贝叶斯优化

周志健

1 贝叶斯优化(bayesian optimization)

1.1 简介

贝叶斯优化是一种全局黑盒优化方法。该方法利用先验数据建立目标函数的代理模型，不需要依赖目标函数的梯度，这是此方法可以优化黑盒的根本原因。贝叶斯优化方法主要包含代理模型（高斯过程回归模型）和采集函数这两个核心部分。

1.2 高斯过程(gaussian progress)

当我们有 n 个取样点 $x_1,\cdots,x_n$，以及相应的目标函数值，$y(x_1),\cdots,y(x_n)$，那么我们可以得到下面这个高斯分布，即：

$$\mathbf{y} \sim N(0,\mathbf{K}) \tag{1}$$

其中 $\mathbf{y}=\left(y(x_1)\cdots y(x_n)\right)^{\mathrm{T}}$，$\mathbf{K}_{ij}=k\left(x_i,x_j\right)$，$k\left(x_i,x_j\right)$ 为协方差核函数。本文中，协方差核函数选择常用的 RBF 核:

$$k_{\mathrm{RBF}}\left(x_i,x_j\right)=\sigma^2\exp\left(-\frac{r^2}{2l^2}\right) \tag{2}$$

其中 l 和 σ 表示超参数，它们决定了高斯过程曲线的形状，$r=\left|x_i-x_j\right|$。

因此，训练高斯过程回归模型时，实际是通过梯度法优化其边际似然函数的对数化形式，由此确定参数 l 和 σ 的值，即：

$$\log p\left(\mathbf{y}\middle|x_{1:n}\right)=-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{y}-\frac{1}{2}\log|\mathbf{K}|-\frac{n}{2}\log 2\pi \tag{3}$$

其中 $p\left(\mathbf{y}\middle|x_{1:n}\right)$ 是基于训练集的高斯过程的边际似然函数。

指定一个采样点 x_{n+1}，训练好的高斯过程回归模型可以给出其目标函数值的预测分布，即：

$$y\left(x_{n+1}\right)\sim N\left(\mu\left(x_{n+1}\right),\sigma\left(x_{n+1}\right)\right) \tag{4}$$

其中均值和方差分别由下面两个公式给出：

$$\mu\left(x_{n+1}\right)=\mathbf{K}_{n+1}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{y} \tag{5}$$

$$\sigma\left(x_{n+1}\right)=k\left(x_{n+1},x_{n+1}\right)-\mathbf{K}_{n+1}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{K}_{n+1} \tag{6}$$

其中$\mathbf{K}_{n+1}=\left(k\left(x_{n+1},x_1\right),\ldots,k\left(x_{n+1},x_n\right)\right)^{\mathrm{T}}$。

1.3 采集函数(acquisition function)

采集函数通过采样点的均值和方差指导下一次采样，通过最大化采集函数以尽量少的迭代次数找到目标函数的最优值，即：

$$x_{\mathrm{next}}=\arg\max_{\tau\in\Omega}\mathrm{AC}\left(x\right) \tag{7}$$

其中$\mathrm{AC}\left(\cdot\right)$表示采集函数。

1.3.1 单目标 AC

$$UCB=\mu(x)+k\sigma(x) \tag{8}$$

其中$\mu(x)$指的是均值，$\sigma(x)$指的是方差。

k值人为设定，用以权衡 exploration 和 exploitation。

1.3.2 多目标 AC

Pareto 理论：

目标函数空间中存在成对的被支配点和支配点，支配点在任一方向上都优于或等于被支配点。

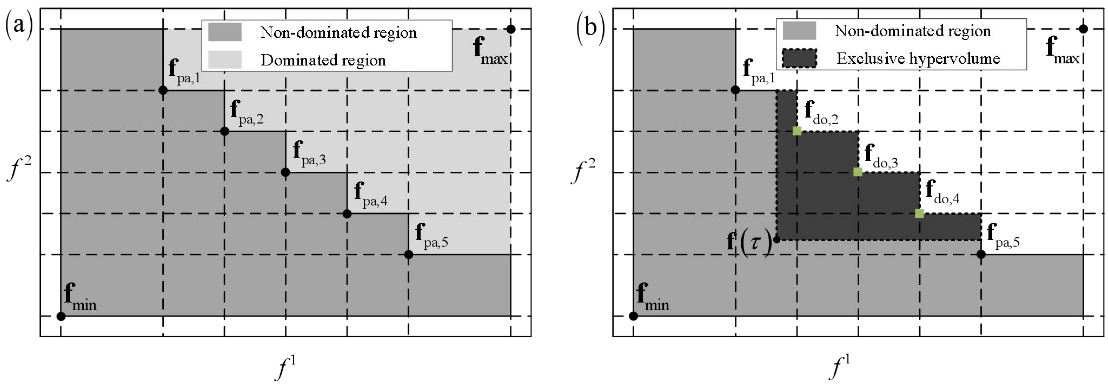

图 1 两个目标函数的帕累托前沿示意图

Emmi(the expected maximin improvement)AC 函数：

$$\mathrm{E}[I(\mathbf{f}(x))] \tag{9}$$

其中 $I(\mathbf{y}(x))$ 如下：

$$\begin{aligned}\boldsymbol{I}\left(\mathbf{f}\left(\boldsymbol{x}\right)\right) \equiv -\max_{\boldsymbol{x}_i \in \mathcal{P}_x} \min_{j=1,\ldots,m}\left(f_j\left(\boldsymbol{x}\right)-f_j\left(\boldsymbol{x}_i\right)\right) \\ \times 1\left[-\max_{\boldsymbol{x}_i \in \mathcal{P}_x} \min_{j=1,\ldots,m}\left(f_j(\boldsymbol{x})-f_j(\boldsymbol{x}_i)\right)>0\right]\end{aligned} \tag{10}$$

