ulm university universität ulm

Model-Agnostic Meta-Learning
in the domain of Deep Reinforcement Learning

a Replication Study

Carolin Schindler | 27.02.2022 |
Deep Reinforcement Learning Project

Motivation

Human Intelligence
- efficient
- general-purpose

Artificial Intelligence
- huge amount of data and hardware resources
- task-specific

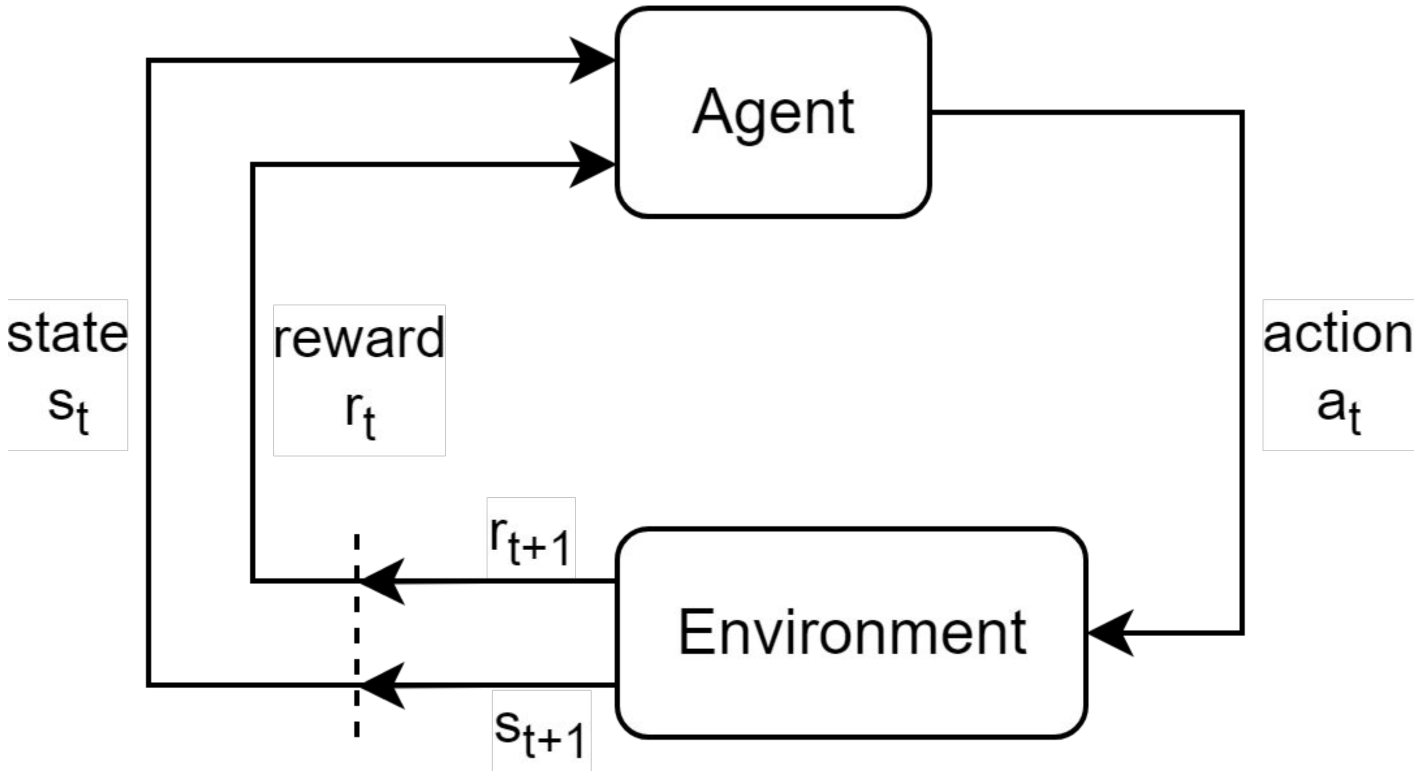
Solutions
- meta-reasoning
- meta-learning

Thomas L. Griffiths, Frederick Callaway, Michael B. Chang, Erin Grant, Paul M. Krueger, and Falk Lieder. “Doing more with less: meta-reasoning and meta-learning in humans and machines”.
In: Current Opinion in Behavioral Sciences 29 (2019). Artificial Intelligence, pp. 24-30.
DOI: https://doi.org/10.1016/j.cobeha.2019.01.005.

