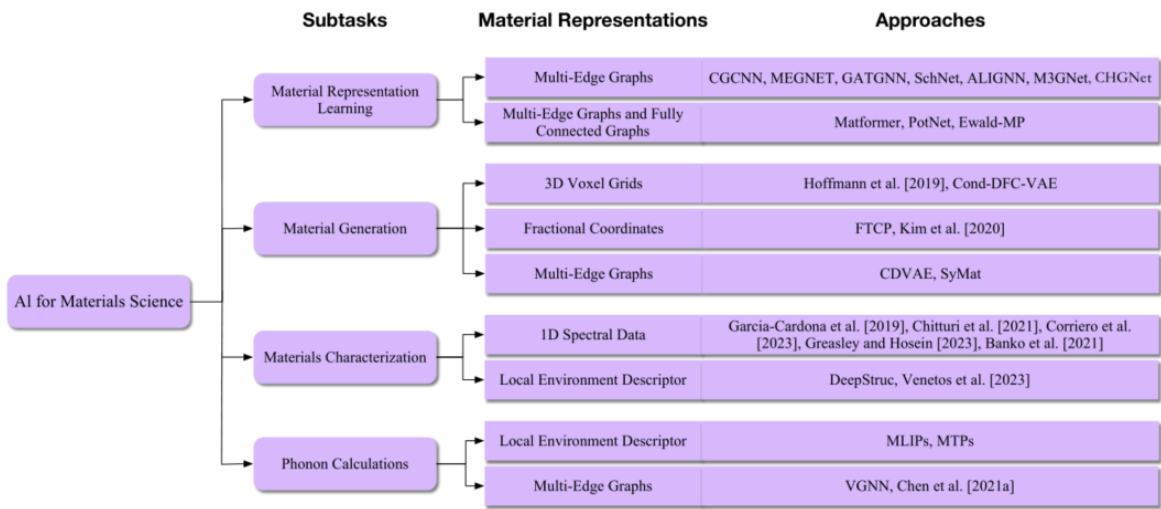


# AI4Materials的计算任务



## 1.材料表示学习

### 1.1 多边形

- 回归任务：MAE, MSE, RMSE, MAD
- 分类任务：ROC, AUC

#### CGCNN

- 晶体
- 对于CGCNN，预测不同的属性就是输入不同的目标值，模型并不改变。
  - 同一个模型和训练过程，输入带隙的ground truth，就能预测带隙；输入形成能的ground truth，就能预测形成能。

#### MP

- 形成能、绝对能量、带隙、费米能、体积模量、剪切模量、泊松比、金属/非金属分类
- 在钙钛矿数据集上做的消融实验

#### Schnet

- 晶体和分子

#### QM9

- $E_{\text{HOMO}}$ 、 $E_{\text{LUMO}}$ 、能隙gap、ZPVE零点振动能量、polarizability极化率 $\alpha$ 、electronic spatial extent  $\langle R^2 \rangle$ 电子空间范围、0K内能 $U_0$ 、内能 $U$ 、焓 $H$ 、定容热容 $C_v$ 、吉布斯自由能 $G$ 、 $\mu$ 电偶极矩。

