Python与机器学习

——机器学习前沿

华算科技 黄老师

2022年1月21日

华算科技
HUASUAN TECHNOLOGY

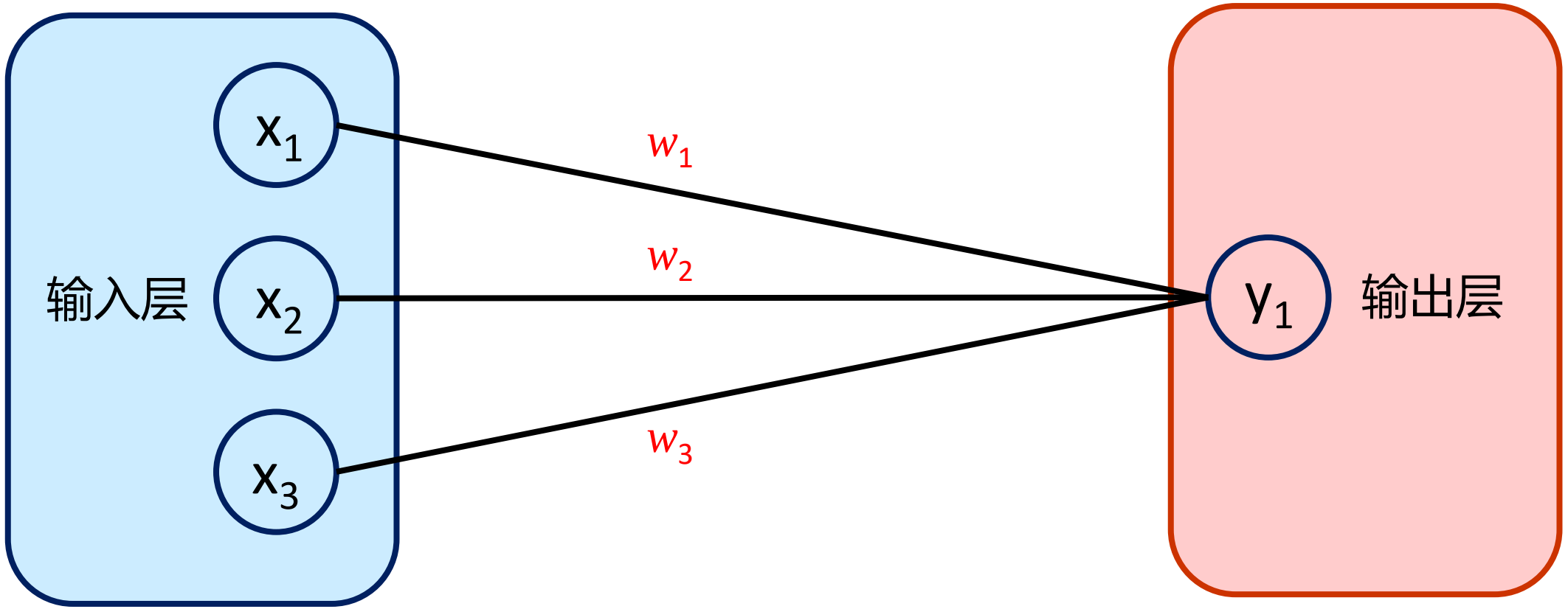
目录

目录

回归算法

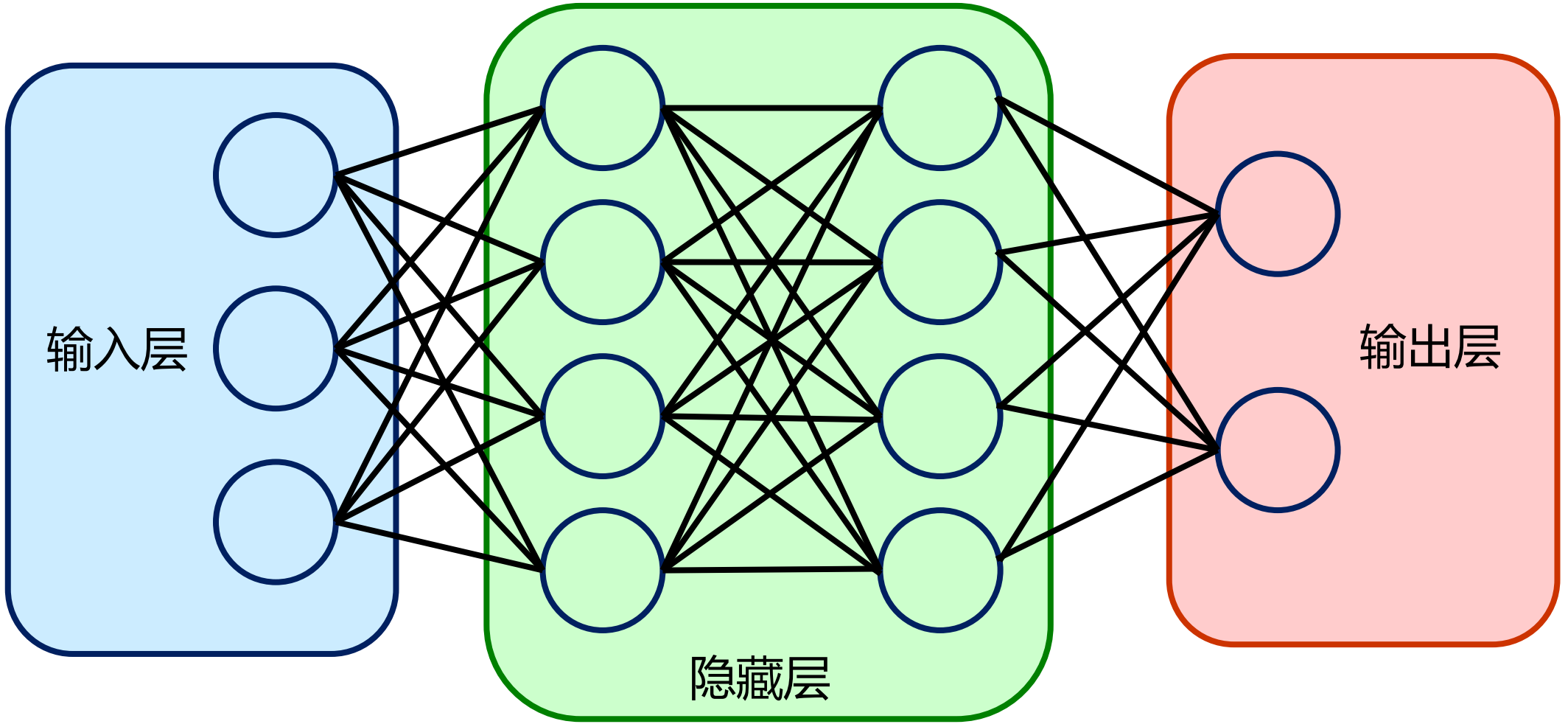
回归算法可以用下图进行简单描述

$$y_1 = \sum w_i x_i + b_1$$

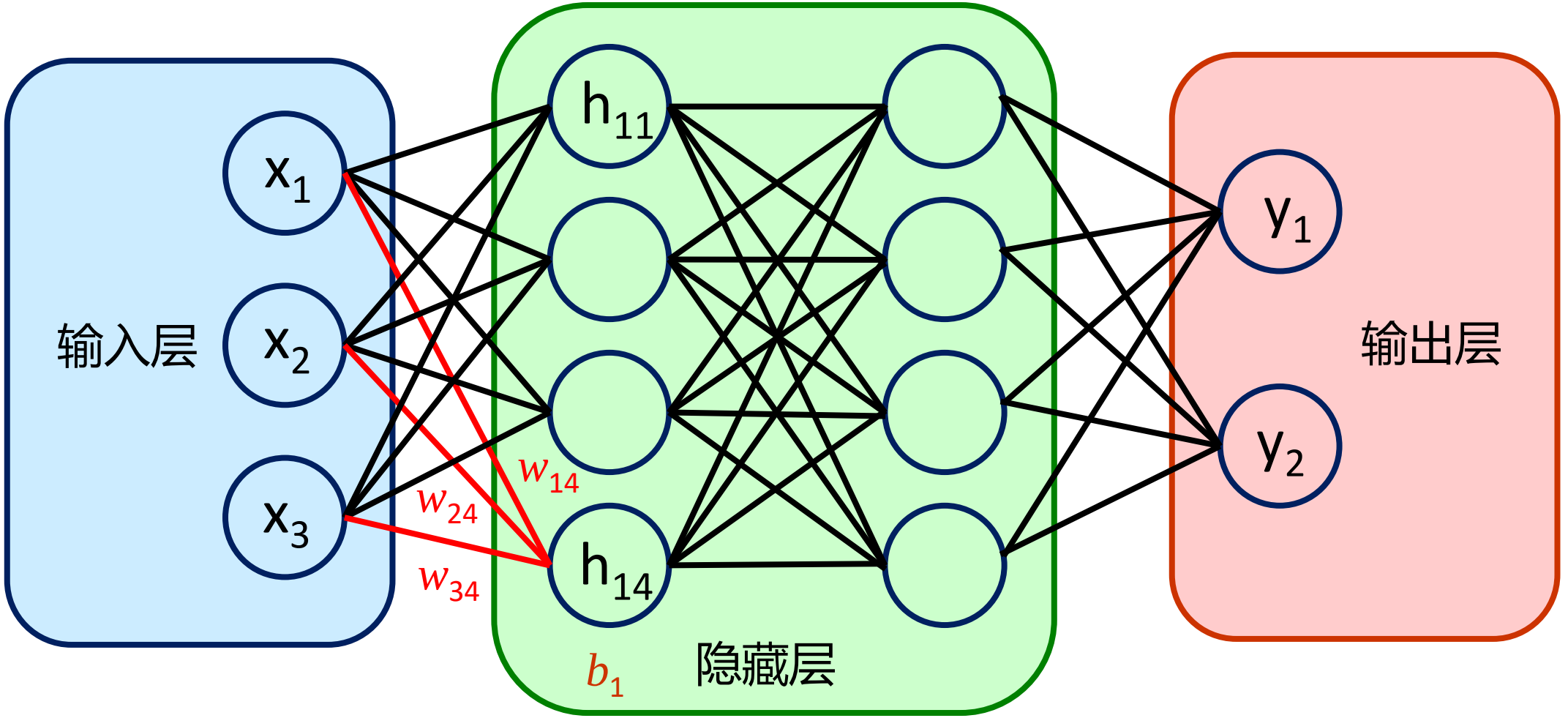