Deep RL and MAML

Methods

Deep Reinforcement Learning (Deep RL)



Richard S. Sutton and Andrew G. Barto. Reinforcement learning: An introduction. 2nd ed. Adaptive computation and machine learning. The MIT Press, 2018. ISBN: 978-0-2620-3924-6.

Yuxi Li. Deep Reinforcement Learning. 2018. arXiv: 1810.06339 [cs.LG].

Meta-Learning: “Learning to Learn”

Meta-Objective

- most distinctive property to related fields
- explicitly takes the distribution over multiple tasks into account

Meta-Training phase

- sample tasks and use these collectively to update the model

Meta-Testing phase

- update the model in a regular fashion for single tasks

Approaches

- metric-based
- model-based
- optimization-based

Mike Huisman, Jan N. van Rijn, and Aske Plaat. “A survey of deep metalearning”.
In: Artificial Intelligence Review 54.6 (2021), pp. 4483-4541.
DOI: 10.1007/s10462-021-10004-4.

Model-Agnostic Meta-Learning (MAML)

Algorithm 1 MAML for Reinforcement Learning (practitioner perspective)
based on Finn et al. [9, Algorithm 3]

Require:

$p(\mathcal{T})$: distribution over tasks
α, β: learning rates (α: inner leaning rate, β meta learning rate)
K, M: batch sizes (K: sample batch size, M: meta batch size)
k: amount of inner gradient steps / base-learner updates to be performed ($k \geq 1$)
$iterations$ and H: amount of iterations to train and horizon
$\mathcal{L}$: loss function
f_θ: reinforcement learner parametrized with θ

randomly initialize θ
for iteration $\in [0, iterations)$ **do**
  Sample batch of M tasks $\mathcal{T}_j \sim p(\mathcal{T})$
  for all $\mathcal{T}_j$ **do**
    Copy f_θ into $f_{\theta_0^j}$
    for $i \in [0, k)$ **do**
      Sample K trajectories $\mathcal{D}_i^j = \{(\mathbf{s}_1, \mathbf{a}_1, ..., \mathbf{s}_H, \mathbf{a}_H)\}$ using $f_{\theta_i^j}$ in $\mathcal{T}_j$
      Evaluate $\nabla_{\theta_i^j} \mathcal{L}_{\mathcal{T}_j}^{(\mathcal{D}_i^j)}(f_{\theta_i^j})$ using $\mathcal{D}_i^j$
      Update $\theta_{i+1}^j := \theta_i^j \leftarrow \theta_i^j - \alpha \nabla_{\theta_i^j} \mathcal{L}_{\mathcal{T}_j}^{(\mathcal{D}_i^j)}(f_{\theta_i^j})$ ⋯⋯⋯ ▷ base-learner update
    end for
    Sample K trajectories $\mathcal{D}_k^j = \{(\mathbf{s}_1, \mathbf{a}_1, ..., \mathbf{s}_H, \mathbf{a}_H)\}$ using $f_{\theta_k^j}$ in $\mathcal{T}_j$
    Evaluate and Store $\nabla_\theta \mathcal{L}_{\mathcal{T}_j}^{(\mathcal{D}_k^j)}(f_{\theta_k^j})$ using $\mathcal{D}_k^j$
  end for
  Update $\theta \leftarrow \theta - \beta \frac{1}{M} \sum_{\mathcal{T}_j} \nabla_\theta \mathcal{L}_{\mathcal{T}_j}^{(\mathcal{D}_k^j)}(f_{\theta_k^j})$ with stored $\nabla_\theta \mathcal{L}_{\mathcal{T}_j}^{(\mathcal{D}_k^j)}(f_{\theta_k^j})$ ⋯ ▷ meta-learner update
end for

Chelsea Finn, Pieter Abbeel, and Sergey Levine. “Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks”.
In: Proceedings of the 34th International Conference on Machine Learning. Vol. 70. PMLR, 2017, pp. 1126-1135.
URL: https://proceedings.mlr.press/v70/finn17a.html.

