

文章编号:1006-9941(XXXX)XX-0001-13

# 机器学习赋能含能材料合成研究:进展和挑战

窦凯乐<sup>1,2</sup>,赵伟波<sup>1</sup>,何春林<sup>1</sup>,张蕾<sup>1,2</sup>,庞思平<sup>1</sup>

(1.北京理工大学材料学院,北京 100081; 2.爆炸科学与安全防护全国重点实验室,北京 10081)

**摘要:** 含能材料因其在国防、航天及特殊工程中的重要作用而备受关注,但其研发过程面临实验代价高、安全风险大和周期漫长等难题,显著制约了新型含能材料的快速迭代与应用落地。近年来,机器学习因其强大的数据建模与预测能力,逐渐成为推动化学与材料研究的重要工具。本研究综述了机器学习在化学合成中的最新应用进展,包括反应预测、合成路径规划及自动化合成等前沿方向,同时重点讨论了其在含能材料合成研究中的潜在价值与局限性。并总结了当前面临的主要挑战,包括数据匮乏与质量不足、安全性评估缺失以及实验验证与模型迭代受限。最后,展望了未来的发展趋势,即建立标准化与可共享的数据库以及发展适用于含能体系的高通量与自动化实验平台,旨在为实现含能材料的高效、智能合成提供理论参考与方法支撑。

**关键词:** 含能材料;机器学习;反应预测;自动化合成

中图分类号: TJ55;TP181

文献标志码: A

DOI:10.11943/CJEM2025237

## 0 引言

含能材料是一类含有爆炸性基团或含有氧化剂及燃料的化合物或混合物,能够在外界特定能量刺激下发生剧烈的氧化还原反应,通过燃烧和爆炸快速释放其储存的化学能,广泛应用于国防军事、航空航天、采矿工程等重要领域,其性能优劣直接关系到武器系统的作战效能与安全性<sup>[1]</sup>。新型含能材料的开发要求在能量、感度、热稳定性、力学强度与环境相容性等多维性能之间实现最优平衡,研发过程不仅复杂而且风险高、成本高。传统研发主要依赖于人工经验与反复实验,研发周期长、资源消耗大,且存在“高能低感”难以调和等固有矛盾<sup>[2]</sup>。

在此背景下,探索更加高效、智能的研发方法成为

含能材料领域的研究焦点。2007年,图灵奖得主 Jim Gary 提出了关于科学研究“四种范式”的演进框架:第一范式为经验观察,第二范式为理论推导,第三范式为计算模拟,第四范式即数据驱动的科学发现(Data-Intensive Scientific Discovery)<sup>[3]</sup>。第四范式强调通过大数据与人工智能技术,从复杂的实验与模拟数据中挖掘规律,从而引导科学研究方向。

机器学习作为数据驱动技术的核心手段,近年来在化学领域取得广泛应用,包括反应预测<sup>[4]</sup>、逆合成规划<sup>[5]</sup>以及自动化合成<sup>[6]</sup>等,均显著提高了研发效率与智能化水平。在材料科学领域,尤其是在有机电子、催化剂、高分子材料等研究中,机器学习已经实现了从结构到性能的精准预测。然而,在含能材料领域,机器学习的应用仍处于初级阶段,研究多集中于性能预测与分子生成,尚缺乏利用机器学习针对含能分子化学合成方面的系统研究。

因此,本研究概述了机器学习在含能材料研发中的应用进展,聚焦于反应预测、合成路径规划以及自动化合成等关键环节,旨在为该领域的后续研究提供系统化的分析思路与方法参考。此外,结合其他材料化学领域的成功经验,深入分析机器学习在含能材料研发中的优势与局限,并对未来数据体系建设和智能化实验平台的发展方向进行了展望。

收稿日期:2025-10-27;修回日期:2025-12-19

网络出版日期:2026-02-02

基金项目:国家自然科学基金优秀青年科学基金(12222204, 12472361),爆炸科学与安全防护国家重点实验室基金(QKKT24-01)

作者简介:窦凯乐(1998-),男,博士,主要从事含能材料合成研究中的机器学习方法与应用。e-mail:kailedou@bit.edu.cn

通信联系人:张蕾(1985-),女,教授,主要从事含能材料理论研究。e-mail:zhang\_lei@bit.edu.cn

何春林(1983-),男,教授,主要从事含能材料合成研究。e-mail:chunlinhe@bit.edu.cn

引用本文:窦凯乐,赵伟波,何春林,等.机器学习赋能含能材料合成研究:进展和挑战[J].含能材料,DOI:10.11943/CJEM2025237.

DOU Kai-le, ZHAO Wei-bo, HE Chun-lin, et al. Machine Learning-Enabled Synthesis of Energetic Materials: Progress and Challenges[J]. Chinese Journal of Energetic Materials (Hanneng Cailiao), DOI:10.11943/CJEM2025237.

## 1 机器学习在化学合成中的应用

随着人工智能技术的快速发展,化学合成正逐步从传统的经验试错模式向数据驱动的智能化模式转变。如图1所示,当前机器学习在化学合成中的应用主要有以下4个方面:首先是大规模反应数据库的构建,包括反应物、试剂、产物及产率等信息在内的海量

数据的挖掘和处理;然后利用这些数据作为输入,通过逆合成分析来反向推导目标分子的合成路线及起始原料;同时开展反应预测,针对产物结构、产率以及反应条件进行预判;最后根据模型预测结果,结合自动化合成平台进行实验验证。这4个环节构成了“数据-模型-实验”的完整闭环,展现了人工智能在化学领域的广泛适用性与巨大潜力。

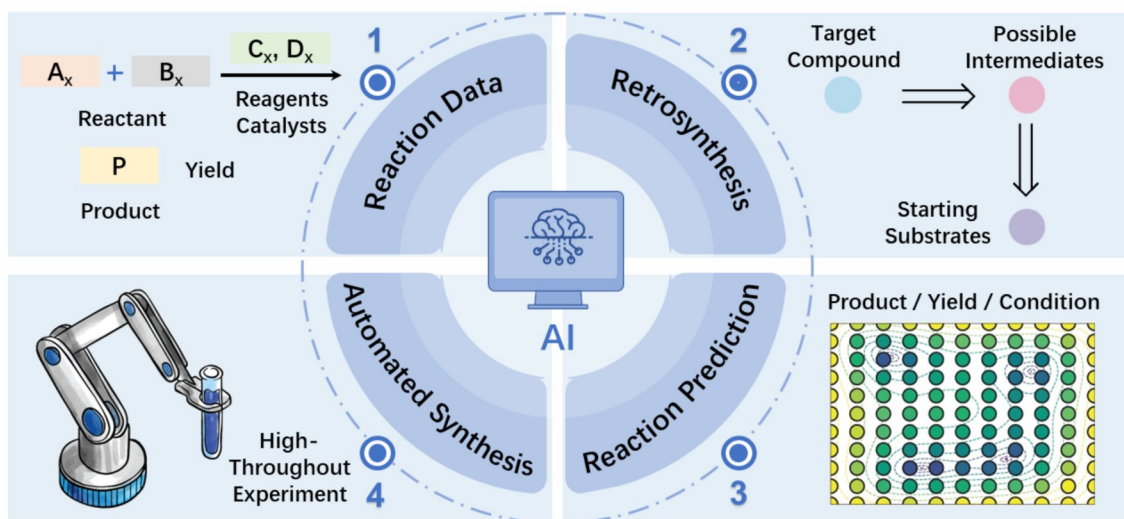


图1 机器学习在化学合成领域的应用

Fig.1 Applications of machine learning in chemical synthesis

### 1.1 反应预测

在有机合成化学领域,发现新反应的“灵感”往往来源于经验、专业知识等对化学机理的深刻理解,甚至来源于直觉。然而,精确预测单一化学反应的结果,无论是对于优秀的化学家还是计算机模型都是巨大的挑战。即使已经建立起的成熟的方法学,也常常难以准确预测复杂底物是否会发生预期的转化,这使得不少化学合成的任务变得挑战性高且费时费力<sup>[7]</sup>。近年来在材料科学、药物分子及有机化学等领域,已有大量研究报道了利用机器学习进行化学反应预测的方法。如正向反应预测中基于反应物、反应条件等预测产物,反应表现中对化学反应的选择性、产率等的预测。

然而,在含能材料领域尚未有相关的研究报道。一方面,含能分子合成中涉及高氮、高能等特殊基团,其反应路径复杂且伴随高风险,难以直接沿用已有模型。另一方面,缺乏大规模开放数据库也使得模型的训练和泛化更加困难。换句话说,即便是当前表现优异的反应预测模型,在面对含能材料体系时,可能出现预测准确性大幅下降的情况。未来亟需发展结合小样本学习、迁移学习的策略,以现有常规有机化学反应数

