双层优化问题：统一GAN，演员-评论员与元学习方法

(Bilevel Optimization Problem unifies GAN, Actor-Critic, and Meta-Learning Methods)

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

之前写过深度学习典型代表——生成对抗网络，写过强化学习典型代表——演员-评论员算法，写过元学习典型代表——MAML算法，现在开始梦幻联动，有没有发现这三个算法有一个共同点，那就是相互博弈(two-player game)，两个优化目标交替执行，最终达到某个平衡(纳什均衡)，停止迭代。而这个问题在运筹学优化问题中有一个术语，叫做双层优化问题(Bilevel Optimization Problem)。以上三个看似毫无关联的算法最终都归结为双层优化问题，可以用一个公共的表示方法来将这三者统一起来。有了这个结论，这三个看似毫无关联的算法以后优化求解就相当于求解双层优化问题，只要双层优化问题有解决方案，这三者的最优解就能获得。可以使用Kriging逼近来求解双层优化问题[1]。

1. Bilevel Optimization (BLO) Problem

Bilevel Optimization Problem: GAN, Actor-Critic, and Meta-Learning

➢ Bilevel Optimization (BLO) Problem

◘ 一般定义

- BLO问题可以表述为(1)，其中x为leader控制的决策变量，y为follower控制的决策变量

$$\min_{x \in X} F(x, y) \quad \text{Leader目标函数}$$

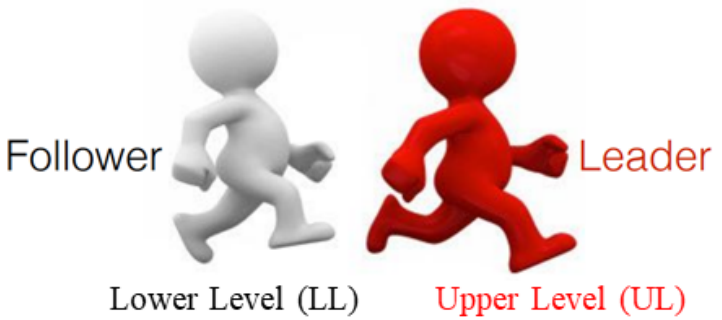
$$s.t.\ G(x, y) \le 0$$

$$y \in \arg\min_{y \in Y} \{ f(x, y) : g(x, y) \le 0 \} \quad \text{Follower目标函数}$$

◘ 等价定义

$$x^* = \arg\min_{x \in X} F(x, y^*(x)) \quad \text{Leader目标函数}$$

$$y^*(x) = \arg\min_{y \in Y} f(x, y) \quad \text{Follower目标函数}$$



Notation(s)	Description
$x \in \mathbb{R}^n$	Decision vector at UL
$y \in \mathbb{R}^m$	Decision vector at LL
$F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$	Objective function at UL
$f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$	Objective function at LL
$G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$	Constraint functions at UL
$g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$	Constraint functions at LL

◘ 解释

- **双层优化问题**本质上是层次优化问题，其中一个优化问题嵌套在另一个优化问题中。外部优化问题和内部优化问题通常分别称为上层优化问题和下层优化问题。
- 在博弈论的背景下，两层优化问题涉及两个参与者：1)上层的参与者是领导者(Leader)，2)下层的参与者是追随者(Follower)。领导者的最优决策依赖于追随者的反应，追随者自身会优化自己内部的决策。这两个层次有各自不同的目标函数、约束条件和决策变量。

A. Sinha and V. Shaikh, "Solving Bilevel Optimization Problems Using Kriging Approximations," IEEE Transactions on Cybernetics, doi: 10.1109/TCYB.2021.3061551.
David Pfau, Oriol Vinyals, "Connecting Generative Adversarial Networks and Actor-Critic Methods", arXiv preprint, 2016.

