

元学习——Meta-Learning in Neural Networks: A Survey

作者：凯鲁嘎吉 - 博客园 <http://www.cnblogs.com/kailugaji/>

这篇博文是对“[Meta-Learning in Neural Networks: A Survey](#)”的阅读理解与总结，此文综述了元学习的最新研究进展，对元学习初学者了解此领域有极大的参考价值。

近年来，人们对元学习的兴趣急剧上升。与传统的AI方法不同，元学习的目的是改进学习算法本身，考虑到多个学习阶段的经验。这个范例提供了一个机会来解决深度学习的许多传统挑战，包括数据和计算瓶颈，以及泛化。本文描述了当代元学习的研究现状。首先讨论了元学习的定义，并在迁移学习和超参数优化等相关领域对元学习进行了定位。然后，提出一个新的元学习分类，提供了一个更全面的细分空间的元学习方法。此文还总结了元学习的一些有前途的应用和成功之处，例如小样本学习和强化学习。最后讨论了突出的挑战和未来研究的前景。

1. 元学习形式化定义——从数据到任务

元学习形式化定义——从数据到任务

传统的机器学习(回顾)

$$(1) \theta^* = \arg \min_{\theta} L(D; \theta, w)$$

模型参数 θ 预设超参 w

从三个角度定义元学习

2. 元学习: 从双层优化看

$$w^* = \arg \min_w \sum_{i=1}^M L^{meta}(D_{source}^{val(i)}; \theta^{*(i)}(w), w)$$

外层

$$s.t. \theta^{*(i)}(w) = \arg \min_{\theta} L^{task}(D_{source}^{train(i)}; \theta, w)$$

内层 训练集(支持集)

3. 元学习: 从前馈模型看

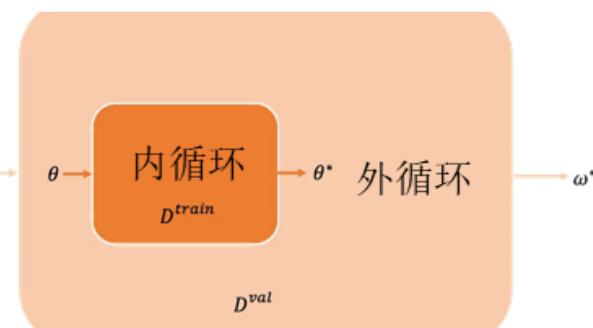
$$\min_w \mathbb{E}_{(D^{tr}, D^{val}) \in \mathcal{T}} \sum_{(x, y) \in D^{val}} [(x^T g_w(D^{tr}) - y)^2]$$

验证集(查询集)

