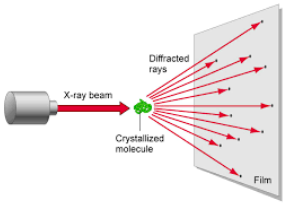
A Machine Learning Method in Computational Materials Science

Computer Network Information Center,
Chinese Academy of Sciences

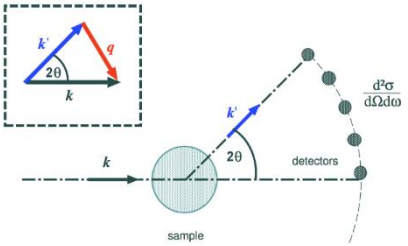
Contributors:
Yangang Wang
Xueyuan Liu
Boyao Zhang
Rongqiang Cao

中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

Experimental Technique



X-ray crystallography



Neutron scattering

Computer Simulation

Based on the Principle of Minimum Energy: For a closed system, the internal energy will decrease and approach a minimum value at equilibrium.

Molecular Dynamics

Fast but Rough

LAMMPS

GROMACS
FAST. FLEXIBLE. FREE.

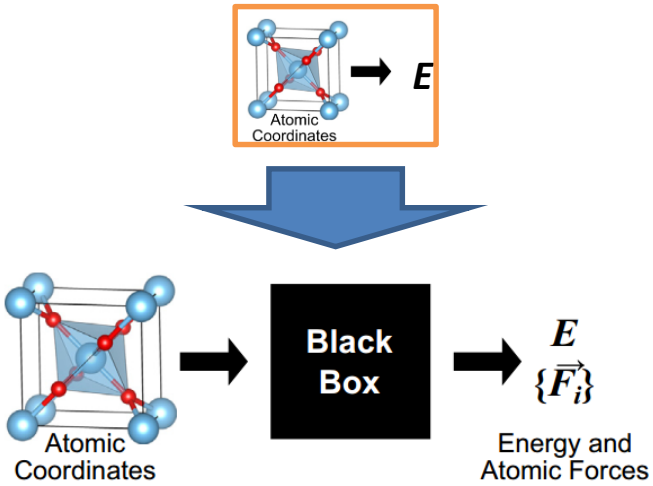
Density Function Theory

Precise but Time-consuming

b-initio
VASP
ienna ackage imulation

QUANTUMESPRESSO

Machine Learning



The basic model of Machine Learning Method

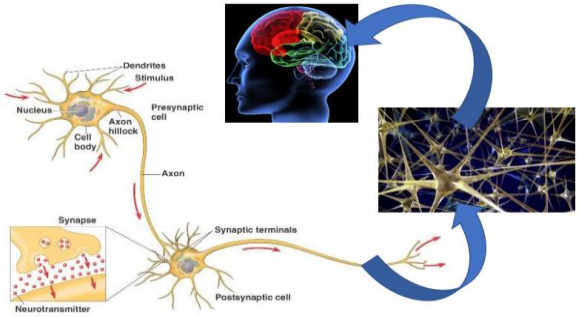
Nongnuch Artrith, Alexander Urban. Computational Materials Science 114 (2016) 135-150

The Goal or Advantage of Machine Learning Potential:

- More precise than molecular dynamics
- Much lower time-consumption than DFT
- Reduce the dependence on the physical model and the human intervention
- Suitable for different molecular systems
- Reuse the data we get during the research

