

智能反应优化：深度学习与优化算法驱动的乙酰水杨酸合成

郑健强, 黄泳斌, 明文灿, 刘英菊*

华南农业大学材料与能源学院, 广州 510642

摘要: 在本科实验设计中, 需要经过大量的试错实验来寻找最优的反应条件, 这一过程需要消耗大量的时间和成本。为了简化复杂的实验设计过程, 同时激发学生对先进技术和跨学科领域的兴趣, 培养多学科交叉的创新型人才, 本研究创新性地提出了智能反应优化的模型框架以预测最优产率下的反应条件组合, 并将这一模块加入到本科有机实验乙酰水杨酸的合成中。本研究收集了1054个符合这一反应机理的反应数据作为模型的训练数据, 主要包括了产物、反应物、催化剂、溶剂、主要反应试剂、反应温度和产率。首先预训练了一个基于化学多模态的产率预测模型, 作为后续反应优化的目标函数; 接着采用贝叶斯优化算法, 以最小化产率的负值(即最大化产率)为目标, 确定最优的反应条件组合; 最后将模型预测的产率和实际产率进行了实验验证, 在100次模型测试下的预测产率都完美地落在了实际产率范围内, 效果优异的同时还给出了产率高达90.1%的反应试剂组合, 为这一本科实验提供了多样化的实验方案。

关键词: 优化算法; 多模态; 机器学习; 乙酰水杨酸; 优化

中图分类号: G64; O6

Intelligent Reaction Optimization: Synthesis of Acetylsalicylic Acid Driven by Deep Learning and Optimization Algorithms

Jianqiang Zheng, Yongbin Huang, Wencan Ming, Yingju Liu*

College of Materials and Energy, South China Agricultural University, Guangzhou 510642, China.

Abstract: Undergraduate experimental design typically requires extensive trial-and-error experimentation to identify optimal reaction conditions, a process that demands considerable time and resources. To simplify this complex experimental design process and to enhance students' interest in advanced technologies and interdisciplinary fields, an intelligent reaction optimization model framework was introduced. This model is designed to predict the optimal combination of reaction conditions for achieving the highest yield, and this framework has been integrated into the undergraduate organic chemistry curriculum, specifically for the synthesis of acetylsalicylic acid. Herein, 1054 reaction data points were collected conforming to this reaction mechanism as training data for the model, mainly including product, reactant, catalyst, solvent, the main reaction reagents, reaction temperature and yield. Initially, a yield prediction model was pre-trained based on a chemical multi-modal transformer, which served as the objective function for subsequent reaction optimization. Then the Bayesian optimization algorithm was utilized to ascertain the optimal combination of reaction conditions, with the aim of minimizing the negative value of the yield (the maximizing yield). Finally, the model-predicted yields were experimentally validated against the actual yields, and in 100 model tests, the predicted yields consistently fell within the range of the actual yields, demonstrating the model's excellent performance. The model also identified a reaction reagent combination that yielded up to 90.1%, providing a variety of experimental options for this undergraduate experiment.

收稿: 2024-11-19; 录用: 2025-02-14; 网络发表: 2025-04-23

*通讯作者, Email: yingju@scau.edu.cn

基金资助: 无机及分析化学广东省课程思政示范团队(粤教高函[2024] 27号); 应用化学国家一流本科专业建设点(教高厅函[2019] 46号); 广东省一流课程-无机及分析化学(粤教高函[2020] 16号); 华南农业大学课程思政示范团队和示范课程-无机及分析化学(华南农教[2023] 1号)

Key Words: Optimization algorithm; Multi-modal transformer; Machine learning; Acetylsalicylic acid; Optimization

1 引言

随着数字化技术和人工智能的快速发展,为各学科的发展提供了新的手段和思路,人工智能与实验设计的结合是目前化学实验创新的有效工具和热门方向。智能反应优化可减少实验设计中不必要的尝试,更加高效地确定反应条件。同时,为研究和改进合成方法提供了有力的支持,促进了化学实验领域的技术发展。Shields等人^[1]提出了一个贝叶斯反应优化的框架,率先利用贝叶斯优化算法来辅助化学家更高效地筛选反应条件。2022年,Luo等人^[2]采用了随机森林模型和深度学习框架直接预测金属有机框架合成的反应条件。Wang等人^[3]将强化学习应用于常见反应的条件优化任务上,成功适应了含有不可见底物的不断变化的环境,并正确识别了多数的反应条件。

传统的实验设计方法是由化学家根据专业知识和经验判断进行选择。当实验影响因素增多、实验参数空间维度升高后,这种人为经验与直觉就可能受限。因此,对化学实验采用数字化设计是非常必要的。这不仅可以更精确地模拟和分析各种因素对实验的影响,确保结果的可靠性,还能够更合理地分配实验资源,避免浪费。朱俊杰等人^[4]提出,当前人工智能技术的迅猛发展促进人工智能技术融入教育教学。唐嘉屿等人^[5]将界面移动法实验装置进行数字化改造,采用电导率-时间曲线测量方法,测定电解质的离子迁移数和电迁移率等性质,为物理化学实验教学提供一种新方法。天津大学魏玉萍等人^[6]将结冷胶/聚丙烯酰胺双网络导电水凝胶这种新兴电子产物的制备及应用引入到本科教学实验中,促进了化学化工与生物医学学科的交叉融合人才的培养。数字化实验设计在分析化学的教学中也得到了广泛的应用,董天乐等人^[7]搭建了手持式色度计,对市场上常见的5种品牌果茶进行了含糖量的测定,并将其用到中学化学教学。温霞等人^[8]则用自制的MXLab-21先进数字化实验系统,开发了碘量法测定葡萄糖含量的数字化实验,帮助学生理解电解和氧化还原反应的微观本质。同时,得益于SciFinder^[9]和Reaxy^[10]等有机化学反应数据库的成功构建,大量有机反应数据得以被记录和保存。人工智能和大数据技术面对海量的有机反应数据具有显著的优势,因此将人工智能应用于有机化学实验教学中也是近年来的研究热点。周成卓^[11]将人工智能预测模型引入当前有机化学实验中的合成实验,选取ASKCOS模型进行正向合成预测和逆合成预测。赖自成等人^[12]构建了基于门控图卷积神经网络用于有机化学反应预测,为有机实验的路线设计提供了新的工具。Li等^[13]基于有机化学中的立体与电子效应,设计了化学信息增强的分子图模型,为反应活性和选择性预测提供新的架构,为合成反应预测和智能设计提供了新的技术,为有机实验教学提供了新的可解释性工具。Coley等人^[14]开发了基于人工智能的自动化合成平台,实现了15种药物或类药物分子的自动化合成,为有机化学实验的条件探索和设计过程提供了便利。