2. Generative Adversarial Networks (GAN)

生成对抗网络(Generative Adversarial Networks, GANs)是通过对抗训练的方式来使得生成网络产生的样本服从真实数据分布在生成对抗网络中，有两个网络进行对抗训练。一个是判别网络，目标是尽量准确地判断一个样本是来自于真实数据还是由生成网络产生；另一个是生成网络，目标是尽量生成判别网络无法区分来源的样本,这两个目标相反的网络不断地进行交替训练当最后收敛时，如果判别网络再也无法判断出一个样本的来源，那么也就等价于生成网络可以生成符合真实数据分布的样本。

Bilevel Optimization Problem: GAN, Actor-Critic, and Meta-Learning

➢ Generative Adversarial Networks (GAN)

◘ 一般定义

- 生成对抗网络将无监督学习问题描述为两个对手之间的博弈——从分布中采样的生成器G和将样本分类为真或假的判别器D。通常，生成器被表示为一个确定性前馈神经网络，通过该网络，固定噪声源$z\sim\mathcal{N}(0,\ I)$，然后，GAN博弈被表述为零和博弈，其中的值是判别器的预测与图像的真实身份(真实或生成)之间的交叉熵损失。我们用y表示：

$$\min_G \max_D \mathbb{E}_{w,y}\left[y\log D(w)+(1-y)\log(1-D(w))\right]=\min_G \max_D \mathbb{E}_{w\sim p_{data}}\left[\log D(w)\right]+\mathbb{E}_{z\sim\mathcal{N}(0,I)}\left[\log(1-D(z))\right]$$

- 为了确保判别器的分类精度很高，生成器能进行梯度更新，生成器的损失函数通常被表示为最大化样本分类为真的概率，而不是最小化其被分类为假的概率。

$$\max_D \mathbb{E}_{w\sim p_{data}}\left[\log D(w)\right]+\mathbb{E}_{z\sim\mathcal{N}(0,I)}\left[\log(1-D(z))\right]$$

$$\max_G \mathbb{E}_{z\sim\mathcal{N}(0,I)}\left[\log(D(z))\right]$$

◘ 修正后的损失仍然很容易表述为双层优化问题(min)

Leader (判别器) $F(D,G)=-\mathbb{E}_{w\sim p_{data}}\left[\log D(w)\right]-\mathbb{E}_{z\sim\mathcal{N}(0,I)}\left[\log(1-D(z))\right]$

Follower (生成器) $f(D,G)=-\mathbb{E}_{z\sim\mathcal{N}(0,I)}\left[\log(D(z))\right]$

David Pfau, Oriol Vinyals, "Connecting Generative Adversarial Networks and Actor-Critic Methods", arXiv preprint, 2016.

3. Actor-Critic (AC) Methods

演员-批评员方法(Actor-Critic, AC)是强化学习中一类长期存在的技术。而大多数强化学习算法要么专注于学习值函数，就像值迭代和时序差分学习一样，要么直接学习策略，就像策略梯度方法一样，AC方法可以同时学习——演员是策略，批评员是值函数。在某些AC方法中，批评员为策略梯度方法提供的方差基线低于从重复值估计的方差基线。在这种情况

下，即使对值函数的错误估计也是有用的。因为无论使用何种基线，策略梯度都是无偏的。在其他AC方法中，根据近似值函数更新策略，在这种情况下，可能导致与GANs中类似的病理学。如果针对错误的值函数对策略进行优化，则可能会导致错误的策略，该策略永远不会充分探索空间，从而阻止找到好的值函数，并导致退化解。

Bilevel Optimization Problem: GAN, Actor-Critic, and Meta-Learning

➢ Actor-Critic (AC) Methods

$\mathcal{S}$：状态
$\mathcal{A}$：动作
$p_0(s)$：初始状态下的分布
$\mathcal{P}(s_{t+1} \mid s_t, a_t)$：转移函数
$\mathcal{R}(s_t)$：回报分布
γ：折扣因子 $\gamma \in [0,1]$
π：策略
$Q^\pi(s,a)$：动作-值函数

◘ 一般定义

- 演员-评论员方法的目的是同时学习预测预期折扣回报的动作-价值函数，并学习对该值函数最优策略

$$Q^\pi(s,a) = \arg\max_Q \mathbb{E}_{s_{t+k}\sim\mathcal{P}, r_{t+k}\sim\mathcal{R}, a_{t+k}\sim\pi}\left[\sum_{k=1}^{\infty}\gamma^k r_{t+k} \,\middle|\, s_t = s, a_t = a\right]$$

$$\pi^* = \arg\max_\pi \mathbb{E}_{s_0\sim p_0, a_0\sim\pi}\left[Q^\pi(s_0,a_0)\right]$$

