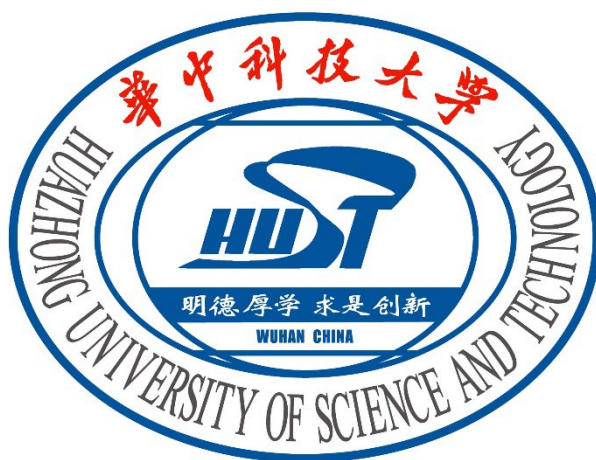


# 《面向催化反应的高通量建模与分析软件（V1.0）》申请说明书



华中科技大学

地址：武汉市洪山区珞瑜路 1037 号

邮编：430074

联系人：单斌

电话：189-7154-7499

## 声 明

华中科技大学材料科学与工程学院微纳材料设计与制造研究中心对面向催化反应的高通量建模与分析软件保留所有权利，在未经华中科技大学材料科学与工程学院微纳材料设计与制造研究中心事先书面许可的情况下，严禁以任何形式复制、传递、分发和存储本文档中的任何内容。

华中科技大学材料科学与工程学院微纳材料设计与制造研究中心会不断对面向催化反应的高通量建模与分析软件进行升级。因此，华中科技大学材料科学与工程学院微纳材料设计与制造研究中心保留在没有预先通知的情况下，对本文档中描述的任何功能进行修改和改进的权利。

本文档的内容按照“现状”提供。除非适用的法律另有规定，否则不对本文档的准确性、可靠性和内容做出任何类型的、明确或默许的保证。华中科技大学材料科学与工程学院微纳材料设计与制造研究中心保留在没有预先通知的情况下随时修订或收回本文档的权利。

## 目录

1 高通量研究方法开发背景 .....	3
2 软件使用指南 .....	4
2.1 运行环境 .....	4
2.2 软件简介 .....	4
2.2.1 基础模块 .....	4
2.2.2 单吸附高通量建模 .....	6
2.2.3 共吸附高通量建模 .....	7
2.2.4 反应过渡态计算过程的自动构建 .....	8
2.2.5 计算结果自动化提取 .....	9
2.2.6 描述符自动提取 .....	10

## 1 高通量研究方法开发背景

目前,传统基于“试错法”的材料研发手段具有“高成本、长周期”的缺点,已无法适应实际需求的更新换代速度。基于大数据的“材料基因组技术”提供了一种超越传统的研发新模式,通过高通量材料计算、高通量材料制备与表征以及材料大数据的融合与协同,开展计算和数据驱动的材料理性设计和构效理论探索研究,为材料创新发展提供了新的契机。因此结合材料数据库、人工智能技术和高通量筛选技术进行特定功能新材料的开发已成为化工与材料科学研究的热点。

在材料基因组思想的带动下,集合高通量技术的材料筛选已经成为催化材料研发的新发展趋势,但目前国内外催化剂研发相关的高通量技术还不成熟,缺乏面向催化反应高通量研究的通用框架,这极大地阻碍了该领域的进一步发展。特别地,典型催化反应过程具有表面中间物种和基元步数目众多,反应网络复杂度高以及基元步数目和网络复杂度随参与物种数增加呈现近指数增加等特点。催化反应的高通量计算要求对催化反应网络的全局枚举,这极大地增大了手动建模与分析的难度。开发面向催化反应的高通量理论计算的自动化研究方法是实现催化反应批量化研究的基础;进一步其与机器学习和跨尺度方法进行耦合研究,也是未来基于材料基因组和大数据技术的材料筛选和设计的关键。