1. 元学习: 从任务分布看

$$(2) \min_w \mathbb{E}_{T \sim p(T)} L(D; w)$$

任务 T 元知识 w



元训练步骤

$$(3) w^* = \arg \min_w \sum_{i=1}^M L(D_{source}^{(i)}; w)$$

$$D_{source} = \{(D_{source}^{train}, D_{source}^{val})^{(i)}\}_{i=1}^M$$

M个源任务

$$D_{target} = \{(D_{target}^{train}, D_{target}^{test})^{(i)}\}_{i=1}^Q$$

Q个目标任务

元测试步骤

$$(4) \theta^{*(i)} = \arg \min_{\theta} L(D_{target}^{train(i)}; \theta, w^*)$$

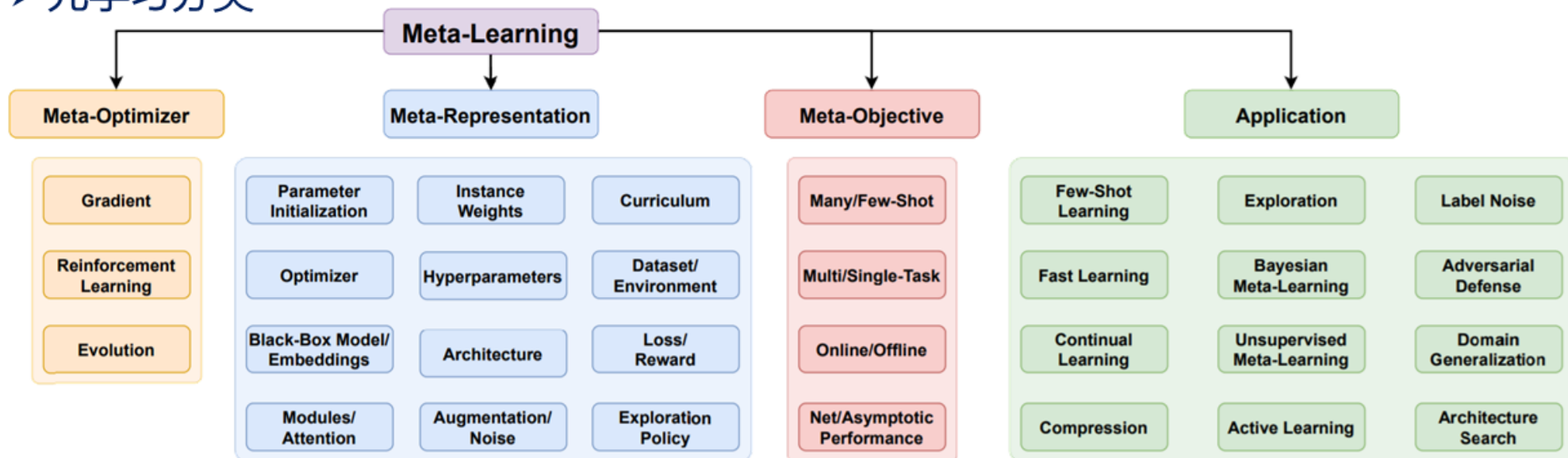
Hospedales, T. M., Antoniou, A., Micaelli, P., & Storkey, A. J. Meta-learning in neural networks: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.
 荐读 Meta-Learning in Neural Networks: A survey. 知乎 - Flood Sung- <https://zhuanlan.zhihu.com/p/133159617>

2. 元学习与其他相关领域的比较

相关领域		元学习
Transfer Learning (TL)	<ul style="list-style-type: none"> 使用源任务的过去经验来提高目标任务的学习(速度、数据效率、准确性) 先验通过对源任务的简单学习来提取，而不使用元目标 	<ul style="list-style-type: none"> 是一种范式，可以用来改进TL和其他问题 相应的先验通过外部优化来定义，该外部优化评估学习新任务时先验的效益，如MAML[19]所示
Domain Adaptation (DA) and Domain Generalization (DG)	<ul style="list-style-type: none"> 域转移是指源问题和目标问题共享同一个目标，但目标任务的输入分布相对于源任务发生转移，从而降低模型性能的情况 DA是TL的一种变体，它试图通过使用来自目标的稀疏或未标记数据来调整源训练模型来缓解这个问题 DG指的是训练源模型对这种域转移具有鲁棒性而无需进一步自适应的方法 DA和DG不使用元目标来优化如何跨域学习 	<ul style="list-style-type: none"> 可用于执行DA和DG
Continual learning (CL)	<ul style="list-style-type: none"> 根据潜在的非平稳分布来学习一系列任务的能力，特别是在加快学习新任务的同时不忘记旧任务 与元学习类似，也考虑了任务分配，其部分目标是加速目标任务的学习，但元目标并没有得到明确的解决 	<ul style="list-style-type: none"> 元学习为推进持续学习提供了一个潜在的框架，一些最近的研究已经开始通过开发元目标来实现这一点，该元目标对持续学习性能进行编码
Multi-Task Learning (MTL)	<ul style="list-style-type: none"> 旨在联合学习多个相关的任务，以受益于参数共享和由此产生的共享表示的多样性带来的正则化，以及计算/内存节省 没有元目标的单一层次优化，其目标是解决固定数量的已知任务 	<ul style="list-style-type: none"> 重点是解决未知的未来任务 元学习可以被引入到MTL中，例如通过学习任务之间的相关性，或者如何在多个任务之间确定优先级
Hyperparameter Optimization (HO)	<ul style="list-style-type: none"> 属于元学习的范畴，学习速率、正则化强度等超参数描述了学习过程 这里定义元目标的HO任务，该元目标通过神经网络进行端到端训练，如基于梯度的超参数学习和神经结构搜索。 但排除了其他方法，如随机搜索[和贝叶斯超参数优化，这些方法很少被认为是元学习 	——
Hierarchical Bayesian Models (HBM)	<ul style="list-style-type: none"> 为元学习提供了一个有价值的视角，它为理解元学习过程提供了一个模型而不是一个算法框架 在实践中，以往的工作主要集中在学习简单的可处理模型θ 	<ul style="list-style-type: none"> 大多数元学习工作考虑复杂的内循环学习过程，涉及许多迭代 一些元学习方法如MAML可以通过HBMs的视角来理解
AutoML	<ul style="list-style-type: none"> 自动化机器学习过程中手动的部分，如数据准备、算法选择、超参调整和架构搜索 经常使用元学习范围之外的启发式方法，并将重点放在诸如数据清理之类的任务上 AutoML有时会对元目标进行端到端的优化 	<ul style="list-style-type: none"> 诸如数据清理之类的任务对元学习来说并不那么重要 可以看作是AutoML的一种特殊化

