

第3章

研究综述

3.1 综述部分的总体结构

近年来,关于贝叶斯优化方法和实践方面已有一些全面综述和教程,每篇综述都专注于特定的重点。例如,文献[184]对克里格插值法在早期研究中的应用进行了回顾,并探讨了其在约束优化方面的扩展。在文献[53]中,研究者提供了一个针对贝叶斯优化的教程,特别侧重于将贝叶斯优化应用于偏好建模和主动用户交互问题。此外,文献[28]全面回顾了贝叶斯优化的基础,并详细阐述了统计建模和流行的适应性函数的应用。同时,文献[185]也探讨了贝叶斯优化领域的一些最新进展,特别关注多目标优化(MFO)和约束优化方面的发展。虽然已有多篇综述论文对贝叶斯优化进行了深入探讨,但尚未涵盖贝叶斯优化领域的所有丰富扩展。自文献[28]发表以来,贝叶斯优化领域不断取得许多新进展,这使得对这一充满活力的研究领域进行更新和全面的综述变得尤为重要。因此,对贝叶斯优化领域进行新的研究和综述将有助于深入了解当前的最新发展,使研究人员和从业者受益。

3.2 相关研究工作

本节将对最先进的贝叶斯优化算法进行简要概述,重点关注其中最重要的研究进展。随后,根据优化问题的特点对现有的研究进行分类和讨论,为当前的研究提供一个清晰的框架。通过系统梳理,可以更好地理解贝叶斯优化领域的快速发展。

3.2.1 高维优化

高维黑盒优化问题是极具挑战性的,在许多应用程序中都很常见^[83,89]。值得注意的是,贝叶斯优化中的维度数可能从几十到数千个,甚至达到10亿^[186]。虽然贝叶斯优化已成功应用于低维昂贵和黑箱优化问题,但在搜索空间的维数大于10~20时,其性能明显下降^[84,86]。因此,对高维问题的扩展仍然是一个关键的开放性挑战。



具体来说,高维问题的贝叶斯优化面临以下主要困难:非参数回归,如高斯过程,在本质上是困难的,因为随着维数增加,搜索空间呈指数级增长。在高维空间中使用常见的基于距离的核函数学习模型变得更加困难,因为搜索空间的生长速度比合理的抽样开销要快得多。此外,超参数的数量通常会随着输入维数的增加而增加,从而使模型的训练变得越来越困难。另外,获取函数通常是多模态的问题,具有较大的平坦表面^[92],因此获取函数的优化并不是平凡的,特别是在高维问题和样本数量有限的情况下。

需要注意的是,上述问题与高斯过程的可伸缩性有所不同。为了在高维空间中构建一个可靠的高斯过程,可能需要更多的观测数据,这由于高斯过程采样复杂度是随数据大小呈指数增长,其可伸缩性面临挑战。尽管近年来对高斯过程的可伸缩性进行了广泛研究,以适应许多观测^[187-188],但这些方法主要关注存在大量数据但维度仍然较小或中等的情况。此外,即使可以将高斯过程用于高维问题,仍然会面临获取函数优化的困难,因为获取函数通常是多模态的问题,需要在高维情况下进行更多的代理模型优化。因此,我们感兴趣的是用可扩展的方法处理高维问题,而不仅仅局限于构建高维的高斯过程。

目前,解决高维贝叶斯优化问题的算法大多基于两个结构假设,但也有例外:①高维目标函数具有低维有效子空间,这推动了基于变量选择和嵌入方法的发展;②原始目标函数可以是几个低维函数的和,从而产生了基于加性结构的方法。解决高维贝叶斯优化问题,尤其是涉及大量数据的情况,通常需要采用代理模型、局部建模和并行批处理选择等方法。接下来,将详细讨论处理高维优化问题的现有工作。

1. 变量选择

为了缓解高维问题的困扰,一种简单的想法是采用降维技术。其中一个重要假设是,原始目标函数只在一个低维子空间内变化,称为主动/有效子空间^[189]。为了确定对感兴趣数量的贡献最大的输入变量,文献[190]利用了一些敏感性分析技术,以评估每个变量相对于感兴趣的数量的相对重要性。文献[189]提出了两种策略,即有限差分序列似然比检验和 GP 序列似然比检验,用于筛选贡献最大的变量。另一个常用的指标是自动相关性确定协方差^[191]的相关长度值。其基本思想是,长度尺度值越大,对应变量的重要性就越小。

2. 线性/非线性嵌入

最近的发展通过定义一个基于线性或非线性嵌入来利用目标函数的潜在空间,而不是去除非活动变量来降低维数。例如,文献[192]指出,对于任意 $x \in \mathbb{R}^D$ 和一个随机矩阵 $A \in \mathbb{R}^{D \times d}$,当概率为 1 时,存在一个点 $y \in \mathbb{R}^d$,使得 $f(x) = f(Ay)$ 。这种观察结果允许我们在低维空间中执行贝叶斯优化算法优化原始的高维函数。因此,该文献提出了一种具有随机嵌入功能的贝叶斯优化算法(BO with random embedding, REMBO)。最近,一





些 REMBO 的变种参见文献[78,83-84]。除了随机嵌入方法的成功应用,人们还提出许多学习内在有效子空间的算法,如基于变分自编码器(VAE)^[193]的无监督学习、监督学习^[81]和半监督学习^[194]。

上述基于结构假设的方法大多采用线性投影将贝叶斯优化扩展到高维。近年来,一些先进的技术可以进一步研究利用非线性嵌入^[195]的搜索空间结构。与线性嵌入相比,非线性嵌入技术,也被称为几何感知的贝叶斯优化^[29],具有相当多的表现力和灵活性。然而,这些方法需要更多的数据来学习嵌入,并假设搜索空间不是欧几里得的,而是各种流形,如黎曼流形^[196]。

3. 可加性结构

上述方法背后的低有效维度假设过度限制,因为所有的输入变量都可能对目标函数产生影响。因此,在高维贝叶斯优化的背景下,另一种显著的结构假设得到了探索,称为可加性结构。可加性结构已经被应用于可加性高斯过程^[197]。在文献[86]中提出了一种名为 Add-GP-UCB 的算法,其假设目标函数 $f(\mathbf{x}): \mathcal{X} \rightarrow \mathbb{R}$, 其中输入空间 $X = [0, 1]^D$ 是由小的、不重叠的维度组合函数构成的,即

$$f(\mathbf{x}) = f^{(1)}(\mathbf{x}^{(1)}) + f^{(2)}(\mathbf{x}^{(2)}) + \cdots + f^{(M)}(\mathbf{x}^{(M)}) \quad (3-1)$$

其中, $\mathbf{x}^{(j)} \in X^{(j)} = [0, 1]^{d_j}$ 表示输入变量的不相交子集。为了避免直接使用复杂的核函数,该方法采用随机生成一组特征空间的潜在分解,并选择具有最高 GP 边际似然值的核函数,每个核函数在输入维度的子集上操作。为了更有效地学习可加结构,该方法还引入马尔可夫链蒙特卡罗法(MCMC)^[198]、Gibbs 抽样^[88]和 TS^[199]等技术。

关于 Add-GP-UCB 算法的另一个主要问题是对输入维度的不相交子集的限制,在随后的文献[26,91]中已经得到解决。通过引入投影可加性假设,对两种结构假设进行了泛化,即低活动假设和可加性结构假设。在文献[91,200]中,通过依赖图或稀疏因子图表示可加分解,允许变量组之间存在重叠。