神经网络算法

由许多简单单元组成的广泛并行互连的网络，能够模拟生物神经系统对真实世界物体做出的交互反应

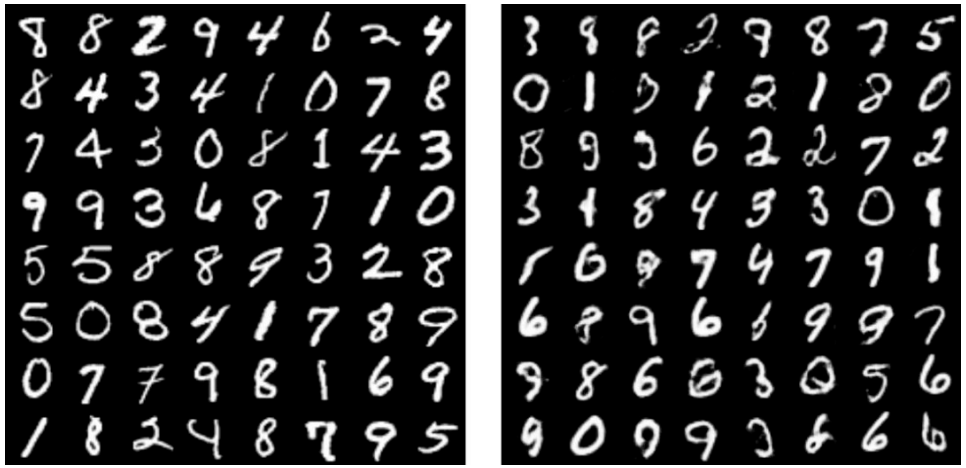


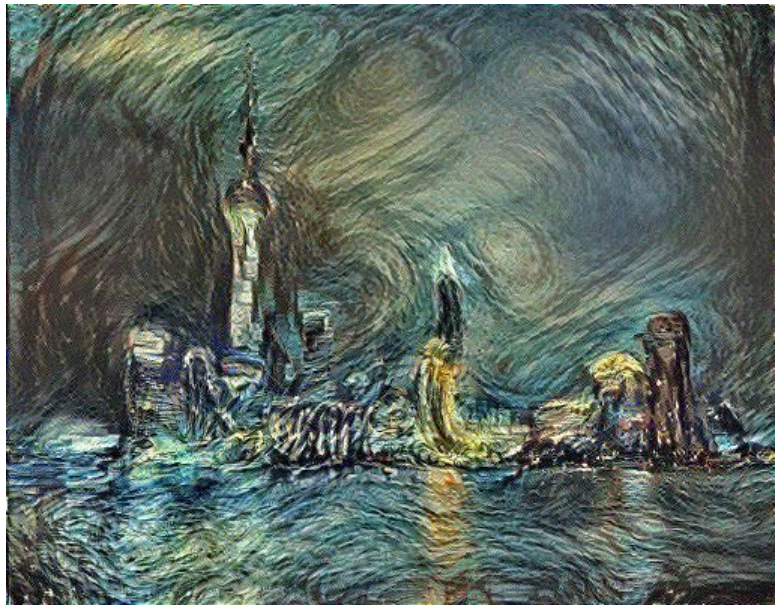
神经网络算法

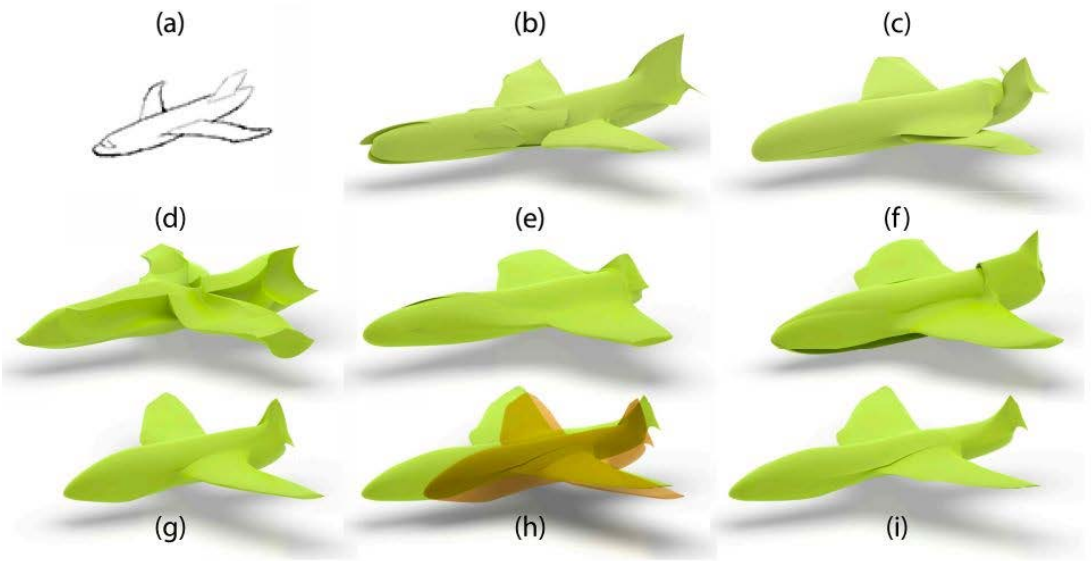


$$h_{14} = f(\sum w_{i4}x_i + b_1)$$ b_1 偏置 f 激活函数