3. 元学习分类及现有论文总结

元学习分类



□ Meta-Representation (“What?”)——元知识 w

要元学习的东西。可以是整个模型，也可以是超参，网络结构，损失函数，数据等等

□ Meta-Optimizer (“How?”)——优化

主要指双层优化的外循环采用的优化方式，即，梯度，强化学习及进化算法，根据需要选用

□ Meta-Objective (“Why?”)——目标

元学习的具体目标，不同的应用会有不同的目标函数

Hospedales, T. M., Antoniou, A., Micaelli, P., & Storkey, A. J. Meta-learning in neural networks: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.
荐读 Meta-Learning in Neural Networks: A survey. 知乎 - Flood Sung- <https://zhuanlan.zhihu.com/p/133159617>

➤ 本文分类——现有论文总结

Meta-Representation	Meta-Optimizer		
	Gradient	RL	Evolution
Initial Condition	MAML [19], [162], MetaOptNet [163], [76], [85], [99], [164]	MetaQ [165], [166], [167] [19], [61], [62]	ES-MAML [168], [169]
Optimizer	GD ² [15], MetaLSTM [14], [91] [16], [76], [103], [104], [170]	PSD [78], [90]	
Hyperparam	HyperRep [20], HyperOpt [66], LHML [68]	MetaTrace [171], [172]	[169] [173]
Feed-Forward model	SNAIL [38], CNAP [107], [44], [83], [174], [175] [176]–[178]	PEARL [110], [23], [112]	
Metric	MatchingNets [87], ProtoNets [22], RelationNets [88], [89]		
Loss/Reward	MetaReg [92], [41] [120]	MetaCritic [117], [122] [179] [120]	EPG [24], [119] [173]
Architecture	DARTS [21], [130]	[27]	[26]
Exploration Policy		MetaCuriosity [25], [180]–[184]	
Dataset/Environment	Data Distillation [151], [152] [155]	Learn to Sim [158]	[159]
Instance Weights	MetaWeightNet [146], MentorNet [150], [147]		
Data Augmentation/Noise	MetaDropout [185], [140] [115],	AutoAugment [139],	[141]
Modules	[135], [136]		
Annotation Policy	[186], [187]	[188]	

TABLE 1

Research papers according to our taxonomy. We use color to indicate salient meta-objective or application goal. We focus on the main goal of each paper for simplicity. The color code is: **sample efficiency** (red), **learning speed** (green), **asymptotic performance** (purple), **cross-domain** (blue).

Hospedales, T. M., Antoniou, A., Micaelli, P., & Storkey, A. J. Meta-learning in neural networks: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.

4. 元学习应用、挑战与开放性问题

► 应用

□ 计算机视觉与图形学

- 小样本学习方法
- 分类
- 目标检测
- 地标预测
- 小样本目标分割
- 行为(识别、预测等)
- 图像与视频生成
- 生成模型与密度估计
- 小样本学习基准
- 数据集的多样性、偏向性和泛化

□ 神经结构搜索

- 热点话题
- 基准

□ 超参数优化

□ 贝叶斯元学习

□ 无监督与半监督元学习

- 有监督学习者的无监督学习
- 有无监督学习者的无监督学习
- 半监督学习

□ 元强化学习与机器人(下一页PPT)

- 问题设置
- 方法
- 探索
- 优化
- 在线元强化学习
- 在/离策略元强化学习
- 其他趋势与挑战

基准

- 离散控制强化学习
- 连续控制强化学习
- 讨论

□ 持续、在线与自适应学习

- 持续学习
- 在线学习与自适应学习
- 基准

□ 域自适应与域泛化

- 域泛化
- 域自适应
- 基准

□ 语言与语音

- 语言建模
- 语音识别

□ 新兴主题

- 环境学习与Sim2Real
- 元学习用于社会福利
- 非反向传播与生物学上合理的学习者
- 网络压缩
- 通信
- 主动学习
- 利用标签噪声学习
- 对抗攻击与预防
- 推荐系统