目前国外学者在高通量方法研究领域已经领先与国内。国际上已经开发出了 Atomic Simulation Environment (ASE) 和 pymatgen 等面向材料理论研究的基础工具包,为材料理论设计提供基础的分析工具。斯坦福大学 SUNCAT Center for Interface Science and Catalysis 成功开发了面向催化结构单吸附建模的工具包 (CatKit)。这些基础工具为国内外的催化材料开发者提供了重要基础工具。但是面向催化反应领域仍然缺乏普适的高通量方法框架。我们站在巨人的肩膀上,进一步开发了面向催化反应领域集合建模、分析和信息提取的通用高通量方法框架。该框架主要包含结构建模、结构分析、批量化计算和信息分析提取等。其中自动化建模可以枚举反应网络中的所有可能中间体的可能构型,进一步批量化的结构分析提取每一种结构的典型特征建立唯一结构指纹,结合批量化的计算快速获取对应的能量等信息,通过自动化的信息分析提取实现对关键结构和能量等信息的快速整合,可以为进一步的机器学习分析、跨尺度分析等提供数据集。该框架可以为催化材料的开发者提供统一工具,降低了理论计算中准备工作和后处理的时间成本,极大地加速了材料理论研究和开发。

## 2 软件使用指南

### 2.1 运行环境

系统环境：面向 linux 操作系统，命令行运行实现各模块的调用

开发软件：Python 3.8.3; ASE 3.22.1; CatKit 0.5.4; Numpy 1.20.3 等

### 2.2 软件简介

基于 python 语言开发的面向催化反应过程研究的高通量的建模、计算和分析框架，可以为催化材料高通量设计和筛选提供了关键工具，基本框架如图 1 所示。该软件主要包括基础模块、单吸附高通量建模模块、共吸附高通量建模模块、过渡态过程自动构建模块、计算结果自动化提取和描述符提取等功能模块。基于这些模块可以建立吸附能计算和分析 workflow、基元反应计算和分析 workflow 等，实现催化基元反应过程吸附能和反应势垒的高通量计算和自动化分析；结合反应动力学方法可以对催化反应性能的定量化分析，结合机器学习方法可进一步实现对吸附能和反应势垒等的自动化预测与分析。因此基于该软件建立起适用于复杂催化反应机理的高通量建模、计算和分析框架。接下来对各个模块进行详细介绍。

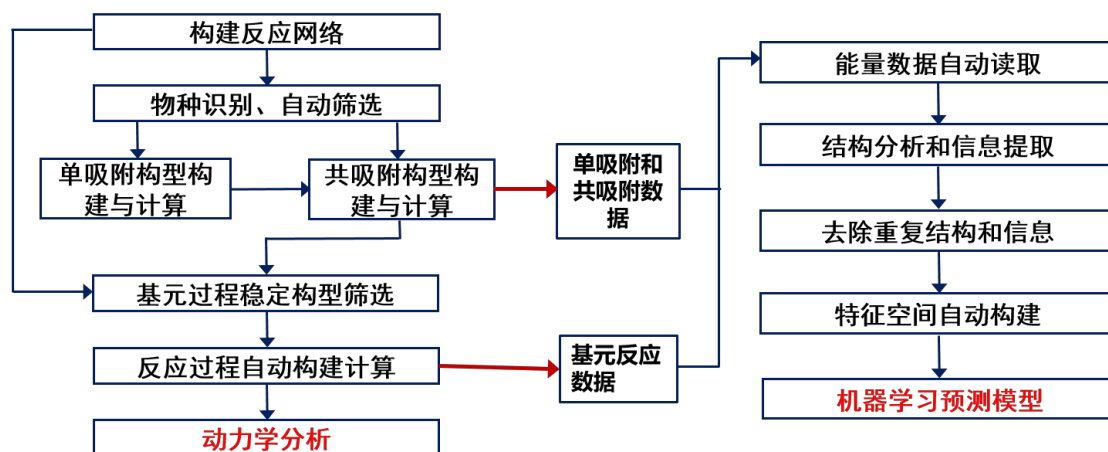


图 1 python 语言的面向催化反应过程研究的高通量的建模、计算和分析框架

#### 2.2.1 基础模块

基础模块主要包含了建模和结构分析等的基础函数，分别包含在 Construct\_model.py 和 Extract\_info.py 两个函数集中。具体函数如下：