4. 高维贝叶斯优化中的大规模数据

虽然已经有许多贝叶斯优化研究致力于解决大规模观测和高维输入空间的问题,但很少有人考虑到带有大量训练数据的高维问题。然而,该情况是不得不考虑的,因为高维空间中构建代理模型需要更多的数据。一些先进的技术如贝叶斯神经网络等被提出来取代传统的高维 GPs,以获得更好的可扩展性和灵活性。采用并行的局部建模和批量选择的方法,如集成贝叶斯优化和异构集成模型,也被应用于解决大规模高维问题。同时,置信域方法用一种局部概率方法(TuRBO)处理高维空间中的大规模数据。

5. 讨论

虽然上述结构假设对于高维空间中的 GP 建模非常有利,但在现实应用中,目标函数





或搜索空间可能是不可分解的,可能不符合这些假设。因此,未来的研究方向之一是如何有效地学习低维潜在空间。近年来,高维组合优化和图结构目标函数的存在给贝叶斯优化带来了挑战,值得进一步研究。此外,尽管大多数研究工作都考虑了高维搜索空间,但对高维多输出贝叶斯优化的研究仍然相对缺乏。这些问题都值得在未来的研究中深入探讨。

3.2.2 组合优化

在现实世界的应用中,对于组合空间上的黑盒函数进行优化是一个普遍存在但具有挑战性的任务,比如涉及整数、集合、序列、分类或图结构等输入变量的情况。假设昂贵的黑盒目标函数 $f: \mathcal{H} \rightarrow \mathbb{R}$, 组合优化的目标是

$$\mathbf{h}^* = \operatorname{argmax} f(\mathbf{h}) \quad (3-2)$$

其中, $\mathcal{H}$ 为搜索空间。对于混合搜索空间上的问题, $\mathcal{H} = [\mathcal{C}, \mathcal{X}]$, $\mathcal{C}$ 和 $\mathcal{X}$ 分别表示离散搜索空间和连续搜索空间。具体地说,离散变量可以根据给定变量的可能值之间的顺序关系,分为顺序和名义(或定量和定性)变量。例如,分类变量指的是一个无序的集。贝叶斯优化已经成为处理昂贵的评估黑盒问题的典型成熟范例。然而,大多数基于高斯过程的贝叶斯优化方法明确假设连续空间,导致在组合型域中扩展性较差。这主要归因于难以在组合空间上定义内核和距离度量来解释变量之间复杂的相互作用。请注意,基于梯度的优化获取函数的方法并不直接适用于离散变量的存在。此外,贝叶斯优化严重面临候选解的数量随着组合域的参数呈指数增长(称为组合爆炸)的问题。因此,组合贝叶斯优化面临两个主要挑战。一是在组合空间上建立有效的代理模型,二是在组合域上有效地搜索下一个结构。一种简单的方法是通过将离散变量视为连续变量,构造高斯代理模型并优化获取函数,然后通过独热编码策略^[201]识别出的具有实值的下一个样本点的最接近整数。显然,这种方法忽略了搜索空间的性质,可能会重复选择相同的新样本,从而降低贝叶斯优化的效率。另外,许多研究借用了变分自编码器的优雅性,将高维的、离散的输入映射到一个低维的连续空间^[202]上。在贝叶斯优化的背景下,通过引入组合空间的代理模型,人们投入了很多工作处理昂贵的组合优化问题。

1. 固有离散模型

为了避免在基于 GP 的贝叶斯优化中遇到的困难,我们采用了一些固有的离散模型(如神经网络^[203]和随机森林)作为代理模型,其中基于树的模型是应用较广泛的模型。例如,随机森林已被应用于文献[204]中的组合贝叶斯优化。然而,这种方法需要执行不可取的外推效果。因此,通常使用树结构的 Parzen 估计模型替代文献[205]中的 GP,但该方法需要大量的训练数据。另一种想法是使用连续的代理模型保证整数值最优,这激





发了一种称为 IDONE^[206] 的使用分段线性代理模型的方法。

为了提高获取函数在组合优化中的搜索效率,搜索控制知识被引入到分支-定界搜索^[207]中。此外,BOCS 算法用来缓解组合空间^[208]的组合爆炸问题。

2. 具有离散距离度量的核函数

组合贝叶斯优化的另一种常用方法是修改 GP 核计算中的距离度量,从而可以正确地捕获组合空间中的相似性。例如,汉明距离被广泛用于度量离散变量之间的相似度,通常采用进化算法对获取函数^[204]进行优化。最近,组合空间的图表示出现在了最前沿,为 GP 中的图内核做出了贡献。文献[209]提出了 COMBO,它在组合搜索空间上构造了一个组合图,其中图中两个顶点之间的最短路径相当于汉明距离。随后,该方法利用图的傅里叶变换推导出图上的扩散核。为了规避 COMBO 的计算瓶颈,该方法进一步研究了图表示的结构,并提取了一小部分特征^[210]。需要注意的是,基于图的组合贝叶斯优化已被广泛应用于神经架构搜索^[211-212]。

3. 混合搜索空间下的贝叶斯优化

很少有研究考虑混合变量组合问题,其中输入变量包括连续变量和离散变量,如整数和分类输入。在离散空间上具有新距离度量的内核为解决组合优化问题提供了启示。因此,有人尝试用类似的方法解决组合贝叶斯优化问题,即把定义在不同输入变量上的核结合起来^[213]。有趣的是,文献[214]使用了定义在不同域上的内核的乘积解决受限的混合变量问题。接着,类似的核函数被定义用于解决包含不同大小搜索空间的混合变量问题^[215]。在混合变量设置^[206]中,将每个变量视为一个多臂赌博机是指将每个变量视为一个独立的决策问题,通过这种方法将多臂赌博机的方法与贝叶斯优化结合起来。对于混合变量优化问题,这种方法可以在搜索空间中寻找更优解。

4. 讨论

虽然大多数组合贝叶斯优化方法侧重于代理模型的构建,但组合爆炸问题仍然具有挑战性。计算瓶颈和可伸缩性的挑战值得进一步研究。此外,由于组合优化中涉及约束,选择满足约束的新解也非常具有研究价值。

3.2.3 噪声和鲁棒优化

为了构造贝叶斯优化^[216]中的 GP,需要对数据中的噪声进行两个假设。首先,假设输入点的测量是无噪声的。其次,假设观测中的噪声服从一个恒定方差的正态分布,称为同方差高斯白噪声。然而,这两种假设在实践中都不成立,导致优化性能较差。因此,考虑噪声观测、异常值和输入相关噪声的贝叶斯优化方法已经得到深入研究。





1. 贝叶斯优化用于输出噪声

对于有噪声输出的优化问题,目标函数可以用 $f: X \rightarrow \mathbb{R}$ 描述,由噪声观测 $y = f(\mathbf{x}) + \epsilon$ 产生,其中 ϵ 是可加/输出噪声。大多数针对存在输出噪声问题的贝叶斯优化方法都采用标准 GP 作为代理模型,并专注于设计新的获取函数^[217]。首先,无噪声 EI 对噪声观测的扩展得到了广泛的研究。一个主要的问题是,目前的最佳目标值 $f(\mathbf{x}^*)$ 并不完全可知。一种直接的方法是用一些合理的值替换 $f(\mathbf{x}^*)$,这被称为使用“插件”^[217]的期望改进。文献[218]通过替换当前的最佳目标值,提出一个增强的 EI,并随后在标准 EI 中增加了一个惩罚项。此外,该方法用 GP 代理给出的 β 分位数作为参考。该方法进一步定义了基于 β 分位数最低值减少的改进,产生了能够解释异构噪声的预期分位数改进(EQI)。与 EQI 类似,改进由 KG 策略定义,并引入近似知识梯度(Aproximate KG, AKG)^[219]。从根本上说,AKG 是一种基于知识改进的 EI;然而,AKG 的评估是计算密集型的。另一类自然处理输出噪声的获取函数是基于信息的 AFs(用于单目标优化的获取函数),如 PES^[66] 和 TS 算法^[220]。此外,处理输出噪声^[221]的方法利用噪声观测构造克里格回归;并利用克里格提供的预测采样点建立插值克里格插值即再插值,使标准 EI 能够选择新的样本。