#### MP

- 形成能、金属/非金属分类

#### MD17、C20富勒烯的分子动力学模拟

- 势能面（总能量）、力场（原子力）

## MEGNet

- 晶体和分子

QM9 Faber等人处理的

- Schnet的QM9指标 +  $w_1$ 最高振动频率。

MP

- 形成能、带隙、体积模量、剪切模量、金属/非金属分类

## GATGNN

- 晶体

MP

- 形成能、绝对能量、带隙、费米能、体积模量、剪切模量、泊松比

## ALIGNN

- 分子和晶体
- 认为高 MAD:MAE ratio 是良好的预测模型。

MP

- 形成能、带隙

JARVIS-DFT

- 形成能、带隙、总能量、凸包上方能量 (ehull)、Vogit体积(Kv)、剪切模量 (Gv)、太阳能电池效率 (光谱有限最大效率, SLME)、拓扑自旋轨道溢出(Spillage)、介电常数( $\epsilon_x$ (DFPT)、 $\epsilon_x$ (OPT, MBJ))、二维剥离能、电场梯度 (EFG)、最大压电应力 ( $e_{ij}$ )、最大应变 ( $d_{ij}$ ) 张量、n型和p型Seebeck系数和功率因数 (PF)、电子 ( $m_e$ ) 和空穴 ( $m_h$ ) 有效质量的晶体学平均值、平面波截止、K点长度、磁矩(magnetic moments)。

QM9

- $E_{HOMO}$ 、 $E_{LUMO}$ 、能隙gap、ZPVE零点振动能量、polarizability极化率 $\alpha$ 、electronic spatial extent  $\langle R^2 \rangle$ 电子空间范围、0K内能 $U_0$ 、内能U、焓H、吉布斯自由能G、 $\mu$ 电偶极矩

## M3GNet

MP && single-element dataset from Zuo et al

- 能、力、应力

DFT计算

- 声子态密度中心数据、声子色散曲线、德拜温度

ICSD

- 松弛结构的能( $E_{hull}$ )

## CHGNet

Materials Project Trajectory Dataset

- 能、力、磁矩、应力

## 1.2 多边图和全连接图

- 回归任务：MAE

## Matformer

MP

- 形成能、带隙、体积模量、剪切模量

Jarvis

- 形成能、带隙、总能量、凸包上方能量Ehull、**训练/推理时间、模型参数量**

## EwT

- **Energy within threshold**: 测量当绝对误差在某个阈值内时可能实际有用的估计能量的百分比

## PotNet

MP

- 形成能、带隙、体积模量、剪切模量

Jarvis

- 形成能、带隙、总能量、凸包上方能量Ehull、**训练/推理时间、模型参数量**

## Ewald-MP

OC20/OE62

- 能、力、每个结构的**前向传播时间、反向传播时间**。

## 2.材料生成

### 2.1 3D体素网格

#### 生成有效的欧氏距离矩阵

QM9的C<sub>7</sub>O<sub>2</sub>H<sub>10</sub>异构体子集

- 生成欧氏距离矩阵并评估不同类型原子间距离分布与样本基础分布的相似度。
- 生成样本中 对应真实样本和新拓扑结构的数量。
- 筛选、分配、叠加后，生成的独立构象中 已知构象、已知结构新构象和新结构新构象的数量。
- 生成的分子、构象的总能量是否合理。

## Cond-DFC-VAE

MP: 78750个通过计算生成的三元钙钛矿、二元合金和Heusler化合物晶体结构

### 材料生成

- 对比VAE生成的晶体结构和DFT结果，根据形成能的绝对差异、键长的平均绝对百分比变化和晶胞参数来分析结果的合理性。

### 属性预测

- 形成能、能量、带隙、体积模量、剪切模量、泊松比、折射率和介电常数

## 2.2 分数坐标

### GAN for Crystal Structure Prediction

Materials Project数据库中的112种Mg-Mn-O组成

- DFT计算生成的Mg-Mn-O组合的 Ehull、普贝稳定性和HSE带隙，以评估材料稳定性。

### FTCP

MP

- 定义生成材料的三种目标属性（形成能、带隙、TE功率因子），根据FTCP模型设计候选材料，DFT弛豫候选材料，根据第一性原理计算这些候选的属性。
- 用三个指标（有效率、成功率和相对随机的改进）评估模型的性能（生成材料的能力）。