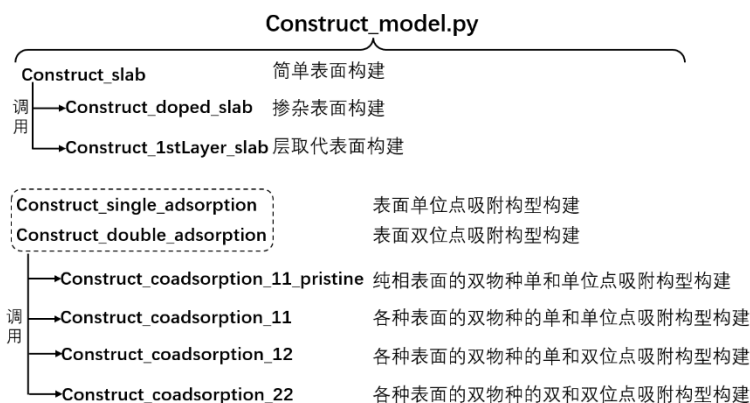


图 2 Construct\_model.py 主要子函数

Construct\_model.py 文件主要包含了表面和吸附构型建模的关键函数，如图 2 所示。其中 Construct\_slab 是面向纯相简单表面构建的基础函数，Construct\_doped\_slab 和 Construct\_1stLayer\_slab 函数则直接调用 Construct\_slab 实现表面特定数目的杂原子掺杂和表面层直接取代模型构建；两个函数支持指定掺杂原子数目和种类实现多样表面模型的订制化建模。另一方面，该文件也包含了在表面模型上进行单位点吸附和双位点吸附模型的构建模块（Construct\_single\_adsorption 和 Construct\_double\_adsorption），基于两模块进一步开发了在纯相和掺杂表面等模型上进行各种共吸附模型的基础函数。具体的应用实例将在 2.2.2 和 2.2.3 具体说明。

