




- 有机无机杂化钙钛矿数据库特征方案
 -  数据库概述
 -  特征体系详细设计
 - 1. 李-辛-商复形特征 (Lie-SympQuotient Complex Transformer)
 - 1.1 0-单纯形特征：原子特征（23维）
 - 1.2 1-单纯形特征：化学键特征（15维）
 - 1.3 2-单纯形特征：三体相互作用特征（18维）
 - 1.4 全局特征：Casimir不变量与结构统计（25维）
 - 2. 有机阳离子图特征【专属特征】
 - 2.1 节点特征：原子特征（11维）
 - 2.2 边特征：化学键特征（8维）
 - 2.3 全局特征：分子特征（16维）
 - 3. 量子化学特征
 - 3.1 基础样本信息（4维）
 - 3.2 B位金属特征（16维）
 - 3.3 X位阴离子特征（15维）
 - 3.4 离子间相互作用特征（6维）
 - 3.5 晶体结构信息（6维）
 - 3.6 结构几何特征（3维）
 - 3.7 全局物理特征（6维）
 - 3.8 工艺参数（5维）
 - 3.9 文本特征（6维）
 - 4. 性能目标特征（9维）
 -  数据库优势
 - 1. 有机阳离子精确建模
 - 2. 杂化界面效应
 - 3. 稳定性预测优势

有机无机杂化钙钛矿数据库特征方案



数据库概述

本数据库专门针对有机无机杂化钙钛矿材料（如 MAPbI_3 、 FAPbI_3 等），整合了四个核心特征设计体系：

- 李-辛-商复形特征 (Lie-SympQuotient Complex Transformer): 基于数学群论的严谨结构特征
- 有机阳离子图特征: 基于图神经网络的分子结构特征 (专属特征)
- 量子化学特征: 基于密度泛函理论的电子结构特征
- 目标性能特征: 实验测量的器件性能指标

总特征维度: 193维



特征体系详细设计

1. 李-辛-商复形特征 (Lie-SympQuotient Complex Transformer)

数学背景说明: 李代数描述连续对称性 (如旋转), 辛代数描述相空间动力学 (如载流子运动), 商代数处理周期性结构的等价关系。这些数学工具能够精确捕捉钙钛矿晶体的对称性、动力学性质和周期性结构特征。

1.1 0-单纯形特征: 原子特征 (23维)

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
atomic_number	原子序数	int	元素表	pymatgen	元素的唯一标识, 决定核电荷数
group_number	族数	int	元素表	pymatgen	价电子构型, 影响化学键合能力
period_number	周期数	int	元素表	pymatgen	电子壳层数, 影响原子半径

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
atomic_mass	原子质量	float	元素表	pymatgen	原子质量，影响晶格振动频率
electronegativity	电负性	float	Pauling	pymatgen	吸引电子能力，决定键的极性
valence_electrons	价电子数	int	电子构型	pymatgen	参与化学键合的电子数
ionization_energy	电离能	float	实验值	pymatgen	失去电子所需能量，影响离子稳定性
electron_affinity	电子亲和能	float	实验值	pymatgen	获得电子释放的能量，影响阴离子形成
oxidation_state	氧化态	int	化学分析	pymatgen	原子的实际电荷状态
covalent_radius	共价半径	float	实验值	pymatgen	共价键长度的一半，预测键长
ionic_radius	离子半径	float	Shannon	pymatgen	离子键长度的组成部分

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
van_der_waals_radius	范德华半径	float	实验值	pymatgen	非键接触距离，影响分子间作用
coordination_number	配位数	int	结构分析	pymatgen	周围配位原子数，决定局部几何
is_metal	金属性	bool	分类	pymatgen	是否为金属，影响导电性
is_transition_metal	过渡金属性	bool	分类	pymatgen	是否为过渡金属，d轨道是否参与
is_lanthanoid	镧系元素性	bool	分类	pymatgen	是否为镧系元素，f轨道效应
tolerance_factor_contrib	容忍因子贡献	float	几何计算	自定义	对Goldschmidt容忍因子的贡献
octahedral_preference	八面体	float	化学规则	化学规则	形成八面体配位的倾向

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
	偏好				
frac_coord_x	分数坐标 x	float	晶体学	pymatgen	晶胞内x方向的相对位置
frac_coord_y	分数坐标 y	float	晶体学	pymatgen	晶胞内y方向的相对位置
frac_coord_z	分数坐标 z	float	晶体学	pymatgen	晶胞内z方向的相对位置
quotient_hash	商等价类散列	int	数学计算	hashlib	周期性结构的等价类标识
avg_site_valence	平均位点价态	float	键价分析	pymatgen	位点的平均键价，反映键合稳定性