Artificial Neural Network

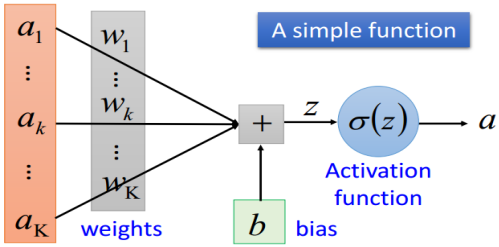
Human Brains



Neural Network

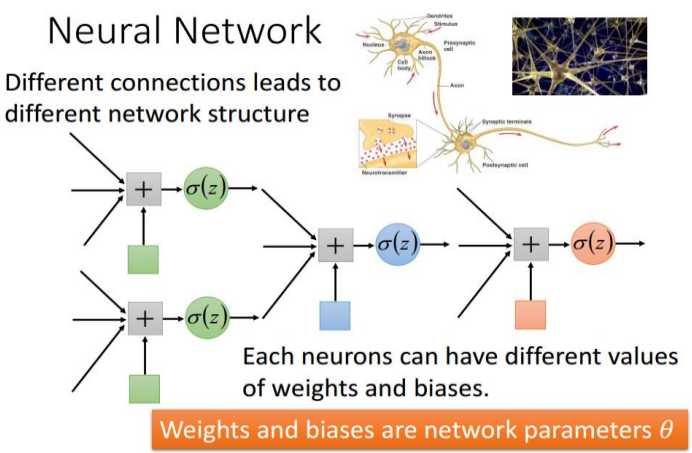
Neuron

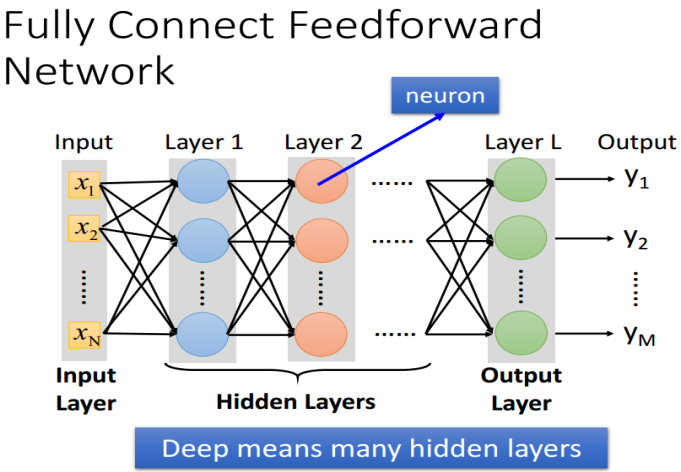
$$z = a_1 w_1 + \cdots + a_k w_k + \cdots + a_K w_K + b$$



Hongyi Li, Open Course: Understanding Deep Learning in One Day

Artificial Neural Network





Hongyi Li, Open Course: Understanding Deep Learning in One Day

Input Node: Description of Atomic Interactions
Output Node: The Energy of Structure

Description of Atomic Interaction

- using directly the Cartesian atomic coordinates as inputs of ANN

$$E(\sigma) \approx E^{\mathrm{ANN}}(\sigma) = \mathcal{N}(\{\mathbf{R}_i\})$$

resulting in highly specialized potentials that are not transferable to systems with different numbers of atoms

- replaced by local structural environment

$$E(\sigma) = \sum_{i}^{\mathrm{atoms}} E_i(\sigma) \approx \sum_{i}^{\mathrm{atoms}} E_i(\sigma_i)$$

the basis set of radial and angular symmetry functions:

$$G_i^{\mathrm{r}}(\sigma_i) = \sum_{j \neq i}^{\mathrm{neighbors}} g^{\mathrm{r}}(R_{ij}) \quad \text{with} \quad R_{ij} = |\mathbf{R}_j - \mathbf{R}_i|$$

$$G_i^{\mathrm{a}}(\sigma_i) = \sum_{k \neq j \neq i}^{\mathrm{neighbors}} g^{\mathrm{a}}(\theta_{ijk}) \quad \text{with} \quad \theta_{ijk} = \angle(\mathbf{R}_j - \mathbf{R}_i, \mathbf{R}_k - \mathbf{R}_i)$$

A.P. Bartók, R. Kondor, G. Csányi, Phys. Rev. B 87 (2013) 184115

Description of Atomic Interaction

No.	η ($Å^{-2}$)	No.	η ($Å^{-2}$)
1–2	0.003214	9–10	0.214264
3–4	0.035711	11–12	0.357106
5–6	0.071421	13–14	0.714213
7–8	0.124987	15–16	1.428426

No.	η ($Å^{-2}$)	λ	ζ	No.	η ($Å^{-2}$)	λ	ζ
17–19	0.000357	−1.0	1.0	44–46	0.000357	1.0	1.0
20–22	0.028569	−1.0	1.0	47–49	0.028569	1.0	1.0
23–25	0.089277	−1.0	1.0	50–52	0.089277	1.0	1.0
26–28	0.000357	−1.0	2.0	53–55	0.000357	1.0	2.0
29–31	0.028569	−1.0	2.0	56–58	0.028569	1.0	2.0
32–34	0.089277	−1.0	2.0	59–61	0.089277	1.0	2.0
35–37	0.000357	−1.0	4.0	62–64	0.000357	1.0	4.0
38–40	0.028569	−1.0	4.0	65–67	0.028569	1.0	4.0
41–43	0.089277	−1.0	4.0	68–70	0.089277	1.0	4.0

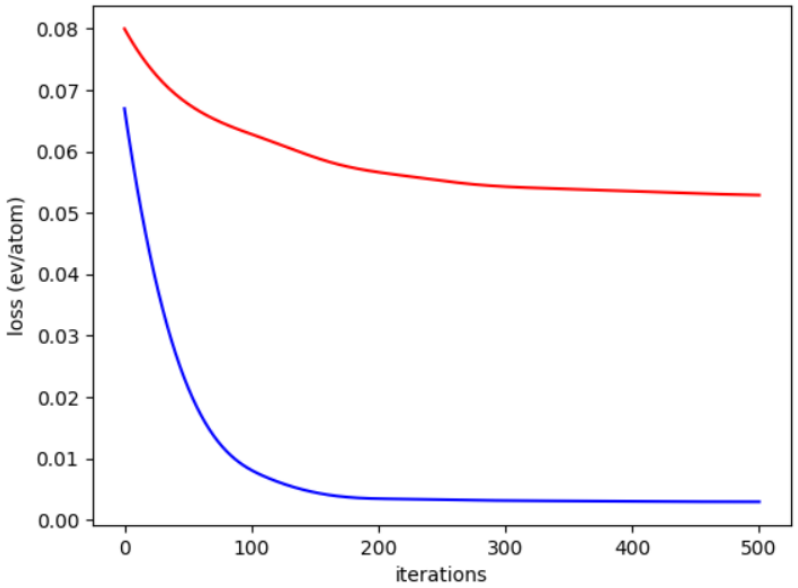
The parameters of symmetry functions for TiO_2

