**Reliable Sequential Learning for Material Optimization**

**via Uniform Sampling and Transfer Sampling**

**Initial Sampling Strategys for Reliable Bayesian Optimization**

**in Material Task**

**Abstract**

基于机器学习的优化策略，如贝叶斯优化，在材料开发任务中显示出令人鼓舞的结果。大多数科学家倾向于模型和优化策略的研究，而忽略了初始采样策略的影响。实际上，材料专家会根据以往经验选择新材料任务的初始实验设置。

在此，我们提出了两种初始采样策略，SE采样和迁移采样，这两种策略可以分别应对优化任务开展前的不同情况，如果我们没有任何关于优化任务的信息，需要beginning from scratch，前者可以保证初始采样数据在全局范围分布更加均匀。如果我们有pre-existing knowledge and databases,后者迁移相关数据知识来帮助目标任务的初始采样，as we konw材料专家会根据以往经验选择新材料任务的初始实验设置。

我们在二维测试函数，钯催化交叉偶联反应数据集，合金蠕变寿命数据集上测试了我们的采样方法。与常规初始采样对比，基于EDS得到数据集构建的GP模型spearman指标平均提高0.1，并且提高了找到全局最优的可能。与标准BO相比，在相关源任务的知识指导下，基于迁移采样的BO有效发现最优解的效率提高2倍以上。

# Introduction

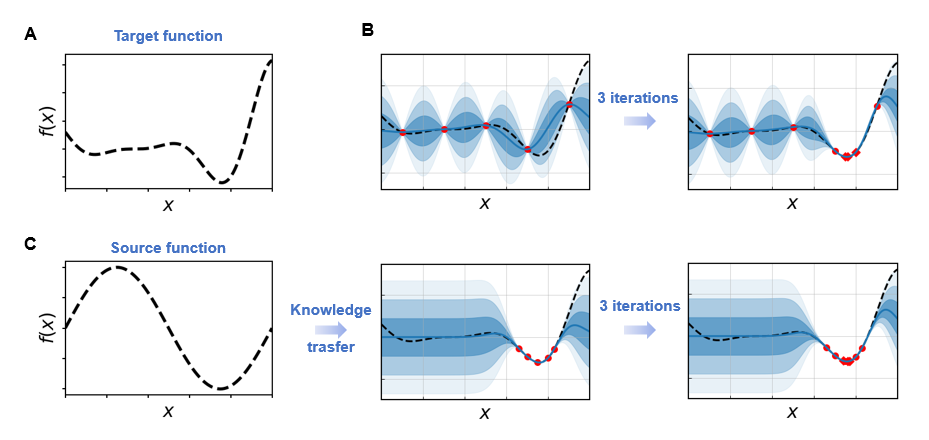
在进行材料实验的开发和设计时，会涉及因高维变量而产生的成百上千种实验组合的系统优化，这种情况很难靠蛮力或者实验设计去解决。 by introducing machine learning (ML) in the loop to guide the experiments of 材料优化任务，可以自适应探索复杂的搜索空间，避免穷列枚举。

贝叶斯优化（Bayesian Optimization, BO）作为一种序贯机器学习方法，凭借其在评估昂贵黑箱函数时的高效取样和优化能力，已被广泛应用于化学反应合成与材料优化的探索中，成为一种有效的策略。

BO框架一般可以划分为三个部分：初始采样策略，代理模型和获取函数。To enhance the effectiveness and applicability of BO in materials development ，当前报道的关于经典贝叶斯优化的研究，更倾向于关注代理模型的选择，和获取函数的调整。Zhe liu 将先前实验数据作为概率约束，并结合主观的人类观察和机器学习见解，实现领域知识与BO的协同工作。Alexander E. Siemenn在传统贝叶斯优化原理的基础上通过缩放内存，迭代放大采样搜索边界，自定义自适应获取函数，以进一步指导新实验的采样走向全局最优。

