

Work Report

Experience Adaptation for Derivative-Free Optimization

Yi-Qi Hu

2019 年 4 月 4 日

1 The Formulation of Experience Adaptation

针对一个优化问题分布 \mathcal{F} , 假设任意两个优化问题 $f_i, f_j \sim \mathcal{F}, i \neq j$ 相似且不同。在工作 [1] 中提出了一种发掘问题之间的相似性, 借助以往优化经验用来加速新优化问题上优化过程的方法。令 $F_s = \{f_1^s, f_2^s, \dots, f_M^s\}$ 表示源优化问题集合 (the set of source optimization problems), 其中任意 $f_i^s \sim \mathcal{F}$ 。令 $F_t = \{f_1^t, f_2^t, \dots, f_N^t\}$ 表示目标优化问题集合 (the set of target optimization problems), 其中任意 $f_j^t \sim \mathcal{F}$ 。

在经验优化中, 对于每一个源优化问题 f_i^s , 我们会事先对其进行求解, 并将求解过程保存下来, 令 \mathcal{D}_i^s 对应 f_i^s 问题的优化经验数据集, 具体的优化经验数据组织方法见 [1]。对于每一个源优化问题均会得到一个优化经验数据集, 将此组成一个集合记为 $D = \{\mathcal{D}_1^s, \mathcal{D}_2^s, \dots, \mathcal{D}_M^s\}$, 在工作 [1] 中, 我们在 D 上训练一个预测器 ψ , 并将 ψ 用于 f_j^t 的问题优化上, 是其优化过程加速。

在以上处理方法中, 我们将得到的所有经验数据均用来训练 ψ 。值得注意的是, 即使 f_i^s 和 f_j^t 出自同一个问题分布, 但是源问题与目标问题之间的相似性是不同的, 即在源问题集合中某些问题与目标问题强相似, 其优化经验可以帮助加速目标问题的优化过程; 源问题中的某些问题与目标问题弱相似, 其优化经验对目标问题优化过程加速效果甚微甚至有负面影响。而且问题之间的相似性很难获得, 甚至很难量化定义相似性。所以我们无法直接将不相似的经验从训练 ψ 的数据集中剔除, 寻找一种自适应的经验使用方法变得非常重要。

2 Experience Adaptation Method

在此方法中, 我们先对每一个经验数据训练一个预测器, 令 $\boldsymbol{\psi} = \{\psi_1, \psi_2, \dots, \psi_M\}$ 。源问题, 经验数据, 预测器的对应关系如下:

$$\begin{aligned} f_1^s &\rightarrow \mathcal{D}_1^s \rightarrow \psi_1 \\ f_2^s &\rightarrow \mathcal{D}_2^s \rightarrow \psi_2 \\ &\dots \\ f_M^s &\rightarrow \mathcal{D}_M^s \rightarrow \psi_M \end{aligned} \tag{2.1}$$

对于任意一个 $\psi_i \in \boldsymbol{\psi}$, 在优化中, 对于一个采样 x , $y_i = \psi_i(x) \in [0, 1]$, y_i 的直观理解为 ψ_i 认为采样 x 能够比当前最优解更优的概率。在我们的自适应算法中, 我们通过 $\boldsymbol{\psi}$ 对 x 做预测的结果与 x 真实的结果之间的比较, 用来淘汰 $\boldsymbol{\psi}$ 中不能正确做出预测的 ψ 从而达到选择的目的。

Algorithm 1 Derivative-Free optimization with Experience Adaptation

Input:

$\boldsymbol{\psi} = \{\psi_1, \psi_2, \dots, \psi_M\}$: 经验预测器集合;
 f : 目标优化问题评估函数;
 B : 评估次数上界;
 K : 预采样次数;
Sample: 优化方法中的采样函数;
Update: 优化方法中的更新模型函数;
Sign: 指示函数, 若输入大于等于 0.5, 输出 1, 若输入小于 0.5, 输出 0。

Procedure:

```

1: 初始化优化方法, 并得到当前最优采样  $\tilde{x}$ 
2: for  $t = 1$  to  $B$  do
3:    $Y = \emptyset, X = \emptyset$ 
4:   for  $k = 1$  to  $K$  do
5:      $x_k = \text{Sample}(), X = X \cup \{x_k\}, \mathbf{y}_k = \emptyset$ 
6:     for  $i = 1$  to  $|\boldsymbol{\psi}|$  do
7:        $y_{k,i} = \psi_i(x_k), \mathbf{y}_k = \mathbf{y}_k \cup \{y_{k,i}\}$ 
8:     end for
9:      $Y = Y \cup \{\mathbf{y}_k\}$ 
10:  end for
11:   $\tilde{k} = \operatorname{argmax}_{k \in \{1, 2, \dots, K\}} \frac{1}{|\boldsymbol{\psi}|} \sum_{i=1}^{|\boldsymbol{\psi}|} y_{k,i}$ 
12:  Update ( $x_{\tilde{k}}, f(x_{\tilde{k}})$ )
13:  if  $f(x_{\tilde{k}}) < f(\tilde{x})$  then
14:     $\ell_{\tilde{k}} = 1, \tilde{x} = x_{\tilde{k}}$ 
15:  else
16:     $\ell_{\tilde{k}} = 0$ 
17:  end if
18:  for  $i = 1$  to  $|\boldsymbol{\psi}|$  do
19:    if  $\text{Sign}(y_{\tilde{k},i}) \neq \ell_{\tilde{k}}$  then
20:       $\boldsymbol{\psi} = \boldsymbol{\psi} - \{\psi_i\}$ 
21:    end if
22:  end for
23: end for
24: return  $\tilde{x}$ 

```

经验自适应的非梯度优化算法如 Algorithm 1所示, 在算法输入中 $\boldsymbol{\psi}$ 为源问题上获得的经验模型, 与非梯度优化算法有关的是 **Sample** 和 **Update** 两个函数, 分别对应采样和模型更新。

在此算法中，每一个外层循环 (line 2 to 23) 对应一个采样评估。每个待评估的采样通过一轮预采样获得 (line 4 to 10)，采样步骤如 line 5 所示，对于每一个预采样，均用所有的经验模型对其做一遍预测 (line 6 to 8)，我们称为预评估。预采样步骤完成后，我们需要从预采样中选出一个采样进行真实函数评估，如 line 11 所示，我们选取预评估值均值最大的采样做真实评估，同时更新优化模型 (line 12)。得到真实评估值之后我们可以给此采样一个真实标记 (line 13 to 17)，利用真实标记，我们可以对经验模型进行选择 (line 18 to 22)，选择方式为若经验模型的预测与真实标记不一致，则直接删除。依次进行循环，直到真实评估次数用尽，返回当前得到的最好的采样 (line 24)。

参考文献

- [1] Yi-Qi Hu, Yang Yu, and Zhi-Hua Zhou. Experienced optimization with reusable directional model for hyper-parameter search. In *Proceeding of the 27th International Joint Conference on Artificial Intelligence*, pages 2276–2282, 2018.