据为“底座”,再针对含能材料进行少量补充训练,从而弥补数据不足带来的问题。

#### 1.1.1 正向反应预测

正向反应预测(Forward reaction prediction)是指预测给定反应物经过化学反应后可能生成的产物。这一过程对于化学研究至关重要,准确的反应结果预测不仅能够节省实验成本和时间,还能为新分子的设计提供重要依据。目前正向反应预测问题的解决方法大致可以分为三类:基于模板的方法、基于序列的方法、基于图的方法<sup>[8]</sup>。

基于模板的方法依赖于人工定义的规则或从反应数据中提取的反应模板(反应中原子和化学键的变化规则),将产物的生成问题转换为模板的选择问题,然后将反应物与排名靠前的模板进行匹配,以产生候选产物。Wei等<sup>[9]</sup>将反应物和试剂的分子指纹拼接作为神经网络的输入特征,通过预测17种不同反应类型的概率,然后将反应物和最可能发生的反应类型匹配,从而预测产物,但预测范围受限于训练集中已定义的反应类型。为了拓展模板法的适用范围,Coley等<sup>[10]</sup>基于美国专利数据库(USPTO)的15000条反应记录自

动提取反应模板,通过模板生成可能的产物集合,再利用神经网络对这些候选产物进行打分排序,从而实现主要产物的预测,Top-1 准确率达到71.8%。

基于序列的方法是近年来随着深度学习特别是自然语言处理(Natural Language Processing, NLP)技术发展而兴起的一类重要方法。该思路将化学反应视为“语言翻译”问题:输入为反应物及试剂的SMILES<sup>[11]</sup>序列,输出为产物的SMILES序列。模型通过学习大规模反应数据库中反应的“语法”和“语义”规律,从而自动捕捉化学反应模式,实现端到端的反应预测,早期的序列模型多采用序列到序列(Seq2Seq)架构<sup>[12]</sup>。Schwaller等<sup>[13]</sup>将自然语言处理中的Transformer架构引入反应预测领域,提出了Molecular Transformer模型。该模型将反应物、试剂与产物的SMILES表示视为“化学语言”,通过多头自注意力机制捕捉反应过程中断键-成键的深层结构关联,无需反应模板或人工特征提取,可直接从大规模反应数据中学习反应规律,在USPTO-MIT数据集上实现了超过90%的Top-1准确率。考虑到传统的SMILES表示仅反映了分子的线性结构,而忽略了反应过程中原子与键的变化,Zhong等<sup>[14]</sup>提出了一种新的分子表示形式Root-aligned SMILES(R-SMILES),通过对齐反应物和产物的关键反应中心来提供更为紧凑的分子表示,增强了模型对反应中心变化的捕捉能力,提高了预测的准确性。与基于模板的方法相比,序列模型摆脱了对预定义反应

规则的依赖,能够自动从反应数据中提取模式,对未知或新颖反应具有更好的泛化性。然而,其主要局限在于缺乏显式的化学结构信息,对反应机理解释性较弱。

与基于序列的模型不同,图方法直接以分子的拓扑结构为输入,将原子视为节点、化学键视为边,通过图神经网络(Graph Neural Network, GNN)学习分子间的结构关联与反应中心特征,从而实现反应产物或反应性能的预测。Coley等<sup>[15]</sup>提出了一种图神经网络模型,通过模拟化学家分析反应的思路,将反应预测任务分为两步:首先通过图匹配机制识别反应中心,然后基于反应中心进行产物生成并评分,该方法在USPTO数据集上的Top-1准确率达到85.6%。Yousung Jung等<sup>[16]</sup>将广义反应模板(Generalized Reaction Template, GRT)与图神经网络相结合,提出了LocalTransform模型,通过原子映射法来描述反应前后原子构型的局部变化,并根据局部化学环境变化和全局注意机制的选择性来识别反应原子,最终反应结果由模板分类器进行预测。该模型在USPTO-480k数据集上的Top-1和Top-2准确率分别为90.8%和94.8%。

在含能材料领域,机器学习在化学反应预测中的应用仍处于探索阶段。2024年,Dou等<sup>[17]</sup>基于公开反应数据,结合高通量计算和机器学习算法,构建了针对六氮杂异伍兹烷笼状化合物的底物反应活性预测模型,如图2所示。其预测准确率达到91.4%,同时揭

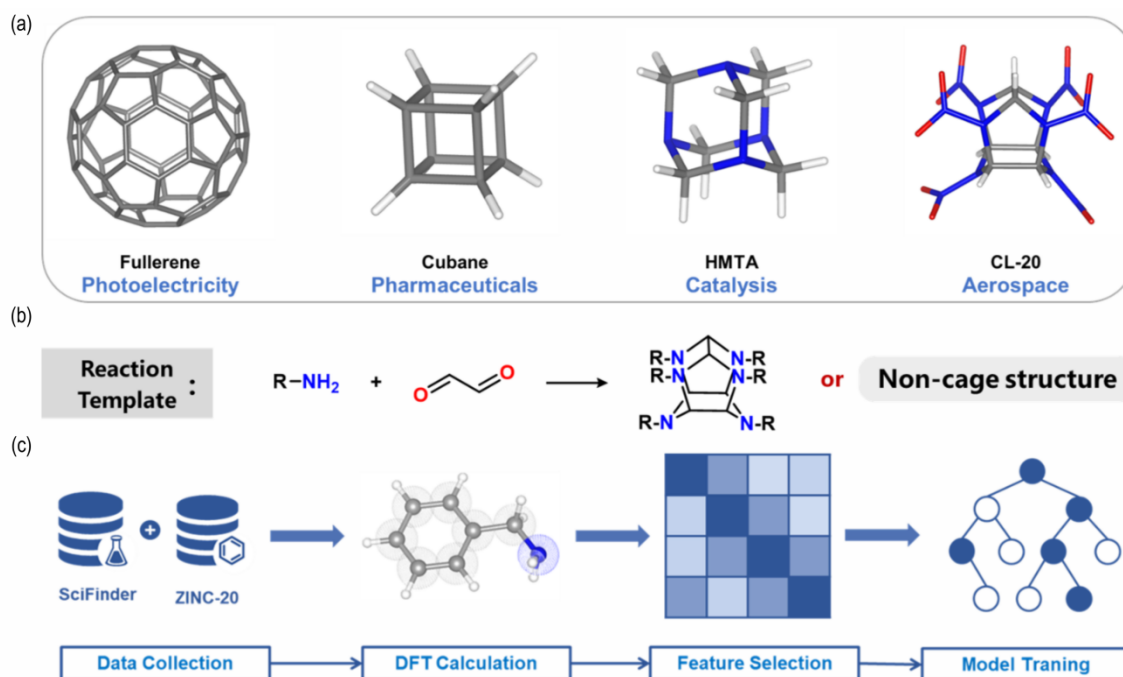


图2 机器学习辅助预测笼状化合物合成<sup>[17]</sup>

Fig.2 Machine learning-assisted synthesis prediction of cage compounds<sup>[17]</sup>

示了成笼反应所依赖的分子层面规律。这一工作展示了机器学习在含能分子反应预测中的应用潜力,但受限于真实反应数据的稀缺,仅使用了91条反应样本进行训练与验证,因此在一定程度上限制了模型的泛化能力和稳健性。未来亟需通过扩大数据量、规范数据收集与表征方法,以构建更为丰富、标准化的反应数据库,从而进一步提升模型在含能材料反应预测中的可靠性。

### 1.1.2 反应产率预测

在实际反应过程中,由于反应不完全或者副反应的发生,往往会导致目标产物产率较低,因此反应产率预测在化学合成中具有重要的实际意义和理论价值。然而随着反应体系复杂性的增加,传统经验或单变量实验优化方法在处理多维反应空间时效率极低,而基于机器学习的数据驱动方法能够在多因素交互作用下实现高维条件空间的快速探索与性能预测<sup>[18-19]</sup>。

近年来,随着高通量实验平台的发展,逐步建立了大规模、高质量的反应产率数据集。2018年,Ahne-man等<sup>[20]</sup>基于自动化实验平台生成了4608条钯催化的Buchwald-Hartwig偶联反应的高通量产率数据集(Buchwald-Hartwig HTE),涵盖不同底物、配体、碱及溶剂条件的组合,并通过DFT计算获取了120个分子描述符作为输入特征,包括HOMO-LUMO、原子电荷、偶极矩以及键长键角等信息,训练了多种机器学习模型来预测产率。其中,随机森林表现最佳,在随机划分下 $R^2$ 达到0.92。Perera等<sup>[21]</sup>利用连续流自动化合成平台,在高温条件下对Suzuki-Miyaura偶联反应的参数空间进行纳摩尔尺度的高通量探索,构建了包含