神经网络算法的应用







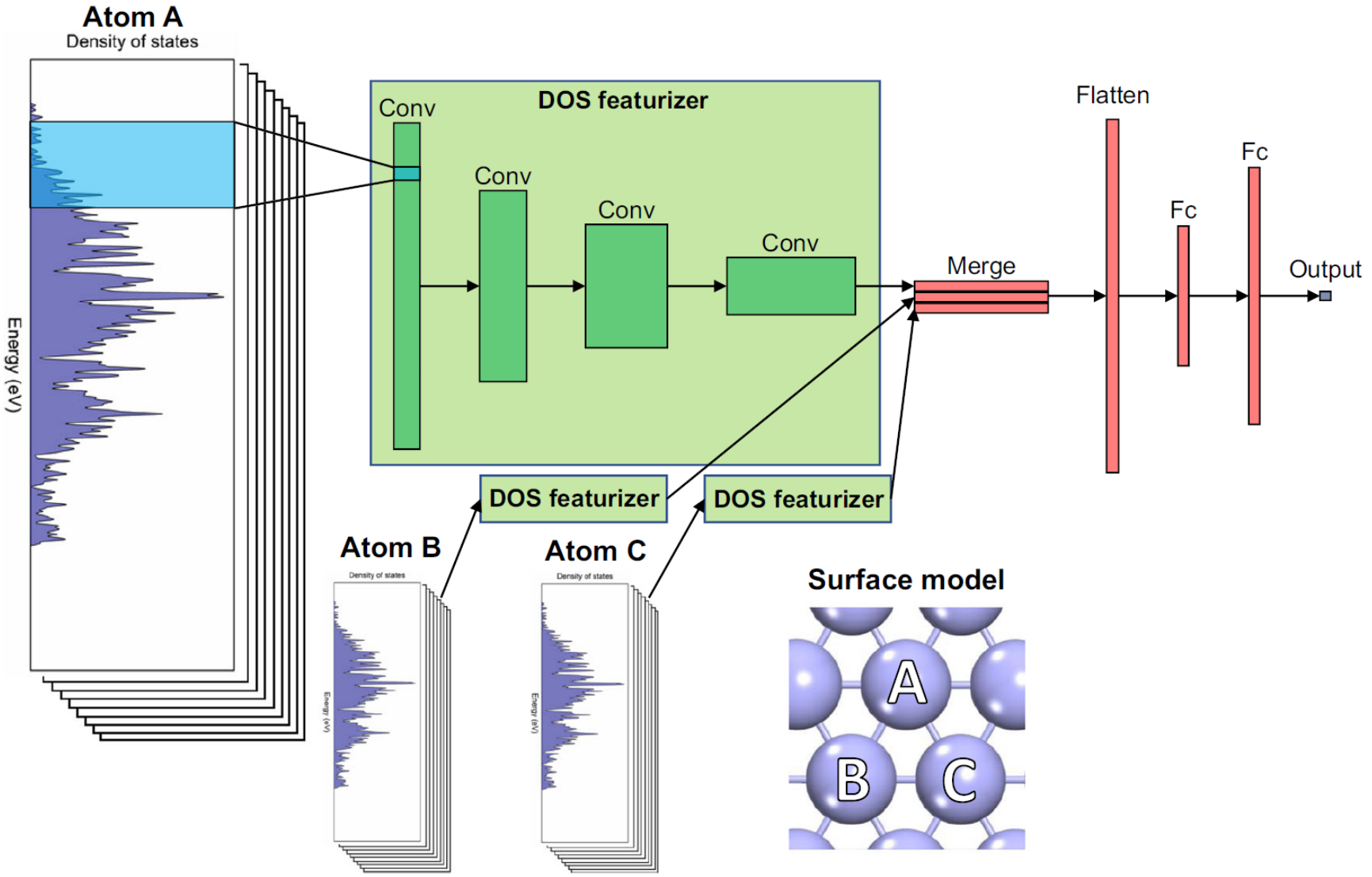
目录

态密度预测吸附能

双金属合金表面吸附小分子、基团

双金属表面：2000个

吸附能数据：37000个

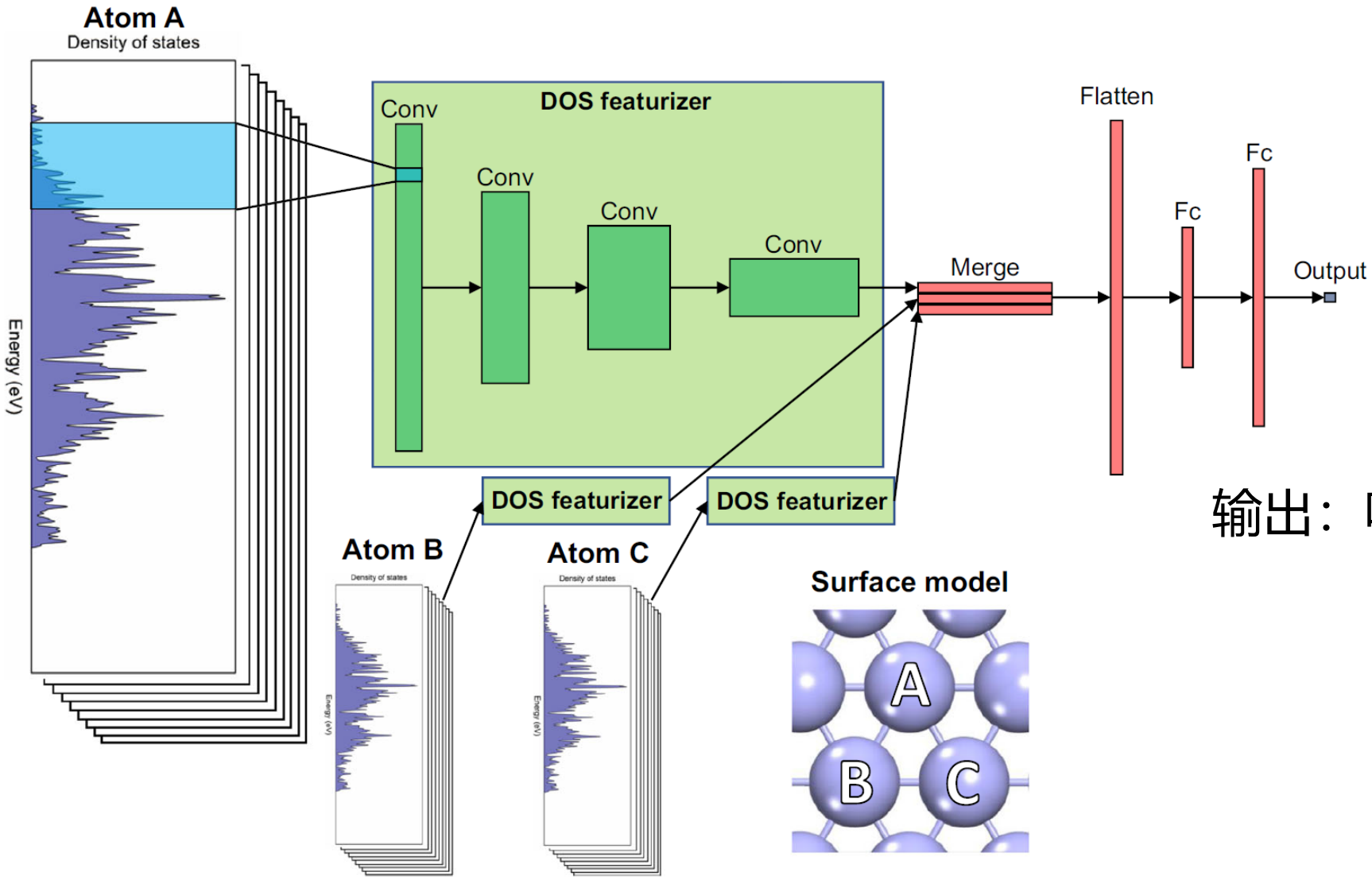


Fung, V., Hu, G., Ganesh, P. and Sumpter, B. G. *Nat. Commun.* **2021**, 12, 88.