2. 异常值的贝叶斯优化

除上述测量/输出噪声外,在实际实验中,由于不规则和孤立的干扰、仪器故障或潜在的人为误差,观测结果经常被异常值/极端观测结果污染。正如在文献[222]中所指出的那样,采用高斯分布作为先验值和似然值的标准 GP 模型对极端观测都很敏感。另一个原因是 GP 是非参数的和插值的。

通常,贝叶斯优化采用对异常值的存在不敏感的稳健 GP 解释异常值。从数学上讲,鲁棒 GP 模型背后的主要思想是,使用一个适当的尾部较重的噪声模型而非假设正常噪声解释离群数据^[223]。最常用的噪声模型是 t 分布^[224-225]。然而,使用 t 似然不允许后验分布的封闭推理形式,因此,通常需要一些近似推理的技术,如拉普拉斯近似^[225]。最近,文献[224]提出了一种离群值处理算法,通过将鲁棒 GP 与 t 似然与离群值检测相结合,将数据点分类为离群值或奇异值。该方法可以去除异常值,因此可建立标准的 GP,从而产生一个更有效的、具有更好收敛性的鲁棒性方法。

3. 受损输入的贝叶斯优化

该类方法在建模 GP^[226] 时首先考虑了输入依赖的噪声,通过允许噪声方差是输入的函数而不是常数,引入了异速噪声。因此,噪声方差被视为一个随机变量,并使用一个独立的 GP 建模噪声水平的对数。异方差 GP 回归中的推理具有挑战性,因为与同方差情况不同,预测密度和边际似然不再具有可分析性。可以使用 MCMC 方法近似后验噪声





方差,但这种方法非常耗时。建议使用其他近似方法,包括变分推理^[227]、拉普拉斯近似^[225]和期望传播^[223]。

上述方法通过将输入测量值保持为确定值,并改变相应的输出方差进行补偿,处理带有输入噪声的数据集。文献[216]指出,输入噪声的影响与输入到输出函数的梯度有关。因此,提出一个噪声输入 GP(NIGP),根据后验的一阶泰勒展开将输入噪声转移到输出。具体来说,NIGP 采用函数的局部线性化,并利用它将不确定性从输入传播到 GP^[216] 的输出。

以上思想的直观想法是将输入噪声传播到输出空间,然而这可能导致不必要的探索。文献[228]通过在有效改进方法中考虑输入噪声解决这个问题,这样输入噪声可以通过所有模型和函数查询进行传播。更确切地说,他们使用无损变换(UT)定义了无损期望改进和无损最优继任者。UT 首先从原始分布确定性地选择一组样本。然后,对每个样本应用非线性函数以生成转换点。因此,变换后分布的均值和协方差可以根据变换点的加权组合形成。

输入/查询不确定性^[229]是与输入相关的噪声密切相关的术语。这意味着对实际查询位置的估计也受环境变量^[230]或受噪声等不确定性的影响。当将贝叶斯优化扩展到具有输入不确定性的问题时,采用了两种经典的问题公式,即概率鲁棒优化和最坏情况鲁棒优化,分别从概率和确定性的角度考虑。在概率鲁棒优化中,假设输入或环境变量的分布是未知的。因此,为了考虑定位噪声,在输入空间上设置一个先验,并通过某些鲁棒性测量的期望值评估性能。文献[229]在贝叶斯优化框架内引入了受噪声干扰的输入,即不确定性。在这种情况下,通过对输入分布的未知函数进行积分,鲁棒优化问题被表述为一个约束问题。因此,噪声因素可以被整合,并引入类似于受约束 EI 的获取函数,在决策空间中完全选择新的查询。

相较而言,最坏情况鲁棒目标的目的是寻找一个对不确定参数的最坏实现具有鲁棒性的解,它被表述为一个最小最大优化问题,即

$$\max_{\mathbf{x}} \min_{\mathbf{c} \in U} f(\mathbf{x}, \mathbf{c}) \quad (3-3)$$

其中, $\mathbf{x}$ 为决策向量, $\mathbf{c} \in U$ 为不确定性,其中 U 为不确定性集合。文献[230]使用一个松弛程序探索如何将 EGO 用于最坏情况下的鲁棒优化,从而对设计变量和不确定性变量进行迭代优化。然而,这种策略效率不高,因为之前的观测结果无法重复使用。文献[231]提出了一种使用新的期望改进的修正 EI。

4. 讨论

新的 AFs 是针对可加输出噪声而设计的,而基于 t 分布的 GP 则是为了适应异常值而开发的。最近,现实场景中对鲁棒性有新要求的更复杂问题设置引起越来越多的关注。





例如,如何应对对抗性破坏^[232]是一个很有前景的研究方向。此外,批量优化^[232]中的鲁棒性也至关重要。

3.2.4 昂贵的约束优化

许多优化问题受到各种类型的约束,对目标函数和约束的评估都可以是计算密集型的,也可以是代价高昂的。这类问题称为昂贵的约束优化问题(ECOPs)。通常 ECOP 可以表述为

$$\begin{aligned} \min_{\mathbf{x}} f(\mathbf{x}) &= (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})) \\ \text{s.t. } c_j(\mathbf{x}) &\geq a_j, j = 1, 2, \dots, q \\ h_i(\mathbf{x}) &= b_i, i = 1, 2, \dots, r \\ \mathbf{x} &\in X \end{aligned} \quad (3-4)$$

其中, $\mathbf{x} = (x_1, x_2, \dots, x_d)$ 是具有 d 个决策变量的决策向量, X 表示决策空间, $c_j(\mathbf{x})$ 和 $h_i(\mathbf{x})$ 分别表示不等式和等式约束。由于同时考虑单目标和多目标问题,目标向量 f 由 m 个目标 ($m = 1, 2, \dots, N$) 组成。

基于约束优化问题的贝叶斯优化可以大致分为两类: ①在利用高斯过程的情况下, 提出新的获取函数考虑贝叶斯优化框架内的约束问题, 这被称为受限制的贝叶斯优化(CBO)。最近, CBO 变得流行起来, 尤其是在求解单目标约束问题的情况下。根据 CBO 中不同的获取函数, 将各种 CBO 算法分为 3 个子类别, 即基于可行性概率的方法、基于期望体积减小的方法和多步骤前瞻方法。②为了规避 ECOP 中遇到的计算负担, 现有的约束处理方法(通常是进化算法)采用了贝叶斯优化。将这些称为代理辅助约束处理方法。下面进行介绍和讨论。

1. 可行性概率

现有获取函数与约束可行性指标(如可行性概率)的结合, 为约束优化提供了原则性方法。最具代表性的工作是对完善的 EI 的扩展, 称为带约束的 EI(EI with Constraints, EIC)^[8]。以前的 EIC 方法之一称为约束 EI(constrained EI, cEI)或约束加权 EI, 旨在最大化当前最佳可行观察点的期望可行改进。通常, cEI 是 EI 乘以约束满足概率, 公式如下, 即

$$\text{cEI}(\mathbf{x}) = \text{EI}(\mathbf{x}) \prod_{j=1}^q \text{Pr}(\hat{c}_j(\mathbf{x}) \leq a_j) \quad (3-5)$$