Symmetry functions (G^2)			Symmetry functions (G^2)		
No.	Neighboring element	η ($Å^{-2}$)	No.	Neighboring element	η ($Å^{-2}$)
1	O	0.0009	15	Cu	0.0350
2	Au	0.0009	16	H	0.0350
3	Cu	0.0009	17	O	0.0600
4	H	0.0009	18	Au	0.0600
5	O	0.0100	19	Cu	0.0600
6	Au	0.0100	20	H	0.0600
7	Cu	0.0100	21	O	0.1000
8	H	0.0100	22	Au	0.1000
9	O	0.0200	23	Cu	0.1000
10	Au	0.0200	24	H	0.1000
11	Cu	0.0200	25	O	0.2000
12	H	0.0200	26	Au	0.2000
13	O	0.0350	27	Cu	0.2000
14	Au	0.0350	28	H	0.2000

Symmetry functions (G^4)			
No.	η ($Å^{-2}$)	λ	ζ
29-38	0.0001	1.0	1.0
39-48	0.0001	-1.0	1.0
49-58	0.0001	1.0	4.0
59-68	0.0001	-1.0	4.0
69-78	0.0030	1.0	1.0
79-88	0.0030	-1.0	1.0
89-98	0.0030	1.0	4.0
99-108	0.0030	-1.0	4.0
109-118	0.0080	1.0	1.0
119-128	0.0080	-1.0	1.0
129-138	0.0080	1.0	4.0
139-148	0.0080	-1.0	4.0

The parameters of symmetry functions for $Cu_aAu_bO_cH_d$

Description of Atomic Interaction



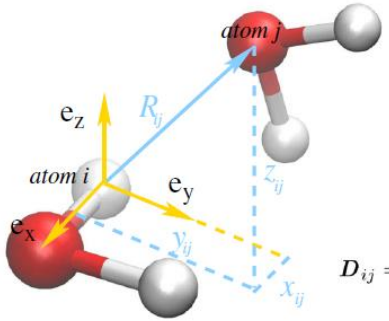
SiAu

TiO_2

The convergence curve using the same functional parameters as TiO2 in SiAu

Description of Atomic Interaction

A new method for getting descriptors without designing different functional parameters



$$\boldsymbol{D}_{ij} = \{1/R_{ij}\}$$

only the radial information

or

$$\boldsymbol{D}_{ij} = \{1/R_{ij}, x_{ij}/R_{ij}^2, y_{ij}/R_{ij}^2, z_{ij}/R_{ij}^2\}$$

the full radial and angular information of atom i, j

Linfeng Zhang, Jiequn Han, Han Wang, etc. Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics, Physical Review Letters 120, 143001 (2018)

Description of Atomic Interaction

get the new coordinate based on its local framework of centered atom i

$$\mathcal{R}(R_{ia(i)}, R_{ib(i)}) = \begin{pmatrix} e[R_{ia(i)}] \\ e[R_{ib(i)} - (R_{ia(i)} \cdot R_{ib(i)} R_{ia(i)}] \\ e[R_{ia(i)} \times R_{ib(i)}] \end{pmatrix}^T, \qquad e[x] \equiv \frac{x}{||x||}$$

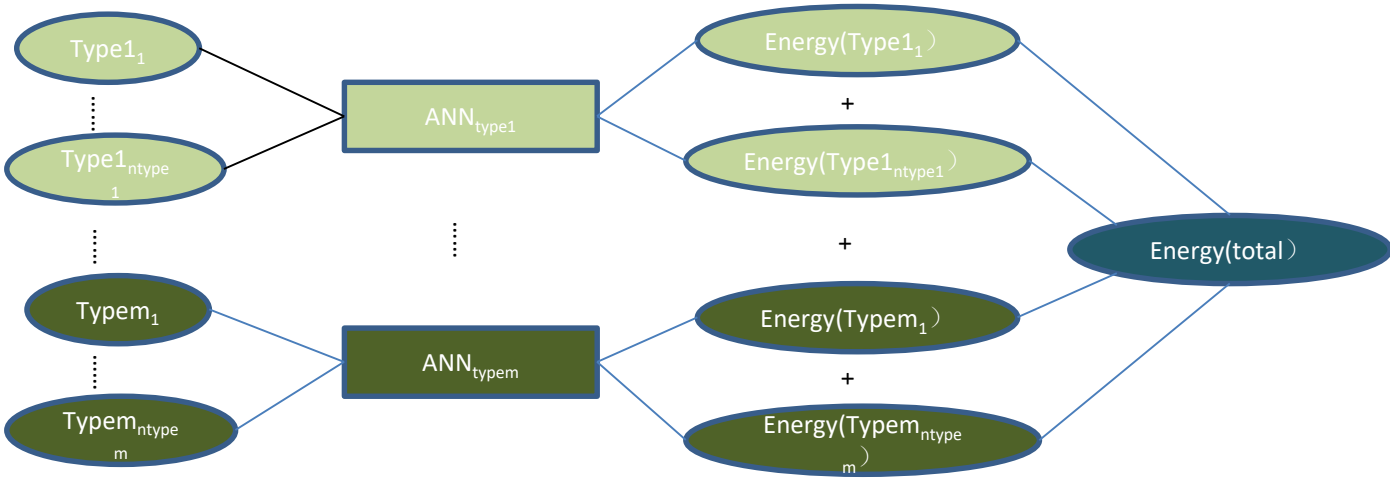
$$R'_{ij} = \{x'_{ij}, y'_{ij}, z'_{ij}\} = \{x_{ij}, y_{ij}, z_{ij}\}\,\mathcal{R}(R_{ia(i)}, R_{ib(i)})$$