- Q可以表示为一个极小化问题的解:

$$Q^\pi = \arg\min_Q \mathbb{E}_{s_t,a_t\sim\pi}\left[\mathcal{D}\left(\mathbb{E}_{s_{t+1},r_t,a_{t+1}}[r_t + \gamma Q(s_{t+1},a_{t+1})] \,\|\, Q(s_t,a_t)\right)\right]$$

- 其中$\mathcal{D}$可以是除两者相等外的任何正散度。

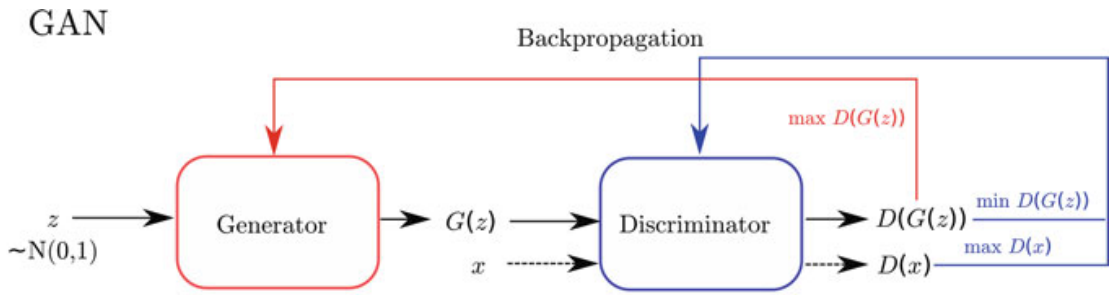
◘ 行动者-批评者问题也可以表示为一个双层优化问题

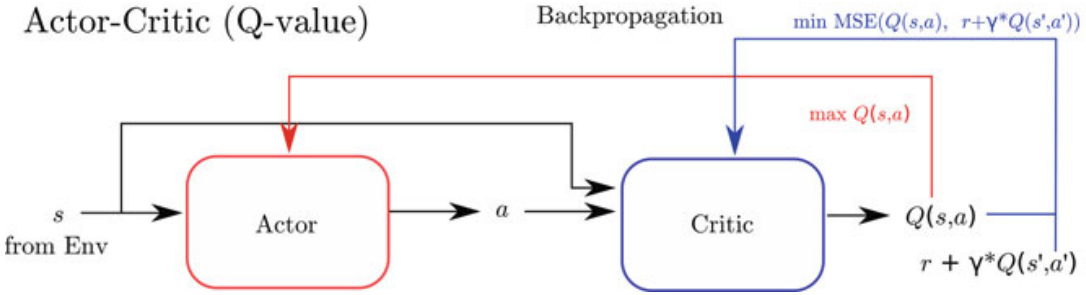
Leader (评论员) $F(Q,\pi) = \mathbb{E}_{s_t,a_t\sim\pi}\left[\mathcal{D}\left(\mathbb{E}_{s_{t+1},r_t,a_{t+1}}[r_t + \gamma Q(s_{t+1},a_{t+1})] \,\|\, Q(s_t,a_t)\right)\right]$

Follower (演员) $f(Q,\pi) = -\mathbb{E}_{s_0\sim p_0, a_0\sim\pi}\left[Q^\pi(s_0,a_0)\right]$

David Pfau, Oriol Vinyals, "Connecting Generative Adversarial Networks and Actor-Critic Methods", arXiv preprint, 2016.

补充：生成对抗网络 vs 演员-评论员





方法	生成对抗网络	演员-评论员
冻结学习	有	有
标签平滑	有	无
历史平均	有	无
小批量判别	有	无
批量归一化	有	有
目标网络	不适用	有
经验回放	无	有
熵正则化	无	有
兼容性	无	有

4. Meta-Learning

元学习(Meta Learning)通常可以理解为学会学习(Learn to Learn)；在多个学习事件中改进学习算法的过程。相比之下，传统的机器学习改进了对一组数据样本的模型预测。在基础学习过程中，内部(或下层/基础)学习算法解决了由数据集和目标定义的任务，如图像分类。在元学习过程中，外部(或上层/元)算法更新内部学习算法，使其学习的模型改进外部目标。例如，这个目标可能是泛化性能或内部算法的学习速度。

