Exploration of thermodynamic stability of two-dimensional materials based on construction of feature space map

Junfeng Zhao[1,2], Jian Wu[2] and Lixin Tang[1]

[1] National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, Shenyang, 110819, China
[2] Key Laboratory of Data Analytics and Optimization for Smart Industry (Northeastern University), Ministry of Education, Shenyang, 110819, China
zhaojunfenglv@outlook.com

Abstract. With the discovery of graphene with many excellent properties, there has been an upsurge in the exploration of two-dimensional (2D) materials in the scientific and industrial circles, and has promoted the development of many 2D materials with excellent properties in the research fields such as energy, photocatalysis, microelectronics and photonics. In order to continue to explore and discover new 2D materials with novel properties, using big data and machine learning methods to screen promising candidate materials is a practical means to accelerate the discovery of new materials. In this paper, we exploit deep learning techniques based on convolutional neural networks (CNN) to identify thermodynamically stable 2D materials, which is the first fundamental requirement for any application. The proposed method does not need to know the spatial structure information of the materials. It only constructs the feature space map based on the properties and structural symmetry of the constituent elements of the compound, and then inputs it into a specially designed 2DMs-ResNet network to realize the discrimination and prediction of the thermodynamic stability. Experimental results show that the proposed method has strong generalization performance and satisfactory prediction accuracy.

Keywords: 2D materials, Machine learning, CNN, Thermodynamic stability, Feature space map.

1 Introduction

With the successful preparation of graphene [1] nano two-dimensional (2D) materials with many excellent characteristics, and won the Nobel Prize in Physics in 2010, the research and exploration of 2D materials ushered in a blowout development. Subsequently, a series of other 2D materials with excellent performance have been developed, such as BN, MoS2, WS2, MoSe2, WSe2, MXene materials, and transition metal carbides. They can be not only conductors, semiconductors, or insulators, but even superconductors, opening new research paths and directions for many existing fields, especially energy, photo/electrocatalysis, microelectronics, and photonics. For the

discovery of new 2D materials and the establishment of related databases, many excellent works have emerged in recent years. Starting from 108,423 experimentally known 3D compounds, Mounet et al. [2] used van der Waals density functional theory (DFT) for high-throughput calculations and identified 1,825 compounds that are either easily or potentially exfoliable. Haastrup et al. [3] created the Computational 2D Materials Database (C2DB), which contains a variety of structural, thermodynamic, electronic, magnetic, and optical properties for around 4,000 2D materials. Zhou et al. [4] proposed 2DMatPedia, a large dataset with structural, energy, and electronic properties of more than 6,000 2D materials, through top-down and bottom-up approaches. Based on the high-throughput calculation of DFT, Fukuda et al. [5] constructed a database of AB_2-type monolayer structure maps for 3,844 compounds and obtained the comprehensive structural trends of these compounds in 2D structures through symmetry-unconstrained geometry optimizations, which provided a new perspective for finding unknown 2D materials. Fabian et al. [6] proposed a Quantum Point Defect Database (QPOD) including 503 intrinsic point defects in 82 different 2D semiconductors and insulators, which provides a reference for the application of promising defects in quantum technology. Yao et al. [7] selected and evaluated 19 2D materials with high mobilities at room temperature and good dynamic stability based on the MatHub-2d database, using bandgap, magnetism, elasticity, and deformation potential as search criteria, which provided an opportunity for the preparation of new semiconductor electronic devices.

However, the above works show that many 2D materials are mainly calculated by DFT, and the exploration of new 2D materials is usually only selected and tried from a limited number of possible categories according to the required material properties, which makes the exploration space very limited. After obtaining a promising new material, it is also necessary to evaluate whether the material's properties meet the expected requirements and its thermodynamic stability, which is the first fundamental requirement for most possible applications [8]. Unfortunately, the number of 2D materials candidates obtained by high-throughput screening is limited and may not be stable.