除此之外,将AI+反应优化的手段应用到本科实验教学中,同样也能够激发学生对先进技术和跨学科领域的兴趣,让学生了解到人工智能等前沿技术在化学领域的应用,拓宽学生的学术视野。同时,还能够鼓励学生思考如何将不同领域的知识和技术结合起来解决实际问题,培养创新能力和综合能力。随着各高校对计算机教育的重视,学生在进行有机化学实验之前已具备一定的数字化基础。

2 数字化设计方案

2.1 数字化设计的应用与预期

在化学实验教学过程中,节约课堂时间、提升教学效率以及降低实验试错成本是至关重要的。鉴于乙酰水杨酸的制备过程对于反应条件要求较为苛刻,因此有必要将此实验的设计过程数字化。建立智能优化反应模型,精确地确定实验的最佳反应条件并预测产率,减少实验次数和材料浪费。

产率预测模块根据输入的条件组合能够输出溶剂的物理化学性质、可视化催化剂分子的三维结构和主要反应试剂的电子密度分布,这不仅有助于学生更高效地预习实验,还能促进学生对实验所

需药品特性的深入理解。此外，任课教师和实验设计人员可以借助模型的条件优化模块设计不同的教学方案，提高课堂效率，丰富教学内容，提升整体教学效果。同时还为教师提供了评价学生实验操作的新标准，降低了实验方案本身的产率误差，更加合理地评估学生的实验操作水平。

2.2 原型实验

2.2.1 实验背景

乙酰水杨酸(俗称阿司匹林)的制备是一个重要的本科有机化学实验。该反应制备产物乙酰水杨酸是一种历史悠久的解热镇痛药，用于普通感冒的治疗。近年来还发现乙酰水杨酸能抑制血小板凝集，防止血栓的形成，临床上多用于预防心脑血管疾病的发作。乙酰水杨酸在实验教学和人类健康中都扮演着重要的角色，因此不断寻找最优的反应条件组合，对于本科实验教学以及人类生命健康有着举足轻重的作用。与此同时，由于该合成反应的反应机理为较为典型的O-酰基化反应，现存大量的反应数据供模型训练，具备反应优化的前提条件。

2.2.2 实验原理

(1) 水杨酸与醋酸酐在浓硫酸或磷酸催化下，发生O-酰基化反应生成乙酰水杨酸，反应式如图1所示。

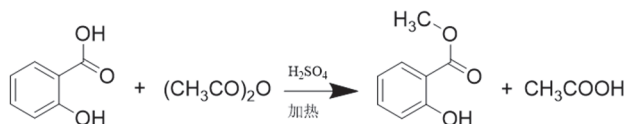


图1 乙酰水杨酸合成反应式

由于乙酰化反应不完全，或产物在分离步骤中发生水解，最终产物中可能混有少量水杨酸，残留的水杨酸可以用三氯化铁水溶液检出，通过纯化及重结晶过程除去^[15]。

(2) 水杨酸缩合形成聚合物为该反应的主要副反应，其反应式如图2所示^[15]。在乙酰水杨酸合成中，浓硫酸作为催化剂，能促进亲核反应，提高反应活性。碳酸氢钠因其能溶解阿司匹林钠盐而不溶解副产物聚合物，可作为溶剂纯化阿司匹林。而控制反应温度可以防止水杨酸缩合生成聚合物。总的来说，选择最优的试剂和反应温度可提高产率。因此，选择了该合成反应的催化剂种类、溶剂种类、主要反应试剂种类以及反应温度作为反应优化模块的优化参数。

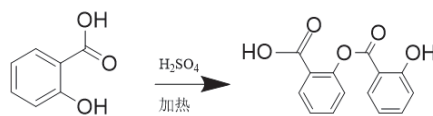


图2 乙酰水杨酸合成的副反应

2.2.3 实验步骤

实验具体操作步骤如下^[15]：

- (1) 加入3 g干燥水杨酸、4.5 g新蒸馏醋酸酐和5滴浓硫酸于50 mL圆底烧瓶，摇匀至固体溶解；
- (2) 安装球型冷凝管，通入冷凝水，水浴加热至80–85 °C，搅拌反应45 min后冷却至室温；
- (3) 加入38 mL饱和NaHCO₃溶液，搅拌反应直至无气泡产生，随后抽滤保留滤液；
- (4) 向滤液中滴加4 mol·L⁻¹的HCl溶液，调节pH至1–2，待大量沉淀析出后冰水冷却结晶，随后抽滤，冷水洗涤，所得白色晶体转移至表面皿烘干、称重，计算产率；
- (5) 取几粒产物分散于一定体积的水中，加入1–2滴1% FeCl₃溶液，根据溶液颜色是否变紫来判断是否仍有未反应的水杨酸。