5760个数据点的高通量数据集(Suzuki-Miyaura HTE),系统遍历了不同芳基卤化物、硼酸酯、催化剂/配体组合、碱类型及溶剂体系等关键反应组成。同时,连续流微通道反应器可对反应温度、停留时间、流速、试剂摩尔比以及催化体系组成等工艺变量进行精确调控,从而实现了传统批式体系中难以达到的高参数密度与可重复性。这两个数据集目前已成为反应产率预测模型常用的重要基准,为后续产率预测模型的性能评估提供了有力支撑<sup>[22]</sup>。

随着模型结构与分子表示方法的不断发展,研究者进一步引入了图神经网络、Transformer等深度学习框架,以实现复杂分子体系中电子效应与空间效应的捕捉。例如Hong等<sup>[23]</sup>提出了空间和电子嵌入分子图(Spatial and Electronic Molecular Graph, SEMG)模型,通过在传统分子图的节点与边特征中引入来源于量子化学的电子结构信息(如原子电荷分布、轨道能量等),同时结合外部化学知识(如电子诱导效应、共轭结构、空间位阻参数等)来提高模型对局部化学环境的区分能力,从而显著提升了反应产率与对映选择性的预测精度( $R^2=0.969$ ,  $RMSE=4.81\%$ ),如图3所示。除了出色的预测能力外,空间和电子效应的物理意义编码还提供了原子级的化学解释。

相较于常规有机反应体系,含能材料合成反应在产率预测中面临更高的复杂性与不确定性。一方面,含能分子通常包含高能官能团(如 $-NO_2$ 、 $-N_3$ 、 $-NHNO_2$ 等),反应路径多样且伴随副反应或分解风险,使得实验产率受反应条件的非线性影响显著;另一方面,该领

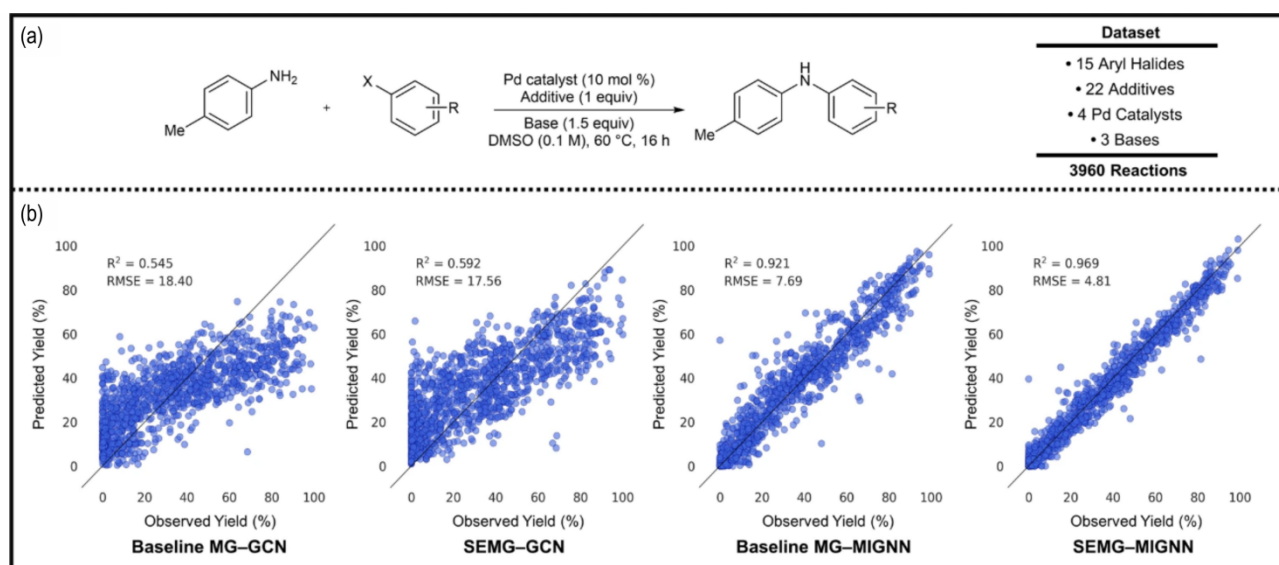


图3 SEMG-MIGNN模型预测反应产率<sup>[23]</sup>

Fig.3 Prediction of reaction yield using the SEMG-MIGNN model<sup>[23]</sup>

域的实验数据极为有限,模型难以通过传统的监督学习获得稳定泛化性能。陈锐等<sup>[24]</sup>以环四亚甲基四硝胺(HMX)为对象,基于实验工艺参数构建了支持向量回归(SVR)、随机森林(RF)和极限梯度提升树(XGBoost)等多种机器学习模型,实现了对HMX合成产率的定量预测。

### 1.1.3 反应条件优化

在准确预测反应性能的基础上,可以利用预测模型指导实验设计实现更高效的反应条件优化。通过精确调节催化剂、溶剂、试剂及反应温度等条件,使得反应的效率和可重复性得到显著改善<sup>[25-26]</sup>。

Shields等<sup>[27]</sup>在高通量钯催化的芳基化反应基准数据集上,系统评估了贝叶斯优化(Bayesian optimization, BO)与实验人员在反应条件优化任务中的表

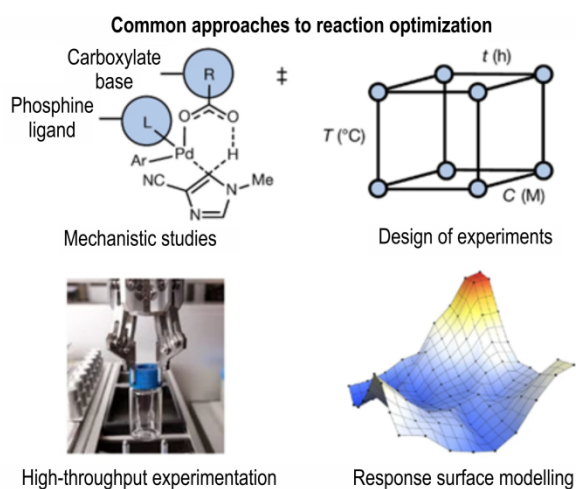


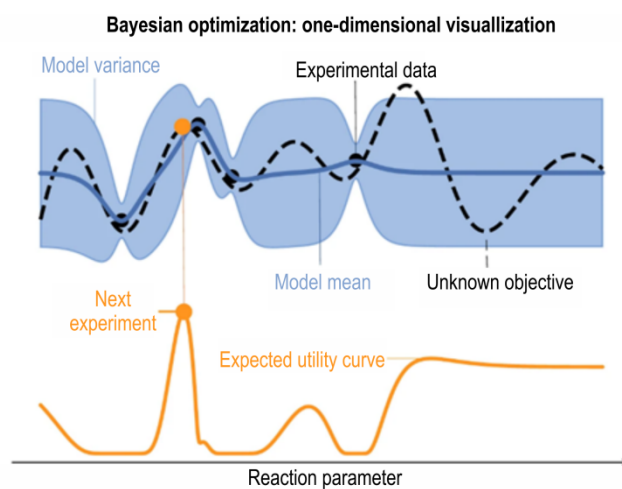
图4 贝叶斯优化反应条件策略<sup>[27]</sup>

Fig.4 Bayesian optimization strategy for reaction conditions<sup>[27]</sup>

## 1.2 逆合成预测

逆合成的概念最初由Corey在20世纪60年代提出,用来描述通过断键将一个复杂的目标分子还原为一个简单前体的迭代过程,即从产物出发,搜索可能的前体,最终找到可商购获得的反应物<sup>[29]</sup>。研究人员几十年来一直在寻求基于计算机辅助合成规划(Computer-assisted synthesis planning, CASP)的高效和准确的方法<sup>[30-32]</sup>。最初的逆合成预测主要依赖于经验丰富的化学家的已有知识或经验,缺乏系统的方法,所以难以满足一直不断快速增长的新有机化学反应的需求。近年来,得益于快速发展的数据驱动模型以及大型反应数据库,化学家和计算机科学家在逆合成预测方面取得了显著的进步。现有的逆合成预测方法可以分为三类:基于模板的方法,无模板的方法和半模板的方法<sup>[33]</sup>。