注意此处 $\mathbf{f}=\left(f_1\left(x\right)\cdots f_n\left(x\right)\right)^{\mathrm{T}}$，$P_x$ 指训练数据中的 pareto 集合。

1.4 伪代码和流程图

Algorithm 1 单目标贝叶斯优化

随机选取n个样本构成训练集$D_{0:0}$
for $i = 1, 2...$ **do**
  获取训练集$X_{1:N}$和训练集对应的目标函数值y
  优化似然函数：$logp(\mathbf{y}|X_{1:n}) = -\frac{1}{2}\mathbf{y}^T K^{-1}\mathbf{y} - \frac{1}{2}log|K| - \frac{n}{2}log2\pi$
  求解$x_i = argmax(UCB)$
  将 x_i 代入黑匣子中获得返回值 $y(x_i)$
  将$y(x_i)$ 和 x_i 加入训练集 $D_{0:i}$ 中构成新的训练集 $D_{0:i+1}$
end for

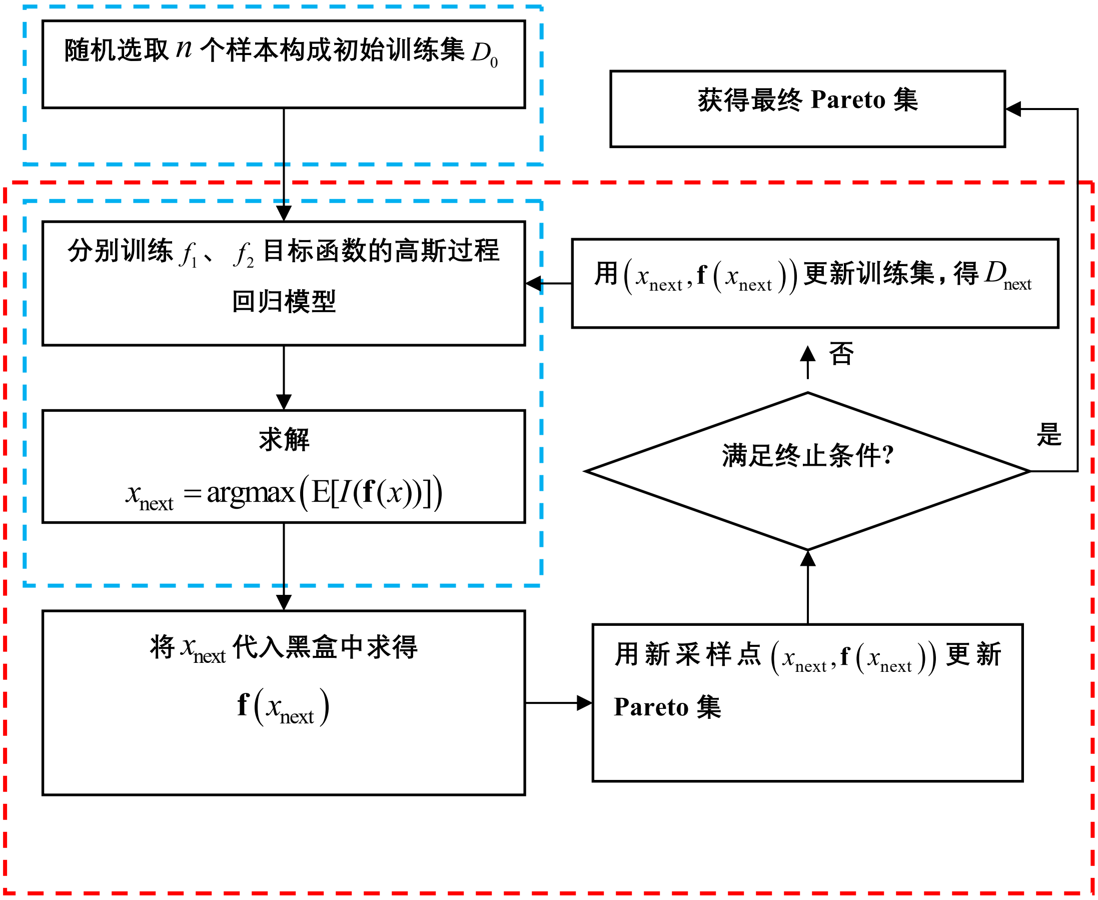

图 2 两个目标函数的贝叶斯优化

2 均方误差

$$\mathrm{MSE} = \frac{1}{\mathrm{m}} \sum_{\mathrm{i=1}}^{\mathrm{m}} (f(x_i) - y_i) \tag{10}$$

其中 $f(x_i)$ 为模型预测值， y_i 为真实值。

3 总结

MSE 表

手动实现 GP	Sklearn 库 GP
24963.5	41512.1

数据量少时，手动实现的算法效果较好，但算法优化差，时间和空间复杂度都很高。特别是未经优化的矩阵运算导致算法无法应对大数据集。