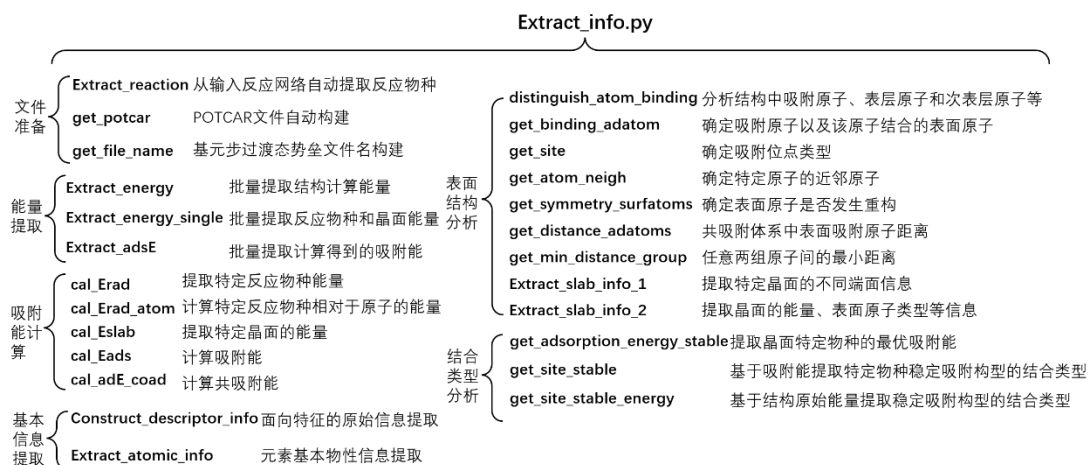


图 3 Extract\_info.py 主要子函数

Extract\_info.py 文件主要是面向高通量计算中的文件准备和后处理，主要包含了文件准备，计算能量提取，吸附能相关计算，表面结构分析，结合类型分析以及针对材料本征物性信息的提取等模块。如图 3 所示。首先 Extract\_reaction 函数可以从输入的催化反应网络中子提取出相关的参与反应物种，为进一步的结构建模提供相关信息。get\_potcar 自动构建 vasp 计算中对应结构所需的 POTCAR 文件；get\_file\_name 面向过渡态计算中基元步的计算文

件夹构建。其次，能量提取模块主要面向计算得到的原始能量的提取；吸附能计算是针对原始能量进行再处理得到吸附能等信息。表面结构分析模块是主要针对晶面基本信息提取，吸附构型吸附原子和表面原子的识别，表面结合类型的判定，原子近邻族信息和结合原子类型信息的提取以吸附原子间距离计算等等；例如 `distinguish_atom_binding` 可以分析原子在吸附构型的角色，如图 4 所示，可以识别出吸附基团、吸附原子、表面和次表面原子等。表面结构分析模块中多个函数的配合使用可实现对吸附构型的关键信息提取，如图 5 左侧是  $\text{NH}_2$  和  $\text{O}$  在  $\text{Pt}(111)$  表面的一种共吸附构型，吸附原子列表信息，吸附物种信息，吸附原子结合类型，吸附原子结合表面原子和吸附原子的距离信息等都可以实现自动化提取。结合类型分析模块则进一步实现最优构型的筛选和对应结构信息的提取。另外，基本信息提取模块可以实现对晶面或吸附物的基本元素信息的提取。以上模块将在随后的功能模块中被多次调用去实现特定功能。

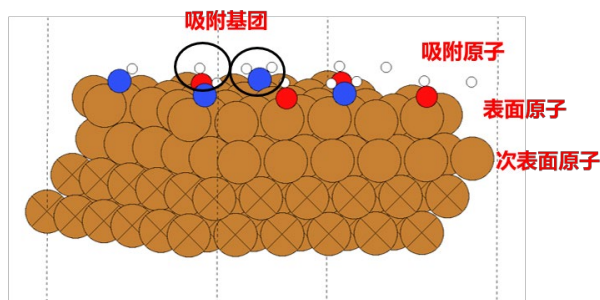


图 4 吸附构型中原子角色分析

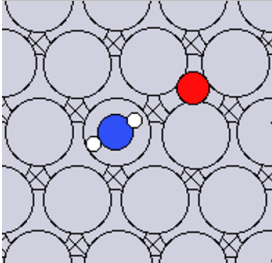
	吸附原子列表信息	N	O
	吸附物种信息	$\text{NH}_2$	O
	吸附原子结合类型	top	hol-hcp
	吸附原子结合表面原子	[Pt]:[24]	[Pt,Pt,Pt]:[39,28,36]
	吸附原子距离信息	N-O:3.352Å	

图 5 表面结构分析模块中实现对吸附构型的关键信息提取

### 2.2.2 单吸附高通量建模

单吸附高通量建模模块通过调用 `Construct_model` 中表面模型和单位点吸附构建模块以及 `Extract_info` 中 `get_binding_adatom` 函数完成对不同表面上不同物种单吸附构型的自动化枚举构建。具体使用流程如下：