1.2 1-单纯形特征：化学键特征（15维）

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
bond_distance	键长	float	结构分析	pymatgen	原子间距离，反映键强度
distance_inverse	距离倒数	float	数学计算	numpy	库仑作用强度，距离越近作用越强
bond_direction_x	键方向x分量	float	几何计算	numpy	键在x方向的单位向量分量
bond_direction_y	键方向y分量	float	几何计算	numpy	键在y方向的单位向量分量
bond_direction_z	键方向z分量	float	几何计算	numpy	键在z方向的单位向量分量
rbf_expansion_1	径向基函数1	float	数学展开	math.exp	短程距离的高斯展开，捕捉近邻效应
rbf_expansion_2	径向基函数2	float	数学展开	math.exp	中程距离的高斯展开
rbf_expansion_3	径向基函数3	float	数学展开	math.exp	长程距离的高斯展开
crosses_boundary	跨越边界	bool	几何判断	pymatgen	键是否跨越晶胞边界
periodic_phase_x	周期相位x	float	相位计算	math	布里渊区中x方向的相位
periodic_phase_y	周期相位y	float	相位计算	math	布里渊区中y方向的相位
wrap_vec_x	周期包装向量x	int	周期性	pymatgen	x方向跨越的晶胞数
wrap_vec_y	周期包装向量y	int	周期性	pymatgen	y方向跨越的晶胞数
wrap_vec_z	周期包装向量z	int	周期性	pymatgen	z方向跨越的晶胞数
lie_bracket_mag	李括号幅值	float	李代数	geomstats	旋转生成元的李括号运算结果

1.3 2-单纯形特征：三体相互作用特征（18维）

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
edge_length_1	三角形边长1	float	几何计算	pymatgen	三体相互作用中的最短边
edge_length_2	三角形边长2	float	几何计算	pymatgen	三体相互作用中的中等边
edge_length_3	三角形边长3	float	几何计算	pymatgen	三体相互作用中的最长边
triangle_area	三角形面积	float	几何计算	trimesh	三体相互作用的几何强度
triangle_perimeter	三角形周长	float	几何计算	numpy	三体相互作用的总几何尺度
shape_factor	形状因子	float	几何计算	自定义	$12\sqrt{3} \times \text{面积} / \text{周长}^2$, 量化三角形形状（等边三角形值为1）
rbf_area_small	小面积径向基函数	float	数学	math.exp	小尺度三体相互作用的敏感性

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
			展开		
rbf_area_medium	中面积径向基函数	float	数学展开	math.exp	中等尺度三体相互作用的敏感性
rbf_area_large	大面积径向基函数	float	数学展开	math.exp	大尺度三体相互作用的敏感性
octahedral_indicator	八面体指示器	float	化学规则	化学规则	是否倾向于形成八面体配位
angle_strain	角度应变	float	几何计算	math.acos	偏离理想键角的程度
coordination_type	配位类型	int	结构分析	pymatgen	配位环境的类型编码
tilt_gen_x	x轴倾斜生成元	float	李代数	geomstats	绕x轴的无穷小旋转生成元
tilt_gen_y	y轴倾斜生成元	float	李代数	geomstats	绕y轴的无穷小旋转生成元
tilt_gen_z	z轴倾斜生成元	float	李代	geomstats	绕z轴的无穷小旋转生成

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
			数		元
casimir_C2	二阶 Casimir 不变量	float	群论	sympy	旋转群的二阶不变量，表征倾斜能量
glazer_cont_param	Glazer连续参数	float	晶体学	numpy	八面体倾斜的连续化描述
mean_bond_angle_variance	平均键角方差	float	统计计算	pymatgen	配位几何的畸变程度

1.4 全局特征：Casimir不变量与结构统计（25维）

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
casimir_2_so3	SO(3)二次 Casimir 不变量	float	群论	sympy	总角动量平方 J^2 ，表征旋转对称性
casimir_2_u1	U(1)二次 Casimir 不变量	float	群论	numpy	电荷平方 Q^2 ，表征电荷守恒
casimir_4_so3	SO(3)四次 Casimir 不变量	float	群论	sympy	四阶角动量不变量，高阶对称性