2.3 数据库建立

(1) 针对产率预测模块, 在SciFinder数据库^[9]里检索了包含与乙酰水杨酸制备反应机理相同的反应数据, 其中明确标有产率和合成方案的数据共有1556个。综合考虑了该合成反应的影响因素、现存的反应数据数量以及模型的计算成本, 本文重点考虑催化剂种类、主要反应试剂种类、溶剂种类和反应温度这四个参数。将1556条反应数据中的反应物名称、产物名称、催化剂名称、溶剂名称、主要反应试剂名称、反应温度以及产率提取出来, 构成了产率预测模块的初始数据集。针对反应物、产物、催化剂、主要反应试剂和溶剂, 用Py2opsin库^[16]进行批量化学名称转换成分子SMILES表示, 并删去缺失值, 剩余1054个反应数据作为模型的训练数据。接着对反应温度和产率进行了标准化处理, 得到了最终用于训练的反应数据集(以下简称“DataSet1”), 并将DataSet1按照8:2划分训练数据集和验证数据集。DataSet1基本信息如表1所示。

表1 训练数据集和验证数据集DataSet1

DataSet	Total	Catalyst_Types	Solvent_Types	Reagent_Types	Yield	Temperature
DataSet1	1054	8	18	24	0.01 < Yield < 1	0 < T < 180

针对所选的原型实验, 在训练结束后对模型的预测性能进行了评估。将教材^[15]中的反应条件(浓硫酸作为催化剂、碳酸氢钠和浓盐酸作为主要反应试剂、反应温度85 °C)作为测试环节的输入参数, 输入到模型中进行产率预测, 与教材给出的产率范围63%–71%进行比对。

(2) 针对条件优化模块, 将DataSet1中的所有催化剂种类、溶剂种类、主要反应试剂种类、反应温度提取出来, 并转化成分子SMILES表示, 构成数据集DataSet2。其中DataSet2内的所有条件构成了贝叶斯优化模块^[17]的参数空间。

2.4 数字化设计原理

2.4.1 多模态特征选择

鉴于化学试剂在反应过程中所扮演的角色各不相同, 本研究针对不同类型的试剂选取了相应的特征(详见表2)。所有的主要反应试剂均采用了smi2vec^[18]对SMILES进行编码, 以获取SF特征。除此之外, 针对反应物和产物, 明确了三种关键的分子指纹: MACCS、Avalon FP和ECFP4。对其编码, 得到了FP特征, 并构建了二维分子图表示, 其中节点特征为27维原子特征, 边特征则根据键类型进行编码, 从而得到Graph特征。针对催化剂、溶剂和主要反应试剂, 除了分子的SF特征外, 还分别选取了三维结构信息、物理化学性质和电子密度分布等特征。

(1) 分子指纹。

分子指纹是连接分子结构、物理化学性质和生物活性的方法, 分为基于字典式指纹、圆形指纹、拓扑指纹、药效团指纹、蛋白质-配体相互作用和基于形状等六种。本研究中选择的MACCS^[19]、Avalon^[20]和ECFP4^[21]指纹分别属于字典式指纹、拓扑指纹和圆形指纹。催化剂则选择了基于药效团^[22]的指纹, 考虑化学和结构信息, 引入几何信息描述催化剂在反应中的结构变化和反应位点原子特征。

(2) 分子图。

分子图主要是将原子视为分子图的节点, 化学键视为分子图的边, 可选择将原子特征和化学键特征编码为节点特征和边特征。这种表示分子的方法包含了更多样的分子信息, 分子图的处理更加符合化学家对于分子在原子层面上存在相互影响的认识。

(3) 分子SMILES。

分子SMILES^[23]是一种具有语法规则的语言结构, 由原子和键符号构成。针对分子SMILES的编码, 采用自然语言编码中最常用的smi2vec方式进行编码。编码过程先遍历全部分子构建分子SMILES

表2 反应特征总表

分类	特征	
反应物和产物	MACCS指纹	
	Avalon指纹	
	ECFP4指纹	
	分子图	
	分子SMILES	
溶剂	分子SMILES	
	物理化学性质	LogP (脂水分配系数的对数)
		MW (相对分子质量)
		NHD (氢键供体数量)
		NHA (氢键受体数量)
TPSA (拓扑极性表面积)		
主要反应试剂	分子SMILES	
	电子密度分布特征	
催化剂	分子SMILES	
	药效团指纹特征	
	三维结构特征	
	电子密度分布特征	

的词汇表，再将每一个符号分别编码为一个向量，采用one-hot编码得到分子表示矩阵。分子表示矩阵(N*X，其中N为最大字符数，X表示词汇表中符号数量)中每一行即对应一个分子SMILES中的一个字符，超出当前字符的用“0”填充。

(4) 物化性质。

本研究计算得到了五种溶剂在反应中比较重要的物理化学性质，具体的物理化学性质及其描述如表2所示。

(5) 三维结构信息。

分子的三维结构常用坐标、网格和形状表示。网格表示将空间划分为网格，标记原子信息。但计算资源需求大，故常用坐标表示。坐标表示能精确描述原子位置，因此设计了TD_Model来提取分子三维坐标特征。

(6) 电荷密度分布。

部分的主要反应试剂在有机反应中通常具有正电性和负电性的位点和基团，这也赋予了它们亲电或者亲核的反应特点。因此，为了更好地将这种特征表示出来，采用rdkit开源包^[24]计算了具有上述反应特点的主要反应试剂的Gasteiger电荷。基于分子的电子结构和相互作用，将电荷分配到分子中各个原子，考虑了原子的电负性、杂化状态和连接情况。