Fortunately, with the development of artificial intelligence technology, data-driven deep learning technology provides a new research paradigm for the field of materials. In view of the primary basic requirement for the application of new materials, which is to have good thermodynamic stability, this paper proposes a deep learning technology based on convolutional neural network (CNN) to identify the thermodynamic stability level of 2D materials. In particular, the method does not need to know the spatial structure information of the materials, only the properties and structural symmetry of the constituent elements of the compounds are used to construct the feature space map, and then input into a specially designed 2DMs-ResNet network to realize the discrimination and prediction of thermodynamic stability. Since this work is still in progress, the existing features of this work are described as follows.

1) Based on the properties of the elements in the compound, we construct a feature space map for the input of the neural network.

2) A lightweight residual neural network structure named 2DMs-ResNet is built to predict the thermodynamic stability of 2D materials.

2 The Proposed Method

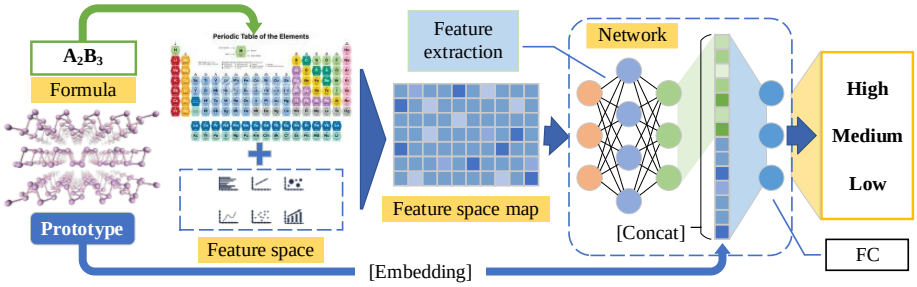

Fig. 1. The framework diagram of the method.

The framework diagram of the proposed method is shown in Fig. 1, which takes the data parsing technique as the core. Firstly, according to the formula of the material, the framework uses the information of each element in the periodic table and the constructed feature space to generate the feature space map and then uses it as the input of the neural network. It should be noted that different 2D materials with different thermodynamic stability may have the same formula, as isomers, their differences are reflected in the 3D structure. In this case, if only the feature space map is used for classification, it will lead to misjudgment. Therefore, the proposed method takes the prototype of materials as an important classification feature and uses the idea of word Embedding [9] in natural language processing to map it into special feature vectors. Finally, the vector concatenated by the prototype feature vector and the image feature vector is mapped to different thermodynamic stability levels through the fully connected layer of the network. The specific implementation details of the proposed method are described in the following.

2.1 Feature space

To construct the material feature space map, from the above, we first need to find the properties that each element of the compound has in the periodic table. Here we selected attributes as shown in Table 1.

Table 1. Properties of atoms.

Property	Description	Property	Description
Z	atomic number	ε, I	electron affinity and ionization potential
P	Pauling electronegativity	η, ζ	highest-occupied and lowest-unoccupied Kohn-Sham eigenvalue
ϑ	periodic group	r	atomic nonbonded radius
v, α	valence, polarizability	r_v	radius of the last occupied valence orbital
ϕ	unfilled valence orbitals	r_s, r_p	extensions of the s and p orbitals

As the core of constructing the feature space map, we define the feature space as shown in Table 2. Next, according to the atomic properties in Table 1, and using a variety of statistical operators given in Table 2, the detailed characteristics of compounds are calculated. Specifically, for a material formed by n_s species of atoms and N atoms in the unit cell. Finally, these characteristics are constructed into a 2D matrix, and the corresponding feature space map of the material is obtained.

Table 2. Feature space, constructed using each of the χ properties in Table 1.

