

1 方法

1.1 环境设计

我们将卡宾分子的构建过程建模为一个有限步长的序列决策问题。构建单元 (blocks) 分为两类: *core* 与 *substructure*。候选集合包含 30 个 *core* 和 852 个 *substructure*。初始状态 s_0 为空图。智能体首先必须选择且仅选择一个 *core*, 随后在其反应位点上执行结构扩展。

在任意状态 s_t , 动作空间由两类操作组成: *add* 与 *combine*。其中, *add* 表示在选定位点接入一个 *substructure*; 新接入片段可引入额外可反应位点, 从而支持后续扩展。*combine* 表示连接两个已有可反应位点 (可来自 *core* 或不同 *substructure*), 用于形成更紧凑的拓扑结构。最终, 包含单一 *core* 且任意数量 *substructure* 的分子作为有效终止状态。

1.2 代理模型

由于真实评估函数 (oracle) 的评估成本非常高, 我们使用代理模型来评估候选分子的性质。代理模型我们使用 GraphGPS [?], 该模型是一个基于图神经网络的代理模型, 能够预测分子的性质。

1.3 多目标生成式模型

我们采用多目标生成范式, 将性质优化表示为向量奖励 $\mathbf{r}(x) = [r_1(x), \dots, r_m(x)]$, 并通过条件化偏好向量 \mathbf{w} 学习一族策略 $\pi_\theta(\cdot|s, \mathbf{w})$ 。该设计使模型能够在单次训练后覆盖不同目标权衡, 并在采样时按需求探索 Pareto 区域。实现上, 我们参考 GFlowNet 的流匹配思想以提升对高奖励区域的覆盖能力 [1]。

1.4 主动学习逻辑

我们采用“生成-评估-回流训练”的主动学习闭环。每轮迭代中, 模型先生成候选分子并进行性质评估; 随后根据不确定性与非支配排序联合选择样本, 优先保留位于或接近 Pareto front 的高信息样本; 最后将新增标注样本并入训练集更新策略模型。该流程在控制评估成本的同时, 持续提升模型在 Pareto 前沿附近的采样效率与解的质量。

参考文献

- [1] E. Bengio, M. Jain, M. Korablyov, D. Precup, and Y. Bengio, “Flow network based generative models for non-iterative diverse candidate generation,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 27 381–27 394, 2021.