现,如图4所示,研究中以配体、溶剂、碱、温度等多维反应条件作为搜索空间,以产率作为目标函数,通过高斯过程作为代理模型并结合基于期望提升的采样策略进行优化。结果表明,BO在优化效率和最终反应产率方面均优于50位化学专家的经验决策,并在Mitsunobu反应和脱氧氟化反应中进一步验证了其普遍适用性。Denmark等<sup>[28]</sup>开发了一种为钯催化C—N偶联反应选择底物适应性条件的工具,为特定的反应物预测合适的配体、溶剂、碱及溶剂组合。通过系统的实验设计过程,神经网络模型主动学习了大量的C—N偶联反应,从而建立起底物结构、电子特征与最适合的催化体系之间的映射关系。模型在实验验证中表现出良好的性能,在一系列偶联反应中以超过85%的收率分离出了10种产物。



### 1.2.1 基于模板的方法

基于模板的逆合成方法依赖于预定义的反应模板或规则,通常可以由专家整理或者从已有的化学反应数据库中自动提取来生成模板,然后将反应规则与目标分子匹配以产生一个或多个候选前体<sup>[34]</sup>。这些模板通过识别分子中潜在的断键位点,指导如何将目标分子拆解为简单的前体分子<sup>[35-37]</sup>。其优点在于,使用明确的化学知识进行指导,具有较高的准确性和可解释性。然而,它们依赖于已有的反应模板库,对新颖或未被充分记录的反应类型预测能力有限。Chen等<sup>[38]</sup>将逆合成预测问题转化为局部模板的匹配和应用问题,提出了一种基于图神经网络的局部逆合成框架LocalRetro。通过仅关注断键点附近的原子和分子局部结构,模型能够更好地处理复杂分子,避免了全局模板可能无法处理的复杂化学环境。当反应类别已知

时,模型在 USPTO-50K 数据集上的 top-1 和 top-5 准确率达到了 63.9% 和 92.4%。

对于含能分子而言,模板法的局限尤为突出。因为许多含能分子反应路径较少见,在已有的有机合成数据库中几乎缺乏相关模板。比如从常见芳香胺到多硝基芳香化合物的转化,或高氮杂环骨架的构建,目前的模板库很难覆盖。针对这一问题, Fortunato 等<sup>[39]</sup>提出了一种“模板适用性预测”的机器学习方法,用于提升稀有反应模板在含能体系中的识别能力。该模型以 USPTO 数据库为基础,自动提取了约 25 万个反应模板,并通过子图同构算法进行模板匹配,构建了包含稀有反应类型的增强训练集。模型首先利用神经网络预测目标分子的最可能反应模板,随后进一步评估各模板的适用性。结果表明,该方法在异环成环和硝化等含能典型反应的识别中表现优异,稀有模板的精确率达到 0.83。未来,若要将模板法应用于含能材料合成,亟需建立专门的含能反应模板库,同时结合安全性参数(如爆热、撞感),以确保推荐的路线不仅“可行”,而且“安全”。

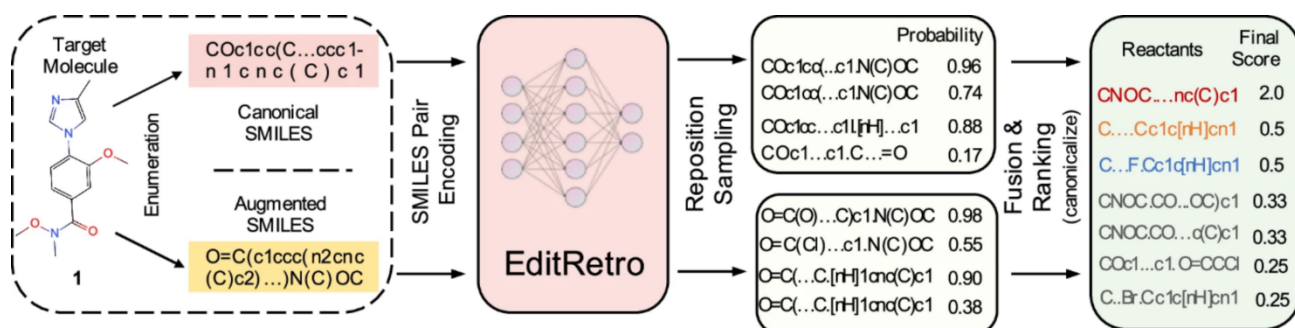


图5 EditRetro 推理模块示意图<sup>[42]</sup>

Fig.5 Schematic diagram of the EditRetro inference module<sup>[42]</sup>

无模板方法对于含能分子尤其有潜力,因为它们能突破传统模板的局限,直接探索复杂且稀有的反应模式。然而,问题在于首先无模板方法依赖大规模数据训练,而含能体系缺乏类似 USPTO 这样的开源数据库;其次无模板模型可能生成在化学上不合理或高风险的路线,如提出极端条件或危险中间体。

### 1.2.3 半基于模板的方法

基于上述两种方法存在各自的优势和局限,近年来开发了基于半模板的自动化逆合成预测方法<sup>[43-45]</sup>。这类方法既不依赖完整的反应模板,也不直接将产物映射回反应物,而是通过多步生成的中间体或合成子来预测最终反应物,可以兼具基于模板方法的可解释性和无模板方法的灵活性,能够有效提高逆合成预测

### 1.2.2 无模板的方法

无模板的逆合成方法不依赖于预先定义的反应模板,而是采用机器学习、深度学习(如图神经网络、Transformer 等)等技术从大量反应数据中自动学习反应模式,直接将产物转化为潜在的反应物。现有的研究工作将逆合成视为神经机器翻译问题<sup>[14,40]</sup>。它们的优势在于灵活性和对新反应的适应能力,能够处理未知的反应类型。然而,缺点在于模型的可解释性较低,且有时可能会提出不合理的合成路径。Zhang 等<sup>[41]</sup>提出了一种基于 Transformer 的端到端逆合成预测模型 Retroformer,通过引入局部注意力该模型可以联合编码分子序列和图,并在局部反应区域和全局反应之间有效地交换信息。在反应类别已知和未知时的 Top-1 准确率分别达到了 64% 和 53.2%。Hou 等<sup>[42]</sup>提出将单步逆合成预测重新定义为分子串编辑任务,如图 5 所示,通过对产物的 SMILES 进行迭代细化逐步生成前体化合物。提出了基于编辑的逆合成模型 EditRetro,该在标准基准数据集 USPTO-50 K 上取得了出色的性能,Top-1 准确率达到 60.8%。

的准确性和适用性。

Wang 等<sup>[46]</sup>提出了一种基于分子组装的深度学习方法 RetroExplainer,将逆合成任务公式化为分子组装过程,并利用可解释的图神经网络将分子组装建模为子图识别、匹配和连接的多任务学习,如图 6 所示。此外,通过模拟化学反应的发生机制,分子组装过程使模型能够提供透明的决策和定量归因,具有良好的可解释性。Zhong<sup>[47]</sup>提出了一种基于图神经网络的端到端逆合成反应预测模型 Graph2Edits,通过将逆合成反应预测表示为通过一系列相互关联的图编辑,从而完成产物-中间体-反应物推导的过程来学习反应转化的规则。在 USPTO-50 K 上的 top-1 准确率(反应类别已知)为 67.1%(表 1)。

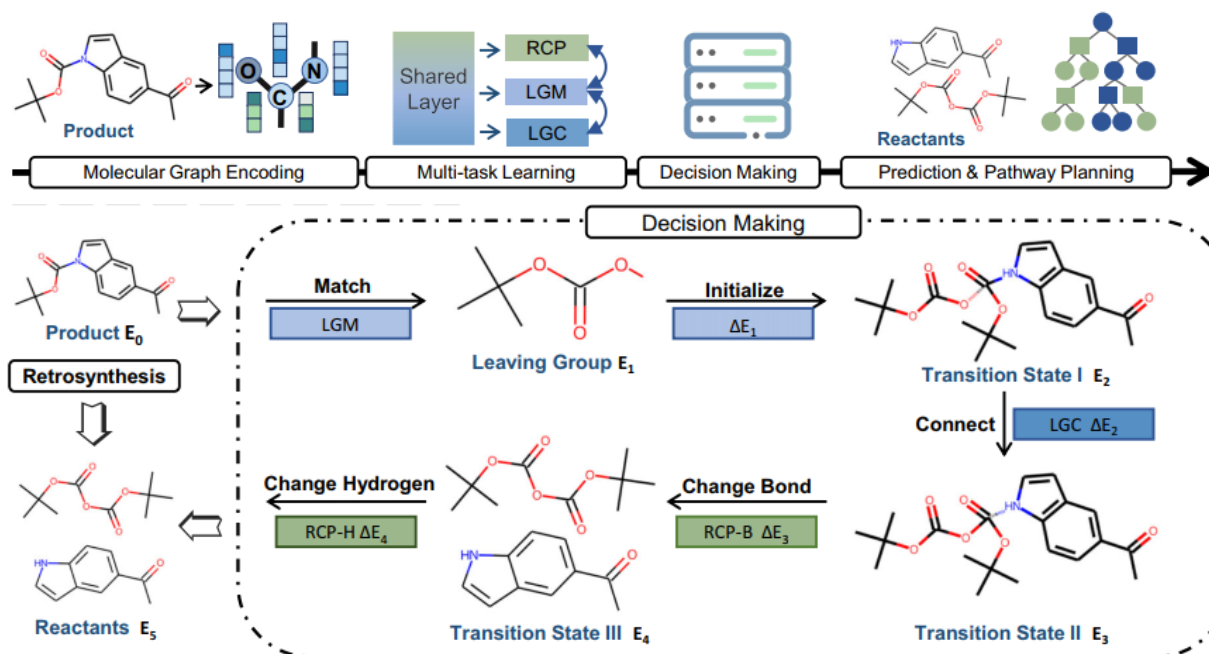
图6 RetroExplainer 预测及决策流程<sup>[46]</sup>Fig.6 Prediction and decision-making process of RetroExplainer<sup>[46]</sup>