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
symplectic_casimir	辛 Casimir 不变量	float	辛 几何	sympy	相空间体积守恒，载流子输运稳定性
symplectic_gen_x	x方向辛生成元	float	辛 几何	geomstats	载流子x方向动态演化的生成元
symplectic_gen_y	y方向辛生成元	float	辛 几何	geomstats	载流子y方向动态演化的生成元
symplectic_weighted_casimir	有效质量加权辛 Casimir	float	辛 几何	sympy	有效质量加权的相空间不变量，输运性质代理
casimir_mixed	混合 Casimir 不变量	float	群论	自定义	角动量与电荷的耦合不变量
mean_bond_length	平均键长	float	统计	numpy	整体键长尺度，影响晶格常数
mean_tilt_angle	平均倾斜角	float	统计	numpy	八面体倾斜的平均程度
octahedral_count	八面体数量	int	计数	pymatgen	结构中八面体配位的数量

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
glazer_mode_ratio	Glazer模式占比	float	晶体学	自定义	不同倾斜模式的分布比例
lie_dielectric_casimir	李介电Casimir不变量	float	群论	sympy	介电响应的旋转不变量
symplectic_dielectric_gen	辛介电生成元	float	辛几何	geomstats	介电响应的相空间动态生成元
lie_polarization_casimir	李极化Casimir不变量	float	群论	sympy	极化各向异性的李代数不变量
lie_energy_casimir	李能量Casimir不变量	float	群论	sympy	晶格能的混合不变量
quotient_volume_metric	商体积度量	float	商代数	pymatgen	晶胞体积的周期等价类表示
quotient_density_hash	商密度散列	int	商代数	hashlib	密度分布的周期等价标识
volume_per_fu	每化学式单元体积	float	几何	pymatgen	结构紧密程度，影响密度
packing_fraction	堆积分数	float	几何	pymatgen	空间利用效率
lattice_anisotropy_ratio	晶格各向异性比	float	几何	pymatgen	晶格参数的各向异

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
					性程度
bond_valence_std	键价标准差	float	统计	pymatgen	键价分布的均匀性
quotient_bartel_tau	商Bartel稳定因子	float	商代数	pymatgen	Bartel稳定因子的商空间表示
quotient_tau_prob	商稳定概率	float	商代数	pymatgen	结构稳定概率的周期等价类表示
symplectic_absorption_gen	辛吸收生成元	float	辛几何	geomstats	光吸收过程的相空间动态生成元，光学性质代理

Bartel稳定性判据说明： $\tau < 4.18$ 表示高概率形成钙钛矿结构，稳定概率 $P(\tau) = 1/[1+\exp(0.5(\tau-4.18))]$

2. 有机阳离子图特征【专属特征】

图神经网络背景说明：有机分子可以表示为图结构，其中原子是节点，化学键是边。图神经网络能够学习分子的拓扑结构、化学环境和空间构型，这对于理解有机阳离子与无机框架的相互作用至关重要。

2.1 节点特征：原子特征（11维）

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
node_formal_charge	形式电荷	int	化学规则	RDKit	原子的形式电荷，影响静电相互作用
node_num_hydrogen	氢原子数	int	分子结构	RDKit	连接的氢原子数，影响空间位阻
node_degree	连接度	int	图拓扑	RDKit	连接的原子数，反映配位环境
node_hybridization_orbitals	杂化轨道类型	int	量子化学	RDKit	轨道杂化类型，决定几何构型
node_is_aromatic	芳香性	bool	化学规则	RDKit	是否参与芳香体系，影响电子结构
node_local_env_density	局部环境密度	float	几何计算	自定义	原子周围的拥挤程度
node_partial_charge	部分电荷	float	量子计算	RDKit	原子上的部分电荷分布
node_polarizability	原子极化率	float	查表	查表	原子的极化能力，影响范德华力
node_chirality	手性编码	int	立体化学	RDKit	手性中心的构型，影响对称性
fraction_aromatic_atoms	芳香原子比例	float	统计	RDKit	分子中芳香原子的比例
ring_atom_ratio	环原子比	float	统计	RDKit	分子中环状原子的比例