态密度预测吸附能

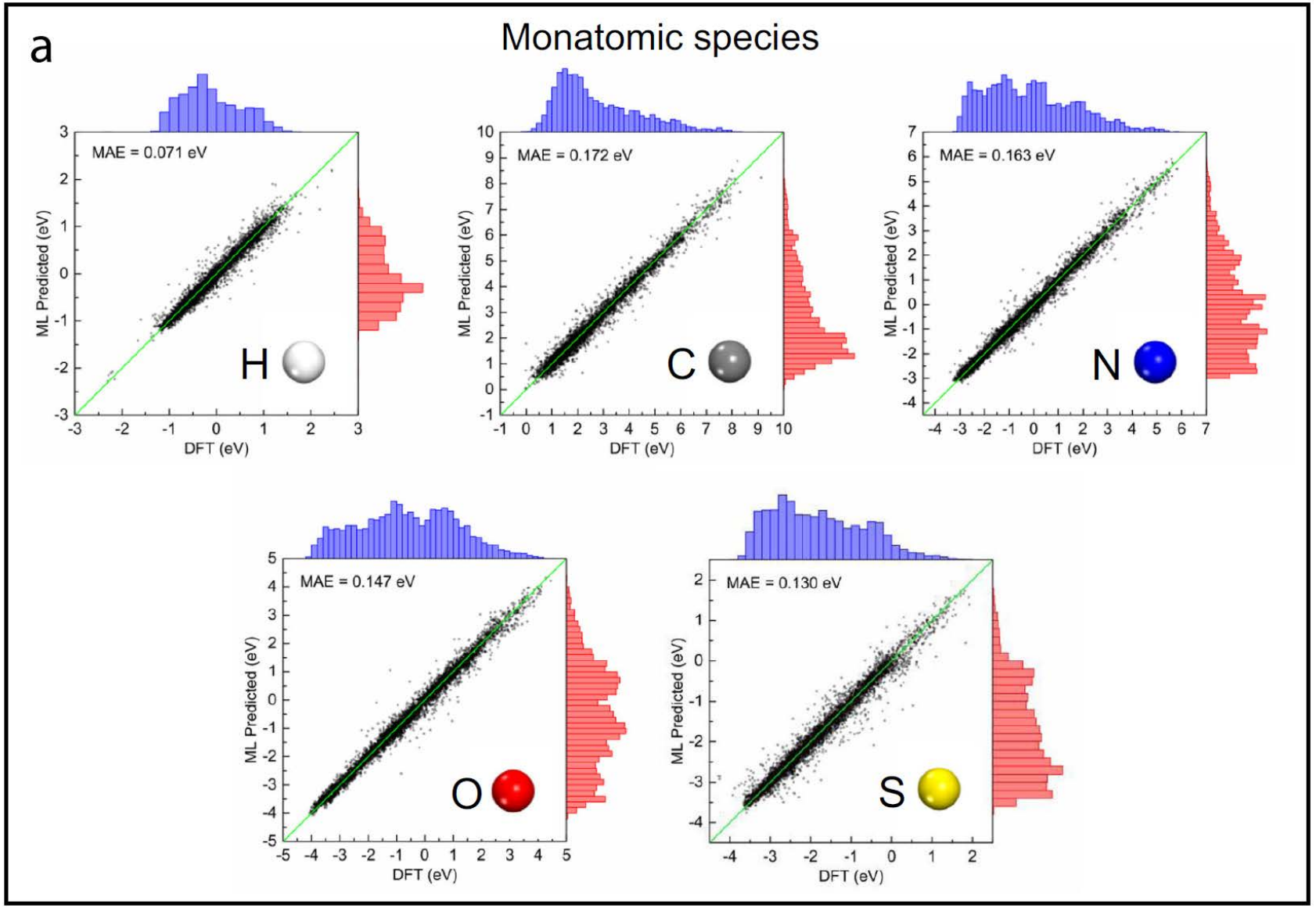
输入：
各原子的PDOS

1个原子对应
9个输入



输出：吸附能

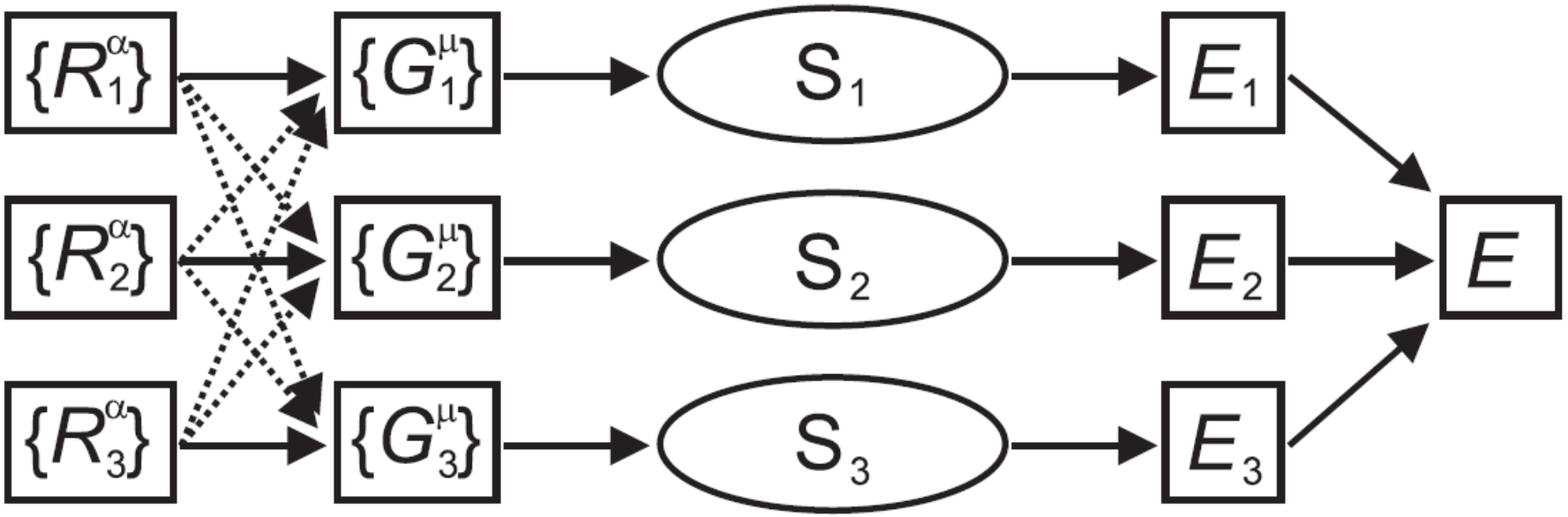
态密度预测吸附能


a
Monatomic species
MAE = 0.071 eV
H
MAE = 0.172 eV
C
MAE = 0.163 eV
N
MAE = 0.147 eV
O
MAE = 0.130 eV
S
ML Predicted (eV)
DFT (eV)