其中每个约束被假定为独立的, 所有评估成本昂贵的函数都由独立的 GP 近似, 并且 $\hat{c}_j$ 表示对第 j 个约束的模型预测。有趣的是, 类似的想法也在文献[234]中讨论过, 而在文献[235]中也进行了重新审视。如式(3-5)所示, cEI 面临几个问题。首先, 需要目前的最





佳观测值,这在一些应用中是站不住脚的,如噪声实验。因此,文献[236]最近的一项工作通过贪婪搜索批量优化直接将 cEI 扩展到噪声观测。其次,对于高维约束的问题,cEI 可能很脆弱,因为可行性概率的乘积在最优值所在的可行性边界附近趋于零,导致在有趣的区域^[233]内的 cEI 值非常小。

2. 期望体积减小

另一类获取函数是通过减少基于观察的感兴趣数量的特定类型的不确定性度量适应约束的,这被称为逐步不确定性减少^[237]。正如在之前的研究^[237]中所建议的,根据不同类型的不确定性度量,可以导出许多获取函数推断任何感兴趣的数量。在文献[238]中,定义了一个基于 PI 的不确定性测度,其中通过结合可行性概率进一步解释约束。使用同样的原理,文献[239]中的集成期望条件改进定义了约束满足概率下 EI 的期望减少,允许不可行的区域提供信息。另一个流行的不确定性度量是受信息论启发的熵,文献[66, 240]对此进行了探索。文献[67]通过引入条件预测分布,假设目标和约束具有独立的 GP 先验,将 PES 扩展到未知的约束问题。后续文献[241]进一步研究了在解耦约束存在下 PES 的使用,其中目标函数和约束函数的子集可以独立评估。然而,PES 因为计算困难,促使在最近的一项工作[240]中使用 MES 处理约束问题。

3. 多步预测方法

大多数获取函数是短视的,称为一步前瞻方法,因为它们贪婪地选择下一个真实评估的位置,忽略了当前选择对未来步骤的影响。相比之下,很少有非短视的获取函数被提出,这些函数通过最大化长期回报选择样本。例如,文献[154]将前瞻贝叶斯优化表述为一个动态规划(DP)问题,并通过一种称为 rollout 的近似 DP 方法解决。随后,这项工作随后通过将阶段奖励重新定义为满足约束条件的目标函数的减少而扩展到 CBO^[242]。由于 rollout 导致了计算负担,文献[243]提出了一种称为 2-OPT-C 的受限两步前瞻获取函数。

4. 辅助替代约束处理方法

上述提到的约束处理技术主要关注贝叶斯优化框架内的获取函数,其中高斯过程模型通常作为全局模型。在进化计算社区中,针对受约束的昂贵问题,已经尝试将两者结合。其中一种方法是使用多目标进化算法(Multi-Objective Evolutionary Algorithms, MOEAs)同时优化目标和约束。例如,可以将 EI 和可行性概率的乘积替换为两个目标,并通过 MOEA 进行优化,然后从得到的帕累托最优候选集中随机选择一组新样本^[244]。

5. 讨论

大多数约束处理的贝叶斯优化方法是通过引入新的获取函数(AFs)实现的,也有少





数尝试采用增广 Lagrange 松弛法将受约束的优化问题转化为简单的无约束问题^[245]。对于高度约束的问题,由于可行样本有限或甚至不可用,很难在整个搜索空间上构建具有良好质量的代理模型。一种有前途的方法是首先搜索可行区域,然后逐步逼近最佳可行解。例如,进行局部和全局搜索以加速对可行点的搜索^[85]非常有前景,但需要进一步的研究。在许多应用中,评估成本、用户偏好和公平性等因素可被定义为约束条件^[246],这是未来一个有趣的研究方向。

3.2.5 多目标优化

许多现实世界的优化问题都有多个可能相互冲突的目标需要同时进行优化,这类问题统称为多目标优化问题(Multi-objective Optimization Problems, MOPs)^[247]。在数学上,一个 MOP 可以被表述为

$$\begin{aligned} \min_{\mathbf{x}} \mathbf{f}(\mathbf{x}) &= (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})) \\ \text{s.t. } \mathbf{x} &\in \mathcal{X} \end{aligned} \quad (3-6)$$

其中, $\mathbf{x} = (x_1, x_2, \dots, x_d)$ 是具有 d 个决策变量的决策向量, $\mathcal{X}$ 表示决策空间, 目标向量 $\mathbf{f}$ 由 m ($m \geq 2$) 个子目标组成。请注意, 对于许多目标问题(Many-objective Optimization Problems, MaOPs)^[248], 目标问题的数量大于 3。这里的目标是找到一组最优解, 在不同的目标之间进行权衡, 被称为帕累托最优解。决策空间中的整个帕累托最优解集称为帕累托集(Pareto Set, PS), 目标空间中的映射称为帕累托前沿(Pareto Front, PF)。多目标优化的目的是找到 PF 的一个代表性子集, 而 MOEAs 已被证明能够成功地求解 MOPs^[247]。

MOP 中的目标函数可能是非常耗时或评估代价昂贵的。因此, 只有少量的适应性评估是负担得起的, 这使得普通的 MOEAs 几乎不实用。回想一下, 贝叶斯优化中的 GP 和 AFs 是为单目标黑盒问题设计的, 因此当贝叶斯优化扩展到 MOP 时, 出现新的挑战, 其中需要确定多个目标函数的采样, 并且必须考虑所获得的解集的准确性和多样性。为了应对这些挑战, 多目标贝叶斯优化得到广泛研究, 即将贝叶斯优化嵌入 MOEAs 或将 MOP 转化为单目标问题。多目标贝叶斯优化主要可分为 3 类: 贝叶斯优化与 MOEAs 的组合、基于性能指标的 AFs 和基于信息论的 AFs。请注意, 其中一些可能会重叠, 因此不能完全分离。

1. 贝叶斯优化与 MOEAs 的组合

由于多目标进化算法(MOEAs)在求解多目标优化问题(MOPs)方面取得了成功, 所以将贝叶斯优化与 MOEAs 结合是一种直接的方法。这样, 高斯过程(GPs)和现有的用于单目标优化的获取函数可以直接应用于 MOPs 中的每个目标函数。根据贝叶斯优化和进化算法协同工作的方式, 这些组合可以进一步分为两类, 即进化贝叶斯优化





(Evolutionary Bayesian Optimization, EBO) 和贝叶斯进化优化 (Bayesian Evolutionary Optimization, BEO)^[249], 如图 3.1 所示。在 EBO 中(见图 3.1(a)), 贝叶斯优化是基本框架, 其中使用进化算法优化获取函数。相比之下, 在 BEO 中(见图 3.1(b)), 以进化算法为基本框架, 采用 AF 作为选择后代个体进行采样的标准。然而, MOEA 环境选择的目标函数可能与 AF 不同。这些方法的区别在于采用的 MOEA 和选择新样本的策略。通常, 基于分解的 MOEAs 使用标量化函数(如切比雪夫标量化函数或加权和)生成一组单目标问题。ParEGO^[38] 是这类 EBO 的一个早期例子: 采用增广的切比雪夫函数和一组随机生成的权重向量构建多个单目标优化问题, 传统的 AFs 可以直接应用于采样。相比之下, MOP 可以分解为多个单目标子问题, 如同在基于分解的多目标进化算法(MOEA/D)^[15] 和参考向量引导的进化算法(RVEA)^[250] 中那样。然后, 就可以用贝叶斯优化解决这些子问题。例如, 一种 EBO 方法, 即 MOEA/D-EGO^[251], 使用 Tchebycheff 标量化函数将 MOP 分解为一组单目标子问题, 并通过优化 EI 从种群中选择一组新样本。另一种 BEO 方法, 即 Kriging 辅助的 RVEA(K-RVEA)^[50], 使用参考向量将 MOP 分解为多个子问题。然后, 如果需要促进整体种群的多样性, 则为每个子问题选择最不确定的解进行采样; 否则, 根据预测的目标值选择具有最佳惩罚角距离的解进行采样。在文献[252]中, RVEA 也被用作优化器求解昂贵的 MOPs, 其中预测的目标值和不确定性被加权组合成为一个获取函数, 并且调整以平衡利用和探索。