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
例					

手性编码说明：根据记忆，使用 0 表示非手性，-1 表示 S 构型（左旋），+1 表示 R 构型（右旋）

2.2 边特征：化学键特征（8维）

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
edge_bond_type	键类型	int	化学键	RDKit	单键、双键、三键等键级
edge_is_conjugated	共轭性	bool	化学规则	RDKit	是否参与共轭体系，电子离域
edge_bond_length	键长	float	几何	RDKit	化学键的长度，反映键强度
edge_is_in_ring	环内键	bool	拓扑	RDKit	是否在环结构中，影响分子刚性
edge_ring_membership	环成员特征	int	拓扑	自定义	所属环的大小和类型
edge_connectivity_index	连接性指数	float	拓扑	自定义	键的连接性重要程度
edge_stereo_config	立体构型	int	立体化学	RDKit	键的立体化学构型
rotor_bond_energy	旋转键能	float	分子力学	RDKit	键旋转所需的能量

2.3 全局特征：分子特征（16维）

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
global_molecular_volume	分子体积	float	几何计算	Mordred	分子占据的真实体积
global_mcgowan_volume	McGowan 体积	float	经验公式	Mordred	基于原子贡献的分子体积
global_polarizability	分子极化率	float	量子计算	Mordred	分子的整体极化能力
global_topological_charge	拓扑电荷	float	拓扑计算	Mordred	基于拓扑结构的电荷分布
global_dipole_moment	偶极矩	float	量子计算	RDKit	分子的电偶极矩，表征极性
global_num_rotatable_bonds	可旋转键数	int	拓扑分析	RDKit	分子的柔性程度
global_num_hbd	氢键给体数	int	化学规则	RDKit	可以提供氢键的基团数
global_num_hba	氢键受体数	int	化学	RDKit	可以接受氢键

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
			规则		的基团数
global_molecular_weight	分子量	float	原子质量	RDKit	分子的相对分子质量
global_tpsa	拓扑极性表面积	float	拓扑计算	RDKit	分子表面的极性区域面积
global_logp	脂水分配系数	float	经验公式	RDKit	分子在脂相和水相中的分配倾向
kappa_shape_index1	κ形状指数1	float	拓扑计算	RDKit	分子形状复杂度的一阶描述
kappa_shape_index2	κ形状指数2	float	拓扑计算	RDKit	分子形状复杂度的二阶描述
asphericity	非球形度	float	几何计算	RDKit+numpy	分子偏离球形的程度

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
surface_area_vdw	范德华表面积	float	几何计算	RDKit	基于范德华半径的分子表面积
graph_hotspot_density	图热点密度	float	图分析	RDKit+numpy	芳香区域的极化率密度，电子活跃程度

图热点密度计算：`graph_hotspot_density = sum(node_polarizability where is_aromatic) / num_nodes`

3. 量子化学特征

量子化学背景说明：基于密度泛函理论(DFT)计算和实验测量的电子结构、几何结构和热力学性质，这些特征直接关联材料的基本物理化学性质。

3.1 基础样本信息（4维）

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
sample_id	样本标识符	str	人工编号	-	样本的唯一标识
chemical_formula	化学分子式	str	化学分析	-	样本的化学组成
composition_string	组成描述	str	化学分析	-	详细的组成信息

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
crystal_structure_file	结构文件路径	str	CIF文件	-	晶体结构数据文件的路径

3.2 B位金属特征（16维）

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
b_site_ionic_radius	B位离子半径	float	Shannon半径	查表	B位金属离子的有效半径
b_site_oxidation_state	B位氧化态	int	电荷平衡+BV	分析	B位金属的氧化态
b_site_coordination_number	B位配位数	int	CIF+CrystalNN	分析	B位金属的配位原子数
b_site_d_electron_count	B位d电子数	int	电子构型	查表	B位金属d轨道的电子数
b_site_bader_charge	B位Bader电荷	float	DFT+Bader分析	DFT	基于电子密度分析

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
					的实际电荷
b_site_d_band_center	B位d带中心	float	DFT能带分析	DFT	d电子能带的重心位置，影响催化活性
b_site_d_band_width	B位d带宽度	float	DFT态密度	DFT	d电子能带的宽度，反映d轨道离域程度
b_site_crystal_field_splitting	B位晶体场分裂	float	DFT+配体场	DFT	配体场导致的d轨道能级分裂
b_site_covalency_parameter	B位共价性参数	float	DFT键合分析	DFT	B-X键的共价