表1 部分逆合成预测模型在 USPTO-50k 数据集上的 Top-k 准确率

Table 1 Top-k accuracy of selected retrosynthesis prediction models on the USPTO-50k dataset %

Model	Top-k accuracy							
	Reaction class known				Reaction class unknown			
	1	3	5	10	1	3	5	10
Template-Based Methods								
RetroSim <sup>[48]</sup>	52.9	73.8	81.2	88.1	37.3	54.7	63.3	74.1
NeuralSym <sup>[49]</sup>	55.3	76.0	81.4	85.1	44.4	65.3	72.4	78.9
GLN <sup>[50]</sup>	64.2	79.1	85.2	90.0	52.5	69.0	75.6	83.7
LocalRetro <sup>[38]</sup>	63.9	86.8	92.4	96.3	53.4	77.5	85.9	92.4
Template-Free Methods								
SCROP <sup>[51]</sup>	59.0	74.8	78.1	81.1	43.7	60.0	65.2	68.7
Retroformer <sup>[41]</sup>	64.0	82.5	86.7	90.2	53.2	71.1	76.6	82.1
Dual-TF <sup>[52]</sup>	65.7	81.9	84.7	85.9	53.6	70.7	74.6	77.0
EditRetro <sup>[42]</sup>	60.8	80.6	86.0	90.3	50.0	64.3	68.6	71.1
Semi-Template-Based Methods								
GraphRetro <sup>[33]</sup>	63.9	81.5	85.2	88.1	53.7	68.3	72.2	75.5
RetroPrime <sup>[43]</sup>	64.8	81.6	85.0	86.9	51.4	70.8	74.0	76.1
RetroExplainer <sup>[46]</sup>	66.8	88.0	92.5	95.8	57.7	79.2	84.8	91.4
Graph2Edits <sup>[47]</sup>	67.1	87.5	91.5	93.8	55.1	77.3	83.4	89.4

目前来看半模板方法或许是最适合含能材料的逆合成预测方法。一方面,它能提供明确的中间体和化学路径,使研究人员能评估危险中间体的稳定性和风险;另一方面,具备一定的灵活性,能够处理新颖的含

能分子合成路线。未来的突破点在于如何将半模板方法与风险评估模型相结合,使得推荐路线既可解释,又能动态评估潜在危险。这样才能真正推动含能材料从“经验驱动”走向“智能规划”。

### 1.3 自动化合成

自动化合成平台在加快潜在含能分子的筛选识别过程中具有极高的价值。化学发现本质上是一个耗时且具有高度不确定性的过程,主要依赖于研究者在实验室中大量的人工尝试,尤其是在含能材料等高风险化合物的合成领域,传统方法面临着实验危险性高、重复性差以及操作规范性难以保障等诸多挑战。因此现代机器人系统通过与实时分析技术(如原位傅里叶变换红外光谱 (FTIR)、液相色谱-质谱联用 (LC-MS) 等)集成,现已能够实现全自动的化合物合成与性能表征<sup>[53]</sup>。当前自动化合成主要分为高通量试验 (High-throughput experimentation, HTE) 和自动化学合成系统两大方向<sup>[54]</sup>。其中,高通量实验是数据的“生成器”,利用微流控或微孔板技术在短时间内并行完成成百上千次反应,侧重于大规模数据采集和条件筛选,其核心目的是解决机器学习面临的“数据饥渴”问题。而自动化学合成则是物理层面的“执行者”,利用机械臂或者连续流技术,精准进行合成操作,将化学家从重复劳动中解放出来,其更侧重于全流程的自动化执行与集成,并保证操作的一致性与安全性;在此

基础上,智能合成可视为自动化化学合成的进一步发展,通过将反应预测、合成路线规划与自动化平台深度结合,实现数据驱动的自主化学发现,代表了未来化学合成的终极目标。

### 1.3.1 高通量实验

高通量实验是一种能够在极短时间内对大量反应组合进行并行测试、快速分析与批量记录的前沿实验技术,广泛应用于底物筛选、反应条件优化及催化剂开发等多个领域<sup>[55-57]</sup>。该方法通过实现化学实验的微量化、标准化与自动化,极大地提高了实验效率和数据获取速度,是现代合成化学与材料科学中不可或缺的技术手段。

Dreher等<sup>[55]</sup>将微阵列技术与低分辨质谱(LRMS)分析相结合,开发了一种高通量化学反应筛选平台,可对纳摩尔量级别的Buchwald-Hartwig偶联反应的底物和反应条件进行筛选,其效率可达到每天1536个反应。但是这种基于微孔板的系统反应中只能使用高沸点极性溶剂DMSO,反应温度也仅限于室温,因此一定程度上限制了在其它反应体系中的适用性。目前在含能材料合成反应中尚缺乏专门的高通量实验体系,但针对该领域关键反应类型的高通量实验和反应预测研究已有报道。Kerackian等<sup>[58]</sup>针对芳烃及杂环芳烃(如吡啶、苯并咪唑)的硝化,设计了包含12种硝化剂和7种活化剂的HTE矩阵,共进行了864个反应,值得注意的是,该实验产生了大量通常被忽略的“负样本”,约56%的反应未能生成可定量的产物。利用这864个包含正负结果的数据点构建了预测反应成功与否的二分类模型。另外,通过对不同正负样本比例的训练

集进行分析,作者证实了包含约40%~60%成功率的“平衡数据集”所训练出的模型准确率最高,从而经验性地证明了保留“失败实验数据”对于构建高鲁棒性化学预测模型的关键价值。Zheng等<sup>[59]</sup>利用高通量实验平台针对电化学C-H键氧化反应构建了覆盖582个底物的数据集,并基于此训练了可精确预测底物反应性(准确率>90%)及原子级位点选择性的机器学习模型,从而实现了大规模商业化合物库的虚拟筛选。此外,该工作还将大语言模型(LLM)和贝叶斯优化算法相结合来执行主动学习,在少量迭代内即成功为8种药物分子确定了高产率的最佳反应条件(涵盖催化剂、电解质及浓度组合),显著提升了实验探索效率。

在含能材料研究中,高通量技术尤其适用于对多种前驱体、取代基及合成条件的快速组合测试,通过实验数据的快速积累与分析,可为后续的反应路径建模和分子设计提供关键指导。此外,结合机器学习方法进行数据驱动分析,有助于实现反应趋势的预测与实验设计的优化,形成实验-模型-优化的闭环工作流。

### 1.3.2 自动化化学合成

相较于HTE专注于实验条件的快速筛选和数据获取,自动化化学合成则更侧重于合成全流程的智能化执行,致力于构建一个“从分子设计到实验验证”的端到端合成框架,涵盖了合成路线设计、反应条件筛选、实验验证等多个环节,能够在最小化人工干预的情况下自主完成复杂的多步合成任务。