然而，很少有研究关注初始采样方法的选择。初始数据的质量不仅决定了初始代理模型的性能，还对后续优化的方向产生深远影响，对于BO的整体优化效果至关重要。

目前BO常用的初始采样方法主要有RANDOM和LHS。然而，这两种方法有两种局限。（1）无法均匀探索全局信息。随机采样由于高度随机性，很可能导致样本分布不均匀；LHS的原理是对每个维度进行独立的分层采样，然后将这些层组合成采样点，这意味着采样的均匀性仅在单个维度上得到保证。（2）无法利用相关数据知识。一般的初始采样方法适用于对目标任务完全未知的情况。然而，人类专家通常不会随机选择初始实验，而是依赖从之前的优化活动中获得的经验来指导新的优化活动的初始探索。对于BO来说，如果能将之前优化活动中的相关数据中进行知识迁移，指导目标任务的采样过程，可进一步提高采样效率。



**Figure 1. 均匀采样和迁移采样原理图。（a）待优化的目标函数，（b）在没有知识借鉴情况下，通过均匀初始采样，更快找到全局最优解，（c）在源函数的知识借鉴下，通过迁移采样，更快找到全局最优解。**

此工作中，我们探究并讨论了两种BO场景下的初始采样策略：（a）在缺乏相关数据参考的情况下，提出了一种全局均匀的初始采样策略——Euclidean Distance Sampling（EDS）。与随机采样和拉丁超立方采样（LHS）相比，基于SE的采样策略能够在全局范围进行均匀的初始采样，得到目标函数的大致信息，从而避免陷入局部最优区域，找到全局最优解(Fig. 1b)。（b）在有相关数据可供利用的情况下，我们采用迁移学习策略，将源函数的相关知识迁移到目标黑盒函数的优化任务中，帮助BO直接定位到全局最优区域(Fig. 1c)。值得注意的是，上述工作提到的两种方法分别针对不同数据情境下的初始采样策略，具有并列关系。在实际优化过程中，可根据具体数据情况选择适宜的初始采样方法。

本研究的组织结构如下：第二节介绍了我们提出的EDS采样方法和迁移采样策略；第三节利用Goldstein-Price函数对这些方法进行了演示验证；第四节通过两个真实的材料数据集（Buchwald-Hartwig交叉偶联反应数据集和合金蠕变寿命数据集）进行了模拟优化和算法测试。最后，我们提供了证据，证明在不同的数据情境下，EDS采样在初始化代理模型时，能够显著提高模型性能的稳定性，并降低在BO过程中陷入局部极值的风险。迁移采样策略能够在贝叶斯优化（BO）的迭代过程中分析源模型与目标任务之间的相关性，并根据这种相关性动态调整知识迁移的程度。这种方法能够在源任务与目标任务相关时，有效地辅助目标任务的优化。

# Methdology

在本文中，我们进行了两个主要的探究：（1）SE初始采样策略（2）迁移采样贝叶斯优化方法。通过这两种方法，可以适应在不同情况下启动贝叶斯优化策略，实现更稳定，更高效率的优化。

我们演示了上述两种方法在Goldstein-Price函数数据集， Buchwald-Hartwig交叉偶联反应数据集，合金蠕变寿命数据集上的实现。对于SE初始采样，我们和将其与LHS和随机采样进行对比，比较不同采样方式下的得到数据对建立初始模型和后续优化的影响。对于迁移采样，我们将其与标准贝叶斯优化进行对比，比较了不同源数据的迁移BO与标准BO的优化情况。通过在一个函数数据集，两个材料数据集上的测试，我们展示了这两种采样策略稳定的有效性和广泛的适用性。

## 欧氏距离采样

为了保证全局搜索空间采样的均匀性，我们提出了基于欧氏距离的采样策略，简称为SE。该策略通过以下步骤实现：

1.获得启动采样点：对于参数空间X，LHS采样3个点作为启动采样点集。

2.计算未采样点欧氏距离：对于X中每个未采样的点，计算它与中每一个已知点的欧氏距离，并找到这些距离中的最小值。

3.选择下一个采样点：对于X中所有未采样的点，选择与已采样点集 中的点的最小欧氏距离最大的点作为下一个采样点。

4.将加入到中，重复步骤2和3，直到获得需要的采样数量。

## 迁移贝叶斯优化

在有一些相关数据数据作为源数据的情况下，我们可以利用源数据的知识帮助我们进行目标优化任务的采样。该过程的难点在于a)如何判断源数据与目标任务的相关性。b)如何利用源数据的知识帮助目标任务采样。c)如何随着迭代过程自动调节源数据的影响。在本研究，我们结合了Rank-Weighted Gaussian process ensemble (RGPE)和Transfer Acquisition Function（TAF）方法，实现了随着迭代的进行计算源任务和目标任务的相关性，自动更新源数据的权重，实现了可靠的迁移贝叶斯优化。通过以下步骤：