► 挑战与开放性问题

- 多样化和多模式的任务分布
- 元泛化
- 任务族
- 计算成本和多样本

► 应用——元强化学习

- 强化学习是一种典型的学习控制策略，它能使智能体在环境中完成一个连续的动作任务后获得高回报。由于报酬稀少、探索的需要和优化算法的高方差，强化学习通常存在极端的样本效率低下问题。然而，应用程序通常自然包含元学习可以利用的任务族，例如移动到或到达不同的位置，在不同的环境中导航，穿越不同的地形，驾驶不同的汽车，与不同的竞争对手智能体竞争，以及处理不同的故障，例如单个机器人肢体的故障。因此，强化学习提供了一个丰富的应用领域，其中任务分布的元学习在提高样本效率方面取得了显著的成功，超过了标准的强化学习算法。人们可以直观地理解这些方法的效用。例如，迷宫布局的元知识对于所有需要在迷宫中导航的任务都是可转移的。

• 问题设置

传统的强化学习是标准的马尔可夫决策过程 M ，策略： π_θ ，动作： a_t ，状态： s_t ，回报： r_t

一个回合： $\tau \sim M$ ， $\tau = (s_0, a_0, \dots, s_T, a_T, r_T)$ ，时间步长： T

最大化期望累计回报： $\max R_T = \sum_t^T \gamma^t r_t$ $\theta^* = \arg \max_{\theta} E_{\tau \sim M, \pi_\theta} [R_\tau]$

元强化学习假设为马尔科夫决策过程上的一个分布 $p(M)$ ，策略 θ 学习过程的某些方面被元知识 w 参数化，例如初始条件，或替代奖励/损失函数，元强化学习的目标是训练元表示 w ，从而使马尔科夫决策过程分布 $p(M)$ 内的学习期望回报最大化：

$$w^* = \arg \max_w \mathbb{E}_{M \sim p(M)} \mathbb{E}_{\tau \sim M, \pi_{\theta^*}} [R_\tau]$$

这就是公式2中元训练问题的强化学习版本，求解同样基于双层优化或前馈方法。最后，一旦训练完成，元学习者 w^* 也可以部署到一个新的马尔科夫决策过程 $M \sim p(M)$ 中快速学习，类似于公式4中的元测试步骤。

Hospedales, T. M., Antoniou, A., Micaelli, P., & Storkey, A. J. Meta-learning in neural networks: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.

► 应用——元强化学习

• 方法

- 在强化学习中已经探讨了几个先前讨论的元表示，包括学习初始条件、超参数、步长方向和步长大小。这使得基于梯度的神经策略学习具有较少的环境交互作用，并训练快速卷积或递归黑盒模型，通过嵌入环境经验来合成策略。最近的工作为这些任务开发了改进的元优化算法，并为元强化学习提供了理论保障。

□ 探索

探索策略是强化学习独有的元表示。强化学习是复杂的，因为数据分布不是固定的，而是根据智能体的行为而变化。此外，稀疏奖励可能意味着一个智能体必须采取许多动作，才能实现奖励，可以用来指导学习。因此，在任何强化学习算法中，如何探索和获取用于学习的数据是一个至关重要的因素。传统上，探索是基于抽样随机动作，或手工启发式。一些元强化学习研究明确地将探索策略或好奇心函数视为元知识；并将他们的收益建模为一个元学习问题——通过“学习如何探索”来提高采样效率。

□ 优化

强化学习优化较为困难，学习到的策略往往远不是最优的，总是在“训练集”的回合。这意味着，与元监督学习相比，元强化学习方法更常用于改善渐近性能以及采样效率，并且能带来显著更好的整体解决方案。许多元强化学习框架的元目标是智能体在一个完整事件中的净收益，因此样本有效学习和渐近性能学习都得到了回报。优化难度还意味着在学习损失（或回报）方面有相对更多的工作，而不是传统的稀疏回报目标，或者除了传统的稀疏回报目标之外，强化学习智能体应该对其进行优化。与真实目标相比，这种已知损失可能更容易优化（更密集、更平滑）。这也与探索作为奖励学习有关，可以认为是学习内在动机元学习的实例。

□ 在线元强化学习

元强化学习研究中有相当一部分是针对单个任务设置的，在学习单个任务时，元知识如损失、奖励、超参数或探索策略与基策略一起在线训练。因此，这些方法不需要任务族，直接提高各自基础学习者的表现。

Hospedales, T. M., Antoniou, A., Micaelli, P., & Storkey, A. J. Meta-learning in neural networks: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