$$D_{ij} \equiv \begin{cases} \{D^0_{ij}, D^1_{ij}, D^2_{ij}, D^3_{ij}\} = \left\{\frac{1}{R'_{ij}}, \frac{x'_{ij}}{R'^2_{ij}}, \frac{y'_{ij}}{R'^2_{ij}}, \frac{z'_{ij}}{R'^2_{ij}}\right\}, \\ \{D^0_{ij}\} = \left\{\frac{1}{R'_{ij}}\right\}, \end{cases}$$

$$R'_{ij} = ||R'_{ij}||$$

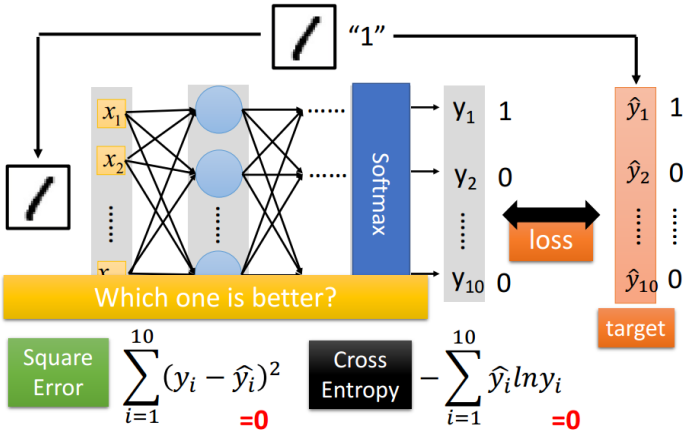
Linfeng Zhang, Jiequn Han, Han Wang, etc. Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics, Physical Review Letters 120, 143001 (2018)

The Structure of Artificial Neural Network



Loss Function

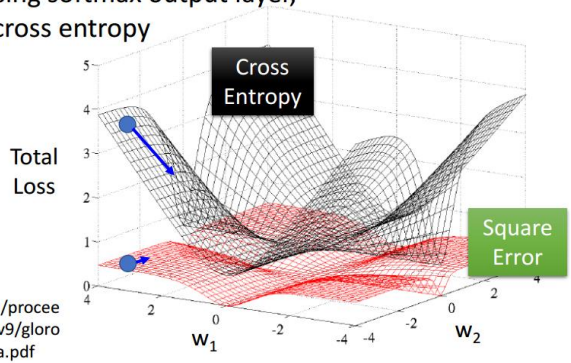
Choosing Proper Loss



Hongyi Li, Open Course: Understanding Deep Learning in One Day

Choosing Proper Loss

When using softmax output layer, choose cross entropy



http://jmlr.org/proceedings/papers/v9/glorot10a/glorot10a.pdf

Loss Function

Apart from energy, force and virial are considered in loss function as well

$$L(p_\epsilon, p_f, p_\xi) = p_\epsilon \Delta\epsilon^2 + \frac{p_f}{3N}\sum_i |\Delta \boldsymbol{F}_i|^2 + \frac{p_\xi}{9}||\Delta\xi||^2$$

The learning rate and the weight of energy, force, virial vary throughout the training procedure

$$r_l = r_{l0} d_r^{-c_s/d_s}$$

How the learning rate varies

$$p = p_{limit}(1 - \frac{r_l}{r_{l0}}) + p_{start}(\frac{r_l}{r_{l0}}),$$

How the weight of different factors vary

$$\begin{cases} p_{estart} = 1, & p_{elimit} = 400; \\ p_{fstart} = 1000, & p_{flimit} = 1, \end{cases} \quad \begin{cases} p_{vstart} = 1, p_{vlimit} = 400, & \text{for liquid water and ice (b);} \\ p_{vstart} = 0, p_{vlimit} = 0, & \text{for ice (c) and (d) and the molecules.} \end{cases}$$