Coley等<sup>[60]</sup>开发了一个系统整合了合成路线规划,反应条件优化以及自动化合成的实验平台,如图7

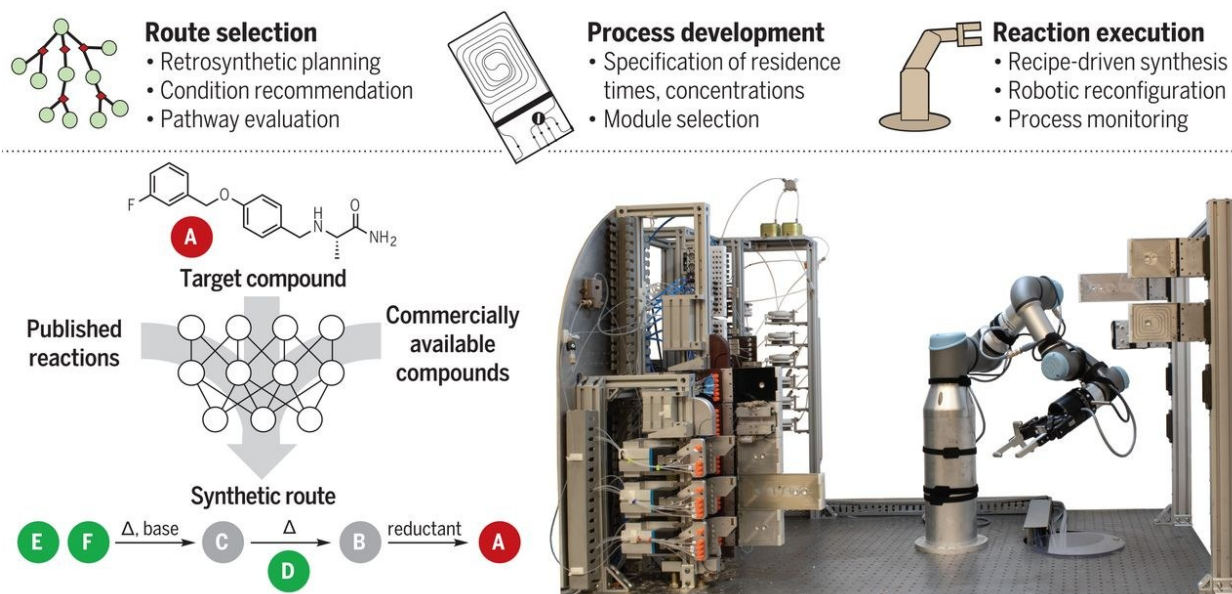


图7 AI规划合成路线和机器人执行的自动化合成平台<sup>[59]</sup>

Fig.7 Automated synthesis platform integrating AI route planning and robotic execution<sup>[59]</sup>

所示。基于USPTO和Reaxys数据库中的数百万反应的训练,能够为给定的目标分子提出多条合成路线,并综合预测产率与反应步骤进行优选,最后通过机械臂执行所有合成操作。成功实现了15个小分子药物的合成路线设计和自动化合成。Wen等<sup>[61]</sup>通过集成现有的商业化硬件、自定义软件和贝叶斯优化算法,开发了一个名为RoboChem的机器人平台,实现了光催化反应的自优化、过程强化和放大生产,通过实时分析与动态参数调整,平台能够在无人干预下自动探索最优反应条件,以氢原子转移机理的光催化烃类活化为模型体系,仅通过19次实验,即在4 h内获得超过90%的NMR收率。

## 2 含能材料合成面临的关键难题

尽管机器学习在药物研发、催化剂筛选等领域已展现出显著优势,但其在含能材料合成中的应用仍面临诸多挑战。这些挑战不仅来源于数据层面的稀缺与质量不足,也深受含能材料自身高能、高敏感特性的制约。总体而言,当前机器学习在含能材料合成研究中面临的挑战主要有以下三个方面。

### 1) 高能反应体系的安全性

相比于其它材料体系,含能材料由于其固有的高能特性,在合成过程中通常涉及到高活性官能团的引入,如硝基、叠氮、高氮杂环等,其反应过程可能伴随显著放热或瞬间能量释放,因此具有更高的安全风险。许多中间体或产物对温度、摩擦、撞击等极为敏感,即便在相对温和的反应条件下,外界的热、光或电等微小刺激也可能引发其剧烈反应甚至爆炸。因此,含能材料体系中往往需要优先满足绝对安全性要求,这在客观上大幅压缩了可探索的反应空间。这种受限条件下的反应行为,叠加到本身具有高能特性的化学体系,使其在动力学和热力学上呈现出高度非线性响应特征,进而显著放大了预测与控制的难度。

### 2) 数据数量和质量的的双重限制

数据数量不足:含能材料研究属于材料科学中的一个敏感和特殊领域,受到国家安全、军工保密等因素的制约,其实验数据的公开程度显著低于常规有机化学领域。现有公开文献中,与含能材料相关的合成路线、反应条件、产率及物化性质等数据总体数量有限,且集中于少数经典体系,导致数据分布不均衡、反应空间覆盖不足。另外,现有的公开文献绝大多数仅记录了成功的合成案例,而极少包含反应失败或发生安全

事故的负样本,使模型难以学习成功与失败的决策边界,限制了其泛化能力和安全风险预测能力。

数据质量低下:即使是已获取的实验数据,也常常存在记录信息不完整、结构格式不一致的问题。例如,反应条件描述不规范(如温度范围模糊)、实验操作细节缺失、文献中同一反应条件下实验结果差异显著等,这些都会对模型输入产生干扰。此外,实验设备和环境条件的微小差异,也可能造成数据波动甚至系统偏差。因此,高质量的数据采集和标准化的数据处理是实现有效建模的前提,也是当前该领域面临的核心瓶颈之一。

### 3) 实验验证与模型迭代的闭环瓶颈

当前机器学习辅助化学合成的方法多采用离线建模思路,即基于既有数据训练模型并输出预测结果,缺乏模型与实际实验过程的实时数据反馈。在含能材料这一特殊体系中,受限于实验周期和安全风险等因素,模型预测结果往往面临验证滞后性,无法通过快速、规模化的实验反馈来及时修正预测偏差。另外,尽管高通量实验和自动化合成平台已在部分前沿领域得到广泛应用,但由于含能材料的高敏感性和对设备安全性的特殊要求,通用的实验设备难以直接适配,现阶段模型验证仍高度依赖人工实验,反馈效率较低,阻碍了“预测——验证——优化”这一数据闭环的有效运行。

## 3 结论与展望

人工智能作为新兴的科学工具,正在深刻重塑科学研究范式,加速了从经验探索到数据驱动发现的转变,为化学研究提供了全新的思维方式和技术路径。本文概述了机器学习在反应预测、合成路线规划以及自动化合成等方面的重要进展,并讨论了其在含能材料领域的应用现状和挑战。尽管机器学习在化学合成上展现出巨大的应用潜力,但受限于实验数据稀缺、反应体系安全约束及模型可迁移性不足等原因,其在含能材料领域的发展仍面临诸多挑战。未来,为推动机器学习在含能材料化学合成中的深入应用,应着重从以下三个方面开展研究:

### 1) 建立开放共享的高质量反应数据库

高质量、可共享的反应数据是构建机器学习模型和提升泛化能力的基础。目前含能材料领域的反应数据量有限,且来源分散、格式不统一,严重制约了模型性能与可迁移性。为解决这一瓶颈,未来需要从数据质量提升和数据数量扩展两个方面来协同推进。数据

质量方面,应制定统一的数据格式和元数据描述,明确反应物结构、产物结构、反应条件、实验步骤、表征方法、产率等关键字段的记录规范,减少由于描述模糊或测量方法差异带来的系统偏差。数据数量方面,应充分利用现有的化学反应数据库,结合化学信息学与自然语言处理(NLP)技术,基于现有已报道的含能分子,通过相似性分析、高能基团指纹及关键反应类型进行定向挖掘。同时整合文献、专利及实验室历史数据,通过结构化处理实现多源异构数据融合,快速填补数据空白。同时,发挥自动化合成和表征平台的优势,通过高通量反应筛选和自动化数据采集,实现大规模、结构化、可重复反应数据的快速生成,从而建立覆盖广泛反应空间、反应信息完备的高质量数据库。

### 2) 发展小样本学习和迁移学习策略

针对当前含能材料反应数据稀缺的问题,单纯依赖实验数据积累难以满足快速研发需求,因此亟需构建适应小样本学习的算法策略。在数据层面,可利用数据增强技术,根据反应机理和化学规则生成合理的虚拟样本,或引入量子化学计算特征(如活化能、原子电荷)作为物理先验知识嵌入模型,提升模型泛化能力;在模型层面,可利用元学习算法优化模型初始化参数,使其在有限样本条件下快速适应新反应类型。此外,还可采用“预训练-微调”的迁移学习范式,先在海量通用有机反应数据库(如USPTO)上训练深度学习模型,使其掌握化学键转化的普适规律,再利用少量的含能材料反应数据进行微调,将通用化学空间中学习到的潜在特征有效映射至高能化学空间。