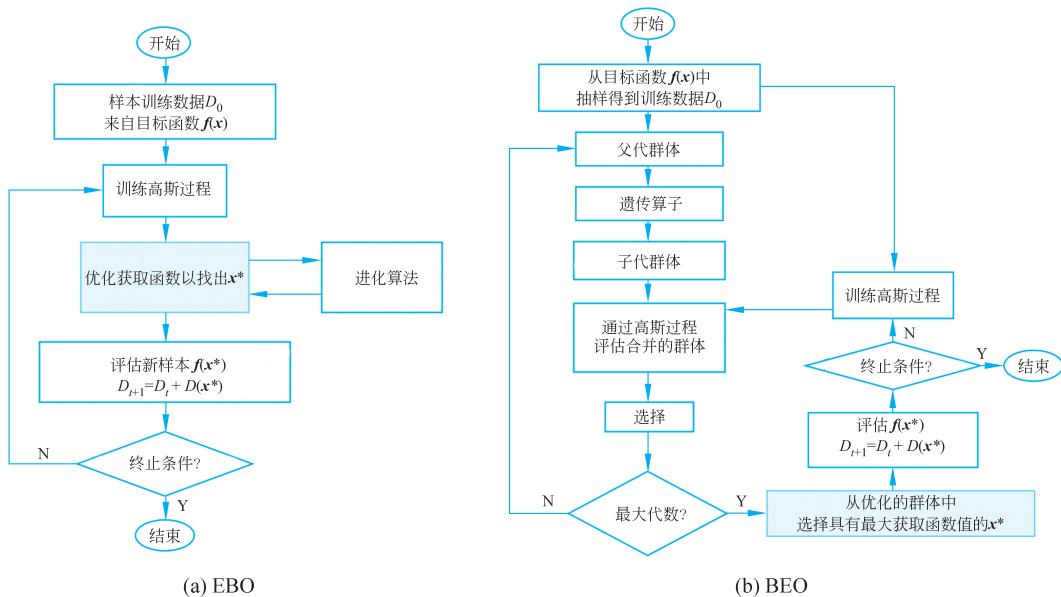


图 3.1 结合进化算法与贝叶斯优化的两种主要方法





非支配排序是 MOEAs 中广泛采用的另一种方法。例如,Shinkyu 等提出一种基于非支配排序的多目标进化优化方法(Multi-EGO),将其作为拓展的高斯过程优化的一部分。Multi-EGO 同时最大化所有目标的期望改进,因此采用非支配排序选择新样本。在最近的研究中,非支配排序也被用于基于代理模型选择一个代价相对较低的帕累托前沿^[253-254]。同样,在某些研究中,非支配排序也与贝叶斯优化结合在一起,如多目标粒子群优化^[255-256]。

2. 基于性能指标的获取函数

性能指标最初是为了评估和比较不同算法得到的解集合的质量(而不是单个解)。已经提出的各种质量指标包括反世代距离(Inverted Generational Distance)^[257]和超体积(Hypervolume, HV)^[258]。HV 是由一组非支配解 P 支配并由参考点 r 限定的目标空间的体积,即

$$HV(P) = \text{VOL}(\cup_{y \in P} [y, r]) \quad (3-7)$$

其中, $\text{VOL}(\cdot)$ 表示常用的勒贝格测度, $[y, r]$ 表示以 y 和 r 为边界的超矩形。因此,获得更大的 HV 值的算法会更好。

有趣的是,性能指标可以以不同方式融入 MOEAs 中。它们可以被采用作为环境选择中的优化准则,因为它们提供了将多目标问题缩减为单目标问题的替代方法。因此,人们开发了各种具有基于性能指标的 AF(获取函数)的多目标贝叶斯方法,其中 HV 是最常用的性能指标。早期的工作之一是基于 S 度量或 HV 度量的选择性多目标高效全局优化(SMS-EGO)^[39]。SMS-EGO 为每个目标构建一个 Kriging 模型,然后优化 HV 以选择新样本,其采用 LCB 计算适应度值。类似地, TSEMO^[55] 使用 GP 后验上的 TS 作为 AF,使用 NSGA-II 优化多个目标,然后通过最大化 HV 选择下一批样本。

实际上, EI 和 HV 的组合被称为期望超体积改进(EHVI),在昂贵多目标优化问题中更为常见。给定当前 PF(帕累托前沿)的近似集 P , 非支配解 (x, y) 对 HV 的贡献可以用下述公式计算,即

$$I(y, P) = HV(P \cup \{y\}) - HV(P) \quad (3-8)$$

EHVI 用于量化非支配区域上的超体积的期望。因此, EHVI 的一般形式可以表示为

$$EHVI(x) = \int_{\mathbf{R}^m} I(y, P) \prod_{i=1}^m \frac{1}{\sigma_i(x)} \phi\left(\frac{y_i(x) - \mu_i(x)}{\sigma_i(x)}\right) dy_i(x) \quad (3-9)$$

EHVI 最初在文献^[54]中引入,用于提供对预筛选解的改进的标量度量,并且后来在处理昂贵的多目标优化问题中变得流行^[57,260]。文献^[261]研究了用于多目标优化问题的不同 AFs,表明 EHVI 具有理想的理论性质。EHVI 与其他准则(如 EI 和目标值估计^[262])的比较表明, EHVI 在保持代理模型精度和优化探索之间保持良好平衡。尽管性





能良好, EHVI 本身的计算由于涉及积分而计算量很大, 限制了其在 MOP/MaOP 中的应用。为了提高 EHVI 的计算效率, 人们进行了各种研究。在文献[54]中, 采用蒙特卡罗积分近似 EHVI。文献[263]引入了一种直接计算 EHVI 的方法, 该方法将积分区域划分为一组区间盒。然而, 区间盒的数量至少随着帕累托解和目标数量呈指数增长。在后续工作中, 文献[264]通过减少区间盒的数量引入一种高效的方法。

另一个常用的指标是基于距离的, 尤其是欧几里得距离。预期欧几里得距离改进 (EEuI)^[114] 定义了概率改进函数和基于欧几里得距离的改进函数的乘积, 用于表示双目标优化问题的闭合形式表达式。文献[264]使用“行鱼群”算法提出了 EEuI 的快速计算方法。另外, 在文献[265]中采用最大最小距离改进作为改进函数。