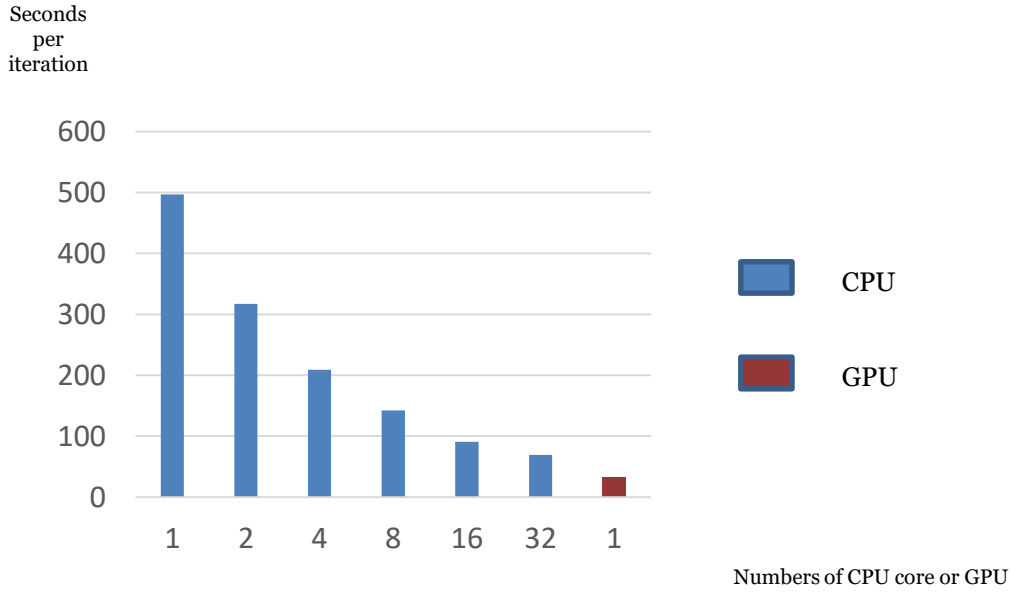
Linfeng Zhang, Jiequn Han, Han Wang, etc. Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics, Physical Review Letters 120, 143001 (2018)

Training with Deep Learning Framework

We want to use PyTorch to implement something as follows:

- Tensor computation with strong GPU acceleration
- Various optimizers for different systems
- Save the model and retrain it at any point

Tensor computation with strong GPU acceleration



Various optimizers for different systems

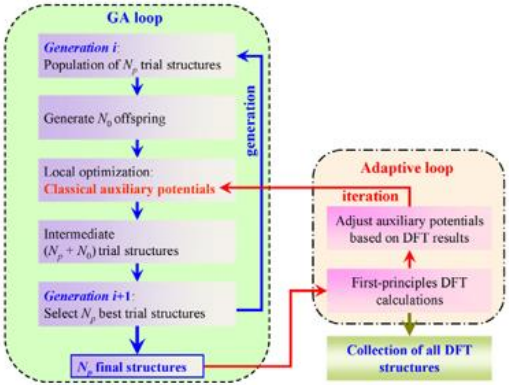
Optimizer	Best Loss
Adadelta	0.0155
Adam	0.0072
Adamax	0.0110
ASGD	0.0135
SGD	0.1676
Rprop	0.1159
RMSprop	0.1037

train TiO_2 system

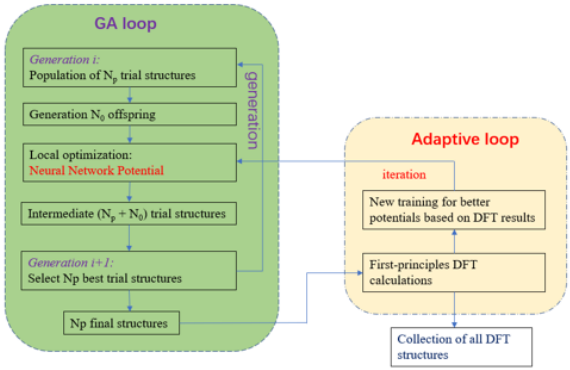
Optimizer	Best Loss
Adadelta	0.0341
Adam	0.0139
Adamax	0.0172
ASGD	0.0196
SGD	0.0294
Rprop	0.0172
RMSprop	0.0083