目录

机器学习势

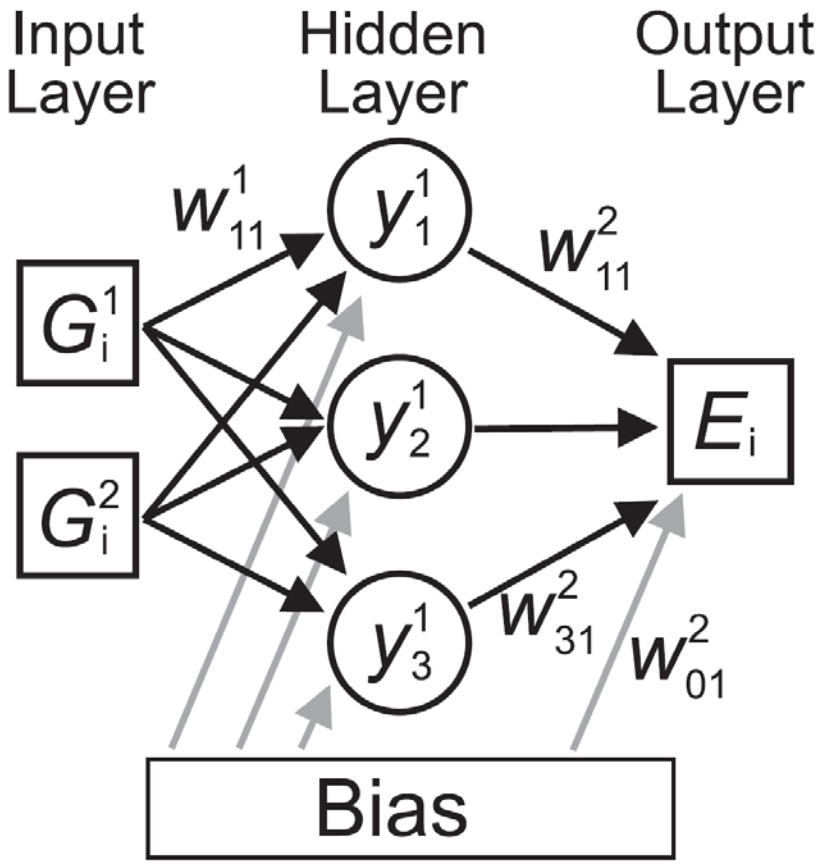
使用第一性原理计算的结果作为训练数据创建分子力场，并用于分子动力学计算



$$E = \sum_i E_i.$$

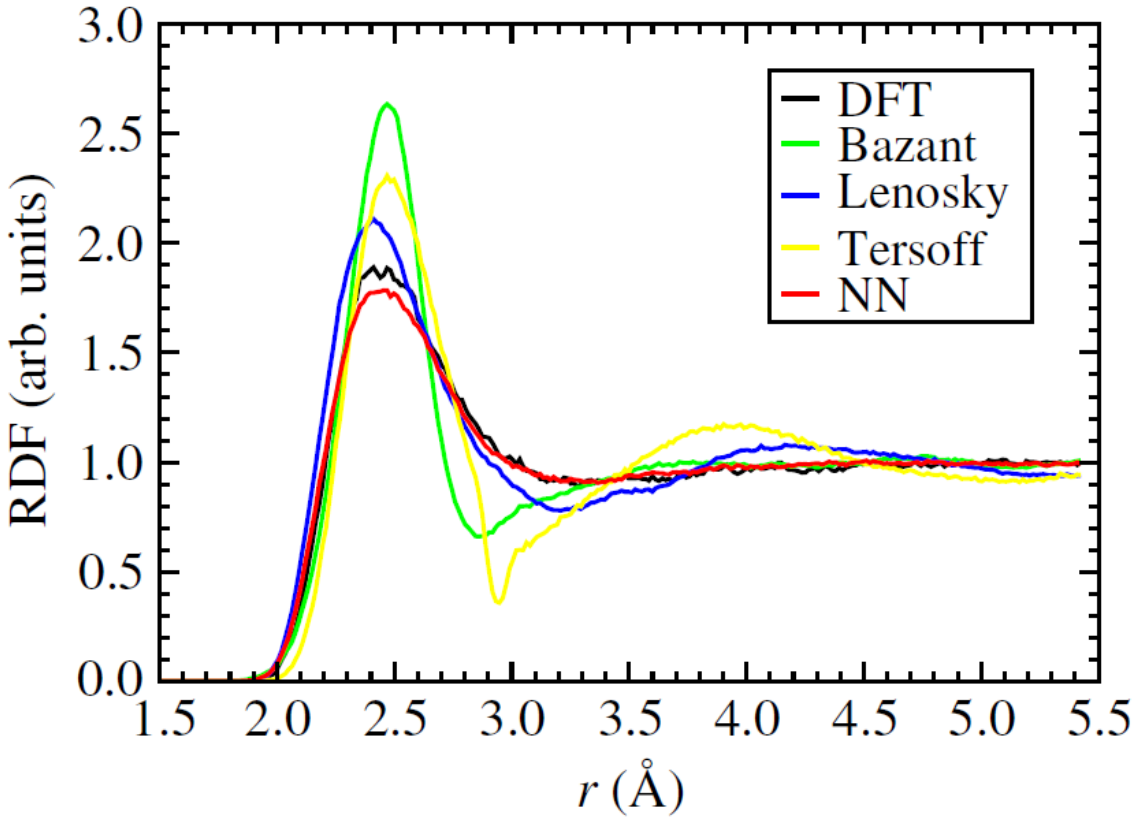
Jorg Behler and Michele Parrinello. *Phys. Rev. Lett.* **2007**, 98, 146401.

机器学习势

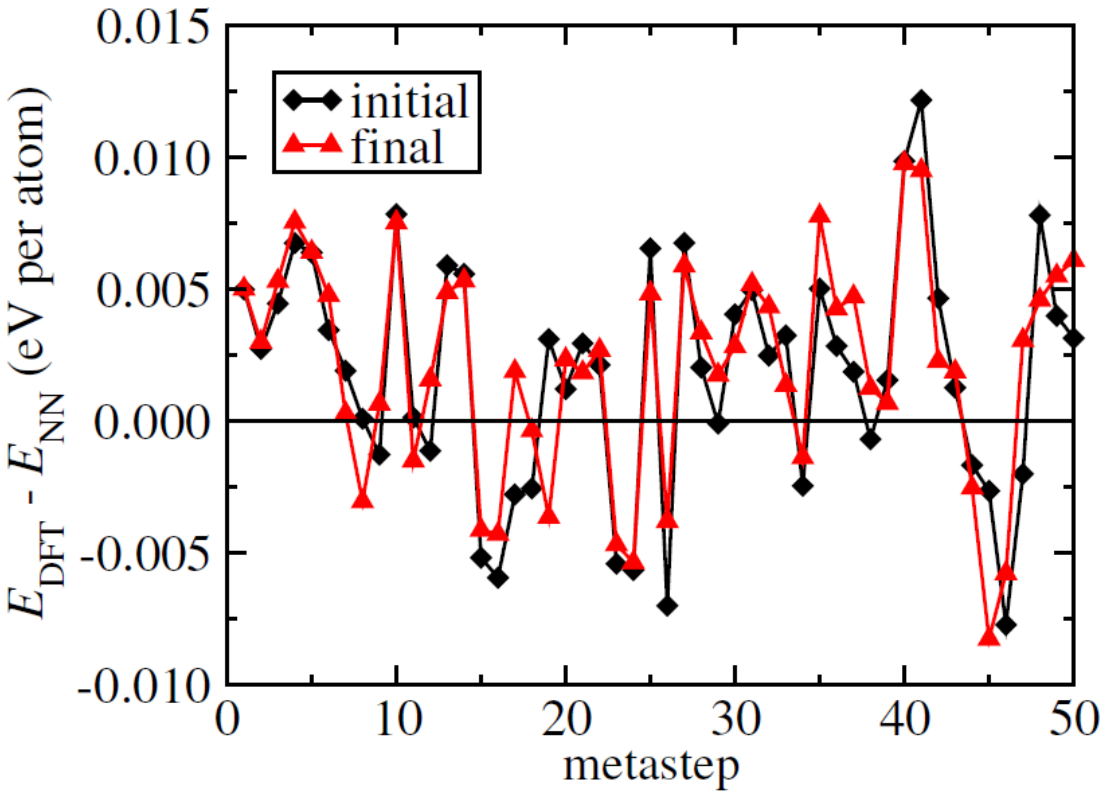


$$E_i = f_a^2\left[w_{01}^2 + \sum_{j=1}^{3} w_{j1}^2 f_a^1\left(w_{0j}^1 + \sum_{\mu=1}^{2} w_{\mu j}^1 G_i^\mu \right) \right].$$