► 应用——元强化学习

• 方法

▣ 在/离策略元强化学习

传统强化学习中的一个主要二分法是在在策略和离策略之间学习，例如PPO和SAC。离策略方法通常更有效。然而，离策略方法很难扩展到元强化学习，导致更多的元强化学习方法建立在在策略强化学习方法上，从而限制了元强化学习的绝对性能。早期对离策略元强化学习方法的研究已经取得了很好的结果。离策略学习也提高了元训练阶段的效率，这在元强化学习中可能是昂贵的。它还为通过重放元训练缓冲区样本来加速元测试提供了新机会。

▣ 其他趋势与挑战

值得注意的是文献[63]在实际物理机器人上成功地演示了元强化学习。机器人技术中的知识迁移通常是最好的组合研究。例如，行走、导航和物体拾取/放置可能是房间清洁机器人的子程序。然而，开发具有有效成分知识迁移的元学习者是一个开放式问题，模块化元学习是一种选择。无监督元强化学习变体的目标是在没有人工指定奖励的情况下进行元训练，或者在元测试中适应变化的环境，但没有新的奖励。与持续学习类似，持续适应使智能体能够适应一个元测试事件中的一系列任务。最后，元学习也被应用于模仿和逆向学习。

• 基准

- 强化学习的元学习基准通常定义一个族来求解，以便训练与评估一个智能体学习“如何学习”。这些可能是要实现的任务(奖励函数)，或者领域(不同的环境或马尔科夫决策过程)。

▣ 讨论

视觉驱动元强化学习的一个复杂问题是将视觉泛化（如计算机视觉）与控制策略的快速学习相分离。例如，CoinRun的评估显示，标准视觉技术（如批量标准）带来了巨大的好处，这表明感知是一个主要的瓶颈。

➤ 应用——元强化学习

• 基准

□ 离散控制强化学习

视觉驱动控制的早期元强化学习基准是街机学习环境，它定义了一组分为元训练和元测试的经典Atari游戏。协议是在元测试环境中计算固定时间步数后的回报。挑战之一是跨游戏的多样性($p(T)$ 宽)，使得成功的元训练很困难，并导致知识迁移带来的收益有限。另一基准是基于将Sonic-hedgehog水平分成元训练/元测试。这里任务分配比较窄，有益的元学习相对容易实现。Cobbe等提出了两种专门设计的视频游戏，用于元强化学习的基准测试。CoinRun游戏提供了 2^{32} 程序生成的不同难度和视觉外观的关卡。他们表明，要可靠地推广到新的关卡，大约需要10000个关卡的元训练实验。CoinRun主要用于测试直接泛化，而不是快速自适应，并且可以被视为在马尔科夫决策过程环境中提供分布来测试泛化，而不是在任务上测试自适应。为了在更广泛的任务分布中更好地测试快速学习，ProcGen提供了包括CoinRun在内的16款程序生成游戏。

□ 连续控制强化学习

虽然像gym这样的通用基准极大地促进了强化学习研究，但是对于元强化学习基准的共识却很少，这使得现有的工作很难进行比较。大多数连续控制元强化学习研究都提出了自制的基准，这些基准是特定任务的低维参数变体，例如导航到不同的位置或速度，或穿越不同的地形。最近提出了几个多马尔科夫决策过程基准，但这些基准主要测试不同环境扰动下的泛化，而不是不同的任务。元世界基准提供了一套50个基于状态驱动的连续控制任务，不同于简单的参数变量，如杠杆拉拽和开门。该基准应能够进行更具可比性的评估，并调查任务分布内和任务分布间的普遍性。元世界评估表明，现有的元强化学习方法难以概括广泛的分布和元训练/元测试迁移。这可能是由于我们的元强化学习模型在任务数量和覆盖范围方面太弱和/或基准太小，无法进行有效的学习。另一个适合元强化学习的最新基准是PHYRE，它提供了一组50个基于视觉的物理任务模板，可以用简单的动作来解决，但可能需要基于模型的推理来有效地解决。它们还提供了模板内和跨模板的泛化测试。

Hospedales, T. M., Antoniou, A., Micaelli, P., & Storkey, A. J. Meta-learning in neural networks: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.

6. 参考文献

[1] Hospedales, T. M., Antoniou, A., Micaelli, P., & Storkey, A. J. [Meta-learning in neural networks: a survey](#). IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.

[2] 荐读 Meta-Learning in Neural Networks: A survey. 知乎 - Flood Sung- <https://zhuanlan.zhihu.com/p/133159617>