train SiAu system

Search for reasonable crystal structures

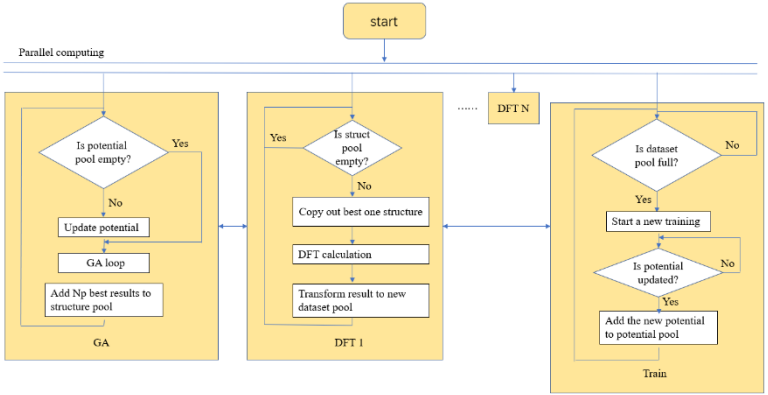


adaptive genetic algorithm



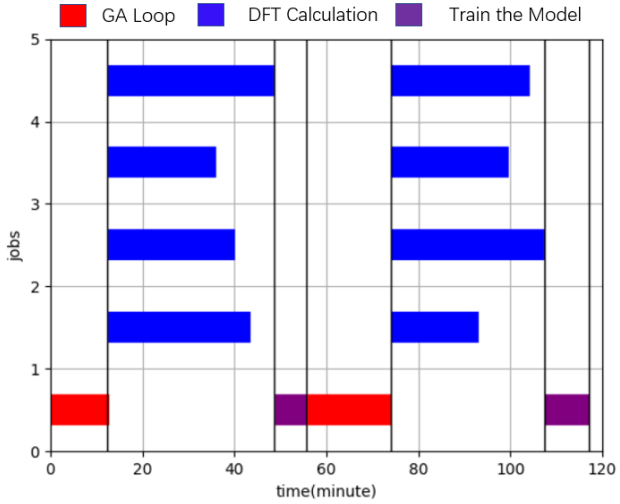
adaptive genetic algorithm using NNP

Performance Optimization of AGA

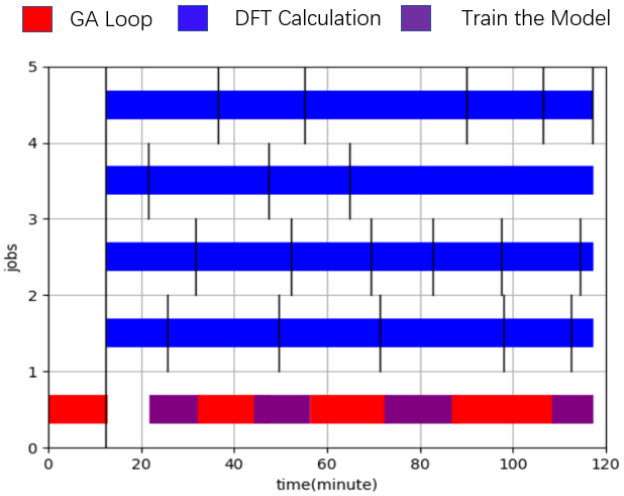


Parallel framework for GA, DFT and retrain module of AGA

Performance Optimization of AGA



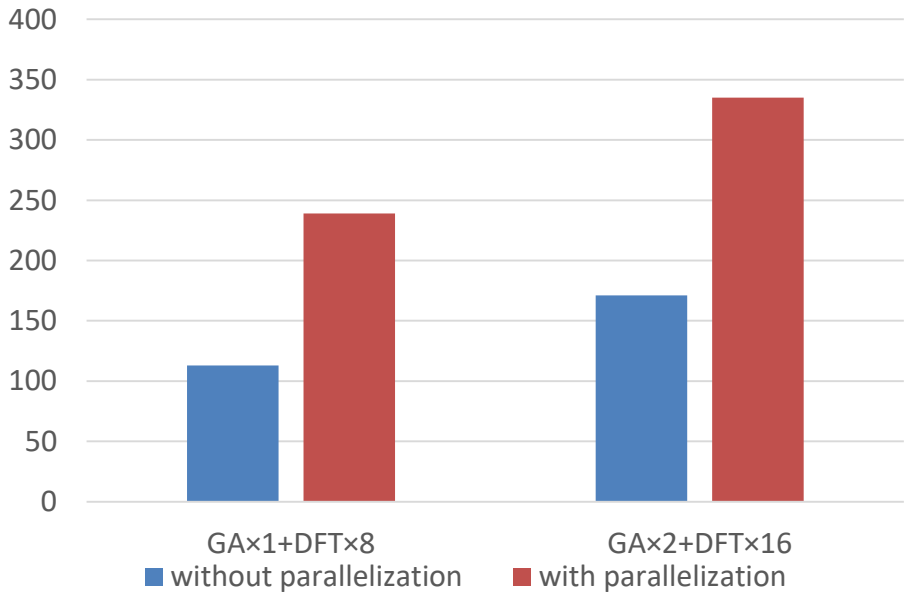
without parallelization



with parallelization

Performance Optimization of AGA

The number of new structures saved in 8h



The number of nodes used for GA and DFT

Algorithm Optimization of AGA

The problems encountered in the retraining module after supplementing new data into original dataset

- The data volume of the original dataset is extremely large, while that of the new data are small
- The existing model has fitted the original dataset well already, but difficult to fit the new data

The reason above results in the phenomenon that new data are hard to be learnt in retrain procedure