(1) 构建输入文件，主要包括 `StrucInfo` 和 `Model` 两个文件，其中 `StrucInfo` 是体相晶格和晶面参数，具体格式如下：

***Pt fcc 3.97 111 #元素 晶格类型 晶格参数 晶面***

其中的四项参数分别是体相元素、对应的晶格类型和晶格参数以及要构建的晶面；

另外通过添加更多的行，可以实现更多体系的连续构建。

另一个输入文件是 Model 文件，主要表面掺杂和吸附建模参数，具体格式如下：

**Cu #掺杂元素**

**0 1 1L 2 3 b1 #类型：0、1、2 和 3 对应 1、2 和 3 个表面掺杂，1L 和 b1 对应层取代和体相部分取代**

**H NH<sub>2</sub> NO N<sub>2</sub> N<sub>2</sub>O O<sub>2</sub> H<sub>2</sub>O #吸附物种**

**1 1 1 2 2 1 #吸附类型：1 和 2 分别对应单位点和双位点吸附**

其中，第一行是掺杂元素，第二行是掺杂类型，0 对应不掺杂即是纯相，1、2 和 3 分别对应表面层 1、2 和 3 个原子掺杂，1L 和 b1 代表表面层取代和体相等比例取代。用户可以根据自己研究需要修改相应参数实现定制化建模。

(2) 直接运行 Construct-adsorption.py 函数可以实现对所有可能构型的自动化批量化枚举构建。目前单计算核心上建模效率达到  $>10^2$  个/min。另外注意：最终输出结构文件为 vasp 格式，命名格式是主相元素\_掺杂元素\_晶面\_掺杂原子数目\_吸附物\_编号.vasp，例如 Pt\_Cu\_111\_1\_NH\_0.vasp，其对应的结构如下：

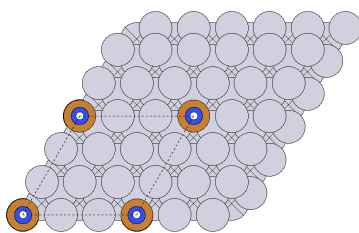


图 6 Pt\_Cu\_111\_1\_NH\_0 对应结构俯视图

另外对于掺杂型表面的吸附模型，我们通过对 get\_binding\_adatom 函数对初步构建构型中吸附物的结合表面原子类型进行判定之后，仅选择吸附物结合原子中有掺杂原子的构型作为最终构型，以此来减少最终输出的结构数量。

### 2.2.3 共吸附高通量建模

共吸附模型构建模块输入文件整体上与单吸附模型构建模块类似，但是由于共吸附模型相较于单吸附模型需要考虑多个吸附物的相互作用，其输入文件格式有所差异。共吸附模型构建模块的两个输入文件依然是 StrucInfo 和 Model，而且 StrucInfo 跟单吸附模型构建模块的此文件格式一样，两种的 Model 文件存在差异。共吸附建模模块的 Model 文件格式如下：

**Cu # 掺杂元素**

**1 1L 2 3 #掺杂类型**

**NH<sub>2</sub> H<sub>2</sub>O;NH H<sub>2</sub>O;N H<sub>2</sub>O;NH<sub>2</sub> H;NH H;N H #共吸附物种**

**1 1;1 1;1 1;1 1;1 1 #共吸附类型**



其中，第一行是掺杂元素，第二行是掺杂类型，第三行和第四行分别是共吸附物种和共吸附类型，不同的共吸附组合以“；”分隔。进一步直接运行 `Construct-Coadsorption.py` 即可实现批量化构建。特别地，通过对单吸附结构和能量的分析 (`get_site_stable`) 可以自动提取每一种物种的单吸附稳定构型，或者手动指定相关的结合类型 (`spec_ads_stable`)，进而实现在共吸附构建中的每一种物种仅使用稳定结合形式。另外考虑到吸附物之间距离的变化导致共吸附模型的不同，这也可能导致大量吸附结构一致但距离不同的相似构型；为了解决这个问题，此函数中通过表面结构的分析引入距离截断参数，在建模中仅考虑在吸附物距离在此截断之内的构型。相似的，共吸附构建模块也引入了 `get_binding_adatom` 函数选择吸附物结合原子中有掺杂原子的构型来减少最终输出的结构数量。目前单计算核心上建模效率可以达到  $>10^2$  个/min。