1.计算模型权重：如果一个模型能够根据观测值的函数值对观测值进行正确排序，那么它对优化是有用的。我们通过构建一个损失函数来衡量模型能够正确对目标观测进行排序的程度，在给定|| = > 1目标任务测量值的情况下，模型i的排序损失被表示为错误排序对的数量：

是指示函数，是异或操作。基于排名的损失确保只有最优的位置是重要的，而不是预测的实际值。目标模型的损失使用留一交叉验证计算，由

其中表示模型拟合除第k点外的所有目标任务观测值。现在，根据每个模型是排名损失最小的集成成员的概率为每个模型分配权重。这个概率是通过在验证集中从模型预测中抽取S个样本来估计的，i.e.模型i的权重计算为

2.迁移获取函数：源模型和目标模型都采用高斯过程模型，通过将上述计算的权重分配到各自模型的获取函数上，得到迁移获取函数

其中

通过迁移获取函数指导目标任务的下一步采样，从而实现将信息从源任务转移到目标任务。

# Demonstration

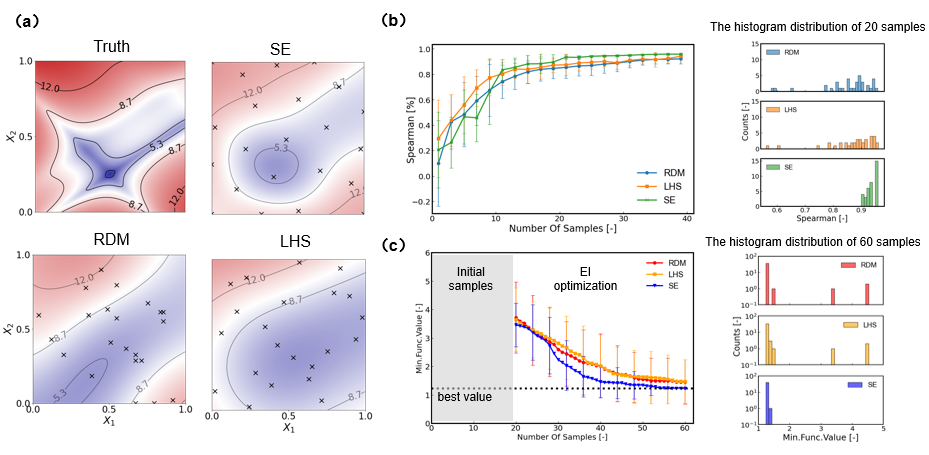
在本章节，我们以Goldstien-Price (2-dimensions)来展示方法的效果，该函数属于公认的测试函数之一，具有多个局部极小值和一个全局最小值。为了使函数优化更接近材料数据优化时的情况，我们对两个维度设置50个间隔采样得到2500组测试函数数据，并对x进行了minmax归一化和对y进行对数化，用于后续实验测试。

## 均匀采样

为了分析随机采样，LHS和SE的建模和优化效果，首先我们用不同方法从数据集中抽取20个样本建立高斯过程回归模型，并对样本点的分布和基于模型的目标函数的映射关系可视化。

随后我们不断增加采样数量，比较不同采样方法下模型的Spearman指标变化，Spearman系数是一个定义在-1到1之间的度量，用于检查是否存在单调相关。

最终我们选择了合适的初始采样点数量建立初始模型，并在此基础上选用EI获取函数进行BO，探究不同采样方式对后续BO优化效果的影响。



**Figure 2. Visualization of the modeling and optimization results under different initial sampling strategies on a two-dimensional function. (a) The left panel shows the true distribution of the objective function, while the other three panels depict the prediction of Gaussian process models based on SE, RDM, and LHS sampling strategies with 20 sampled points. (b) The left panel shows the variation of the Spearman correlation with the number of samples for different strategies, and the right panel displays the distribution of the Spearman correlation coefficients for 20 and 60 samples after 40 repeated experiments. (c) The left panel shows the optimization process using the EI strategy after different initial sampling, and the right panel shows the distribution of the best function values after 40 repeated experiments.**