Feature	Description	Feature	Description
$\chi_{sum} = \sum_{i=1}^{n_s} \chi_i$	sum of values	$\bar{\chi}_\sigma = \sqrt{\sum_{i=1}^{n_s} (\bar{\chi} - \chi_i)^2 / n_s}$	standard deviation with respect to the average
$\chi_{wsum} = \sum_{i=1}^{n_s} \chi_i n_i$	sum weighted by the number of each atom type	$\tilde{\chi}_\sigma = \sqrt{\sum_{i=1}^{n_s} (\tilde{\chi} - \chi_i)^2 / n_s}$	standard deviation with respect to the weighted average
$\chi_{range} = \chi_M - \chi_m$	range of values	$\chi_{coeff} = \bar{\chi}_\sigma / \bar{\chi}$	coefficient of variation
$\bar{\chi} = \chi_{sum} / n_s$ $\tilde{\chi} = \chi_{wsum} / N$	average value, average weighted value	$\chi_{wcoeff} = \tilde{\chi}_\sigma / \tilde{\chi}$	coefficient of variation with respect to the weighted items
$\chi_M = \mathrm{Max}(\chi_i)$	maximum value	$\|\chi\|_1 = \sum_{i=1}^{n_s} \lvert\chi_i n_i\rvert$	L1 norm with respect to the weighted values
$\chi_m = \mathrm{Min}(\chi_i)$	minimum value	$\|\chi\|_2 = \sqrt{\sum_{i=1}^{n_s} \chi_i^2 n_i}$	L2 norm with respect to the weighted values

2.2 2DMs-ResNet

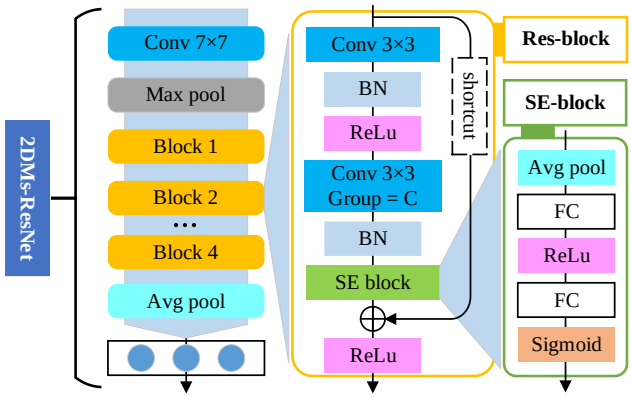

Fig. 2. The structure of 2DMs-ResNet.

Fig. 2 shows the 2DMs-ResNet network designed in this paper. According to the figure, the main body of the network consists of a convolutional layer (Conv 7×7), a Max pooling layer (Max pool), four residual blocks (Res-block), an average pooling layer (Avg pool), and a fully connected (FC) layer. The structure of the Res-block can also be known from the figure, which is mainly composed of two different types of convolutional layers (Conv 3×3) and a SE-block. In addition, it comes with the Batch Normalization Layer (BN) and ReLu activation unit common in CNN. SE-block is an attention mechanism module proposed in SENet [10], which can selectively strengthen important features and compress unimportant features in the network and enhance the extraction of features between channels. Its specific structure can also be clearly obtained from Fig. 2.

3 Experimental Design

To verify the effectiveness of the proposed method, it is a common and convincing means to conduct relevant experiments by selecting widely recognized benchmark datasets. Therefore, based on the above ideas, the experimental design of this paper is described below.

3.1 Benchmark

In this paper, the C2DB with accessed version (2018-12) is selected as the benchmark to realize the construction of 2D material thermodynamic stability prediction model. The database is of considerable size, containing structural, thermodynamic, elastic, electronic, magnetic, and optical properties of about 4000 2D materials distributed over more than 40 different crystal structures, and has been widely used in several excellent works. The properties are calculated by DFT and many-body perturbation theory.