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
					性程度
b_site_charge_transfer_energy	B位电荷转移能	float	DFT光谱计算	DFT	电荷转移跃迁所需的能量
b_site_magnetic_moment	B位磁矩	float	DFT磁性计算	DFT	B位原子的磁矩大小
b_site_spin_density	B位自旋密度	float	DFT自旋计算	DFT	B位原子处的自旋密度
b_site_orbital_mixing	B位轨道混合	float	DFT轨道分析	DFT	不同轨道间的混合程度
b_site_electron_localization	B位电子局域化	float	ELF分析	DFT	电子局域化函数值，反映电子局域程度

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
b_site_bond_valence_sum	B位键价和	float	键价分析	分析	键价方法计算的总键价
b_site_effective_coordination	B位有效配位数	float	键价+几何	分析	考虑键强度的有效配位数

3.3 X位阴离子特征（15维）

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
x_site_ionic_radius	X位离子半径	float	Shannon半径	查表	X位阴离子的有效半径
x_site_oxidation_state	X位氧化态	int	电荷平衡	分析	X位阴离子的氧化态

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
x_site_coordination_number	X位配位数	int	CIF+CrystalNN	分析	X位阴离子的配位原子数
x_site_bader_charge	X位Bader电荷	float	DFT+Bader分析	DFT	基于电子密度分析的实际电荷
x_site_mulliken_charge	X位Mulliken电荷	float	DFT布居分析	DFT	基于轨道布居分析的电荷
x_site_p_band_center	X位p带中心	float	DFT能带分析	DFT	p电子能带的重心位置
x_site_p_band_width	X位p带宽度	float	DFT态密度	DFT	p电子能带的宽度
x_site_electron_affinity_eff	X位有效电子亲和能	float	DFT轨道能级	DFT	获得电子的有

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
					效亲和能
x_site_polarizability_tensor	X位极化率张量	float	DFT响应计算	DFT	极化率张量的迹，整体极化能力
x_site_charge_density_min	X位电荷密度最小值	float	DFT电荷分析	DFT	原子周围电荷密度的最小值
x_site_electrostatic_potential	X位静电势	float	DFT静电计算	DFT	原子处的静电势
x_site_bond_order_to_b	X-B键级	float	DFT键合分析	DFT	X位与B位间的键级强度
x_site_covalency_index	X位共价指数	float	DFT键性分析	DFT	参与共价键的程度
x_site_hardness_parameter	X位硬度参数	float	DFT+HSAB理论	DFT	软硬酸碱理论

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
					中的硬度参数
x_site_electron_localization	X位电子局域化	float	ELF分析	DFT	电子局域化函数值

3.4 离子间相互作用特征（6维）

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
a_b_size_ratio	A-B离子半径比	float	rA/rB	计算	A位与B位离子半径比，影响结构稳定性
a_x_size_ratio	A-X离子半径比	float	rA/rX	计算	A位与X位离子半径比
b_x_size_ratio	B-X离子半径比	float	rB/rX	计算	B位与X位离子半径比

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
electronegativity_variance	电负性方差	float	$\sigma^2(\chi_A, \chi_B, \chi_X)$	计算	三种离子电负性差异的方差
hardness_mismatch_factor	硬度失配因子	float	HSAB分析	计算	软硬酸碱匹配程度
ionic_potential_ratio	离子势比值	float	$(q_A/r_A^2)/(q_B/r_B^2)$	计算	不同离子静电势的比值

有机阳离子半径：使用Kieslich有效半径 $r_{A,eff} = (3V_{mol}/4\pi)^{(1/3)}$

3.5 晶体结构信息（6维）

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
lattice_a	晶格参数a	float	CIF文件	晶体学	晶胞在a方向的长度
lattice_b	晶格参数b	float	CIF文件	晶体学	晶胞在b方向的长度
lattice_c	晶格参数c	float	CIF文件	晶体学	晶胞在c方向的长度
lattice_alpha	晶格角 α	float	CIF文件	晶体学	b和c轴之间的夹角
lattice_beta	晶格角 β	float	CIF文件	晶体学	a和c轴之间的夹角
lattice_gamma	晶格角 γ	float	CIF文件	晶体学	a和b轴之间的夹角

3.6 结构几何特征（3维）

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
tolerance_factor	容忍因子t	float	Goldschmidt公式	计算	结构稳定性的几何判据
octahedral_factor	八面体因子μ	float	几何计算	计算	八面体配位的几何适应性
tolerance_oct_interaction	容忍-八面体交互项	float	几何交互稳定性	计算	两个几何因子的耦合效应