Modify the loss function in order to adjust the weight of each structure in dataset based on the loss in last iteration:

$$FL_i = \left(\frac{n * loss_i}{\sum_{i=1}^{n} loss_i} \right)^{\gamma}$$

Algorithm Optimization of AGA

Loss \ γ	0	0.5	1	2	3
0.015	5	1	2	3	2
0.012	22	10	10	7	9
0.01	45	26	31	21	15
0.009	77	39	49	125	272
0.008	197	103	117	289	
0.007	279	174	256		

Shows the number of iterations needed with different exponent to reach the targeted loss

Algorithm Optimization of AGA

select some unlearnt data into dataset

- Parallel training =>several potentials
- New data + several potentials => several energies
- Calculate the difference between several energies

If the difference is big enough, we can infer that there is few similar structures in dataset and we should put the structure into dataset; otherwise, we leave it away

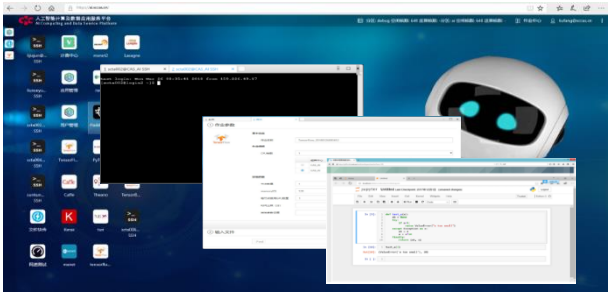
AI Computing and Data Service Platform

中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

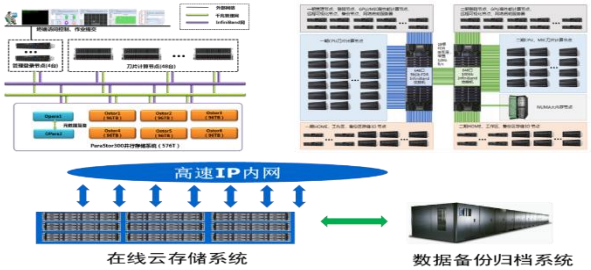
AI Computing and Data Service Platform

Easy-to-use AI platform that supports scientific discovery

- Provide a variety of ways to use
- Various types of artificial intelligence softwares
- Establish standardized public data resources
- Establish platform access standards and evaluation



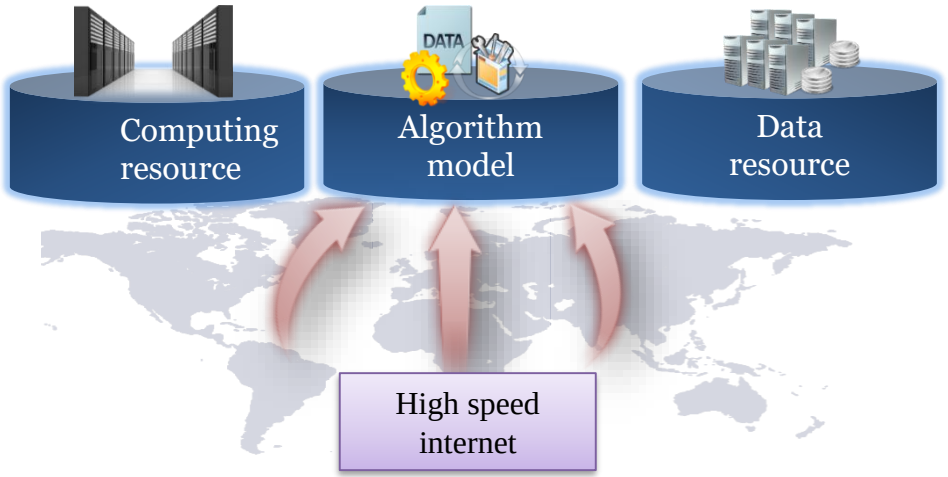
The system is equipped with 380 P100 GPUs, double-precision peak 1.8PF, single-precision peak 3.6PF



Establish 118 service accounts and 200 training accounts

- Institute of High Energy Physics, Chinese Academy of Sciences, Institute of Biophysics, Chinese Academy of Sciences, etc.
- Peking University, China Earthquake Administration and other scientific research institutions
- Caiyun, Yihualu, Yuzhi Technology, Haina Yunfan, Beijing Super Satisfaction and other companies

