逐底竞赛:化学问题的贝叶斯优化

Abstrac

化学问题维度和评估成本很高。本文介绍BO，以及材料中成功案例，讨论挑战。概述多目标并行算法的机会。

## Introduction

高通量化学枚举法不切实际。智能自动化技术可加速材料发现。指导过程可定义为一个优化问题。

优化是系统输入参数（实验条件）以最小或最大化目标函数（运行试验和测量），当考虑复杂高维是，通常存在局部极小值，而使找到全局最大值困难。

为求最值，直觉是对函数求导，使导数为0，这就是梯度下降发。

梯度下降法只适用连续函数。对于不连续，可使用启发式算法，包括模拟退火，遗传算法，粒子群优化，蚁群优化。每种情况，都指定一个预算来限制计算和时间成本，预算耗尽，搜索停止。

机器学习和优化的结合产生了贝叶斯优化。BO是一种主动学习方法，采用顺序策略解决优化问题。

本工作概述BO原理，数学基础，代理模型和获取函数的选择。研究化学中自动化工作流程，讨论处理小型，噪杂，高维数据集新方法。概述发展机遇和高效鲁棒混合算法。

## BO

### 代理模型

GP

RF

### 获取函数

PI

EI

LCB

## 优化演示

### Sine function

### 光伏器件优化

## 化学中的最新应用

### 自动化实验

BO一个应用是驱动输入参数的选择，最大或最小自动化化学实验的观察结果。Liang编译五个数据集，涵盖五个化学领域，BO所有情况下都优于随机采样。

化学反应是复杂高维问题，BO成为有力工具。Shields以量子化学性质作为描述符，在芳基催化数据集进行基准测试，各方面优于人类决策。Lofgren用BO优化合成参数，包括多个实验输出，pareto前沿的构建能同时优化产量和其他化学特性。

BO也用于加速多组分系统的发现。Wahl用BO在8元素相空间发现目标相界面纳米颗粒。

除了合成，化学和物理表表征也是高维问题。操作SEM有三个考虑因素，BO可以帮助。Liu用BO和深度核学习相结合，发现铁电薄膜结构性能关系。

### 加速计算

BO广泛应用于新材料的发现和计算筛选。Deshwal在7w+数据库中搜索最佳COFs。Lampe以最小数据需求提高纳米板的光学质量。Seko搜索5w+中潜在材料，发现低晶格热导系数材料。

大海捞针问题存在广泛领域。Siemenn开发ZOMBI，在针类问题优于标准BO。

机器学习原子间势成为计算表面势能的有效模型，但受精度的限制。Tran证明了密度泛函和BO相结合的力量。

## 挑战和机遇

### 噪声数据

大多数优化任务收到随机噪声的影响。本文将随机波动应用到二维函数上，在三种条件下：零噪声，小噪声和高噪声，说明BO的效用。尽管GP可根据标准差处理无偏噪声，但实验很难确定噪声范围和环境误差。

为提升算法对噪声任务的鲁棒性，Aldeghi开发Golem方法，该方法考虑了不确定性来源并重新加权之前实验。该方法用参数概率分布和对目标的不确定性进行建模，为测量数据分配权重。该目标函数最大化平均性能，惩罚结果的预期方差。对BO效率有较大提升。Noack提出另一种方法，考虑非均匀噪声的重要性，用GP对噪声估计和建模，以纳米棒数据证明他们的方法。

### 并行优化

并行优化可提高优化过程准确性，收敛性和整体性能。与批处理优化相仿，并行优化从搜索空间的不同分区不同起点执行多个优化。最简单的是多起点梯度下降，但该方法仅限于优化函数可微和缺乏全局搜索能力。

更先进方法已经开发出来。其中一种是强盗贪婪选择算法，与汤普森抽样结合使用，根据最优概率选择抽样点。Ueno开发COMBO包，包含各种方法并应用于发现原子结构。Kandasamy结合汤普森抽样和并行BO，强调T个人中进行N次迭代相当于N次顺序优化。

在以上基础，Falkner引入HBBO，将初始起点扩散到参数空间不同分区同时执行多个BO加快搜索。

### 多目标优化

对多个参数优化时，MBO是有吸引力的方法。对多目标问题，确定最佳方案具有挑战性，必须制定排序方案衡量总体得分的各个参数。非支配排序是多目标优化的重要概念。

Agarwal用MBO同时优化氧化还原体的三个目标，性能优于随机选择。

MBO的困难时平衡不同的目标属性。不同性质测量成本相差很大，可以先优化便宜的目标找到合适区域。再结合降维和并行计算可降低成本。

### 化学知识整合

在某些领域，从没有先验信息开始优化是必要的。但在化学，可通过物理定律，相标准和材料数学约束，改进区域的搜索。约束BO可利用已知信息，指导优化过程。

化学物理驱动的模型可帮助导航材料设计。Hase将BO与分类分布的物理化学描述符结合，加速材料搜索。Clancy基于物理模型与BO结合，进一步引入化学知识。Sun演示第一性原理计算的顺序学习框架，来确定最稳定的钙钛矿。Pedersen结合密度泛函和BO，确定高熵合金最有效成分。

## 总结