## 2.3 多边图

### CDVAE

- 数据集是他们自己整理的

Perov-5、Carbon-24和MP-20

- 重建材料任务的性能：评估生成结构和输入结构间的相似性，RMSE和Match rate。
- 材料生成任务的性能：有效性、属性统计数据Coverage(COV)。
  - 有效性：任一对原子间的最短距离大于0.5埃，则认为有效。
  - COV: COV-R（召回率）和COV-P（精度），用于评估测试集中生成的材料与地面真实材料的集合之间的相似性。
  - 属性统计数据：生成材料和测试材料的属性分布之间的earth mover's distance(EMD)。
- 属性优化任务的性能：成功率(SR)
  - SR5/SR10/SR15分别表示生成材料的预测属性 达到目标属性分布的 5/10/15个百分点的百分比。

### SyMat

Perov-5、Carbon-24和MP-20

- 材料生成任务的性能：有效性、属性统计数据Coverage(COV)。
- 属性优化任务的性能：成功率(SR)

## 3.材料表征

### 3.1 一维光谱数据

- MSE、决定系数
- 分类 accuracy、Top-2 accuracy、Balance accuracy、F1 score、confusion matrix、Precision、Recall

#### 利用中子散射数据预测材料结构(2019)

训练集是由GSAS-II软件生成的每种晶格参数下的衍射图案集，约67GB。测试集是不同温度下的BaTiO<sub>3</sub>相图

- 分类器预测材料的晶体学类别
- 回归器预测材料的晶格参数和键角

#### 用于大型 X 射线衍射数据集中可视化和新颖性检测的深度学习(2021)

15000个一维XRD图谱的合成数据集

- 用VAE在X射线衍射(XRD)数据集上做VAE潜在空间和XRD图谱的可视化

#### 根据X射线粉末衍射图案自主预测晶格参数(2021)

#### CrystalMELA(2023)

POW\_COD数据库中收集的有机、无机和金属-有机化合物和矿物的28万多个已发表晶体结构的模拟粉末X射线衍射图

- 晶体系统分类，将材料分类为七种晶体类别之一（三斜晶系 单斜晶系 斜方晶系 四方晶系 三角晶系 六方晶系 立方晶系）。

#### 探索粉末X射线衍射谱多相识别和定量的监督机器学习(2023)

### 3.2 局部环境描述符

- MAE

#### DeepStruc

ASE构建的3742个独特单金属纳米颗粒结构，Au<sub>144</sub>(p-MBA)<sub>60</sub>纳米颗粒/1.8nm的Pt纳米颗粒/2.2/3.4纳米nm的颗粒/Au<sub>144</sub>(PET)<sub>60</sub>纳米颗粒

- 预测简单的单金属纳米颗粒结构

## EGNN预测NMR化学位移张量

Si NMR数据集 (Sun 等人计算的松弛结构的从头算 NMR 化学位移张量的子集)

- 核磁共振张量参数、各向同性化学位移的预测。

## 4.声子计算

- RMSE、百分比误差、MAE

### 4.1 局部环境描述符

#### MLIPs

跨越晶体结构、化学成分、键合的六种元素系统

- 能、力、立方晶格参数、弹性常数、Voigt-Reuss-Hill体积模量、迁移能、空位形成能、空位扩散的激活能垒。
- 松弛表面和松弛滑移面
- 状态方程的准确性
- 分子动力学轨迹的准确性
- 多晶型物能量差异的准确性

#### MTP

密度微扰泛函理论计算出来一些2D材料性质，作为输入训练MTP

- 几种碳基单元素二维晶格的声子色散关系、声子群速度
- 二元二维系统的声子色散关系、声子群速度
- 三元二维系统的声子色散关系、声子群速度
- 依据声子性质进一步得到的自由能、热容和熵

### 4.2 多边图

#### 用e3nn直接预测声子态密度

High-throughput density-functional perturbation theory phonons for inorganic materials 中报告的1521个晶体固体的声子Dos数据

- 预测声子态密度

MP的4346个晶体结构

- 预测声子态密度

#### VGNN

同上

- $\Gamma$ 点声子能带预测

Atsushi Togo声子数据库

- $\Gamma$ 点声子能带预测