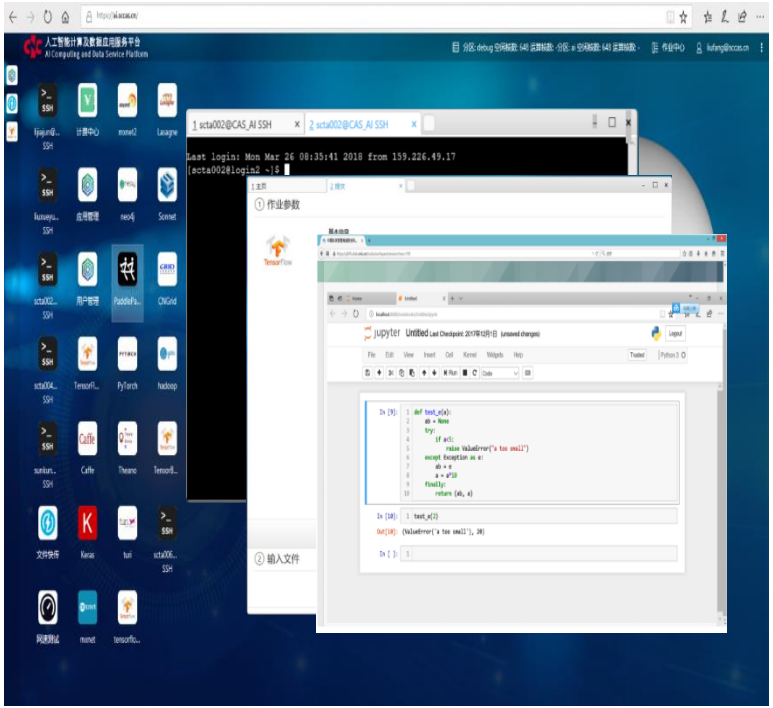
AI Computing and Data Service Platform



- **Academic**
- **Developer**
- **Beginner**

Create an easy-to-use artificial intelligence platform that supports scientific discovery

Artificial intelligence platform construction



Intelligent management

Data and cluster maintenance require information-based intelligent management Increase management efficiency

Parallel Computing

The amount of calculation is huge, and GPU accelerated calculation can greatly speed up the analysis

Fast tool integration

There are many kinds of algorithms related to artificial intelligence, and new algorithms emerge in an endless stream. Need to be able to deploy quickly on the cloud

Interaction is simple

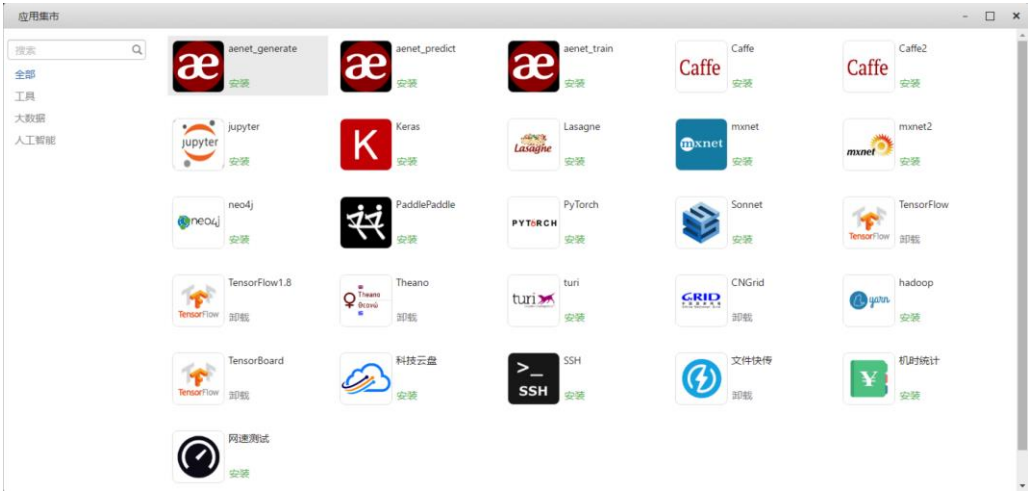
No need to write code in front of the black box, data calculation can be done with simple mouse clicks and settings

Performance visualization

Free users from the ubiquitous data can easily analyze performance status

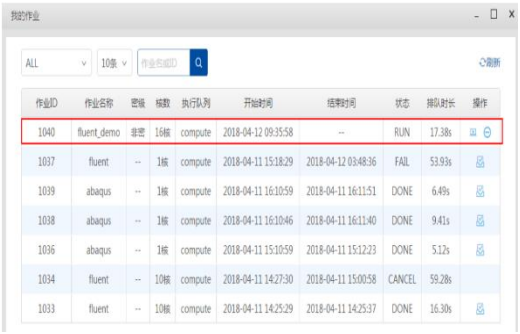
Application Integration: Deep Learning Framework, Industry Applications

- Integrated mainstream deep learning framework
- Integrated parallel application



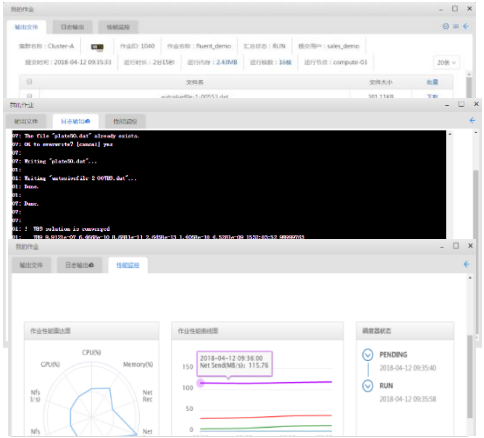
View and manage your jobs in all directions

- **Job files, logs, and performance at a glance**



Output file

Output log



Performance statistics

Hot Resource statistics