2.4.2 多模态产率预测模块

多模态产率预测模块(以下简称MFYP)作为整个智能优化模型的第一个模块，模型架构如图3所示。所有模型均在pytorch^[25]开源深度学习框架下编写。针对不同模态特征的特异性，分别构建了四个模型进行特征的提取，分别是MF_Model、Graph_Model、TD_Model和SF_Model，其中TD_Model作为模块添加到MF_Model中。

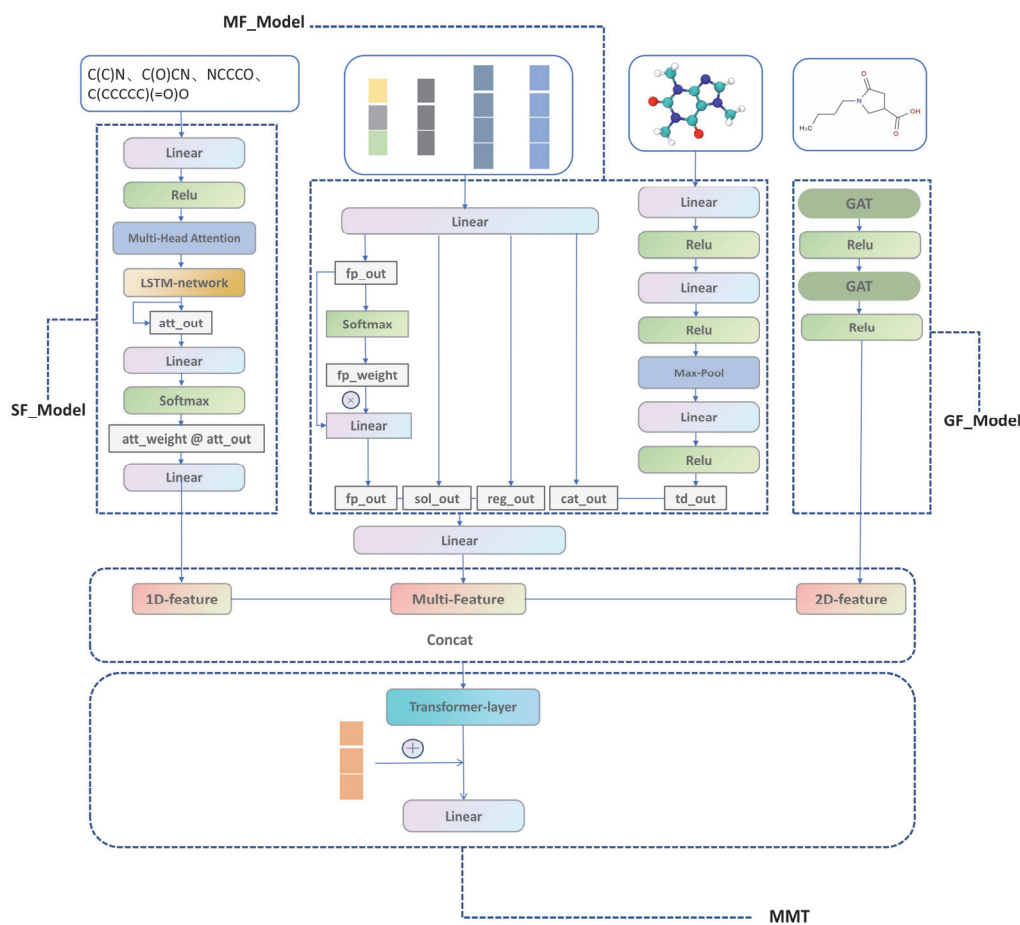


图3 多模态产率预测模块模型架构

(1) MF_Model

MF_Model用于处理二维特征，如指纹特征、溶剂物化特征、催化剂药效团指纹特征和主要反应试剂的电子密度分布特征。具体来说，产物的三种指纹特征先分别通过全连接层 f_{maccs} 、 f_{Avalon} 、 f_{ECP} 和 $RELU$ 激活函数后，进行拼接得到FP_Combine。之后对FP_Combine计算注意力权重 α ，并进行点乘操作更新FP_Combine，再经过一个全连接层 f_{out} 得到产物的指纹特征Pro_FP。具体公式如下所示：

$$\begin{aligned} MACCS_out &= Relu[f_{maccs}(MACCS_FP)] \\ Avalon_out &= Relu[f_{Avalon}(Avalon_FP)] \\ ECFP_out &= Relu[f_{ECP}(ECFP_FP)] \\ FP_Combine &= concat[(MACCS_out, Avalon_out, ECFP_out), dim = -1] \\ \alpha &= Softmax(FP_Combine) \\ FP_Combine &= \alpha * FP_Combine \\ Pro_FP &= f_{out}(FP_Combine) \end{aligned}$$

对于反应物，首先对两个反应物的指纹特征进行拼接，之后重复上述操作，得到反应物的指纹特征Rec_FP。对于溶剂的物化特征Sol_feat、催化剂的药效团指纹Cat_Pharam、催化剂的电子密度特征Cat_quantity以及主要反应试剂的电子密度分布特征Reg_quantity，分别通过全连接层 f_{sol} 、 f_{cat1} 、 f_{cat2} 、 f_{reg} ，得到对应更新后的特征。而对于催化剂的三维结构信息，通过TD_Model进行提取。TD_Model包括六个全连接层和一个最大池化层，具体公式如下所示：

$$Sol_feat = f_{sol}(Sol_feat)$$

$$\begin{aligned} \text{Cat_Pharam} &= f_{cat1}(\text{Cat_Pharam}) \\ \text{Cat_quantity} &= f_{cat2}(\text{Cat_quantity}) \\ \text{Reg_quantity} &= f_{reg}(\text{Reg_quantity}) \\ \text{Cat_structure} &= TD_Model(\text{Cat_structure}) \end{aligned}$$