最终输出结构依然采用 `vasp` 格式，命名方式是主相元素\_掺杂元素\_晶面\_掺杂原子数目\_吸附物 1\_吸附物 2\_编号.vasp，例如 `Pt_Cu_111_2_N_H_0.vasp`，其对应结构如下：

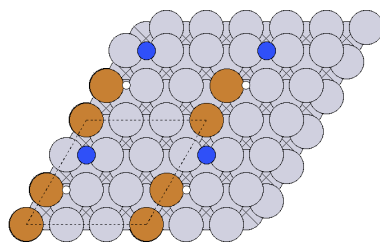


图 7 `Pt_Cu_111_2_N_H_0` 对应结构的俯视图

#### 2.2.4 反应过渡态计算过程的自动构建

催化反应的过渡态计算涉及基元步初态和末态的筛选，中间构型的构建以及计算结果后处理等。因此该模块首先集成了 `Construct_neb.py` 模块，该模块输入参数包括初态和末态信息，插值方式（线性或非线性方法），插入中间态数目等实现中间过程的自动化构建。基于此函数进一步开发了 `Construct_network.py` 功能模块，该模块可以通过读取反应网络文件提取基元步信息，建立对应文件夹，并在每一个文件夹执行过渡态计算的相关流程：根据计算得到的结构和能量文件夹自动提取指定基元步反应物和产物的最稳定吸附结构作为初态和末态，对应能量信息写入日志文件；调用 `Construct_neb.py` 模块生成特定数目中间态；准备相应的计算文件即可高通量计算。具体计算自行选择相应计算软件运行即可，目前开发者主要面向 `vasp` 软件开展相关计算。另外也开发了对应的后处理程序 `post_neb.py`，此函数可以实现计算结果的自动提取、分析和绘图，相关结果写入日志文件以供查询。最终的日志文件主要包含每一个结构原始能量和相对能量基元步势垒和反应热等信息，相关数据可以直接提取用于其他作图软件绘图。如图

8 所示是基于此模块得到的 Pt(100)晶面上吸附的 O 和 H 物种生成吸附 OH 的整个过程的结构和相对能量信息。

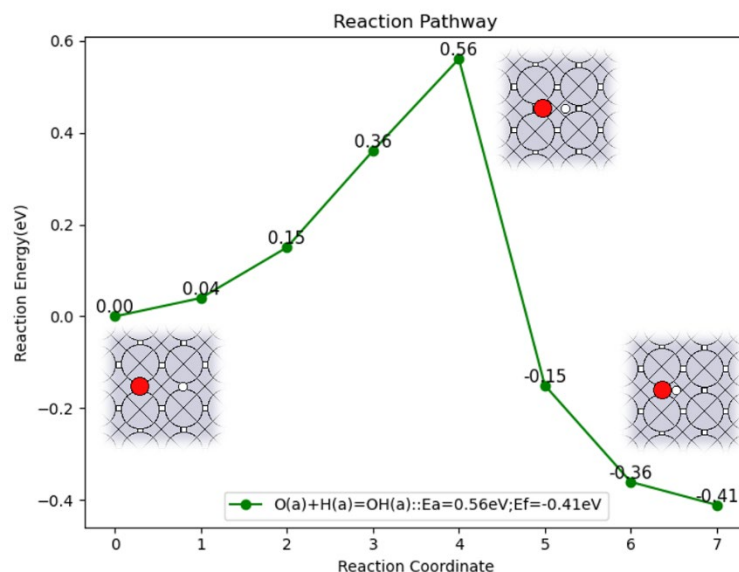
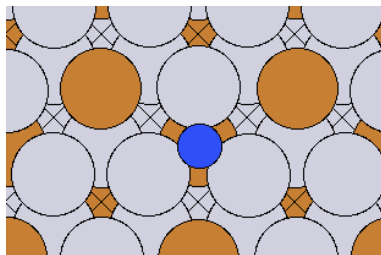