Bilevel Optimization Problem: GAN, Actor-Critic, and Meta-Learning

➢ Meta-Learning

$\mathcal{L}^{meta}$: outer objective

$\mathcal{L}^{task}$: inner objective

$\mathcal{D}_{source}^{train\ (i)}$: training set for meta-training

$\mathcal{D}_{source}^{val\ (i)}$: validation set for meta-training

◘ 一般定义

- 元训练步骤可以看做一个双层优化问题

Leader (外层)

$$w^* = \arg\min_{w} \sum_{i=1}^{M} \mathcal{L}^{meta}(\mathcal{D}_{source}^{val\ (i)}; \theta^{*(i)}(w), w) \qquad F(w,\theta) = \sum_{i=1}^{M} \mathcal{L}^{meta}(\mathcal{D}_{source}^{val\ (i)}; \theta^{*(i)}(w), w)$$

$$s.t.\ \theta^{*(i)}(w) = \arg\min_{\theta} \mathcal{L}^{task}(\mathcal{D}_{source}^{train\ (i)}; \theta, w) \qquad f(w,\theta) = \mathcal{L}^{task}(\mathcal{D}_{source}^{train\ (i)}; \theta, w)$$

Follower (内层)

◘ Model-Agnostic Meta-Learning (MAML)

Require: $p(\mathcal{T})$: distribution over tasks
Require: α, β: step size hyperparameters
1: randomly initialize θ
2: **while** not done **do**
3: Sample batch of tasks $\mathcal{T}_i \sim p(\mathcal{T})$
4: **for all** $\mathcal{T}_i$ **do**
5: Evaluate $\nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$ with respect to K examples
6: Compute adapted parameters with gradient descent: $\theta_i' = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$ Support Set
7: **end for**
8: Update $\theta \leftarrow \theta - \beta \nabla_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta_i'})$ Query Set
9: **end while**

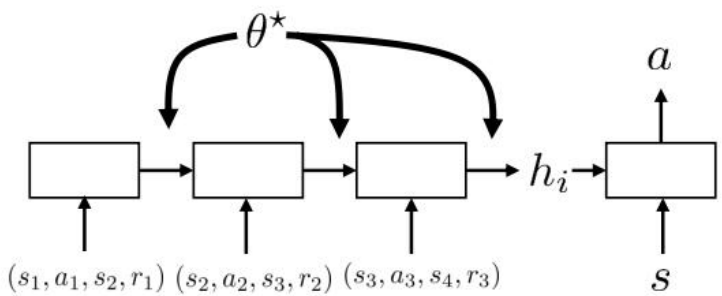
Leader (外层)

Follower (内层)

T. M. Hospedales, A. Antoniou, P. Micaelli and A. J. Storkey, "Meta-Learning in Neural Networks: A Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.

补充(与上述关系不太大，仅供自己学习参考)：从三个角度解释元强化学习，即RNN, 双层优化，以及推断问题。

Perspective 1: just RNN it



+ conceptually simple

+ relatively easy to apply

- vulnerable to *meta-overfitting*

- challenging to optimize in practice

Perspective 2: bi-level optimization

$$f_\theta(\mathcal{M}_i) = \theta + \alpha \nabla_\theta J_i(\theta)$$

MAML for RL

+ good extrapolation ("consistent")

+ conceptually elegant

- complex, requires many samples

Perspective 3: it's an inference problem!