最后将所有提取到的特征在最后一个维度上进行拼接, 经过一个全连接层, 得到MF_Model的输出结果MFP_out。

(2) Graph_Model

Graph_Model用于处理产物和反应物的二维分子图特征。先将分子编码成分子图, 通过两层图注意力层(GAT)^[26]和全连接层后, 用 $\log_Softmax$ 函数计算得到Graph_Model的输出Graph_out, GAT的公式如下所示(输入特征矩阵 $H \in R^{N \times F}$ 、权重矩阵 $W \in R^{F \times F'}$ 、注意力权重 $a \in R^{2 \times F'}$):

$$\begin{aligned} H' &= HW (\text{where } H' \in R^{N \times (H \cdot F')}) \\ e_{ij} &= LeakyReLU(h_i^T a^H + h_j^T a^H) \forall i, j \\ \bar{e}_{ij} &= \begin{cases} e_{ij} & \text{if } adj_{ij} > 0 \\ -\infty & \text{otherwise} \end{cases} \\ \alpha_{ij} &= Softmax(\bar{e}_{ij}) \\ H'' &= \sum \alpha_{ij} h_j \text{ (concatenated or averaged)} \end{aligned}$$

(3) SF_Model

SF_Model用于处理分子的一维特征, 将分子SMILES字符串视为具有语法规则的自然语言, 并将每一个SMILES字符串编码成向量。SF_Model首先将输入通过全连接层 f_1 变换, 然后输入到多注意力层MultiSelfAttention^[27]更新序列特征, 在维度变换之后, 输入到LSTM层中提取序列特征, 将LSTM最后一层隐藏向量 $h_{t=-1}$ 与输出向量SF_out在特征维度进行拼接后, 通过Softmax函数计算序列各位位置的注意力分数。将注意力分数与输出向量进行批矩阵乘法运算后, 将结果展平并通过全连接层得到SF_Model的输出SF_out, 其中SF_Model的公式如下所示:

$$\begin{aligned} SF_feat &= f_1(SF_feat) \\ SF_feat &= MultiSelfAttention(SF_feat) \\ SF_out, (h_t, c_t) &= LSTM(SF_feat) \\ h_{t=-1} &= h_t[-1] \\ SF_out &= concat[(SF_out, h_{t=-1}), dim = -1] \\ \alpha &= Softmax([f_{att}(h_{t=-1})]) \\ att_out &= \alpha @ SF_out \\ SF_out &= f_{out}(att_out) \end{aligned}$$

(4) 多模态融合模块(MMT)

MMT (Multi-Modal Transformer)是一种用于处理多模态数据的Transformer模型^[28]。通过利用Transformer架构的强大能力来处理和融合提取到的SF_out、Graph_out和MFP_out这些数据。首先将三种模态维度的特征在最后一个维度上进行拼接, 然后经过一个Transformer_layer将特征融合, 并将输出与反应温度进行拼接, 经过全连接层变换后输出模型最终的预测值Pred。

2.4.3 基于贝叶斯优化的反应优化模块

贝叶斯优化^[17]是一种通过构建和更新概率分布模型(如高斯过程、多项式回归)来优化目标函数的方法, 作为整个智能优化模型的第二个模块。它利用贝叶斯定理计算后验分布, 并通过acquisition函数指导搜索, 以找到最优解。针对有机化学实验成本高、耗时且危险的问题, 本研究采用训练好的产率预测模型作为贝叶斯优化的目标函数, 并使用高斯过程回归作为代理模型。通过对催化剂种类、主要反应试剂种类、溶剂种类和反应温度进行编码, 模型计算期望改进并选择反应条件, 更新数据集后重复上述步骤, 不断训练和更新模型参数, 直至最大迭代次数。最终, 确定最优的催化剂

种类、主要反应试剂种类、溶剂种类和反应温度。

3 实验部分

3.1 预期教学流程

3.1.1 实验方案设计

(1) 登录智能优化反应网站，如图4所示，点击“产率预测”，输入反应条件(反应物SMILES、产物SMILES、催化剂SMILES、主要反应试剂SMILES、溶剂SMILES、反应温度)，提交后得到预测产率。学生可按需保存下载网站生成的溶剂性质表、分子三维结构和主要反应试剂电子密度图。

(2) 根据模型预测产率值，评估反应条件。如需优化，点击“反应优化”，输入反应物和产物名称，设定迭代次数后提交。随后网站将返回优化后的反应条件、预测产率及贝叶斯优化结果。

(3) 针对模型给出的反应条件，学生通过查阅文献和相关资料，进一步完善实验方案，设计可行的操作步骤。



图4 网站提交页面

3.1.2 实验操作

按照初始的反应条件和优化后的反应条件进行乙酰水杨酸的合成实验，记录两组实验的实验现象、操作步骤以及实际产率。

3.1.3 数据处理与结果分析

(1) 对比两组实验的实验产率，计算实际产率与预测产率的相对误差，基于反应条件和实验过程分析造成产率不同的原因，比较两组实验所选择的催化剂、溶剂、主要反应试剂的性质异同点，从而总结和了解不同类型试剂在化学实验中的重要性。

(2) 分析实际产率与预测产率之间存在误差的原因，进一步分析实验过程中可能存在的错误操作与可改进的方面，撰写实验报告。

(3) 分析网站保存的优化过程图，理解结果图的含义以及绘制过程。学习如何使用计算机将实验结果进行可视化，掌握数据处理与图形绘制的技巧。通过观察和分析优化过程图，深入理解反应的优化过程，并探讨如何将抽象的数学模型应用于实际问题的解决。