**Figure 2 (a)** 图展示了待优化目标函数在不同采样策略下的预测效果。左上角的 "Truth" 图表示真实的目标函数分布，其他三个图分别基于欧氏距离采样（SE）、随机采样（RDM）和拉丁超立方采样（LHS）策略，通过采样 20 个点构建的高斯过程模型的预测情况。基于EDS得到的模型预测更能拟合目标函数的大致关系，而 基于RDM 和 LHS 的模型预测表现出较大的偏差，尤其在目标函数的高变化区域。

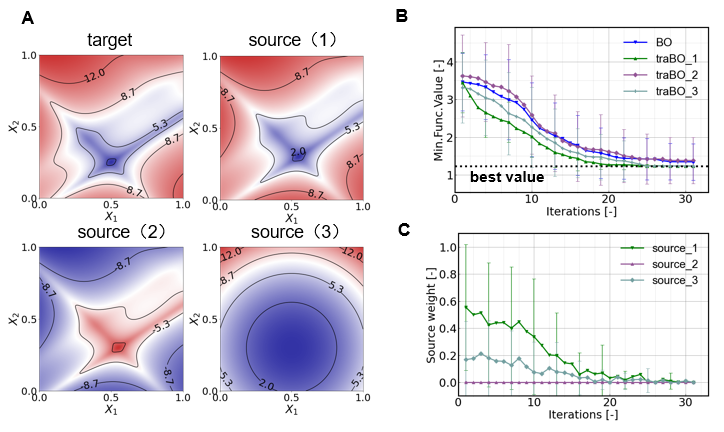
**(b)** 左图展示了不同采样策略下模型的 Spearman 相关系数随样本数量增加的变化情况。可以看到，EDS的 Spearman 系数随着样本数量的增加迅速提高，并且在20个采样点之后高于 RDM 和 LHS，这表明 SE 策略能够更快地构建出准确的模型。右图展示了在 20 个样本和 40 次重复实验下，各采样策略的 Spearman 系数的分布情况。SE 的分布集中在更高的相关系数范围内，进一步验证了其稳定的建模能力。

**(c)** 左图展示了在不同初始采样策略下，再通过期望提升（EI）策略进行采样优化的效果。随着样本数量的增加，各策略的目标函数值逐渐逼近全局最优值。其中，基于EDS的BO在10次EI采样后就有明显的优化效果，并在55次迭代左右基本找到全局最优值，而基于RDM 和 LHS 则需要更多的迭代采样次数。通过对40次独立重复实验中BO优化结果的统计分析，可以发现随机采样和LHS方法下的BO有可能停留在函数值为3.6和4.4的附近，而基于EDS方法的最优值大多集中在1.5以下。这表明，EDS有效降低了陷入局部最优解的风险。

## 迁移采样

为了测试迁移采样方法的迁移优化效果，我们根据目标函数设计了不同类型的源函数，并建立相应的源模型来提供可以访问的已知知识,源函数的等高线图如图3A所示。源函数1通过将目标测试函数的 x1和x2 均移位0.1个单位，并将y 轴缩放至0.8得到。源函数2则是源函数1的 y 值取反。源函数3是我们自行设计的漏斗形函数，用于作为其他类型源函数进行知识迁移，目标函数的蓝色区域使我们期望BO探索的全局最优区域。

我们比较了没有知识迁移的标准BO和不同先验知识的迁移BO的优化曲线，同时追踪了随着迭代过程不同源模型的权重变化。为了尽早在BO中引入源函数的信息，我们随机采样3个样本点后开始各自BO任务，每次迭代新增加1个样本点，



**Figure 3**

**Figure 3 (a)**展示了目标函数和用于迁移的三个源函数的等高线图。左上角的目标函数图显示了优化目标，其中蓝色区域为全局最小区域。**(b)** 展示了与标准BO相比，三种不同源函数的迁移BO的优化性能。经过30次迭代，traBO（1）和traBO（3）可以全部找到全局最优值，这意味着迁移采样可以有效地利用源函数（1）和源函数（3）关于全局最优的知识，帮助完成目标函数的优化，而traBO（2）与标准BO优化性能相当，说明迁移采样能够摆脱源函数（2）的误导信息，回归到标准BO。**(c)** 图显示了三种迁移 BO 过程中源模型的权重变化，在优化前期，源模型权重（1）>（3），源模型（2）的权重基本为0，这与**(b)**图得到的结论相吻合，并且随着迭代进行，所有源模型的权重都在逐步降低，说明随着目标数据的增加，目标模型逐渐完善，目标任务优化逐步摆脱源模型的指导。