3. 基于信息论的获取函数

鉴于信息论方法在单目标优化中的广泛使用, 许多基于信息的获取函数用于求解昂贵的多目标优化问题。例如, 已采用 PES 求解多目标优化问题, 称为 PESMO^[266]。然而, 优化 PESMO 并非易事, 需要进行一系列的近似计算, 因此 PESMO 的准确性和效率可能会降低。随后, 输出空间熵为基础的适应度函数在多目标优化中进行了扩展, 称为 MESMO^[72]。实证结果表明, MESMO 比 PESMO 更高效。然而, 正如在文献[73]中指出的, MESMO 无法捕捉 MOP 目标之间的权衡关系, 其中 PF 中没有点接近每个目标的最大值。为了解决这个问题, 文献[73]提出一个考虑整个帕累托前沿的前沿熵 (Pareto-frontier ES), 其中信息增益的表达式为

$$I(\mathcal{F}^*; \mathbf{y} \mid \mathcal{D}_n) \approx H[p(\mathbf{y} \mid \mathcal{D}_n)] - \mathbb{E}_{\mathcal{F}^*} [H[p(\mathbf{y} \mid \mathcal{D}_n, \mathbf{y} \leq \mathcal{F}^*)]] \quad (3-10)$$

其中, $\mathcal{F}^*$ 是帕累托前沿, $\mathbf{y} \leq \mathcal{F}^*$ 表示 $\mathbf{y}$ 被 $\mathcal{F}^*$ 中至少一个点所支配或与至少一个点相等。

4. 讨论

用于昂贵 MOP 的贝叶斯优化方法主要集中在 AF 的设计上, 由于 GP 的可扩展性问题和一些 AF 的高计算复杂度, 它们的应用通常仅限于低维 MOP。因此, 可能的未来研究方向包括寻求高维多目标优化问题的高效代理模型和有效的 AFs。此外, 由于多目标问题中存在参数空间和帕累托前沿, 在选择新样本时需要更多的努力平衡两者之间的关系。

3.2.6 多任务优化

许多黑盒优化问题不是一次性任务。相反, 可以同时解决几个相关的任务实例, 被称为多任务优化 (Multi-task Optimization, MTO)。假设有 K 个优化任务, $i = \{1, 2, \dots, K\}$, 需要完成。具体而言, 将 T_i 表示为第 i 个要优化的任务, X_i 表示 T_i 的搜索空间。不失一般性, 假设每个任务都是最小化问题, MTO 的目标是找到一组解 $\{x\}$, 满足以下条件, 即





$$x_i^* = \underset{x \in X_i}{\operatorname{argmin}} T_i(x), i = 1, 2, \dots, K \quad (3-11)$$

MTO 和其他一些术语之间存在一些概念上的相似性和重叠,例如多目标优化、多保真度优化(MFO)和迁移/元学习。这些相似性和差异如图 3.2 所示。这 4 种情况下的目标优化任务(用红色矩形表示)各不相同:多目标优化和 MTO 旨在有效且同时地优化多个问题,而 MFO 和迁移/元学习旨在通过利用从低保真模拟或类似源优化任务获得的有效知识(用蓝色矩形表示)加速目标优化任务。在 MTO 中,所有任务都同等重要,并且知识传递发生在任何相关任务之间。最后,多目标优化与 MTO 的区别在于前者处理同一任务的冲突目标,而后的每个任务可以是单目标或多目标问题。而多功能优化和迁移/元学习专注于目标任务(称为非对称依赖结构),MTO 则将所有任务视为平等,并且在任何相关任务之间进行知识传递(称为对称依赖结构)^[267]。

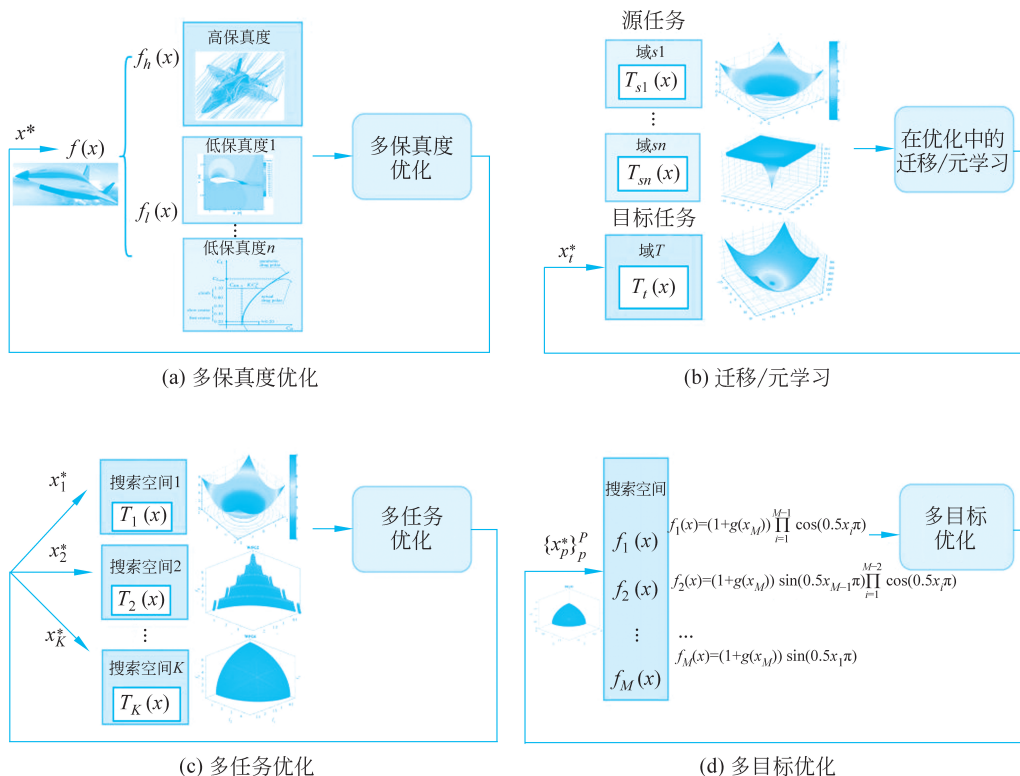


图 3.2 多保真度优化、迁移/元学习、多任务优化和多目标优化之间的相似性和差异(见彩插)

多任务贝叶斯优化旨在同时优化一组相关任务,从而通过利用任务之间的共同信息加速优化过程。为了实现这一目标,需要满足两个要求。首先,需要构建可以学习任务之





间可传递知识的代理模型。其次,获取函数(AF)应该考虑任务之间的相关性,以便通过相关任务之间传递知识进一步提高优化的数据效率。接下来介绍构建多任务 GP 模型和为 MTO 设计特定 AF 的贝叶斯优化方法。

1. 多任务 GP

假设不同任务之间存在一定程度的关联,那么 MTO 就能从不同任务间的知识转移中获益。在统计学领域,共核相关线性模型(Linear Model of Coregionalization, LMC)将输出表示为 Q 个独立随机函数的线性组合,即

$$T_i(\mathbf{x}) = \sum_{q=1}^Q a_{i,q} u_q(x) \quad (3-12)$$

其中,假设潜在函数 $u_q(x)$ 是一个均值为零的高斯过程,其协方差为 $k_q(x, x')$,其中 $a_{i,q}$ 是 $u_q(x)$ 的系数。在机器学习领域,许多贝叶斯多任务模型可以看作是 LMC 的变体,只是参数和约束条件不同而已。其中代表性的工作是多任务高斯过程(Multi-task GP),它使用了内在的共核相关模型核函数。除输入上的协方差函数 $k_X(x, x')$ 外,其还引入任务协方差矩阵 $k_T(t, t')$ 作为共核相关度量模拟任务之间的相似性。因此,乘积核的推导过程如下,即

$$k((x, t), (x', t')) = k_X(x, x') \otimes k_T(t, t') \quad (3-13)$$

其中, $\otimes$ 表示 Kronecker 乘积, $t, t' \in \mathcal{T}$, $k_T(t, t')$ 是一个半正定矩阵,由 Cholesky 分解保证。多任务高斯过程的计算复杂度为 $O(K^3 n^3)$ 。为了解决多任务高斯过程的可扩展性,文献[268]使用 Matheron 规则利用协方差矩阵中的 Kronecker 结构,以实现更快的预测计算。在 LMC 模型中,相关过程通过一组独立过程的线性组合表示。这种方法局限于一个输出过程是另一个输出过程的模糊版本的情况。相反,文献[269]使用卷积过程考虑输出之间的相关性,每个输出可以通过平滑核函数和潜在函数之间的卷积积分表示。