### 3) 建立多任务融合的端到端智能合成平台

含能材料的研发过程复杂,涵盖分子设计、合成路线规划、实验合成、表征测试与条件优化等多个环节。当前研究多集中于单一阶段或局部任务,尚未形成高效的“设计—验证”闭环体系。未来应致力于构建覆盖全生命周期的端到端智能合成系统,实现模型预测、自动化执行与表征分析的深度协同。这不仅需要在模型层面构建多任务学习框架,对合成路线设计、产率预测、反应条件优化等任务进行联合预测,还需在实验层面上深度集成自动化合成与表征平台,使模型预测结果能够快速转化为实验方案,并将实验数据实时反馈至模型,从而实现理论预测与实验结果的高效闭环迭代。

未来,随着高质量数据库、化学推理模型和智能实验平台的逐步完善,含能材料的设计与合成将从“经验驱动”迈向“数据驱动”的新阶段,实现高效、安全、可持续的智能化发展。

### 参考文献:

- [1] 欧育湘,刘进全. 高能量密度化合物[M]. 国防工业出版社, 2005.  
OU Yu-xiang, LIU Jin-quan. High Energy Density Compounds[M]. National Defense Industry Press, 2005.
- [2] 刘锐,刘建,唐岳川,等. 人工智能辅助含能分子设计的应用与展望[J]. 含能材料, 2024, 32(4): 408-421.  
LIU Rui, LIU Jian, TANG Yue-chuan, et al. Applications and prospects of AI-assisted design of energetic molecules[J]. *Chinese Journal of Energetic Materials (Hanneng Cailiao)*, 2024, 32(4): 408-421.
- [3] HEY T, TANSLEY S, TOLLE K M. The fourth paradigm: data-intensive scientific discovery[M]. Redmond, WA: Microsoft research, 2009.
- [4] KAYALA M A, BALDI P. ReactionPredictor: Prediction of complex chemical reactions at the mechanistic level using machine learning[J]. *Journal of Chemical Information and Modeling*, 52 (2012): 2526-2540.
- [5] DE ALMEIDA A F, MOREIRA R, RODRIGUES T. Synthetic organic chemistry driven by artificial intelligence[J]. *Nature Reviews Chemistry*, 2019, 3(10): 589-604.
- [6] GRANDA J M, DONINA L, DRAGONE V, et al. Controlling an organic synthesis robot with machine learning to search for new reactivity[J]. *Nature*, 2018, 559(7714): 377-381.
- [7] ZHANG P, EUN J, ELKIN M, et al. A neural network model informs the total synthesis of clovane sesquiterpenoids[J]. *Nature Synthesis*, 2023.
- [8] 孔祥泰,张润泽,张玮,等. 人工智能技术在化学反应预测中的应用[J]. 中国现代应用药学, 2022, 39(21): 2856-2864.  
KONG Xiang-tai, ZHANG Run-ze, ZHANG Wei, et al. Application of artificial intelligence in chemical reaction prediction[J]. *Chinese Journal of Modern Applied Pharmacy*, 2022, 39(21): 2856-2864.
- [9] WEI J N, DUVENAUD D, ASPURU-GUZIUK A. Neural networks for the prediction of organic chemistry reactions [J]. *ACS Central Science*, 2016, 2(10): 725-732.
- [10] COLEY C W, BARZILAY R, JAACKOLA T S, et al. Prediction of organic reaction outcomes using machine learning[J]. *ACS Central Science*, 2017, 3(5): 434-443.
- [11] WEININGER D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules[J]. *Journal of Chemical Information and Computer Sciences*, 1988, 28(1): 31-36.
- [12] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2, 2014: 3104-3112.
- [13] SCHWALLER P, LAINO T, GAUDIN T, et al. Molecular transformer: A model for uncertainty-calibrated chemical reaction prediction[J]. *ACS Central Science*, 2019, 5(9): 1572-1583.
- [14] ZHONG Z, SONG J, FENG Z, et al. Root-aligned SMILES: a tight representation for chemical reaction prediction[J]. *Chemical Science*, 2022, 13(31): 9023-9034.
- [15] COLEY C W, JIN W, ROGERS L, et al. A graph-convolutional neural network model for the prediction of chemical reactivity [J]. *Chemical Science*, 2019, 10(2): 370-377.
- [16] CHEN S, JUNG Y. A generalized-template-based graph neural network for accurate organic reactivity prediction [J]. *Nature*

- Machine Intelligence*, 2022, 4(9): 772–780.
- [17] DOU K, ZHAO W, WANG C, et al. Switch on amine substrate reactivity towards hexaazaisowurtzitane cage: Insights from a tailored machine learning model [J]. *Chemical Engineering Journal*, 2024, 501: 157677.
- [18] NIELSEN M K, AHNEMAN D T, RIERA O, et al. Deoxyfluorination with sulfonyl fluorides: Navigating reaction space with machine learning [J]. *Journal of the American Chemical Society*, 2018, 140(15): 5004–5008.
- [19] SCHWALLER P, PROBST D, VAUCHER A C, et al. Mapping the space of chemical reactions using attention-based neural networks [J]. *Nature Machine Intelligence*, 2021, 3(2): 144–152.
- [20] Derek T. Ahneman, Jesús G. Estrada, Shishi Lin, et al. Predicting reaction performance in C–N cross-coupling using machine learning [J]. *Science*, 2018, 360(6385): 186–190.
- [21] PERERA D, TUCKER J W, BRAHMBHATT S, et al. A platform for automated nanomole-scale reaction screening and micromole-scale synthesis in flow [J]. *Science*, 2018, 359(6374): 429–434.
- [22] 张玮, 张屹东, 熊嘉诚, 等. 化学与人工智能的交汇时刻: 综述与展望 [J]. *中国科学: 化学*, 2025, 55(8): 2361–2379.  
ZHANG Wei, ZHANG Yi-dong, XIONG Jia-cheng, et al. The converging moment of chemistry and artificial intelligence: A review and outlook [J]. *Science China Chemistry*, 2025, 55(8): 2361–2379.
- [23] LI S, XU L, ZHANG C, et al. Reaction performance prediction with an extrapolative and interpretable graph model based on chemical knowledge [J]. *Nature Communications*, 2023, 14(1): 3569.
- [24] 秦楷丁, 钱石川, 徐银光, 等. 基于数据驱动的 HMX 合成产率预测研究 [C]// 炸药技术学术年会, 2025.  
QIN Kai-ding, QIAN Shi-chuan, XU Yin-guang, et al. Data-Driven Approach for Predicting HMX Synthesis Yield Academic Conference on Explosives and Propellants Technology, 2025.
- [25] ANGELLO N H, RATHORE V, BEKER W, et al. Closed-loop optimization of general reaction conditions for heteroaryl Suzuki–Miyaura coupling [J]. *Science*, 2022, 378(6618): 399–405.
- [26] GAO H, STRUBLE T J, COLEY C W, et al. Using machine learning to predict suitable conditions for organic reactions [J]. *ACS Central Science*, 2018, 4(11): 1465–1476.
- [27] BENJAMIN J S, JASON S, JUN L, et al. Bayesian reaction optimization as a tool for chemical synthesis [J]. *Nature*, 2021, 590(7844): 89–96.
- [28] RINEHART N I, SAUNTHWAL R K, WELLAUER J, et al. A machine-learning tool to predict substrate-adaptive conditions for Pd-catalyzed C–N couplings [J]. *Science*, 2023, 381(6661): 965–972.
- [29] COREY E J, WIPKE W T. Computer-Assisted Design of Complex Organic Syntheses [J]. *Science*, 1969, 166(3902): 178–192.
- [30] IHLENFELDT W D, GASTEIGER J. Computer-assisted planning of organic syntheses: The second generation of programs [J]. *Angewandte Chemie International Edition*, 1996, 34(23–24): 2613–2633.
- [31] SZYMKUĆ S, GAJEWSKA E P, KLUCZNIK T, et al. Computer-assisted synthetic planning: The end of the beginning [J]. *Angewandte Chemie International Edition*, 2016, 55(20): 5904–5937.
- [32] COREY E J, LONG A K, RUBENSTEIN S D. Computer-assisted analysis in organic synthesis [J]. *Science*, 1985, 228(4698): 408–418.
- [33] SOMNATH V R, BUNNE C, COLEY C W, et al. Learning Graph Models for Retrosynthesis Prediction [C]// *Neural Information Processing Systems*, 2020.
- [34] COLEY C W, WILLIAM H G, KLAUS F J. Machine learning in computer-aided synthesis planning [J]. *Accounts of Chemical Research*, 2018, 51(5): 1281–1289.
- [35] ISHIDA S, TERAYAMA K, KOJIMA R, et al. AI-driven synthetic route design incorporated with retrosynthesis knowledge [J]. *Journal of Chemical Information and Modeling*, 2022, 62(6): 1357–1367.
- [36] SEGLER M H S, WALLER M P. Modelling chemical reasoning to predict and invent reactions [J]. *Chemistry—A European Journal*, 2017, 23(25): 6118–6128.
- [37] HAMILTON W L, YING R, LESKOVEC J. Inductive representation learning on large graphs [C]// *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017: 1025–1035.
- [38] CHEN S, JUNG Y. Deep retrosynthetic reaction prediction using local reactivity and global attention [J]. *JACS Au*, 2021, 1(10): 1612–1620.
- [39] FORTUNATO M, COLEY C W, BARNES B, et al. Machine learned prediction of reaction template applicability for data-driven retrosynthetic predictions of energetic materials [J]. *ChemRxiv*, 2020.
- [40] UMIT V U, ISLAMBEK A, JUNSU K, et al. Retrosynthetic reaction pathway prediction through neural machine translation of atomic environments [J]. *Nature Communications*, 2022, 13(1): 1186.
- [41] WAN Y, HSIEH C-Y, LIAO B, et al. Retroformer: Pushing the Limits of Interpretable End-to-end Retrosynthesis Transformer [C]// *International Conference on Machine Learning*, 2022.
- [42] HAN Y, XU X, HSIEH C-Y, et al. Retrosynthesis prediction with an iterative string editing model [J]. *Nature Communications*, 2024, 15(1): 6404.
- [43] WANG X, LI Y, QIU J, et al. RetroPrime: A diverse, plausible and transformer-based method for single-step retrosynthesis predictions [J]. *Chemical Engineering Journal*, 2021, 420: 129845.
- [44] CHEN Z, AYINDE O R, FUCHS J R, et al. G2Retro as a two-step graph generative models for retrosynthesis prediction [J]. *Communications Chemistry*, 2023, 6(1): 102.
- [45] SACHA M, BŁAŻ M, BYRSKI P, et al. Molecule edit graph attention network: Modeling chemical reactions as sequences of graph edits [J]. *Journal of Chemical Information and Modeling*, 2021, 61(7): 3273–3284.
- [46] WANG Y, PANG C, WANG Y, et al. Retrosynthesis prediction with an interpretable deep-learning framework based on molecular assembly tasks [J]. *Nature Communications*, 2023, 14(1): 6155.
- [47] ZHONG W, YANG Z, CHEN C Y-C. Retrosynthesis prediction using an end-to-end graph generative architecture for molecular graph editing [J]. *Nature Communications*, 2023, 14(1): 3009.