图 8 Pt(100)晶面上吸附的 O 和 H 物种生成吸附 OH 的整个过程的结构和相对能量信息

### 2.2.5 计算结果自动化提取

物种吸附能是催化过程研究关键的能量信息，在此软件中，通过直接调用 Extract\_info 中的 cal\_Eads 等函数即可实现，其中 cal\_Eslab 是提取吸附结构对应的干净表面模型的能量，cal\_Erad 和 cal\_Erad\_atom 是计算反应物种的本征能量，两个分别对应物种的计算能量和以原子能量作为基准的形成能，可以具体需要选择相应函数。这些函数的输入是对应结构的能量信息文件。最后将以上函数的输出作为 cal\_Eads 的输入即可输出对应的吸附能信息，此函数在计算过程中会自动判断表面结构和计算能量是否合理对计算结果进行取舍。最后的计算结果会输出指定的文件中。

进一步在庞大的吸附能数据中根据需要进行自动化的筛选是数据处理的关键。本程序主要开发了提取稳定构型的能量和结合类型等信息的模块，主要包括 Extract\_adener.py, Extract\_adener\_alloy.py 和 Extract\_adener\_site.py 等。其中 Extract\_adener.py 和 Extract\_adener\_alloy.py 可以根据输入的反应物种或掺杂类型信息，直接从吸附能数据集中提取对应吸附结构最稳定的能量信息输出到相应文件，Extract\_adener\_site.py 则可以直接提取对应稳定结构的吸附类型以及结合原子种类等信息。例如对于  $\text{Pt}_3\text{Cu}_1$  合金体系的(111)晶面上 N 原子的稳定吸附构型，如图 9 所示，通过 Extract\_adener\_site.py 可以直接得到 N 原子与表面是 3-fold 结合，其结合表面原子是 3 个 Pt 原子。

图 9 Pt<sub>3</sub>Cu<sub>1</sub> 合金体系的(111)晶面上 N 原子的稳定吸附构型

### 2.2.6 描述符自动提取

催化体系的能量是催化结构的本征信息，也是催化反应过程的最关键信息，其由催化局域结构与吸附物种的相互作用决定。因此基于催化结构和吸附物的本征信息建立面向能量预测的特征空间模板以及自动化构建方法是进行机器学习等数据分析的基础，也是加速催化反应研究的关键。在本模块中提出了一种通用的催化能量信息特征空间的基础框架，并构建了自动化的信息提取流程。该基础框架将特征分为 4 大类：催化体系的表面组分信息，吸附物信息，晶面特征以及结合位点信息，如图 10 所示。对于表面组分信息中，主要选取电负性、原子半径、原子质量等原子本征信息；对吸附物而言，则主要取结合原子电负性、配位原子信息等，晶面主要取表面配位信息，位点信息则是结合位点配位和近邻信息。在合金体系中则通过这些本征信息取平均值等方法扩充信息量。我们该模块集成了 `distinguish_atom_binding` 和 `get_binding_adatom` 等结构分析函数以及 `Construct_descriptor_info` 本征信息提取函数等，通过外部输入所需特征名字实现自动化构建描述符信息。

	表面组分	吸附物	晶面	位点
纯金属	电负性 原子半径 价电子 原子质量 原子序数	电负性 未成对电子 平均电负性 配位原子种类 配位原子数目	表面配位数	位点配位数 近邻
合金	表面平均价电子 次表面平均价电子 表面平均原子半径 次表面平均原子半径	电负性 未成对电子 平均电负性 配位原子种类 配位原子数目	表面配位数	位点配位数 结合原子平均价电子 结合原子平均原子半径

图 10 通用的吸附能特征空间的基础框架

对于每一个结构，通过指定需要的特征名字，可以自动化的构建出对应特征指纹，如图 11 所示，当选择原子半径和价电子作为元素特征信息，可以获取

N 原子桥式吸附合金表面上的 9 个维度的信息数组。通过该模块对所有相关构型的批量化分析，便可自动化构建出所需的面向催化能量的机器学习预测的高维特征空间。



图 11 特定结构与其对应的特征空间