2. MTO 中的 AF

虽然已经有很多人尝试提出多任务模型,但直到提出一些多任务贝叶斯优化算法,尤其是在机器学习的超参数优化领域。文献[120]将多任务高斯过程扩展到用于调整超参数的知识迁移的贝叶斯优化中,其中提出一种基于效用和成本考虑的新型获取函数。类似的思路,在文献[270]中采用多任务高斯过程或设计新的 AF,在信息增益和成本最小化之间引入权衡。文献[271]考虑了具有不同数据集特征的深度信念网络的超参数优化,并提出几个问题的协同调整。上下文策略搜索(CPS)学习了上下文-参数空间上的联合高斯过程模型,使得从一个上下文获得的知识可以推广到类似的上下文。最近,通过从后验中采样识别下一个任务和动作,将 TS 扩展到 MTO^[272]在理论上是有保证的。





3. 讨论

关于 MTO 的代理建模,常用的 LMC 模型因其计算复杂性而受到批评。虽然一些简单的模型被提出来缓解这个问题,但它们的预测质量可能会受到影响。因此,开发有效的多任务代理模型是一个有前途的方向。事实上,已经有一些尝试通过定义新的 AF 来解决多任务优化问题,其中大多数都考虑一个目标任务。在未来,同时为所有任务选择新的样本将是非常有益的。

3.2.7 多保真度优化

贝叶斯优化通常假设只有目标昂贵的目标函数可用,这称为单保真度优化。然而,在许多实际问题中,目标函数 $f(x)$ 的评估通常可以在具有不同成本的多个保真度级别上运行,表示为 $f_1(x), f_2(x), \dots, f_M(x)$, 其中保真度 $m \in \{1, 2, \dots, M\}$ 越高,评估将更准确,但成本更高。这被称为多保真度优化(MFO),其可以看作是多任务学习的一个子类,其中相关函数组可以按其与目标函数的相似性进行有意义的排序。

多保真度优化旨在通过从所有保真度模型中共同学习最大量的信息加速目标的优化并降低优化成本。为了实现这一目标,贝叶斯优化进行了两项改变来利用多保真度数据,即多保真度建模和新的样本选择,下面将详细讨论。

1. 多保真度模型

通常,多保真度贝叶斯优化通过学习独立的高低保真度 GP 模型,或者联合建模多保真度数据以捕捉不同保真度数据之间的相关性,如多输出 GP 和深度神经网络。其中,最流行的多保真度模型之一是共克里金(Co-Kriging)模型。文献[273]提出一个自回归模型来近似昂贵的高保真度模拟 $\hat{y}_H(\mathbf{x})$, 其中 $\hat{y}_L(\mathbf{x})$ 是低保真度 Kriging 模型, $\hat{\delta}(\mathbf{x})$ 是离差模型,公式为

$$\hat{y}_H(\mathbf{x}) = \rho \hat{y}_L(\mathbf{x}) + \hat{\delta}(\mathbf{x}) \quad (3-14)$$

其中, ρ 表示缩放因子,使 $\rho \hat{y}_L(\mathbf{x})$ 与公共采样点处的高保真模型之间的差异最小化。因此,可以通过从低保真廉价数据获取信息增强高保真模型。随后,在文献[274]中开发了一个贝叶斯层次 GP 模型,以解释从低保真度到高保真度的复杂尺度变化。为了提高计算效率,在文献[275]中提出一种 Co-Kriging 的递归形式,假设 $\hat{y}_H(\mathbf{x})$ 和 $\hat{y}_L(\mathbf{x})$ 的训练数据集具有嵌套结构,即更高保真度级别的训练数据是较低保真度级别的子集。因此,在式(3-14)中,GP 先验 $\hat{y}_L(\mathbf{x})$ 被替换为相应的 GP 后验,从而提高超参数估计的效率。根据这个思想,通过用非线性映射函数替换缩放因子 ρ ,式(3-14)给出的自回归多保真度模型^[276]得到推广。另外,多保真度深度 GP 模型使用神经网络学习非线性转换^[277],并进一步在参数化形式和维度上进行了扩展,以适用于不同的输入空间^[278]。





2. 多保真度优化的获取函数(AFs)

基于多任务模型^[273,275],在多保真度优化的设置中,设计复杂的AFs来选择输入位置和保真度引起了广泛的研究兴趣。早期的多保真度AFs主要集中在EI的调整上。文献[112]提出了一个增强的EI函数,以考虑填充点的不同保真度水平。具体而言,所提出的EI是期望项、低保真度和高保真度模型之间的相关性、添加新复制后验标准差减少的比率以及不同保真度模型的评估成本之间的比率的乘积。为了提高增强EI的探索能力,文献[279]提出一种样本密度函数,用于量化输入之间的距离以避免样本聚集。

UCB在多保真度优化中得到广泛应用,尤其在老虎机问题中。一个早期基于UCB的多保真度优化算法是MF-GP-UCB^[280]。MF-GP-UCB算法首先为每个保真度设定一个上界,其中最小的上界被最大化用于选择新样本。在选择新点之后,引入一个阈值决定要查询的保真度。在后续的工作中^[281],MF-GP-UCB被扩展到连续保真度空间。文献[282]开发了一种基于分层树状划分的算法,并采用MF-GP-UCB选择叶子节点。该方法的动机是在较低保真度下探索更粗粒度的分区,并在不确定性缩小时以较高保真度进行较精细的划分。在此思想的指导下,文献[283]采用MF-GP-UCB以较低保真度探索搜索空间,然后在相继较小的区域中利用高保真度。

最近,信息论方法在多保真度优化中变得流行。例如,文献[284]中采用带有Co-Kriging模型的ES解决双保真度优化问题。在文献[285]中,具有不同保真度的未知函数被联合建模为卷积GP^[269],然后引入多输出随机特征近似计算PES。由于计算基于ES/PES的多保真度AFs非常复杂,MES因其高计算效率而被推广到MFO中^[68]。

3. 讨论

多保真模型通常需要强大的假设:低保真度和高保真度始终线性相关,并且搜索空间相同。这些假设在实际应用中可能不成立,例如不同保真度的搜索空间维度可能不同。所以应该更加努力探索代理模型。在多保真度优化中的获取函数中,缺乏对连续保真度设置的研究。此外,现有的多保真度优化技术主要处理老虎机问题和单目标问题,因此将它们进一步扩展到多目标问题和鲁棒优化是很有意义的。

3.2.8 迁移学习/元学习

尽管贝叶斯优化为全局黑盒优化问题提供了强大的数据高效方法,但它单独考虑每个任务,并且通常从头开始搜索,这需要足够数量的昂贵评估才能获得高性能解。为了解决这种“冷启动”问题,贝叶斯优化中的迁移/元学习近年来引起人们浓厚的兴趣。给定一组辅助/源域 D_s 和优化任务 T_s ,目标域 D_T 和优化任务 T_T ,贝叶斯优化中的迁移/元学习旨在利用先前相关任务 T_s 的知识来加速目标任务 T_T 的优化。其中一个研究充分的