机器学习势

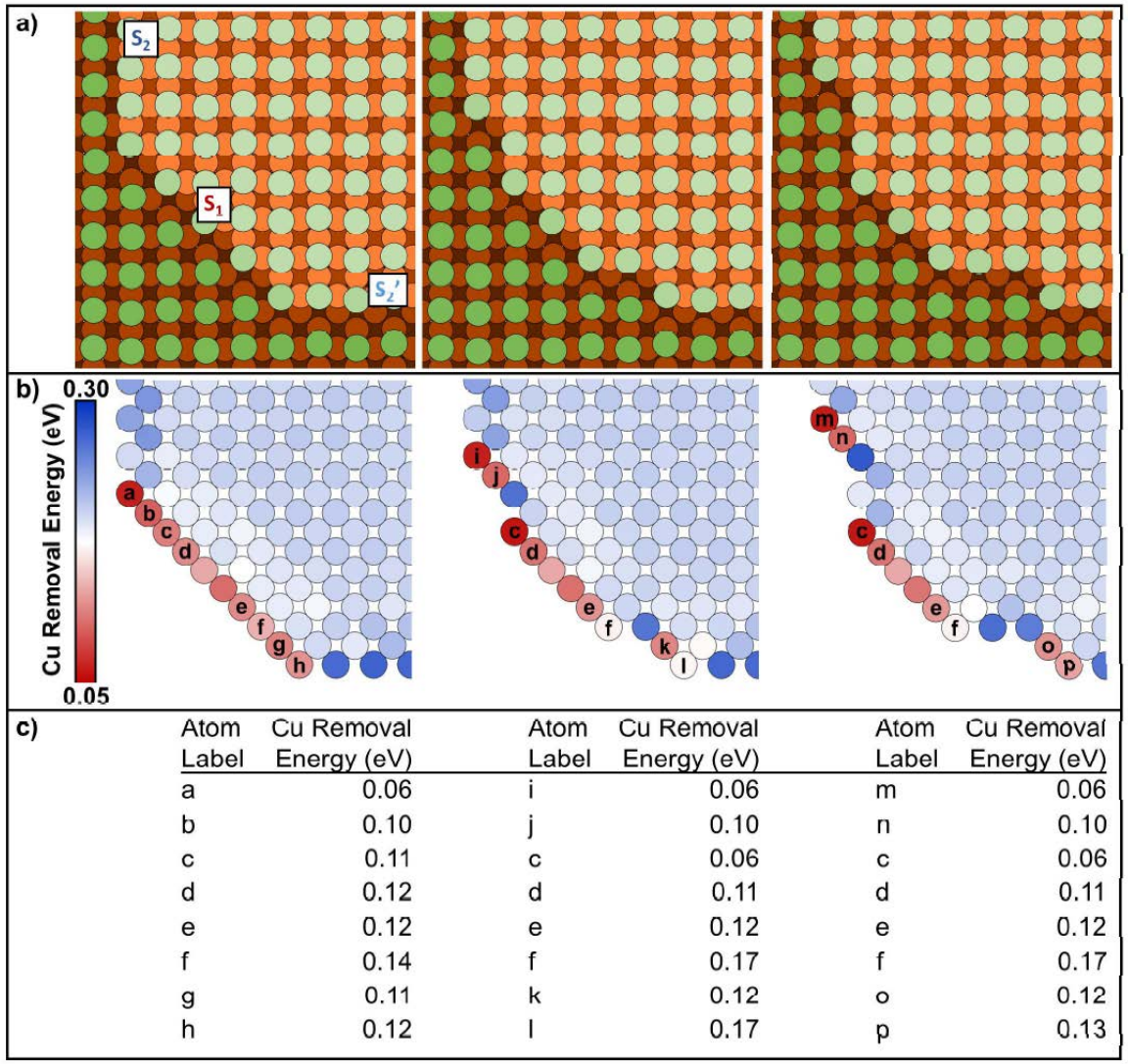


Si, 3000K, 20ps



Si, 300K, 元动力学 (metadynamics), 2ps

机器学习势



c)

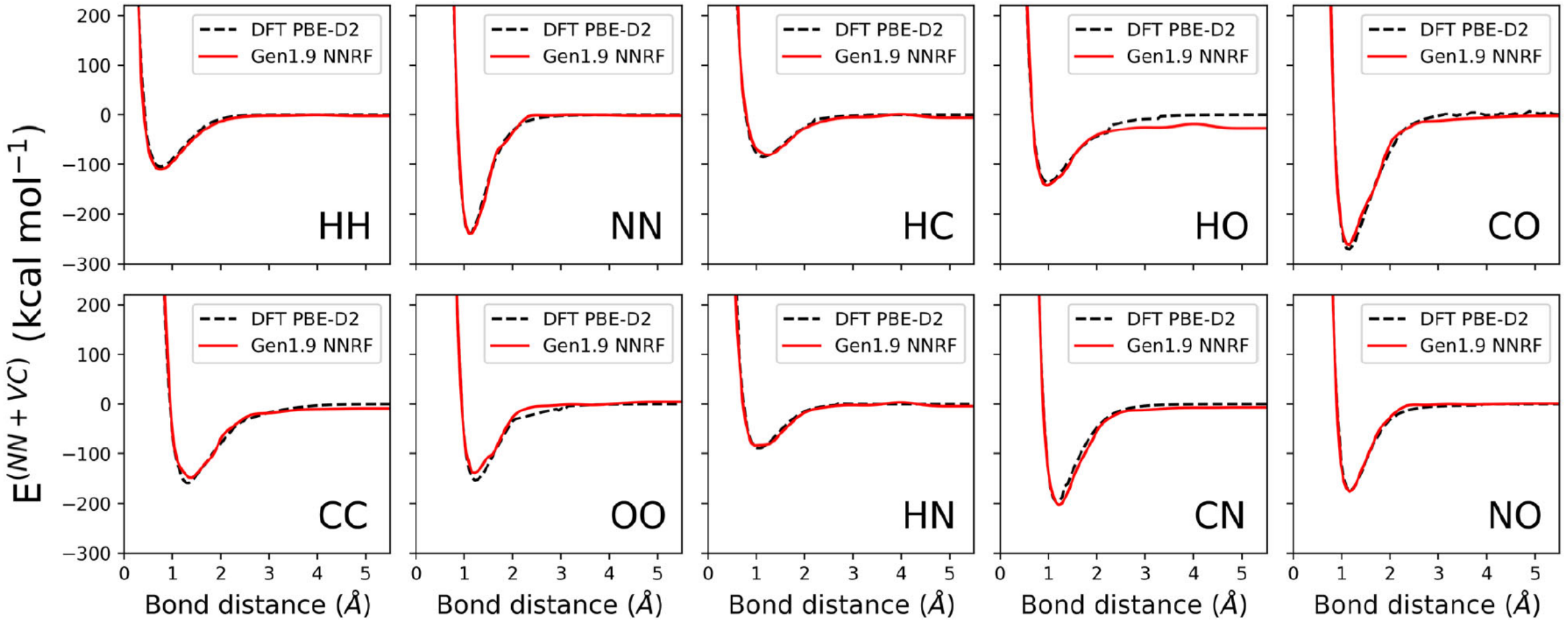
Atom Label	Cu Removal Energy (eV)	Atom Label	Cu Removal Energy (eV)	Atom Label	Cu Removal Energy (eV)
a	0.06	i	0.06	m	0.06
b	0.10	j	0.10	n	0.10
c	0.11	c	0.06	c	0.06
d	0.12	d	0.11	d	0.11
e	0.12	e	0.12	e	0.12
f	0.14	f	0.17	f	0.17
g	0.11	k	0.12	o	0.12
h	0.12	l	0.17	p	0.13