# Experiments

我们选用了两个真实世界的材料数据集来评估来不同采样算法的表现，包括钯催化芳基卤化物交叉偶联反应数据集和高温合金蠕变断裂寿命数据集，他们均来自于公开的材料数据库。

## 模拟交叉偶联反应的优化

Pd催化的Buchwald-Hartwig反应在药物合成中具有广泛的价值，Ahneman等人通过高通量实验获得了15种芳基卤化物对应的实验组合的产率值，每种卤化物的反应空间都包括23种添加剂、4种钯催化剂和3种碱（23\*4\*3=264）。原作者通过93维的原子，分子和振动描述符来对每个选项进行特征描述，本工作在原有特征基础上进行了pca降维和minmax归一化。

**表1，钯催化交叉偶联反应数据集情况**

我们将每个子数据集都视为一个优化任务，从而实现在针对某个芳基卤化物寻找最优反应组合时，可以自由选择其他芳基卤化物的数据知识进行迁移。我们随机选择芳基卤化物1,9,15的优化任务作为目标任务。对于初始采样的探究，通过不同的初始采样方式得到相同数量的初始数据集来建立代理模型，然后再通过EI寻找最大反应产率的实验组合。而对于迁移采样的探究，每种目标任务随机选择3种其他的芳基卤化物的数据进行知识迁移。

**Figure 4. 不同芳基卤化物优化任务下的采样算法对比。**（a）不同初始采样策略下模型spearman系数曲线和BO优化曲线。（b）基于不同源数据的迁移BO和标准BO的优化曲线。折线代表均值，误差棒代表方差，所有实验独立运行40次。

首先，我们在芳基卤化物1的数据集上对不同采样策略进行了评估。将264个数据按照8:2的比例划分为训练集和测试集，并在训练集中通过Random、LHS和EDS获取数据，建立高斯过程回归模型，评估模型在测试集上的预测效果，使用Spearman相关系数作为指标。误差棒表示40次独立实验中指标的方差。当采样数据量超过20个时，基于EDS方法的模型表现优于Random和LHS。随后，我们将初始采样量固定为20个，进一步探讨初始采样对BO的影响。在前40次基于EI的采样中，三种策略表现差异不大；然而，当采样量超过60个时，EDS策略下的BO的误差棒显著减小，推测这是因为前期的全局均匀采样为模型提供了优化任务的大致信息，从而提高了优化的稳定性。最后，我们对比了卤化物2、4、7的反应数据分别作为已知任务，用于迁移BO和标准BO的优化效果。结果显示，两种策略的差异不明显，可能是由于源任务与目标任务之间的相关性较低，导致没有有效的优化信息能够进行迁移。

在芳基卤化物9和芳基卤化物15的测试中，我们同样进行了与之前相同的采样策略评估。结果表明，在模型指标评估方面，EDS方法依然优于Random和LHS，尤其是在BO优化中，EDS方法的方差更小，进一步验证了全局均匀采样有助于提高BO优化稳定性的推测。对于在目标任务开始时引入相关信息的迁移采样，不同源数据的迁移效果存在差异。实验表明，卤化物8和14的数据能够帮助优化卤化物9的反应组合；而对于卤化物15，卤化物8的迁移采样甚至在20次迭代内就找到了最优产率组合。这表明，在存在相关数据知识的情况下，迁移采样策略具有巨大的应用潜力。

**Figure 4** (a) 展示了与Random和LHS相比，SE初始采样下，代理模型的speraman系数更高，这意味着相同采样数量下，基于SE策略采样，得到的数据包含的信息更多，建立起来的代理模型对未知点的预测趋势更准确。（b）图展示了基于Random，LHS和SE初始采样20个点后，再用EI进行60次优化采样，以寻找最优解。三者的优化曲线相差不大，但可以明显看到，SE采样的优化曲线误差棒更窄，这意味这40次独立运行的方差最小，BO的优化效果更加稳定。（c）图为在有其他芳基卤化物的数据作为先验知识进行迁移的情况下，迁移BO和标准BO的优化曲线对比，相同迭代次数下，迁移BO能够找到更高的反应产率组合。

