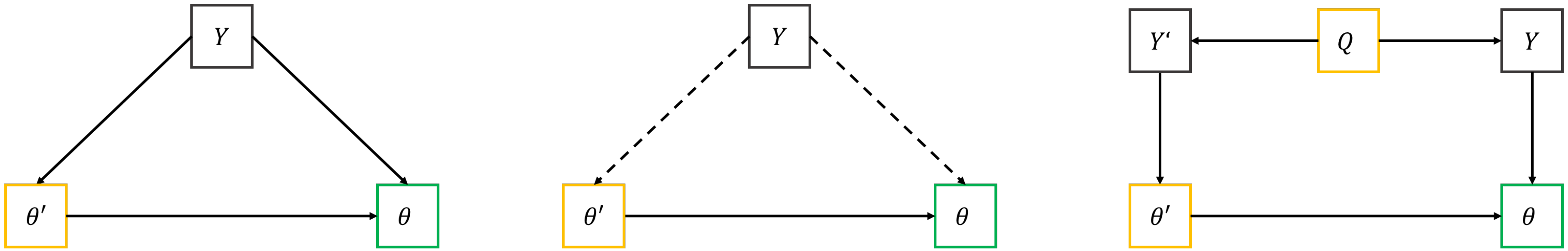
Causal Graph of Meta-learning

- We construct a causal graph according to the workflow of meta-learning.
- We find that memorization is mainly caused by the label space of query set Y, which becomes a confounder during meta-optimization.



Deconfounded Meta-knowledge

- Regularizer-based method [1]
  - To weaken the correlation between Y and θ'
  - Suffering from a trade-off of effectiveness and generalization
- Augmentation-based method [2,3]
  - To randomize the labels of query sets
  - Only partially blocking the correlation

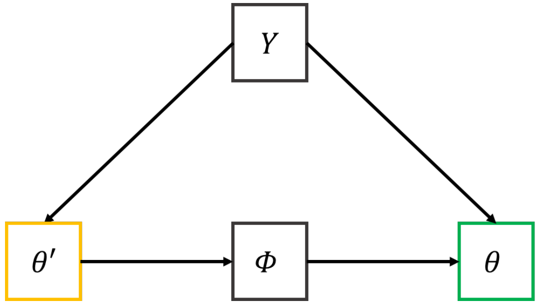


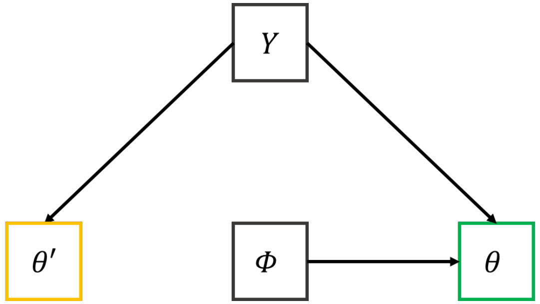
[1] Yin, M., Tucker, G., Zhou, M., Levine, S., & Finn, C. (2019, September). Meta-Learning without Memorization. In *International Conference on Learning Representations*.
[2] Rajendran, J., Irpan, A., & Jang, E. (2020). Meta-learning requires meta-augmentation. *Advances in Neural Information Processing Systems*, *33*, 5705-5715.
[3] Yao, H., Huang, L. K., Zhang, L., Wei, Y., Tian, L., Zou, J., & Huang, J. (2021, July). Improving generalization in meta-learning via task augmentation. In *International Conference on Machine Learning* (pp. 11887-11897). PMLR.

Deconfounded Meta-model

- Under the causal view, we apply front-door adjustment to disconnect ϕ and θ' so that the backdoor path from θ' to θ is blocked.
- The deconfounded meta-learning model is

$$
\begin{aligned}
p(\theta|do(\theta'),S,Q) &= \sum_{\Phi} p(\Phi|\theta',S)p(\theta|do(\Phi),Q) \\
&= \sum_{\Phi} p(\Phi|\theta',S) \sum_{\theta_i'} p(\theta|\Phi,\theta_i',Q)p(\theta_i') \\
&= \sum_{\theta_i'} p(\theta|\Phi,\theta_i',Q)p(\theta_i')
\end{aligned}
$$





How to stratify θ'?

- MAML-Dropout
  - To split θ' into different parts by dropout
- Font-door adjustment is:

$$p(\theta|do(\theta'),S,Q) = \int p(\theta|\Phi,\theta_i',Q)p(\theta_i')\,d\theta_i'$$

$$\approx \frac{1}{N}\sum_{i=1}^{N} p(\theta|\Phi,\theta_i',Q)$$

$$= \frac{1}{N}\sum_{i=1}^{N} p(\theta|\Phi,\theta',Q,z_i)$$

z_i is a set of dropout variables sampled from Bernoulli distribution.