容忍因子计算： $t = (r_A + r_X) / [\sqrt{2}(r_B + r_X)]$ ，其中有机阳离子半径使用Kieslich有效半径

3.7 全局物理特征（6维）

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
formation_energy	形成能	float	DFT/MP	DFT	从单质形成化合物的能量变化
decomposition_energy	分解能	float	DFT/MP	DFT	分解为稳定相所需

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
					的能量
energy_above_hull	相图稳定性	float	DFT/MP	DFT	相对于最稳定相的能量差
bulk_modulus	体模量	float	DFT	DFT	材料的体积压缩模量，反映机械性质
electrostatic_potential_mean	平均静电势	float	DFT+Bader	DFT	整体静电势的平均值
charge_volume_proxy	电荷-体积代理	float	计算	Mordred + pymatgen	微观阳离子部分电荷与宏观体积的耦合，代理多尺

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
					度电荷效应

公式说明： $\text{charge_volume_proxy} = \text{avg_cation_charge} * \text{volume_per_fu}$ ，其中 avg_cation_charge 是 A 和 B 位的平均部分电荷（使用 Mordred/RDKit 从 SMILES 计算有机 A 位，pymatgen 查表 B 位）； volume_per_fu 是每化学式单元体积（从 CIF 计算）。

3.8 工艺参数（5维）

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
annealing_temperature	退火温度	float	实验记录	实验	热处理温度，影响晶体质量
annealing_time	退火时间	float	实验记录	实验	热处理时间，影响晶粒生长
solution_concentration	溶液浓度	float	实验记录	实验	前驱体溶液浓度，影响薄膜厚度
spin_speed	旋涂转速	float	实验记录	实验	旋涂工艺转速，影响薄膜均匀性
spin_time	旋涂时间	float	实验记录	实验	旋涂工艺时间，影响薄膜质量

3.9 文本特征（6维）

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
synthesis_method	合成方法	str	文献记录	文本	材料制备的工艺路线

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
solvent_type	溶剂类型	str	实验记录	文本	使用的溶剂种类
additive_type	添加剂类型	str	实验记录	文本	添加的改性剂种类
surface_treatment	表面处理	str	实验记录	文本	表面修饰或处理方法
device_architecture	器件结构	str	器件设计	文本	太阳能电池的器件架构
encapsulation	封装方式	str	器件处理	文本	器件的封装保护方法

4. 性能目标特征（9维）

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
power_conversion_efficiency	光电转换效率	float	自动J-V扫描仪测试（机器人合成后）	测试	太阳能电池的能量转换效率
open_circuit_voltage	开路电压	float	自动电流-电压曲线测试	测试	无外部负载时的最大电压
short_circuit_current	短路电流	float	自动电流-电压曲线测试	测试	短路条件下的最大电流
fill_factor	填充因子	float	自动电流-电压曲线测试	测试	J-V曲线的方形度，反映器件质量
band_gap	带隙	float	DFT计算或自动UV-Vis光谱	计算	价带顶与导带底之

特征名称	中文名称	数据类型	数据来源	计算库	物理意义
			(高通量光谱仪)		间的能量差
thermal_stability_index	热稳定性指数	float	自动TGA/DSC测试或加速老化实验	测试	材料在高温下的稳定性
moisture_stability_index	湿度稳定性指数	float	自动湿度chamber测试	测试	材料在潮湿环境下的稳定性
exciton_binding_energy	激子结合能	float	DFT计算	计算	电子-空穴对的结合能，影响载流子分离
charge_carrier_lifetime	载流子寿命	float	DFT估算或自动TRPL测量	测试	载流子的平均寿命，影响器件性能

数据库优势

1. 有机阳离子精确建模

- 图神经网络特征：35维分子图特征，涵盖节点、边、全局三个层次的完整描述
- 化学精确性：基于RDKit的精确分子描述符，确保化学意义的准确性
- 构象敏感性：捕捉旋转键能和分子柔性，反映有机分子的动态行为

2. 杂化界面效应

- 界面相互作用：有机分子与无机框架的相互作用通过氢键、范德华力等特征量化

- **氢键网络**：氢键给体/受体特征精确描述有机-无机界面的氢键相互作用
- **范德华相互作用**：分子表面积和极化率特征量化弱相互作用

3. 稳定性预测优势

- **分子运动性**：可旋转键数和构象自由度反映有机阳离子的热运动
- **热力学稳定性**：Bartel τ 因子专门针对有机阳离子优化，提高稳定性预测精度
- **降解机制**：有机分子的化学稳定性指标预测材料在使用环境中的降解行为