The materials in the C2DB are divided into different prototype categories. In our experiment, we selected the AuSe, BN, BiTeI, C2F2, CdPS3, CdI2, CuI, CoH2O2, FeOCl, FeSe, GaS, GeS2, MoH2, MoS2, MoSSe, PbSe, PdSe2, ReS2, Ga2S2, TiS3, WTe2, and HfSCl. For the levels of thermodynamic stability, the database identifies them as numbers 1, 2, and 3, representing the three levels of low, medium, and high, respectively. The criterion of this level is defined by the corresponding formation energy (ΔH_f) and the energy above the convex hull (ΔH_{hull}) of the material, that is, the material with non-negative ΔH_f has low thermodynamic stability. For ΔH_f is negative and $\Delta H_{hull} > 0.2$eV/atom, the stability is considered medium, while stability is considered high if ΔH_{hull} is negative and $\Delta H_f < 0.2$eV/atom. Moreover, since magnetic order is an additional degree of freedom that affects the learning of structural stability, we restrict our study to non-magnetic materials. Finally, compounds with extreme behavior are considered during training, that is, those that are clearly defined within the limits of the low, medium, and high levels. Thus, we separate compounds with ΔH_f between -0.1 and 0.1 eV/atom and with ΔH_{hull} between 0.1 and 0.3 eV/atom. After the above screening process, we obtained 1,310 compounds for experiments and divided them into training, validation, and test data sets according to the ratio of 8:1:1.

3.2 Parameter Settings

In the 2DMs-ResNet structure, we first set the number and size of filters for Conv 7×7 to 32 and 7×7 with a step size of 2, and for Max pool to 3×3 with a step size of 2. Next, the number of filters for the next four Res-blocks is set to 32/64/128/256 with a step size of 1/2/2/2. Moreover, for each Res-block, the size of both filters is 3×3, and the group parameter of the second filter is set to 8. Finally, the number of neurons in the FC layer is set to 512 because its input is the concatenation of the Avg pool output and the prototype vector (the dimension of the embedding layer is set to 256). Based on the conventions of the deep learning community, during the training process of the 2DMs-ResNet, we use the AdamW optimizer and Cross-Entropy loss function; the batch size is set to 8; the learning rate (*lr*) is set to 1×10^{-3}; the learning rate update strategy is set as OneCycleLR, the maximum learning rate in the strategy is set as 5 times of the initial *lr*, the total number of iteration rounds is set as 270, and the proportion of the rising learning rate is 20%; the total number of network training epochs is set to 250.

4 Experimental Results and Analysis

Since this work is still in progress and the experimental design is insufficient in terms of adequacy, this section only presents and analyzes the experimental results of the preliminary design, which are presented and analyzed as follows.

4.1 Effectiveness of 2DMs-ResNet

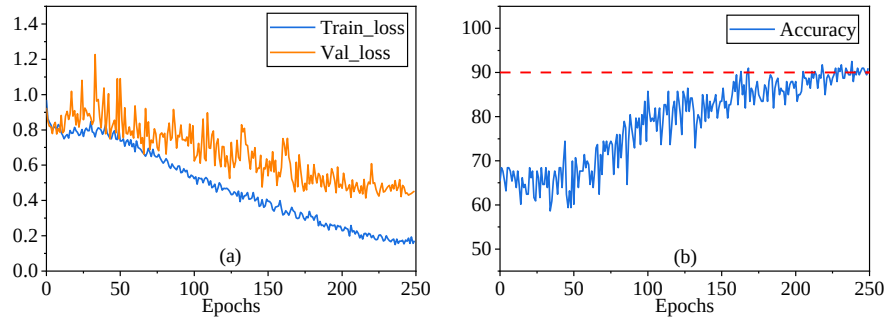

Fig. 3. Loss curve and accuracy curve.

Fig. 3(a) shows the Train_loss and Val_loss curves of 2DMs-ResNet during the training process. By analyzing the trend of the loss curves, we can see that 2DMs-ResNet not only fits the training set very well, but also performs quite well on the validation set. This shows that the method of constructing the space feature map and then using 2DMs-ResNet to predict the thermodynamic stability of materials has considerable feasibility and effectiveness. By the way, it also reflects the good feature extraction ability and generalization performance of the designed 2DMs-ResNet. In addition, during the training, we also continuously evaluate the performance of the model ob-

tained from each training epoch on the test set. From the accuracy curve shown in Fig. 3(b), we can conclude that the trained model has a good effect on predicting the thermodynamic stability of unknown 2D materials, and its prediction accuracy eventually exceeds 90%.