How to stratify θ'?

- MAML-Bins
  - To generate several feature groups which are stratifications of θ'.
  - Feature groups are classified by unsupervised methods.
- Font-door adjustment is:

$$p(\theta|do(\theta'), S, Q) = \frac{1}{M}\sum_{i=1}^{M} p(\theta|\Phi, \theta_i', Q)$$

In i-th group, the feature $\mathbf{feat}_i = f_{\theta_i}(x)$, where x is the input and θ_i indicates the parameters that lead to this feature group.

- The output of the model is an average result of these feature groups.

Experimental Results

Report our experiments and conclusions.

Performance of Regression

Table 3: Performance (MSE ± 95% confidence interval) of pose prediction.

MODEL	10-SHOT	15-SHOT
WEIGHT DECAY	2.772 ± 0.259	2.307 ± 0.226
CAVIA	3.021 ± 0.248	2.397 ± 0.191
META-DROPOUT	3.236 ± 0.257	2.425 ± 0.209
META-AUG	2.553 ± 0.265	2.152 ± 0.227
MR-MAML	2.907 ± 0.255	2.276 ± 0.169
IFSL	3.186 ± 0.256	2.482 ± 0.231
TAML	2.785 ± 0.261	2.196 ± 0.163
ANIL	6.746 ± 0.416	6.513 ± 0.384
ANIL-METAMIX	6.354 ± 0.393	6.112 ± 0.381
ANIL-OURS	$\mathbf{6.289 \pm 0.416}$	$\mathbf{6.064 \pm 0.397}$
MAML	3.098 ± 0.242	2.413 ± 0.177
MAML-METAMIX	2.438 ± 0.196	2.003 ± 0.147
MAML-OURS	$\mathbf{2.396 \pm 0.209}$	$\mathbf{1.931 \pm 0.134}$
METASGD	2.803 ± 0.239	2.331 ± 0.182
METASGD-METAMIX	2.390 ± 0.191	1.952 ± 0.154
METASGD-OURS	$\mathbf{2.369 \pm 0.204}$	$\mathbf{1.926 \pm 0.112}$
T-NET	2.835 ± 0.189	2.609 ± 0.213
T-NET-METAMIX	2.563 ± 0.201	2.418 ± 0.182
T-NET-OURS	$\mathbf{2.487 \pm 0.212}$	$\mathbf{2.402 \pm 0.178}$

Table 2: Performance of drug activity prediction.

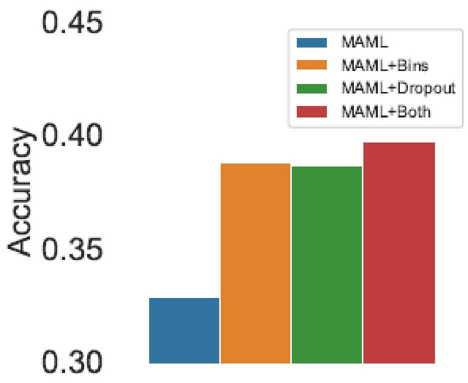
MODEL	GROUP 1			GROUP 2			GROUP 3			GROUP 4		
	MEAN	MED.	>0.3	MEAN	MED.	>0.3	MEAN	MED.	>0.3	MEAN	MED.	>0.3
ANIL	0.357	0.294	50	0.300	0.245	45	0.327	0.301	50	0.338	0.302	50
ANIL-OURS	0.394	0.321	53	0.312	0.284	46	0.338	0.271	48	0.370	0.297	50
MAML	0.366	0.317	53	0.312	0.239	44	0.321	0.258	43	0.348	0.280	47
MAML-OURS	0.410	0.376	60	0.320	0.275	46	0.355	0.257	48	0.370	0.337	56
METASGD	0.388	0.306	51	0.298	0.236	41	0.326	0.237	46	0.353	0.316	52
METASGD-OURS	0.390	0.342	57	0.316	0.269	43	0.358	0.339	56	0.360	0.311	50

Performance of Image Classification

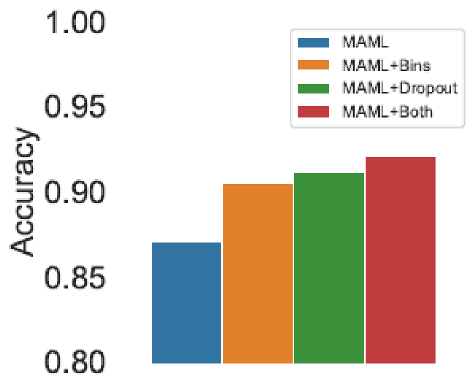
Table 4: Performance (accuracy ± 95% confidence interval) of image classification on Omniglot and MiniImagenet.