3.1.4 结果返回与模型迭代

学生将优化后反应条件下的实验结果整理成文档上传到网站，实现模型的更新迭代。同时让学生参与到与AI模型的实际交互当中，进一步加深学生对于AI工作原理的理解，模型的工作流程如图5所示。

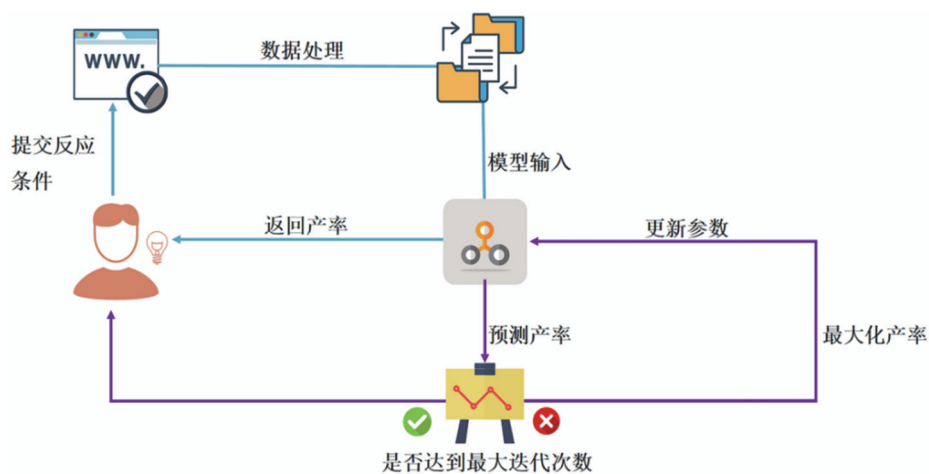


图5 模型工作流程

3.2 实验现象与模型结果分析

3.2.1 模型结果

(1) 多模态产率预测模块训练结果。

训练过程中，使用Adam优化器^[29]以 $1e-4$ 的学习率训练模型，每次迭代的样本数量为275，训练周期为200。训练过程中采用均方误差(MSE)作为损失函数，并使用均方根误差(RMSE)进行模型评估，训练结果如图6(a)、(b)所示。结果表明，模型在40个epoch以后逐渐收敛，MSE收敛值逼近于0，RMSE收敛值也处于合理范围，由此说明模型的训练情况良好，并没有出现梯度爆炸或者梯度消失的异常情况。同时，为了评估模型预测性能，将课本中的反应条件输入模型预测，根据课本产率范围评估准确性。如图6(c)，课本产率0.63–0.71，模型在50个epoch后收敛至实际产率范围，表明模型可泛化，能优化原型实验反应条件。

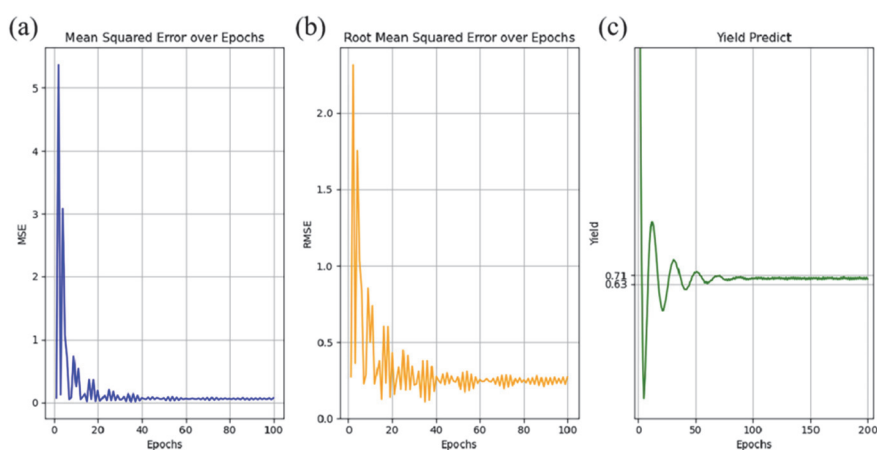


图6 训练结果

(a) MSE曲线；(b) RMSE曲线；(c) 原型实验条件下产率预测曲线

(2) 网站保存结果的可视化。

图7展示了网站呈现的部分结果，学生可以直接获得不同试剂在实际反应中较为重要的性质，并进行可视化展示。这不仅有助于学生在预习阶段更好地了解试剂的性质，提高预习的质量和效率，还能在实验结束后，帮助学生对比试剂性质的差异，进一步分析实验结果和原因。

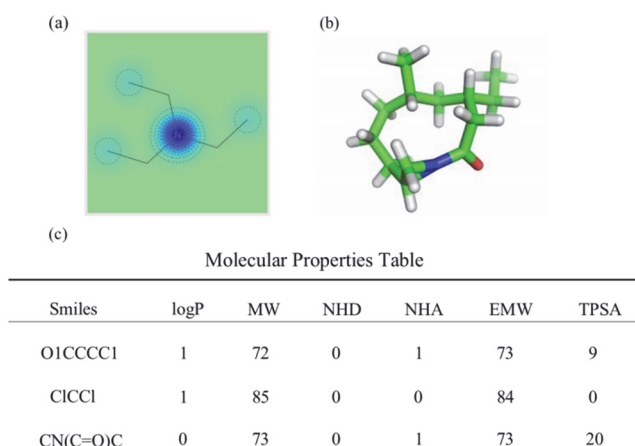


图7 网站部分结果

(a) 电子密度分布图; (b) 分子三维结构图; (c) 溶剂物化性质表

3.2.2 实验操作现象

根据原始反应条件,进行了乙酰水杨酸的合成实验。如图8所示,最终产品为白色粉末,质量为2.49 g,产率为63.6%。与预测产率的相对误差为0.63%,表明模型预测效果十分准确。最后,以 FeCl_3 试剂检测纯度,溶液呈微黄色,表明未检测到水杨酸残留,产品纯度较高。

