关于元学习

元学习（meta-learning）是过去几年被重点关注的机器学习研究领域之一，各种基于元学习的研究在近些年频繁出现。

伴随着深度学习的流行，对深度神经网络模型的训练调试成为人工智能研究的关键工作，但由于深度模型训练所需的计算量十分庞大，因此训练需要耗费大量时间，对计算机的硬件要求较高，并且在某个任务下利用大量数据训练的模型，在切换到另一个任务后，往往出现性能的极大下降，需要针对性地进行微调甚至重训练，这使得模型的部署训练往往成为一个复杂的工程问题。如何快速地进行模型训练，并提升最终模型结果的泛化性，成为深度学习领域关注的重点问题，而这正是元学习的应用背景。

元学习提出利用深度学习的方法提升模型训练的效率，人们希望通过元学习提高模型“学习”的能力，最终实现利用小样本的数据即可快速训练出适配任何任务的模型算法。如图1所示，传统机器学习研究如何优化函数以解决某一任务领域的问题，元学习则是研究如何优化生成函数的函数以提升最终模型生成的效率，对于深度学习而言，正是模型的训练调参过程。

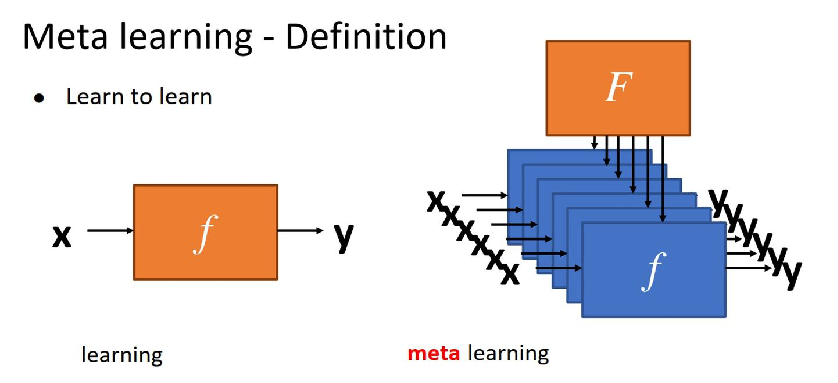


图1 - 元学习的定义

一、元学习介绍

元学习希望获取一种能够高效生成模型的模型，该模型可以在已有任务知识的基础上快速学习新的任务。

传统的机器学习是人为调整超参数，之后在特定任务下训练深度模型，使得模型参数能最好地适应该任务。元学习则是先通过某些任务训练出一个较好的超参数，使得超参数能令模型训练以最快的速度收敛到最优。

在机器学习中，训练单位是样本数据，通过数据对模型进行优化，数据可以分为训练集、验证集和测试集，最终训练得出的模型参数应能较好地处理测试集中的数据。在元学习中，训练单位是任务，并且将任务分为训练任务（Train Tasks）和测试任务（Test Task），利用训练任务训练得出超参数，该超参数应能在测试任务上使模型以较快的速度收敛。

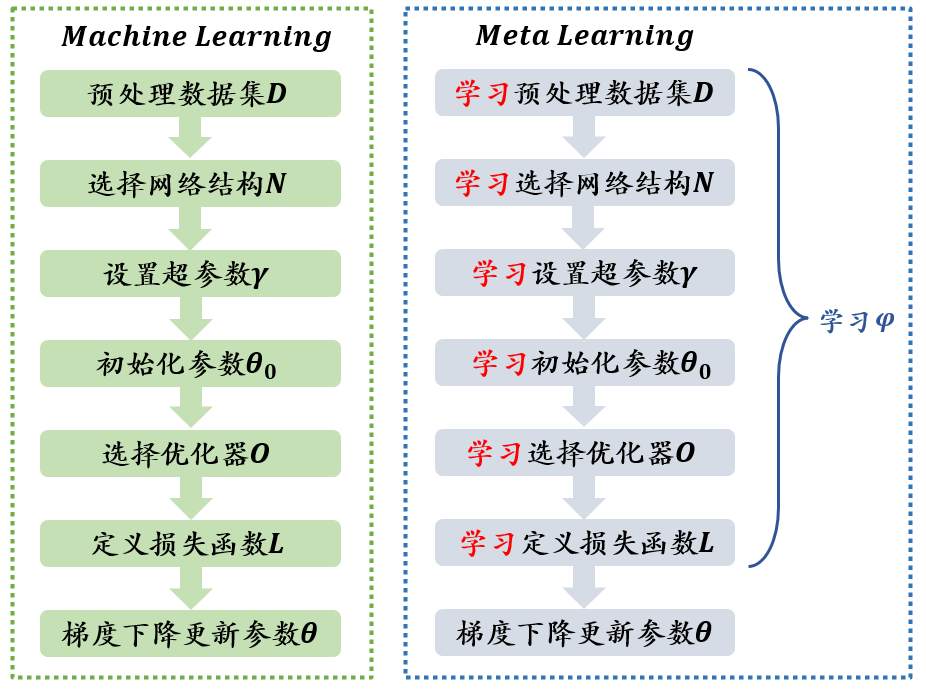


图2 - 机器学习与元学习的学习对象

如图2所示，当训练一个神经网络时，具体步骤有：预处理数据集，选择网络结构，设置超参数，初始化参数，设置优化器，定义损失函数，梯度下降最优化参数，机器学习一般关注于最后的更新参数部分，而元学习则关注于前半部分的超参数。