Model	Omniglot 20-way 1-shot	Omniglot 20-way 5-shot	MiniImagenet 5-way 1-shot	MiniImagenet 5-way 5-shot
Weight Decay	$86.81 \pm 0.64\%$	$96.20 \pm 0.17\%$	$33.19 \pm 1.76\%$	$52.27 \pm 0.96\%$
CAVIA	$87.63 \pm 0.58\%$	$94.16 \pm 0.20\%$	$34.27 \pm 1.79\%$	$50.23 \pm 0.98\%$
DropGrad	$87.69 \pm 0.57\%$	$94.21 \pm 0.20\%$	$34.42 \pm 1.70\%$	$52.92 \pm 0.98\%$
MR-MAML	$89.28 \pm 0.59\%$	$96.66 \pm 0.18\%$	$35.00 \pm 1.60\%$	$54.39 \pm 0.97\%$
Meta-dropout	$85.60 \pm 0.63\%$	$95.56 \pm 0.17\%$	$34.32 \pm 1.78\%$	$52.40 \pm 0.96\%$
TAML	$87.50 \pm 0.63\%$	$95.78 \pm 0.19\%$	$33.16 \pm 1.68\%$	$52.78 \pm 0.97\%$
ANIL	$88.35 \pm 0.56\%$	$95.85 \pm 0.19\%$	$34.13 \pm 1.67\%$	$52.59 \pm 0.96\%$
ANIL-MetaMix	$92.24 \pm 0.48\%$	$98.36 \pm 0.13\%$	$37.94 \pm 1.75\%$	$59.03 \pm 0.93\%$
ANIL-ours	$\mathbf{92.82 \pm 0.49\%}$	$\mathbf{98.42 \pm 0.14\%}$	$\mathbf{38.09 \pm 1.76\%}$	$\mathbf{59.17 \pm 0.94\%}$
MAML	$87.40 \pm 0.59\%$	$93.51 \pm 0.25\%$	$32.93 \pm 1.70\%$	$51.95 \pm 0.97\%$
MAML-MetaMix	$92.06 \pm 0.51\%$	$97.95 \pm 0.17\%$	$39.26 \pm 1.79\%$	$58.96 \pm 0.95\%$
MAML-ours	$\mathbf{92.89 \pm 0.46\%}$	$\mathbf{98.03 \pm 0.15\%}$	$\mathbf{39.89 \pm 1.73\%}$	$\mathbf{59.32 \pm 0.93\%}$
MetaSGD	$87.72 \pm 0.61\%$	$95.52 \pm 0.18\%$	$33.70 \pm 1.63\%$	$52.14 \pm 0.92\%$
MetaSGD-MetaMix	$93.59 \pm 0.45\%$	$98.24 \pm 0.16\%$	$40.06 \pm 1.76\%$	$60.19 \pm 0.96\%$
MetaSGD-ours	$\mathbf{93.93 \pm 0.40\%}$	$\mathbf{98.49 \pm 0.12\%}$	$\mathbf{40.22 \pm 1.78\%}$	$\mathbf{60.24 \pm 0.91\%}$
T-Net	$87.71 \pm 0.62\%$	$95.67 \pm 0.20\%$	$33.73 \pm 1.72\%$	$54.04 \pm 0.99\%$
T-Net-MetaMix	$93.27 \pm 0.46\%$	$98.09 \pm 0.15\%$	$38.33 \pm 1.73\%$	$59.13 \pm 0.99\%$
T-Net-ours	$\mathbf{93.54 \pm 0.49\%}$	$\mathbf{98.27 \pm 0.14\%}$	$\mathbf{38.38 \pm 1.77\%}$	$\mathbf{59.25 \pm 0.97\%}$

Performance of Image Classification



(c) MiniImagenet



(d) Omniglot

Table 7: Comparison with MetaMix on image classifications.

Model	Omniglot		MiniImagenet	
	20-way 1-shot	20-way 5-shot	5-way 1-shot	5-way 5-shot
MAML	$87.40 \pm 0.59\%$	$93.51 \pm 0.25\%$	$32.93 \pm 1.70\%$	$51.95 \pm 0.97\%$
MAML + MetaMix	$92.06 \pm 0.51\%$	$97.95 \pm 0.17\%$	$39.26 \pm 1.79\%$	$58.96 \pm 0.95\%$
MAML + ours	$92.89 \pm 0.46\%$	$98.03 \pm 0.15\%$	$39.89 \pm 1.73\%$	$59.32 \pm 0.93\%$
MAML + MetaMix + Ours	$\mathbf{93.02 \pm 0.68\%}$	$\mathbf{98.07 \pm 0.22\%}$	$\mathbf{39.92 \pm 1.77\%}$	$\mathbf{59.37 \pm 0.95\%}$

Thank You!