Model-Agnostic Meta-Learning with Open-Ended RL

Anonymous Author(s)
Affiliation
Address
email

Abstract

This paper builds on the Open-Ended Reinforcement Learning with Neural Reward Functions proposed by Meier and Mujika [1] that uses reward functions encoded by neural networks. One key limitation of their paper is the necessity of re-learning for each new skill learned by the agent. Consequently, we propose integrating meta-learning algorithms to tackle this problem. We, therefore, study the use of MAML, Model-Agnostic Meta Learning that we believe could make policy learning more efficient. MAML operates by learning an initialization of the model parameters that can be fine-tuned with a small number of examples from a new task which allows for rapid adaptation to new tasks.

1 Model-Agnostic Meta-Learning

Model-agnostic meta-learning (MAML) [2] is a meta-learning approach to solving different tasks from simple regression to reinforcement learning but also few-shot learning. The key idea of MAML is to mitigate few-shot learning since there is simply too little data for too many parameters, leading to overfitting. We learn not only from the data regarding our exact tasks but also from data on similar tasks. To incorporate this, we make an additional assumption, namely that τ comes from some distribution of tasks $p(\tau)$ and that we can sample freely from this distribution. Eventually, we want to use the data available from the other tasks in the distribution to be able to converge to a specific task $\tau_i \sim p(\tau)$, which we can express in terms of an expectation over the distribution. τ is now a random variable and $p(\tau)$ is a set of parameters for task τ. We may use different parameters for each task, use the same parameters for every task, or do something in between.

Additionally, we will not simply use the data from other tasks to find parameters that are optimal for all tasks, but keep the option to fine-tune our model, i.e., take additional optimizer steps on data from the new task τ_i. Afterward, we want to converge to τ_i and reuse the pre-fine-tune-version of the model for each new task. Thus, we can express our optimization objective as

$$min_\theta E_\tau[L_\tau(U_\tau(\theta))] \tag{1}$$

where $U_\tau : \phi \to \phi$ is an optimization algorithm that maps θ to a new parameter vector $U_\tau(\theta)$, being the result of fine-tuning θ on data from task τ, using optimizer U_τ.

In conventional machine learning settings, we consider trainable parameters that are tied to our task. However, the θ in the above objective is learned concerning a variety of tasks. This, together with the fact that it can further be regarded as the initialization of the optimizer U_τ, enables us to interpret θ to be above task level and thus acquire the status of a meta-parameter. Consequently, optimizing such a meta-parameter corresponds to meta-learning.

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

References

[1] Meier, R., Mujika, A. (2022). Open-Ended Reinforcement Learning with Neural Reward Functions. ArXiv. /abs/2202.08266

[2] Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks Finn, C., Abbeel, P. and Levine, S., 2017. Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017, Vol 70, pp. 1126–1135. PMLR..