## 模拟合金蠕变寿命优化

第二个实验数据集是合金蠕变寿命数据集，它来自于Refs.以及 NIMS公共数据库，前者包括745个样品，后者包括88个样品，每个样品特征包括化学成分(22个元素)、热处理工艺(9个参数)、实验条件(2个参数)等共33个，目标特征为蠕变断裂寿命，因为蠕变断裂寿命偏小，我们使用该属性的对数，即lg(蠕变断裂寿命)作为输出变量。我们的目标就是最大化蠕变断裂寿命。本工作在上述特征上用matminer描述符对化学成分进行描述，并将成分描述符与工艺变量组合一起进行minmax归一化和pca降维。

**表2，高温合金蠕变寿命数据集情况**

同样，对于初始采样的探究，我们以两个数据集分别作为BO任务，去寻找最蠕变断裂寿命的参数组合。通过不同的初始采样方式得到相同数量的初始数据集来建立代理模型，然后再通过EI寻找最大反应产率的实验组合。而对于迁移采样的探究，我们以目标任务外的另一个数据集作为源数据，进行知识迁移。

**Figure 5. 不同合金数据集优化任务下的算法对比。（a）基于合金数据集1的初始采样对比和迁移采样对比。（b）基于合金数据集1的初始采样对比和迁移采样对比。**

图5（A）展示了在合金数据集1上的测试结果。数据集1包含745组数据，经过8:2划分为训练集和测试集后，我们仅使用从训练集中初始采样得到的几十组数据进行建模预测，尽管此时的Spearman指标并不会很高，但相比其他方法，EDS仍表现出更好的建模效果。在如此多的数据组合中寻找最优实验组合非常困难，虽然三种初始采样方法下的BO优化曲线的均值相当，但EDS在150次采样后展现出更小的方差，说明其在优化过程中的稳定性更强。关于迁移采样的表现，由于数据集2仅包含88组数据，可供迁移的知识有限，因此迁移BO和标准BO的优化效果相当，且都优于随机采样，这证明了BO在优化问题中的有效性。图5（B）展示了在合金数据集2上的测试结果。即便在小样本数据集上，EDS也展现了优秀的建模效果。尤其是在BO的初始采样后，EDS的优化效果极为出色。这是由于EDS的原理决定了它更倾向于在各维度边界进行采样，如果目标变量在某个维度呈线性变化，EDS能够更快地探索到极值点。最后，在数据集1（大样本）的支持下，迁移BO在仅10次迭代后，其优化效果就超过了标准BO和随机采样，展示了迁移算法在大样本数据向小样本数据进行知识迁移中的强大能力。

# Summary ＆ Conclusions

在本研究中，我们提出了基于贝叶斯优化（BO）框架的两种初始采样策略：欧氏距离采样（Euclidean Distance Sampling，EDS）和迁移采样（Transfer Sampling）。这两种策略分别针对不同的优化任务背景进行优化：（1）当目标任务缺乏任何先验信息时（即对任务一无所知），我们通过EDS策略进行初始采样，旨在确保数据点在高维空间中尽可能保持最大欧氏距离，从而实现初始采样的均匀分布和全局覆盖。这种方式能够有效地描绘黑盒函数的大致轮廓，提供优化的初步方向。（2）当已有相关的实验数据可供参考时（这些数据来自与目标任务平行的源任务，且无法直接用于目标任务的建模与优化），我们采用迁移采样策略。该策略首先评估源任务与目标任务之间的相关性，并根据这一相关性为采样分配不同的权重，以此指导下一轮的采样。通过这一方式，源任务的数据能够向目标任务传递知识，促进优化过程的收敛。值得注意的是，迁移采样的权重随着迭代过程的进行而不断更新，当目标任务数据逐渐增多并且目标模型不断完善时，复合采样函数将逐步减小源数据的影响，从而保证了迁移优化过程的鲁棒性。