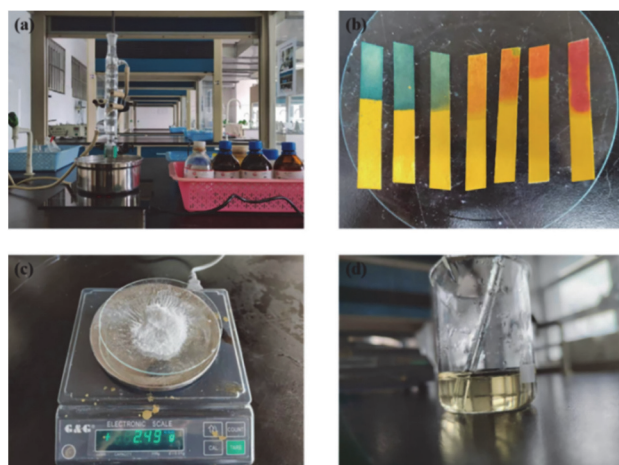


图8 实验过程

(a) 反应装置图; (b) pH条件控制; (c) 乙酰水杨酸的产量; (d) 产品水溶液经 FeCl_3 溶液验纯

3.2.3 贝叶斯优化反应条件结果

通过对原型实验进行了200次迭代优化,最终得到了多组高于原型实验产率的反应条件组合,最高预测产率达到90.4%,模型目标函数的变化曲线以及优化过程如图9、图10所示。由图9,反应优化模块成功执行了目标函数的优化过程,目标函数值持续下降,并经历了三个平台期,最终收敛至最优值-0.904。图10中横坐标与纵坐标分别表示不同反应条件的取值或编号。据此分析,模型的搜索范围广泛覆盖了取值空间的大部分区域,并识别出具有极高产率的反应条件组合,有效完成了反应条件优化的任务。

4 结语

综上,本实验成功构建了一套智能反应优化模型,该模型融合了深度学习技术和优化算法,在

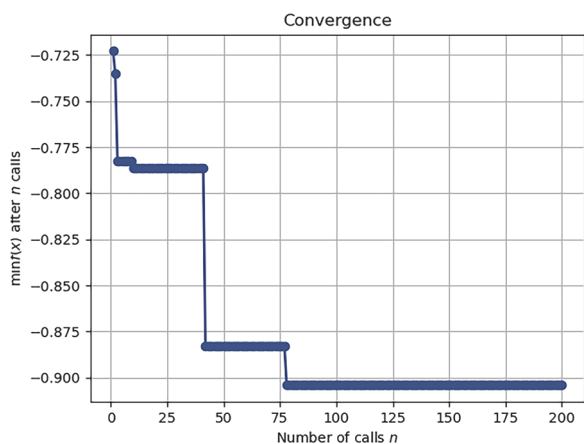


图9 贝叶斯优化目标函数的变化曲线

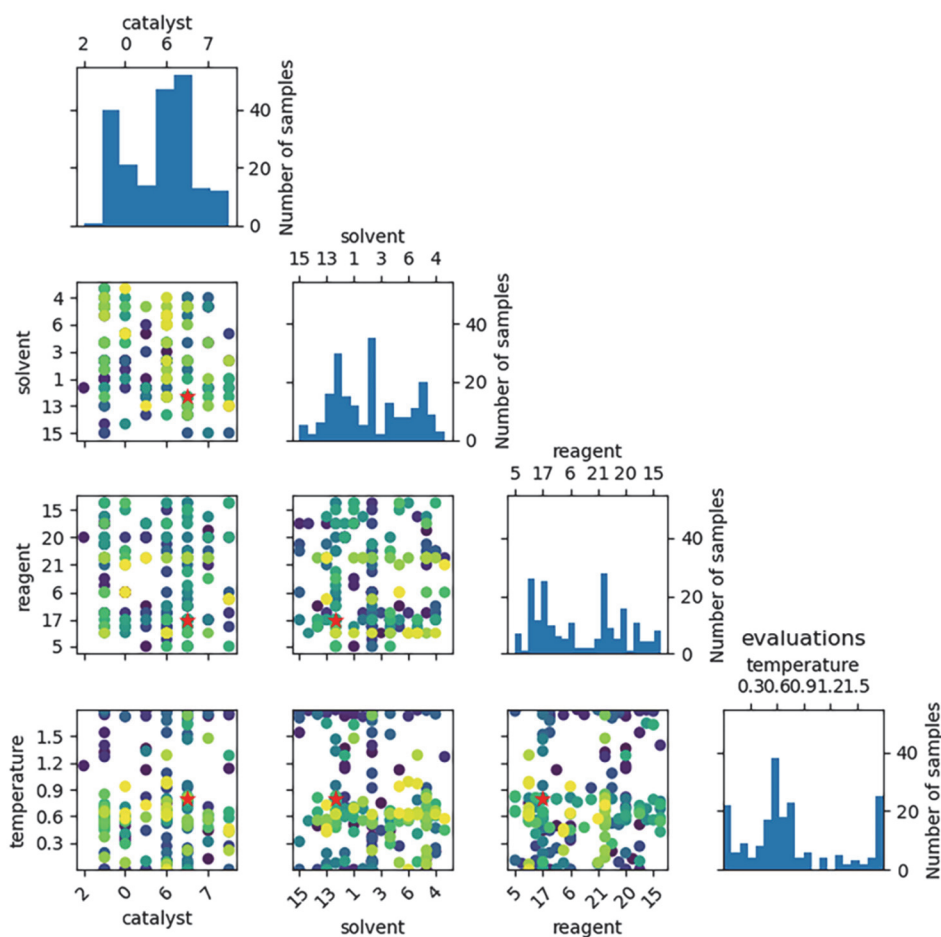


图10 贝叶斯优化的结果分布

散点图可视化了探索的区域，星标为最佳的反应条件；直方图显示了不同反应条件的搜索次数

数字化实验设计上表现出卓越的性能，但也存在一定的局限性。此外，该模型集成了多模态产率预测模型与贝叶斯优化算法，能够有效避免昂贵的试错实验，大幅降低研究成本。为了促进知识的共享和教育的便捷性，该模型可被开发成易于访问的网络平台，不仅方便了师生的使用，还通过保存