Cl/Cu(100)表面

M. C. Groenenboom, T. P. Moffat, and K. A. Schwarz. *J. Phys. Chem. C* **2020**, 124, 12359-12369.

神经网络反应力场

3100分子，11941构型　15凝聚体系，32973构型



P. Yoo, A. Strachan, et. al. *NPJ Comput. Mater.* **2021**, 7, 1-10.

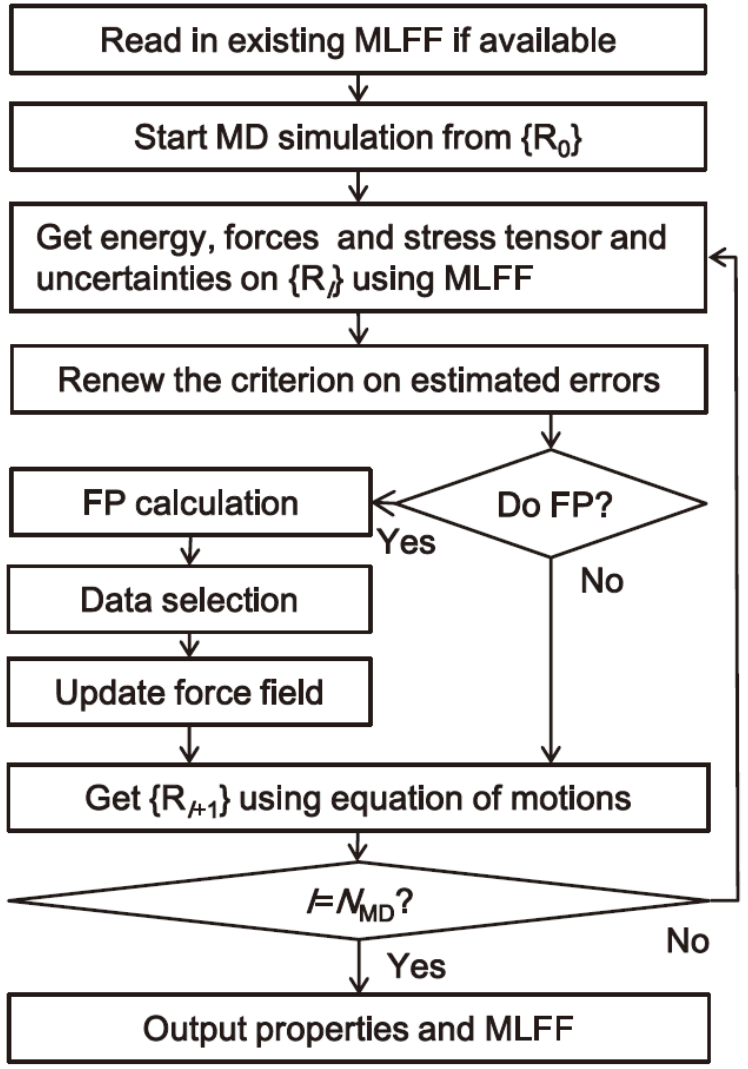
目录

实时从头算分子动力学

机器学习势的缺陷：需要事先计算大量的DFT数据，并训练好力场

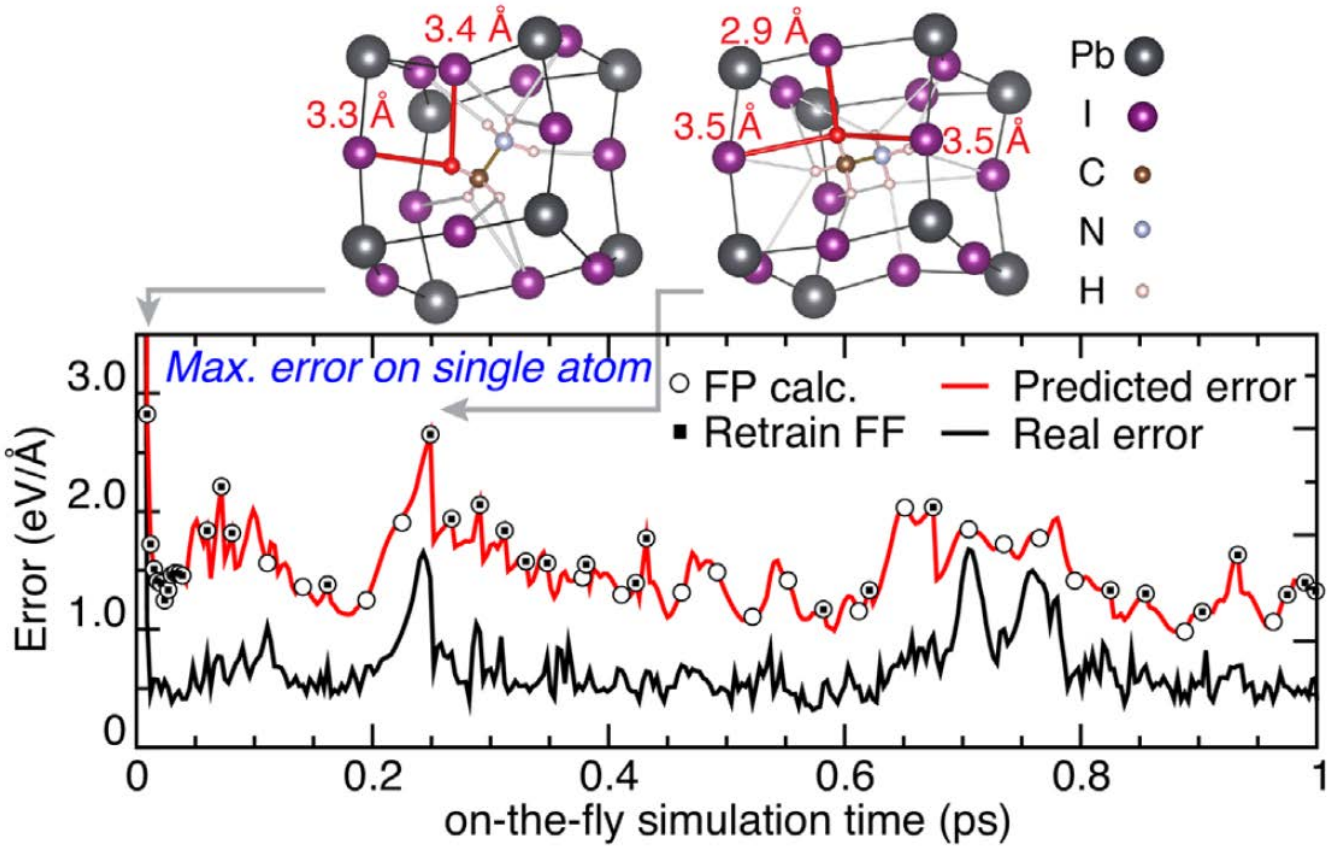
解决办法：“边算边用”

实时从头算分子动力学（on-the-fly AIMD）



R. Jinnouchi, F. Karsai, and G. Kresse. *Phys. Rev. B* **2019**, 100, 014105.

实时从头算分子动力学



$(MA)PbI_3$ 的模拟

R. Jinnouchi, M. Bokdam et. al. *Phys. Rev. Lett.* **2019**, 122, 225701.

目录

pymatgen

Python Materials Genomics (pymatgen)

用于材料分析的python库（Python 3.7以上）

1. 高度灵活的类（Element，Site，Molecule，Structure）
2. 支持各种输入输出（VASP，ABINIT，CIF，Gaussian，XYZ等）
3. 分析工具（生成相图，扩散分析，反应，态密度等）
4. 结合Materials Project，Crystallography Open Database与其它数据库

pymatgen

In [1]:
```
from pymatgen.ext.matproj import MPRester
```

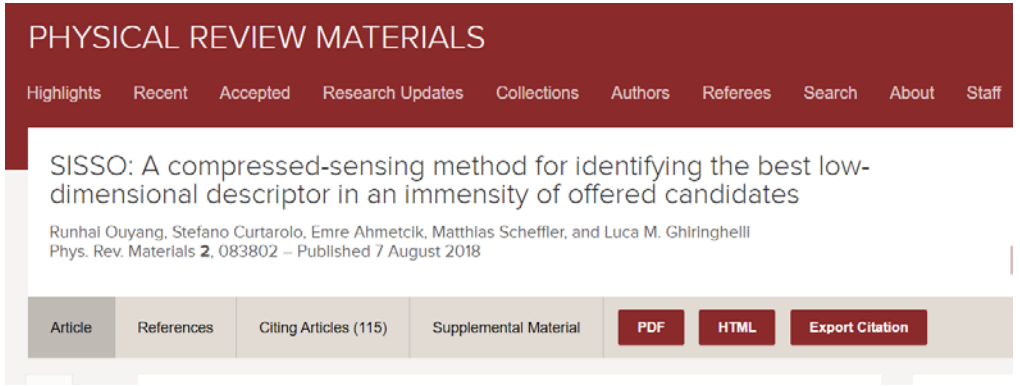
In [2]:
```
with MPRester("YOUR API KEY") as m:
    struct = m.get_data('Fe2O3')
print(len(struct))
print(struct[0])
for i in struct:
    print(i['unit_cell_formula'],i['material_id'])
```

```
25
{'energy': -516.78134568,
sites': 80, 'unit_cell_for
s': 2, 'e_above_hull': 0.2
e': 'spglib', 'symbol': 'F
-1844496'], 'band_gap': 0.
en\ndata_Fe2O3\n_symmetry_
47\n_cell_angle_alpha   87
```

In [4]:
```
with MPRester("YOUR API KEY") as m:
    structure = m.get_structure_by_material_id("mp-1234")
    data = m.get_data('Li-Fe-O',prop = 'energy')
print(structure)
print(data)
```

```
Full Formula (Lu2 Al4)
Reduced Formula: LuAl2
abc   :   5.488740   5.488740   5.488740
angles:  60.000000  60.000000  60.000000
Sites (6)
  #  SP        a       b       c    magmom
```