Model-Agnostic Meta-Learning (MAML)

Assumptions

- model is parametrized
- parameters can be learned with gradient-based learning techniques

Meta-Objective:

$$\min_{\theta} \sum_{\mathcal{T}_j \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_j}^{(\mathcal{D}_k^j)}(f_{\theta_k^j})$$

Meta-Gradient contains second-order derivatives:

$$\nabla_\theta \mathcal{L}_{\mathcal{T}_j}^{(\mathcal{D}_k^j)}(f_{\theta_k^j}) = \nabla_{\theta_k^j} \mathcal{L}_{\mathcal{T}_j}^{(\mathcal{D}_k^j)}(f_{\theta_k^j}) \cdot \prod_{i=0}^{k-1} \left(I - \alpha \nabla_{\theta_i^j} \left(\nabla_{\theta_i^j} \mathcal{L}_{\mathcal{T}_j}^{(\mathcal{D}_i^j)}(f_{\theta_i^j}) \right) \right)$$

Chelsea Finn, Pieter Abbeel, and Sergey Levine. "Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks".
In: Proceedings of the 34th International Conference on Machine Learning. Vol. 70. PMLR, 2017, pp. 1126-1135.
URL: https://proceedings.mlr.press/v70/finn17a.html.

Replicate MAML results in the 2D Navigation environment

Experiments and Results

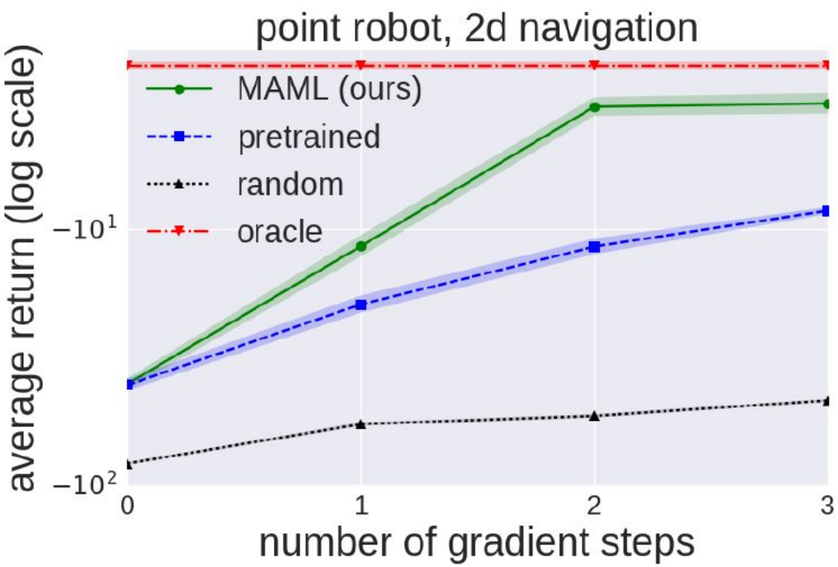
Experimental Setup

Baseline methods

- oracle
- random
- pretrained

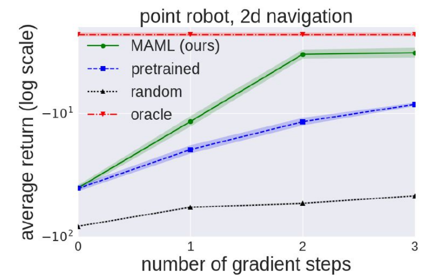
Meta-Learning methods

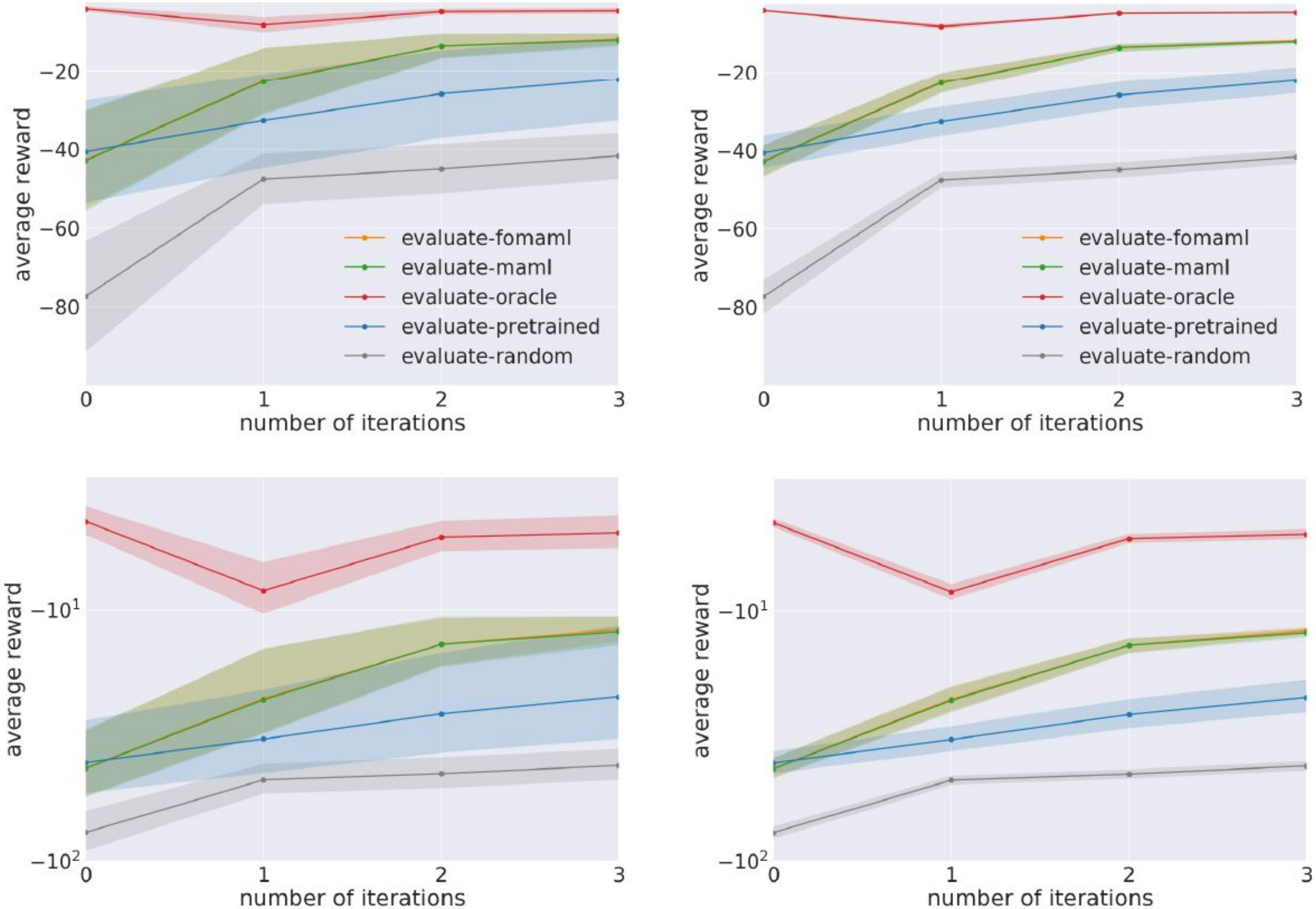
- MAML
- FOMAML



Chelsea Finn, Pieter Abbeel, and Sergey Levine. “Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks”. In: Proceedings of the 34th International Conference on Machine Learning. Vol. 70. PMLR, 2017, pp. 1126-1135. URL: https://proceedings.mlr.press/v70/finn17a.html.

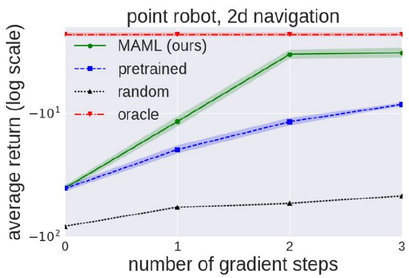
Results with our Meta-Learning Setup



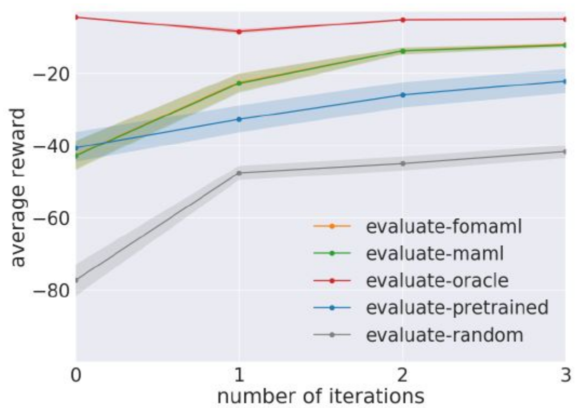


(a) Mean and standard deviation.