可视化结果来提升学习效率和质量。展望未来, 将进一步拓展该模型的应用范围, 包括植物保护、药物合成等重要领域, 以增强模型的性能和可解释性, 为科学实验和研究工作带来革命性的进步。

参 考 文 献

- [1] Shields, B. J.; Stevens, J.; Li, J.; Parasram, M.; Damani, F.; Alvarado Martinez, J. I.; Janey, J. M.; Adams, R. P.; Doyle, A. G. *Nature* **2021**, *590* (7844), 89.
- [2] Luo, Y.; Bag, S.; Zaremba, O.; Cierpka, A.; Andreo, J.; Wuttke, S.; Friederich, P.; Tsotsalas, M. *Angew. Chem. Int. Ed.* **2022**, *61* (19), e202200242.
- [3] Wang, J. Y.; Stevens, J. M.; Kariofillis, S. K.; Tom, M. J.; Golden, D. L.; Li, J.; Tabora, J. E.; Parasram, M.; Shields, B. J.; Primer, D. N.; *et al.* *Nature* **2024**, *626* (8001), 1025.
- [4] 朱俊杰, 薛永飞, 周国雄, 郑志安. 实验室研究与探索, **2024**, *43* (10), 142.
- [5] 唐嘉屿, 庞继川, 肖少华, 许新华, 吴梅芬. 大学化学, **2024**, *39* (5), 193.
- [6] 魏玉萍, 王佚婷, 姜佳良, 邓锦轩, 张宏, 马晓飞, 李俊杰. 大学化学, **2024**, *39* (10), 261.
- [7] 董天乐, 王永森, 杨露辉, 魏建业. 化学教育(中英文), **2024**, *45* (19), 89.
- [8] 温霞, 徐琦, 王海军, 杜钦之, 张明晓, 陈时洪. 化学教育(中英文), **2024**, *45* (19), 83.
- [9] SciFinder. [2024-08-15]. <https://www.cas.org/>
- [10] Reaxy. [2024-08-16]. <https://www.reaxys.com/>
- [11] 周成卓, 谢召军. 大学化学, **2025**, *40* (2), 320.
- [12] 赖自成, 张玉萍, 马燕. 计算机应用, **2021**, *41* (10), 3070.
- [13] Li, S. W.; Xu, L. C.; Zhang, C.; Zhang, S. Q.; Hong, X. *Nat Commun.* **2023**, *14* (1), 3569.
- [14] Coley, C. W.; Thomas D. A.; Lummiss, J. A. M.; Jaworski, J. N.; Breen, C. P.; Schultz, V.; Hart, T.; Fishman, J. S.; Rogers, L.; Gao, H.; *et al.* *Science* **2019**, *365* (6453), eaax1566.
- [15] 张淑婷, 杨卓鸿. 有机化学实验. 北京: 中国农业出版社, 2017: 94-96.
- [16] Lowe, D. M.; Corbett, P. T.; Murray-Rust, P.; Glen, R. C. *J. Chem. Inf. Model.* **2011**, *51* (3), 739.
- [17] Brochu, E.; Cora, V. M.; Freitas, N. D. **2010**, arXiv:1012.2599. [2025-08-29]. <https://doi.org/10.48550/arXiv.1012.2599>
- [18] Goh, G. B.; Hodas, N. O.; Siegel, C.; Vishnu, A. **2017**, arXiv:1712.02034. [2025-08-29]. <https://doi.org/10.48550/arXiv.1712.02034>
- [19] Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (6), 1273.
- [20] Gedeck, P.; Rohde, B.; Bartels, C. *J. Chem. Inf. Model.* **2006**, *46* (5), 1924.
- [21] Rogers, D.; Hahn, M. *J. Chem. Inf. Model.* **2010**, *50* (5), 742.
- [22] Watson, P. *J. Chem. Inf. Model.* **2008**, *48* (1), 166.
- [23] David, W. *J. Chem. Inf. Comput. Sci.* **1988**, *28* (1), 31.
- [24] RDKit: Open-source cheminformatics. [2025-08-29]. <https://www.rdkit.org>
- [25] Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; *et al.* **2019**, arXiv:1912.01703. [2025-08-29]. <https://doi.org/10.48550/arXiv.1912.01703>
- [26] Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. **2017**, arXiv:1710.10903. [2025-08-29]. <https://doi.org/10.48550/arXiv.1710.10903>
- [27] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; Polosukhin, I. **2017**, arXiv:1706.03762. [2025-08-29]. <https://doi.org/10.48550/arXiv.1706.03762>
- [28] Huang, Z.; Zeng, Z.; Liu, B.; Fu, D.; Fu, J. **2020**, arXiv:2004.00849. [2025-08-29]. <https://doi.org/10.48550/arXiv.2004.00849>
- [29] Kingma, D. P.; Ba, J. **2014**, arXiv:1412.6980. [2025-08-29]. <https://doi.org/10.48550/arXiv.1412.6980>