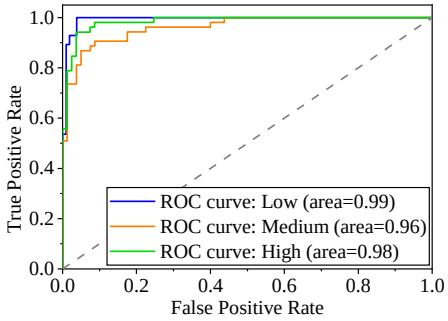

Fig. 4. ROC curve.

Receiver Characteristic Operator (ROC) curve is a common indicator used to verify the performance of a classifier model, which can well evaluate the overall classification effectiveness of the model. The ROC curve of the model constructed by the proposed method with respect to the test set is shown in Fig. 4. Through the analysis, it can be concluded that the trained model has good predictive ability for different levels of thermodynamic stability. However, from a subtle point of view, the prediction performance of the medium level is at a disadvantage compared to the other two levels. This indicates that the material with medium stability level has greater uncertainty, and it may also have similar characteristics with the other two levels, so it is more difficult to distinguish.

5 Summary

Aiming at the prediction of 2D material thermodynamic stability, this paper proposes a method to construct the material's feature space map and designs a lightweight residual neural network structure 2DMs-ResNet for realizing the prediction of material thermodynamic stability level. Preliminary experimental results show that the proposed method has excellent performance.

This work is still in progress. In the future, we will focus on improving the feature space we have constructed so far, as well as trying to incorporate some mathematical methods and optimization methods into our work. In addition, collecting and searching multiple different 2D material databases to expand the data set used in our experiments is also a feasible way for us to continue to improve the proposed method and enhance its robustness.

6 Funding

This research was supported by the Major Program of National Natural Science Foundation of China (72192830, 72192831), and the 111 Project (B16009).

References

1. Geim A K, Novoselov K S.: The rise of graphene. Nature materials 6(3), 183-191 (2007).
2. Mounet N, Gibertini M, Schwaller P, et al.: Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds. Nature nanotechnology 13(3), 246-252 (2018).
3. Haastrup S, Strange M, Pandey M, et al.: The Computational 2D Materials Database: high-throughput modeling and discovery of atomically thin crystals. 2D Materials 5(4), 042002 (2018).
4. Zhou J, Shen L, Costa M D, et al.: 2DMatPedia, an open computational database of two-dimensional materials from top-down and bottom-up approaches. Scientific data 6(1), 86 (2019).
5. Fukuda M, Zhang J, Lee Y T, et al.: A structure map for AB_2 type 2D materials using high-throughput DFT calculations. Materials Advances 2(13), 4392-4413 (2021).
6. Fabian B, Sajid A, Manti S, et al.: Quantum point defects in 2D materials-the QPOD database. NPJ Computational Materials 8(1), 56 (2022).
7. Yao M, Ji J, Li X, et al.: MatHub-2d: A database for transport in 2D materials and a demonstration of high-throughput computational screening for high-mobility 2D semiconducting materials. Science China Materials, 1-9 (2023).
8. Schmidt J, Shi J, Borlido P, et al.: Predicting the thermodynamic stability of solids combining density functional theory and machine learning. Chemistry of Materials 29(12), 5090-5103 (2017).
9. Li, Y., Yang, T.: Word Embedding for Understanding Natural Language: A Survey. In: Srinivasan, S. (eds) Guide to Big Data Applications. Studies in Big Data, vol 26, pp. 83-104. Springer, Cham (2018).
10. Hu J, Shen L, Albanie S, et al.: Squeeze-and-Excitation Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence 42(8), 2011-2023 (2019).