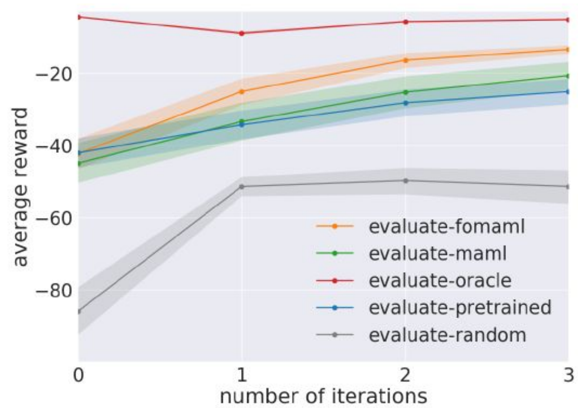
(b) Mean and 95% confidence interval.



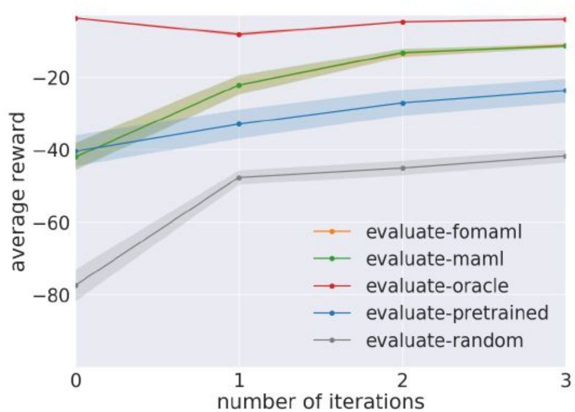
Results of Supplementary Experiments



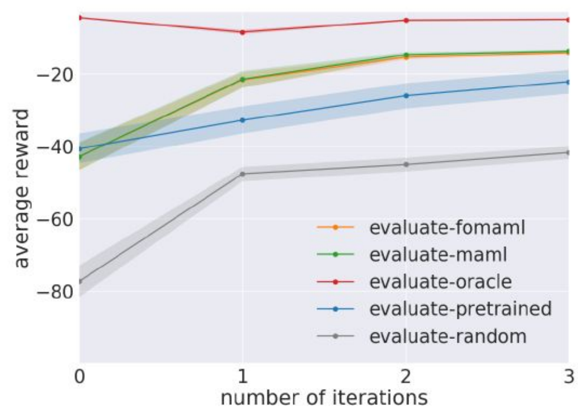
(a) Our meta-learning setup.



(b) Changed training seed (to 2).



(c) Increased meta batch size ($M = 40$).



(d) Applied PPO to the base-learner update.

Were the results reproducible?

Conclusion

Conclusion

MAML

- our meta-learning setup confirms hypotheses of the original work
  - MAML enables fast learning of new tasks.
  - MAML is applicable to the domain of RL.
  - MAML-trained model improves during the meta-testing phase.
- concerns in terms of generalisability
- reason: instability of MAML with respect to hyperparameters

FOMAML

- performs as well as MAML or even better
- reason: MAML is only an approximation as well if second-order derivatives are implemented primitively in the domain of RL

Chelsea Finn, Pieter Abbeel, and Sergey Levine. “Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks”. In: Proceedings of the 34th International Conference on Machine Learning. Vol. 70. PMLR, 2017, pp. 1126-1135. URL: https://proceedings.mlr.press/v70/finn17a.html.

Jakob Foerster, Gregory Farquhar, Maruan Al-Shedivat, Tim Rocktäschel, Eric P. Xing, and Shimon Whiteson. DiCE: The Infinitely Differentiable Monte-Carlo Estimator. 2018. arXiv: 1802.05098 [cs.LG].