- [48] COLEY C W, ROGERS L, GREEN W H, et al. Computer-Assisted Retrosynthesis Based on Molecular Similarity [J]. *ACS Central Science*, 2017, 3(12): 1237–1245.
- [49] SEGLER M H S, WALLER M P. Neural-Symbolic Machine Learning for Retrosynthesis and Reaction Prediction [J]. *Chemistry—A European Journal*, 2017, 23(25): 5966–5971.
- [50] DAI H, LI C, COLEY C W, et al. Retrosynthesis Prediction with Conditional Graph Logic Network [J]. *arXiv*, 2020, abs/2001.01408.
- [51] ZHENG S, RAO J, ZHANG Z, et al. Predicting retrosynthetic reactions using self-corrected transformer neural networks [J]. *Journal of Chemical Information and Modeling*, 2020, 60(1): 47–55.
- [52] SUN R, DAI H, LI L, et al. Towards understanding retrosynthesis by energy-based models [J]. *Neural Information Processing Systems*, 2021, 34: 10186–10194.
- [53] BURGER B, MAFFETTONE P M, GUSEV V V, et al. A mobile robotic chemist [J]. *Nature*, 2020, 583(7815): 237–241.
- [54] SHEN Y, BOROWSKI J E, HARDY M A, et al. Automation and computer-assisted planning for chemical synthesis [J]. *Nature Reviews Methods Primers*, 2021, 1(1): 23.
- [55] BUITRAGO SANTANILLA A, REGALADO E L, PEREIRA T, et al. Nanomole-scale high-throughput chemistry for the synthesis of complex molecules [J]. *Science*, 2015, 347(6217): 49–53.
- [56] ROBBINS D W, HARTWIG J F. A simple, multidimensional approach to high-throughput discovery of catalytic reactions [J]. *Science*, 2011, 333(6048): 1423–1427.
- [57] COLLINS K D, GENSCHE T, GLORIUS F. Contemporary screening approaches to reaction discovery and development [J]. *Nature Chemistry*, 2014, 6(10): 859–871.
- [58] KERACKIAN T, WESPISER C, DANIEL M, et al. Investigation of arene and heteroarene nitration supported by high-throughput experimentation and machine learning [J]. *Digital Discovery*, 2025, 4(7): 1662–1671.
- [59] ZHENG Z, FLORIT F, JIN B, et al. Integrating machine learning and large language models to advance exploration of electrochemical reactions [J]. *Angewandte Chemie International Edition*, 2025, 64: e202418074.
- [60] COLEY C W, THOMAS D A, LUMMISS J A M, et al. A robotic platform for flow synthesis of organic compounds informed by AI planning [J]. *Science*, 2019, 365(6453): eaax1566.
- [61] SLATTERY A, WEN Z, TENBLAD P, et al. Automated self-optimization, intensification, and scale-up of photocatalysis in flow [J]. *Science*, 2024, 383(6681): eadj1817.

## Machine Learning-Enabled Synthesis of Energetic Materials: Progress and Challenges

DOU Kai-le<sup>1,2</sup>, ZHAO Wei-bo<sup>1</sup>, HE Chun-lin<sup>1</sup>, ZHANG Lei<sup>1,2</sup>, PANG Si-ping<sup>1</sup>

(1. School of Materials Science and Engineering, Beijing Institute of Technology, Beijing 100081, China; 2. State Key Laboratory of Explosion Science and Safety Protection, Beijing 100081, China)

**Abstract:** Energetic materials have attracted significant attention due to their critical roles in national defense, aerospace, and specialized engineering applications. However, their research and development are hindered by high experimental costs, safety risks, and lengthy synthesis cycles, which greatly limit the rapid iteration and practical deployment of novel energetic compounds. In recent years, machine learning (ML) has emerged as a powerful tool in chemistry and materials science owing to its strong capabilities in data modeling and prediction. This review summarizes the latest advances in machine learning – assisted chemical synthesis, focusing on three major aspects: reaction prediction, synthesis route planning, and automated synthesis. Particular emphasis is placed on the potential value and limitations of applying ML techniques to energetic material synthesis. The key challenges—such as data scarcity and inconsistency, lack of safety evaluation frameworks, and limited experimental validation and model retraining—are also discussed. Finally, the review outlines future perspectives, including the establishment of standardized and shareable databases, and the development of high-throughput and automated experimental platforms tailored for energetic systems. This work aims to provide theoretical insights and methodological support for achieving efficient and intelligent synthesis of energetic materials.

**Key words:** energetic materials; machine learning; reaction prediction; automated synthesis

**CLC number:** TJ55; TP181

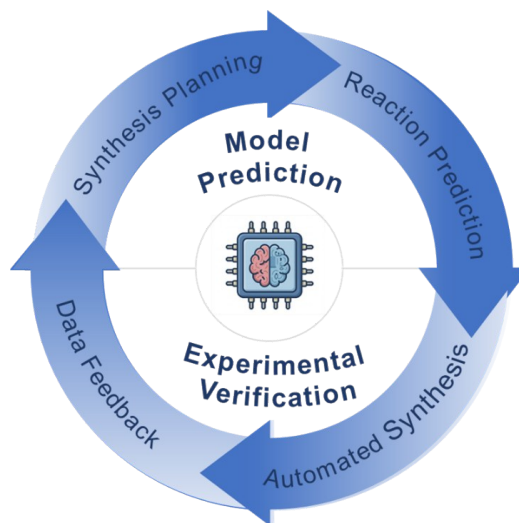
**Document code:** A

**DOI:** 10.11943/CJEM2025237

**Grant support:** National Natural Science Foundation of China (No.12222204, 12472361), Foundation of State Key Laboratory of Explosion Science and Safety Protection (No. QKKT24-01)

(责编:姜梅)

图文摘要:



An intelligent closed-loop workflow synergizes machine learning algorithms with automated experimentation. Synthesis planning and reaction prediction guide automated synthesis, while real-time data feedback continuously refines computational models, enabling a self-optimizing cycle for autonomous and data-driven chemical discovery.