Supplementary Material for "Self-supervised graph neural networks for accurate prediction of Néel temperature"

Jian-Gang Kong[1], Qing-Xu Li[1,2], Jian Li[1,2,3], Yu Liu[4], Jia-Ji Zhu[1,2,3]

[1]School of Science, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[2]Institute for Advanced Sciences, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[3]Southwest Center for Theoretical Physics, Chongqing University, Chongqing 401331, China
[4]Inspur Electronic Information Industry Co., Ltd, Beijing 100085, China

In the supplementary material, we provide more detailed information about magnetic material datasets used in this work, and the equivalence between the self-supervised training of graph neural networks(GNN) and classification tasks.

1 Datasets

MAGNDATA[1, 2] is a high-quality experimental database with more than 1500 published commensurate and incommensurate antiferromagnetic materials, which contains rich information on the magnetic structure and experimentally measured Néel temperatures.

The Néel temperature full dataset constructed from MAGNDATA, excluding the disordered and incommensurate magnetic structures which can't be handled appropriately by pymatgen[3], contains only 1,007 entries, which is relatively small. While the Néel temperature sub-dataset of size 748 is obtained by keeping the materials with a unit cell of fewer than 60 atoms in the entire dataset.

The magnetic moment dataset of size 1816 is constructed from the Néel temperature dataset. Transition metals account for 67% of the total number of atoms in the dataset, and the most frequent elements of transition metals are Mn, Fe, Co, Ni, Cu, and Cr. In contrast, lanthanides account for 25% of the total atoms in the dataset, and the most frequent lanthanide elements are Nd and Tb. The average size of magnetic moments in the dataset is 3.05 μ_B, while the minimum is 0.04 μ_B and the maximum is 12.11μ_B, and 90% of the magnetic moments are smaller than 6 μ_B.

2 The self-supervised training of GNN

In CGCNN[4], the element information and distance between atoms are one-hot encoded in the node vectors and edge vectors, respectively. More specifically, for properties represented by discrete values, such as the period number and group number of elements, the vector is encoded according to the category to which the value belongs. On the other hand, for a property represented by continuous value, for example, the electronegativity of an element, the value will be discretized into ten bins, and the vector will be constructed accordingly.

Therefore, the self-supervised learning tasks of reproducing period number and group number of randomly masked atoms correspond to classification problems of 8 and 18 categories, respectively. We also randomly mask the edge information, the distance between atoms of the masked nodes, and train the neural network to predict the distance information, which corresponds to a classification problem of 41 categories.

We group the magnetic moments represented by continuous values into 64 discrete data bins to reduce the difficulty of the training and extract more information. As a result, the prediction of magnetic moments can also be regarded as a classification problem of 64 categories. The numbers of training epochs, mask ratio, and discrete bins are hyperparameters that can be fine-tuned for the best model.

References

[1] Gallego S V, Perez-Mato J M, Elcoro L, Tasci E S, Hanson Robert M, Momma K, Aroyo M I and Madariaga G 2016 J. Appl. Crystallogr. 49 1750

[2] Gallego S V, Perez-Mato J M, Elcoro L, Tasci E S, Hanson Robert M, Momma K, Aroyo M I and Madariaga G 2016 J. Appl. Crystallogr. 49 1941

[3] Ong S P, Richards W D, Jain A, Hautier G, Kocher M, Cholia S, Gunter D, Chevrier V L, Persson K A and Ceder G 2016 Comput. Mater. Sci. 68 314–319

[4] Xie T and Grossman J C 2018 Phys. Rev. Lett. 120 145301