SISSO


PHYSICAL REVIEW MATERIALS

Highlights Recent Accepted Research Updates Collections Authors Referees Search About Staff

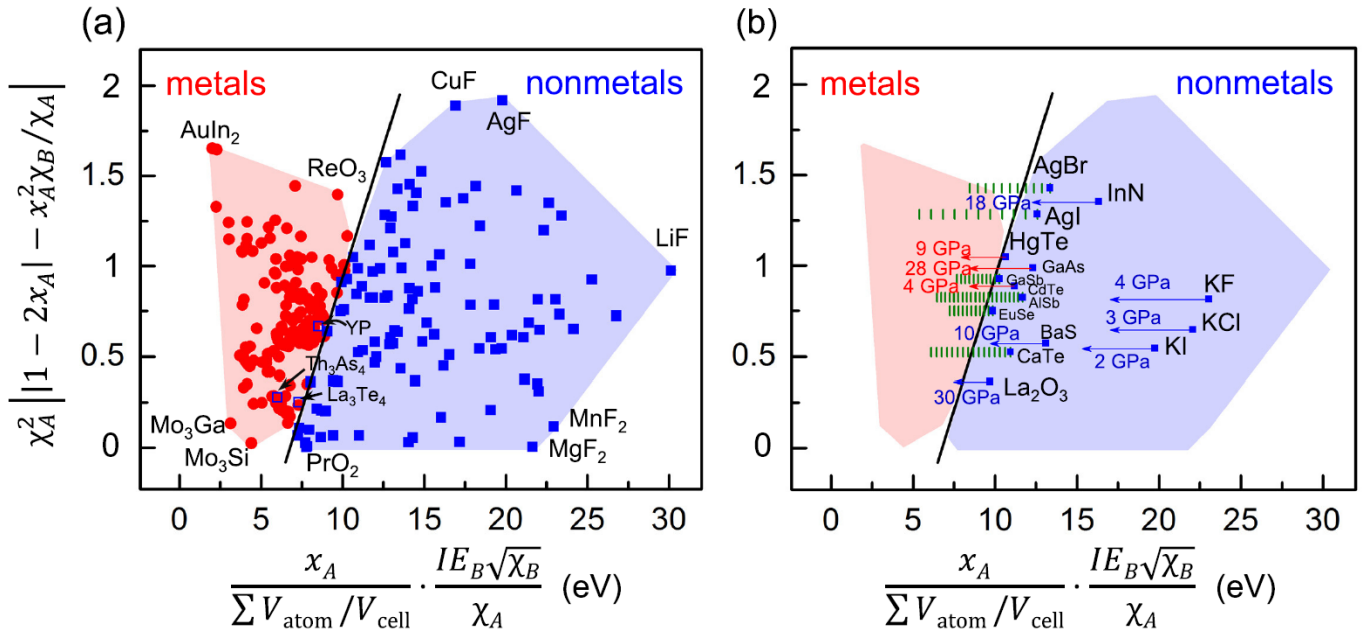
SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates

Runhai Ouyang, Stefano Curtarolo, Emre Ahmetcik, Matthias Scheffler, and Luca M. Ghiringhelli
Phys. Rev. Materials **2**, 083802 – Published 7 August 2018

Article References Citing Articles (115) Supplemental Material PDF HTML Export Citation

SISSO（Sure Independence Screening and Sparsifying Operator）

用于从高维描述符中识别出低维描述符



χ Pauling电负性

IE 电离能

x 原子组分

V 体积

R. Ouyang, L. M. Ghiringhelli, *et. al. Phys. Rev. Mater.* **2018**, 2, 083802.

MLFF相关软件

MAML

https://github.com/materialsvirtuallab/maml

n2p2 - A neural network potential package

"v2.1.4-9-g72f58c0"

n2p2

https://github.com/CompPhysVienna/n2p2

SIMPLE-NN

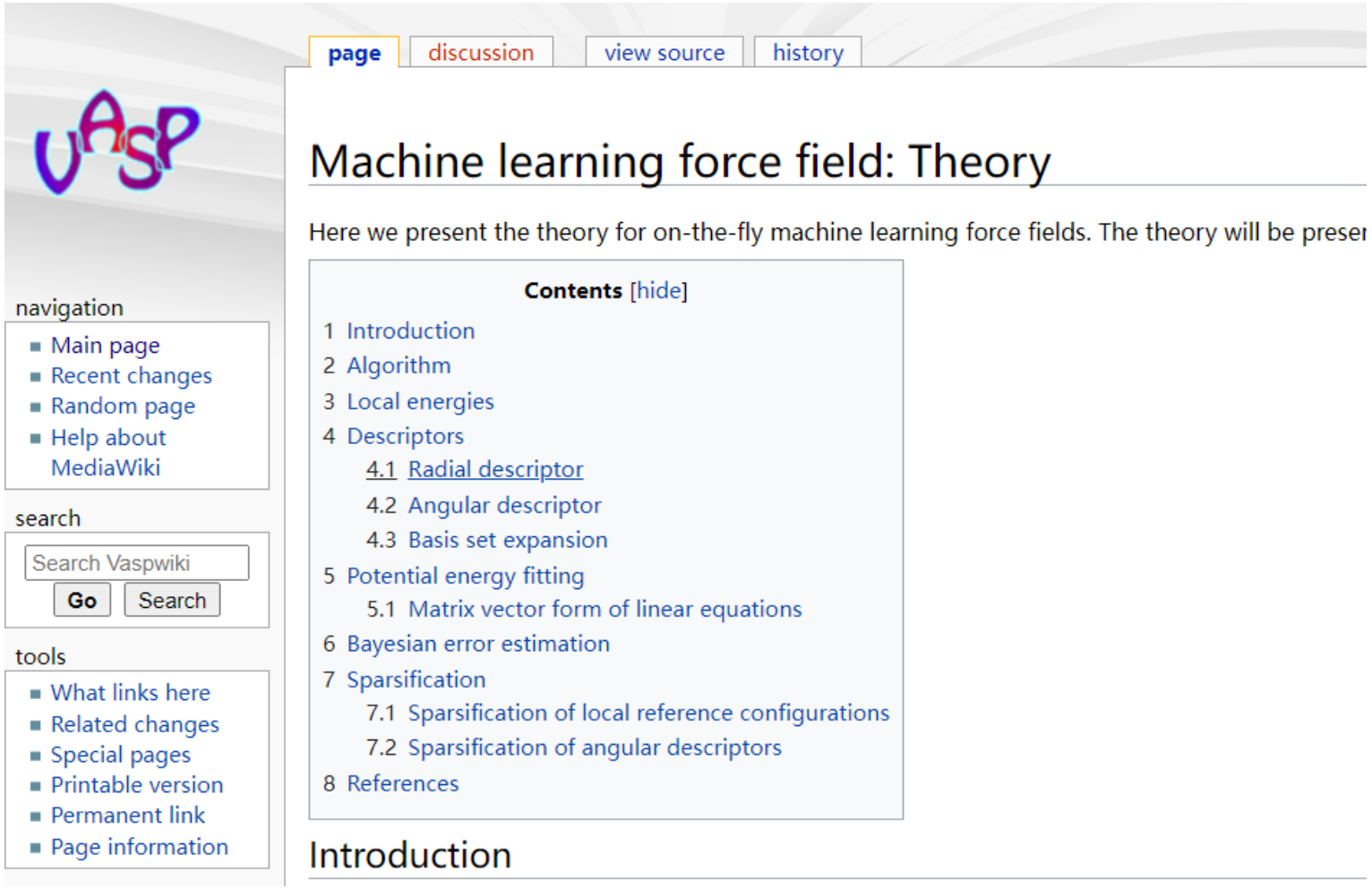
https://github.com/MDIL-SNU/SIMPLE-NN

NequIP

https://github.com/mir-group/nequip

用于训练神经网络势，后续可使用 LAMMPS 或 ase进行MD模拟

从VASP 6.3开始，支持on-the-fly AIMD



写在最后

练习永远是提升自己数据科学水平的最佳途径

三个推荐的网站

GitHub:
https://github.com

LeetCode:
https://leetcode.com
https://leetcode-cn.com

kaggle

kaggle:
https://www.kaggle.com