具体来说，我们在Goldstein-Price测试函数、交叉耦合反应数据集和高温合金数据集上进行了贝叶斯优化（BO）模拟优化和采样策略的基准测试。首先，在Goldstein-Price测试函数上，我们分别基于RDM、LHS和EDS进行20次初始采样，并对采样点分布及代理模型的预测结果进行了可视化。结果表明，EDS下的采样点分布更均匀且离散，且代理模型的预测与目标函数的拟合效果较好。接下来，我们比较了随着采样数量增加，三种采样策略下代理模型Spearman相关系数的变化。基于40次独立实验的结果，EDS策略下，当样本量超过20时，模型的Spearman相关系数具有更大的均值和更小的方差。最后，我们固定初始采样数量为20，使用常用的EI获取函数进行BO优化，结果表明，EDS策略下BO优化效率更高，并且更容易找到全局最优解。对于迁移优化的基准测试，我们设计了三种不同的源函数，作为已知的源任务，进行知识迁移至Goldstein-Price函数的目标优化任务。同时，我们与没有源函数迁移的标准BO进行了对比。结果表明，当源函数与目标函数相似时，迁移BO显著提高了优化速度；而当源函数与目标函数差异较大时，迁移BO与标准BO的优化效果相当。通过追踪迭代过程中源模型权重的变化，我们发现，优化效果较好的迁移BO在初始阶段具有较大权重，表明相关源函数的知识得到了有效迁移。

在交叉耦合反应数据集（包含15种芳基卤化物）中，我们随机选择了编号为1、9和15的芳基卤化物数据集作为优化任务，并分别测试了RDM、LHS和EDS三种策略的建模和优化效果。同时，我们还随机选择了其他三个编号的芳基卤化物数据集作为源数据，用于迁移采样优化测试。在这三个优化任务中，EDS下代理模型的Spearman相关系数显著高于RDM和LHS。在Aryl halide 15数据集中，抽取20个样本时，EDS的Spearman系数比LHS高出0.2。在Aryl halide 1和Aryl halide 9的优化任务中，经过EDS初始采样后，BO优化结束时模型的方差降为0，意味着在40次独立重复实验中，均找到了全局最优点。在迁移采样测试中，不同源数据的迁移BO表现各异。在Aryl halide 1的优化任务中，三种迁移BO与标准BO的优化曲线差异不大；而在Aryl halide 9的优化任务中，选择Aryl halide 2和8作为源模型的迁移BO，在10次采样后，分别提高了1.3%和2.9%的最优产率。特别是在Aryl halide 9的优化任务中，选择Aryl halide 8作为源模型的迁移BO表现最佳，仅需16次采样便能找到264种实验组合中的最佳产率组合，而标准BO在50次采样后仍未找到。

在两个规模不同的高温合金子数据集上，我们同样进行了上述测试，并分别建立模型进行迁移优化。在包含745个样本的数据集1上，EDS的建模优势尤为明显，代理模型的Spearman相关系数相比RDM和LHS最高提升了0.3。在88个样本的数据集2上，EDS初始采样下发现的蠕变寿命显著高于其他两种策略，我们认为这得益于EDS的边界采样特性，使其更易探索变量范围两端的极值点。在迁移采样方面，数据量较大的数据集1的知识对数据集2的优化任务有一定帮助，而数据集2对数据集1的优化任务支持有限，说明数据集的规模对迁移效果有显著影响。

综上所述，我们提出了两种适用于不同背景的初始采样策略，分别支持贝叶斯优化（BO）在不同情境下的应用。EDS能够在全局范围内生成离散且均匀的采样点，从而更好地覆盖黑盒函数的大致轮廓，提升代理模型的预测性能并为后续BO优化指明方向。而迁移采样则在目标任务数据较少时，利用已有的相关任务数据辅助优化，并在目标任务数据逐渐丰富时逐步减少对源数据的依赖。迁移优化效果依赖于源任务与目标任务的相关性，每次迭代中的权重调整确保了优化的鲁棒性，使迁移优化在缺乏相关源任务时，能够回归到标准BO。

我们的观察表明，合适的初始采样策略能够显著提升BO算法的性能，减少所需实验数量和成本，并更快速地实现高性能材料的优化。这项工作不仅适用于单一材料，还可以推广至材料优化的普遍实验设计问题。我们希望通过本研究的测试和结果，与材料发现领域分享见解，并推动机器学习与材料科学界之间的深度合作。

# Reference