根据关注的学习对象不同，元学习被分为许多类，下面为简单的分类介绍：

**学习预处理数据集。**深度学习在模型训练前，通常对数据进行增强以增加数据的多样性，使模型训练更加泛化，一般的数据增强包括旋转、伸缩、更改对比度等操作，利用元学习对其进行训练，可以生成更加合理的数据增强操作。相关代表作可参考DADA。

* 论文名称：DADA: Differentiable Automatic Data Augmentation
* 论文链接：<https://arxiv.org/pdf/2003.03780v1.pdf>
* 论文详情：ECCV 2020

**学习初始化参数。**权重参数的初始化很大程度上决定着最终的模型训练结果，目前对于模型参数初始化有一些基础的研究，如著名的Kaiming初始化，元学习的提出使该问题上升到了一个新的高度。该领域的代表作是MAML，该论文提出元学

习应专注于提升模型的整体学习能力，并希望获取一种适用于多种训练任务的初始化参数，使得在多种任务上从该初始化参数出发能够最快地得出最好的收敛结果。

* 论文名称：Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks
* 论文链接：https://arxiv.org/pdf/1703.03400.pdf
* 论文详情：ICML2017

**学习网络结构N。**神经网络的结构是深度学习领域的关键，如何设计合理高效的网络结构一直是相关领域人员研究的关键，网络的深度、网络的宽度、卷积网络中的卷积核、激活函数、Dropout的设定等等细节都需要神经网络设计者的深入思考，由此元学习提出利用学习的方法确定一个最优化的神经网络结构，代表论文是神经网络结构搜索NAS。

* 论文名称：Neural Architecture Search with Reinforcement Learning
* 论文链接：https://arxiv.org/abs/1611.01578
* 论文详情：ICML2017

**学习优化器。**基于梯度下降的学习方法中优化策略发挥着关键作用，常用的优化器有Adam，AdamW，SGD，NAG，RMsprop等，而元学习可以帮助我们在训练任务前确定一个较好的优化器，代表作为：

* 论文名称：Learning to learn by gradient descent by gradient descent
* 论文链接：https://arxiv.org/pdf/1606.04474.pdf
* 论文详情：NIPS 2016

**学习损失函数。**损失函数在梯度下降的优化中至关重要，合理的损失函数能加快训练的速度，提升最终结果的性能，而不合理的损失函数甚至不能使模型收敛到稳定的结果，该领域的代表作可参考如下：

* 论文名称：Meta-learning via learned loss
* 论文链接：https://arxiv.org/pdf/1906.05374.pdf
* 论文详情：ICPR 2020

二、元学习的训练

元学习可分为两个阶段，阶段一是对训练任务的训练，阶段二是对测试任务的训练，分别对应机器学习中的训练和测试阶段。

（一）阶段一：对训练任务的训练

给定个训练任务，每个训练任务的数据集分别为Support set 和 Query set，首先利用这个训练任务的Support set 训练，分别得出各自任务的模型参数，然后用各自任务的Query set测试模型的性能，得出训练任务的损失，根据这些损失构建元学习任务的损失：



最后利用梯度下降法更新参数，从而寻找到最优的超参数设置。

（二）阶段二：对测试任务的训练

元学习的最终目的是获取泛化的模型生成能力，因此最终视角应聚焦到对测试任务的训练上，根据在阶段一中得出的参数，在训练任务上进行模型的训练。

该训练过程即为正常的机器学习过程，所不同的是我们现在拥有通过元学习获得的最佳的模型训练超参，因此理论上可以在最短时间内训练获得性能最佳的模型结果，该模型的训练结果同时也作为元学习的测试集，用以验证元学习最终结果的性能。

以上为元学习的基本训练框架，而基于不同的元学习方法，元学习的训练可以分为基于度量的方法、基于模型的方法和基于优化的方法三种：基于度量的方法学习度量函数，通过不同任务中样本的训练，使模型遇到新任务时也能比较两个样本间的相似性；基于模型的方法学习网络结构，利用来自不同任务的信息帮助预测当前输入样本的标签；基于优化的方法学习优化策略，来自不同任务的信息指导模型参数的更新，寻找对多个任务都有潜力的优化策略。

可以看出，元学习的训练是一个双重训练的过程，即在机器学习训练参数的流程以外嵌套一个大的训练超参的流程，而机器学习的训练结果作为元学习训练的评价指标参与外层训练过程，把握这两层的训练过程，才能深入理解元学习相关研究的基本原理。

三、元学习的未来发展

元学习对于强人工智能的实现有很强的推动作用，当前阶段的深度学习模型一方面朝着大模型的方向前进，以求获得最优的模型结果，另一方面朝着小模型的方向前进，以求获得最高效的模型结果，元学习在后者的研究中占据着关键地位。

目前元学习在许多领域有着具体应用，如计算机视觉中的小样本学习（few-shot learning）、仿真（sym2real）、神经结构搜索（NAS）、无监督元学习领域、自适应和领域泛化、自然语言处理（NLP）等，未来发展前景十分广阔。

当然我们也要看到元学习目前也有许多发展的桎梏，例如元学习的训练速度问题，双重训练的方式使得元学习与深度学习的组合在训练时耗费极大的运算量，训练时间远超传统的深度学习，以及目前元学习的实际应用价值并不高，许多元学习的训练结果并没有显著的性能优势（相比之下，迁移学习反而更有实用价值）。

因此，未来元学习的发展仍任重而道远，如何将其与业界相结合，真正实现一定的经济价值，使其得到更加广泛地应用，是相关研究者需要思考的问题。