$$\pi_\theta(a|s, z) \qquad z_t \sim p(z_t|s_{1:t}, a_{1:t}, r_{1:t})$$

everything needed to solve task

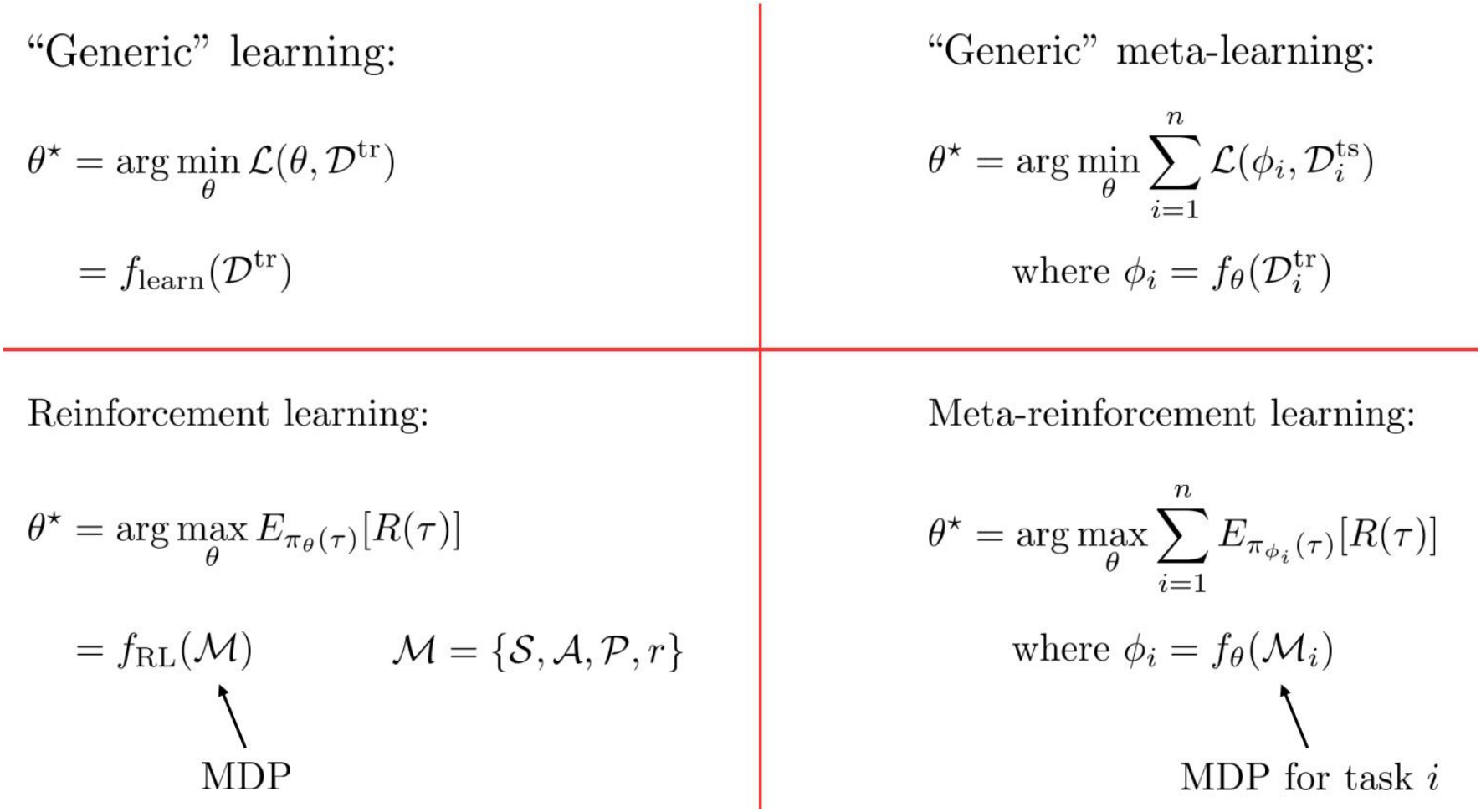
+ simple, effective exploration via posterior sampling

+ elegant reduction to solving a special POMDP

- vulnerable to *meta-overfitting*

- challenging to optimize in practice

学习，元学习，强化学习，元强化学习四种方法总结：



图源：CS 285 Meta-Learning 2020 http://rail.eecs.berkeley.edu/deeprlcourse-fa20/static/slides/lec-22.pdf

5. 参考文献

[1] A. Sinha and V. Shaikh, "Solving Bilevel Optimization Problems Using Kriging Approximations," IEEE Transactions on Cybernetics, doi: 10.1109/TCYB.2021.3061551.

[2] David Pfau, Oriol Vinyals, "Connecting Generative Adversarial Networks and Actor-Critic Methods", arXiv preprint, 2016.

[3] T. M. Hospedales, A. Antoniou, P. Micaelli and A. J. Storkey, "Meta-Learning in Neural Networks: A Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.