例子是在新数据集(目标)上对机器学习算法进行超参数优化,并在其他数据集(源/元数据)上观察到超参数性能。超参数优化中先前相关任务的元数据的可用性促使元初始化根据类似数据集的最佳超参数配置初始化超参数搜索。在贝叶斯优化的上下文中,通常将迁移/元学习这两个术语互换使用。值得注意的是,在贝叶斯优化研究领域中,对知识传递也进行了多方面的研究,包括多任务学习和多保真度优化,这些研究可能与广义的迁移学习领域有所重叠。根据捕捉相似性的方法,我们将与转移学习技术相结合的贝叶斯优化算法分为以下3种。

1. 层次模型

在整个数据集上学习的分层模型成为利用相关源领域知识的一种自然解。例如,文献[271]指出,不同数据集上的损失值可能在尺度上有所不同,因此提出一个排序代理,将所有运行的观测映射到相同的尺度。然而,这种方法的缺点是排序算法导致计算复杂度很高。为了解决这个问题,文献[286]建议通过减去每个数据集的均值并通过标准差进行缩放来重构响应值,而文献[287]则提出了一种高效的分层高斯过程模型,使用源后验均值作为目标的先验均值。

2. 多任务 GP

由于多任务 GP 模型对于捕获源任务和目标任务之间的相似性非常有用,文献[120]采用了一个直接的多任务 GP 进行知识转移。同时,多任务 GP 中的正半定矩阵(见式(3-13))已被修改以提高计算效率^[286,288]。另外,文献[289]假设源数据是目标任务的带噪观测值,因此源数据和目标数据之间的差异可以通过噪声方差建模。在此基础上,文献[290]通过使用多臂赌博机算法识别最优源,进一步提高了知识转移的效率。

3. GP 加权组合

贝叶斯优化中的知识转移也可以通过 GP 的加权组合实现。文献[291]建议不在大型训练数据集(即历史数据)上训练单个代理模型,而是使用多个 GP 的乘积来提高学习性能。具体而言,在每个不同的数据集上学习一个单独的 GP。这样,由这些单独 GP 组合得到的对目标数据的预测是各个均值的加权和,权重是根据 GP 的不确定性进行调整的。人们提出不同的策略来调整组合中的权重^[292]。在多目标优化中,文献[293]建议通过优化样本外预测的平方误差确定权重。

在一个互补的方向上,一些尝试致力于在 AF 中利用元数据,类似于加权组合 GP。其中一个代表性工作称为迁移 AF^[294],它由目标数据集和源数据集上的期望改进的加权平均值定义。最近,文献[295]利用强化学习实现了这一目标。

4. 讨论

直观上,如果学习到的知识降低了性能,目标任务的优化可能会受到负迁移的影响。





因此,捕捉目标任务和辅助任务之间相似性的代理模型以及如何缓解负迁移的代理模型仍然是活跃的研究领域。通常,隐含假设是源域和目标域共享相同的搜索空间,这极大地限制了它们的应用。未来,应该研究异构搜索空间。此外,在知识转移过程中保护数据隐私也是一个有趣的研究方向。

3.2.9 并行/批次贝叶斯优化

标准的贝叶斯优化本质上是一个顺序搜索过程,因为每次迭代中只采样一个新数据点,这在许多可以并行采样多个数据点的应用中可能效率较低^[296]。顺序贝叶斯优化的优点是,由于立即更新的 GP,使用最大可用信息选择新数据点,因此同时搜索多个查询点更具挑战性。随着并行计算的日益普及,越来越多的研究涉及批量贝叶斯优化,可以大致分为两类:一种是将现有的自适应采样准则扩展到批量选择,另一种是问题重构。

1. 对现有 AFs 的扩展

一种开创性的多点自适应采样准则是 EI 的并行版本,称为 q 点 EI(q -EI)^[123-124]。 q -EI 直接定义为 q 点超出当前最佳观察值的期望改进。然而, q -EI 的精确计算取决于 q 维高斯密度的积分,因此随着 q 的增加,计算变得棘手和复杂。因此,文献[123]通过使用 Kriging Believer 或 Constant Liar 策略顺序地识别 q 个点,以替换最后选定点的未知输出,从而便于基于 q -EI 的批量选择。文献[124-125]对 q -EI 棘手计算的处理方法进行了研究。此外,文献[297]中提出了 q -EI 的异步版本。

并行扩展的 GP-UCB 因其理论保证,即累积遗憾的次线性增长而得到广泛研究。文献[311]提出了一个扩展的 GP-UCB 方法,其利用更新的方差促进更多的探索^[311]。类似地,在文献[299]中提出了一个纯探索的 GP-UCB 方法,该方法通过 GP-UCB 确定第一个查询点,而其余的点则通过最大化更新的方差选择。由于多目标进化算法可以提供一组非支配解作为推荐,因此它们非常适合通过同时优化预测均值和方差确定剩余的点。通过行列式点过程(DPPs)^[88]中采样,可以探索更多样化的批量采样点。

随着对批量贝叶斯优化的研究兴趣迅速增长,更多的 AF 已经扩展到并行设置。例如,并行 PES^[300]和 KG^[126]共同识别下一次迭代中要探测的一批点,但是在批量大小上的可扩展性较差。有趣的是,可信最大化 ES 算法通过引入可信最大化器简化信息度量,其利用基于信息的 AF,可以很好地扩展到批量采样的情况。TS 还可以通过采样 q 函数扩展到并行设置。最近,TS 引起了人们的广泛关注,因为 TS 的固有随机性自动实现了开发和探索之间的平衡。需要注意的是,TS 的性能不一定比传统的 AF(如 EI 和 UCB)更好。

2. 问题重构

在并行贝叶斯优化中,许多工作致力于通过重新定义 AF 的优化问题开发新的批处





理方法。一个有趣的方向是,开发新的批处理 AF 来选择与顺序方法的预期推荐非常接近的输入批次。例如,在文献[301]中定义了一个批量目标函数,用于最小化顺序选择与批量之间的损失,它对应于加权 k 均值聚类问题。鉴于顺序选择的输入彼此之间足够不同,通过向 AF 添加局部惩罚引入最大化惩罚策略^[302]。文献[303]应用多次启动策略和基于梯度的优化器来优化 AF,旨在识别 AF 的局部最大值。此外,多目标优化器是一种很有前途的查找一批查询点的方法,特别适用于求解昂贵的多目标优化问题^[252]。类似地,顺序优化多个 AF 可以生成批量查询点^[304]。同理,为了更好地平衡利用和探索,可以组合不同的选择指标^[305]。

3. 讨论

在批量选择中设计新的 AF 的主要挑战是需要在最大化信息增益的同时避免冗余。此外,批处理大小的可扩展性也值得进一步研究。由于批量贝叶斯优化可以应用于许多实际应用中,因此考虑更实际的问题设置(如高维搜索空间和异步并行设置)也是很有意义的。

3.3 本章小结

本节对贝叶斯优化进行了系统的文献综述,重点研究了构建 GP 模型的新技术,并设计了新的 AF,旨在将贝叶斯优化应用于各种优化场景。根据优化中的挑战将这些场景分为几类,包括高维决策和目标空间、不连续搜索空间、噪声、约束和高计算复杂度,以及提高贝叶斯优化效率的技术,如 MTO、MFO、知识迁移和并行化。对于每个类别,提出了在构建代理模型和获取函数的适应方面的主要进展。希望通过上述介绍能够帮助读者清楚地了解贝叶斯优化的研究,包括其动机、优势和局限性,以及未来值得进一步研究的发展方向。

