Robustness Test of the spacegroupMining Space Group Determination Model for atomic pair distribution functions

Ling Lan, a Chia-Hao Liu, a* Qiang Du a,b* and Simon J. L. Billinge a,c*

a Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY 10027, United States, b Data Science Institute, Columbia University, New York, NY 10027, United States, and c Condensed Matter Physics and Materials Science Department, Brookhaven National Laboratory, Upton, NY 11973, United States. E-mail: sb2896@columbia.edu

Robustness Test, Machine Learning, Data Mining, Space Group, Pair distribution function

Abstract

Machine learning models based on convolutional neural network have been used for predicting space groups of crystal structures from their atomic pair distribution function (PDF). However, the PDFs used to train the model are calculated with respect to a fixed set of parameters, and the accuracy on PDFs generated with different choices of these parameters is unknown. In this paper, we report that the results of the top-1 accuracy and top-6 accuracy are robust when applied to PDFs of different choices of experimental parameters \( r_{\text{max}} \), \( Q_{\text{max}} \), and atomic displacement parameters.

1. Introduction

Recently it was shown (Liu et al., 2019) that a convolutional neural network (CNN) machine learning model could predict the space group of a material from its atomic...
pair distribution function (PDF) (Billinge, 2014; Egami & Billinge, 2012) with good accuracy.

The atomic pair distribution function (PDF) method is a total scattering technique for determining local order in nanostructured materials. Theoretically, the PDF gives the scaled probability of finding two atoms in a material a distance \( r \) apart and is related to the density of atom pairs in the material (Billinge, 2014; Egami & Billinge, 2012).

The model of Liu et al. (Liu et al., 2019) was trained, as shown in the red section in Figure 1, using calculated PDFs, denoted by \( G(r, \Omega) \) here, where \( \Omega \) indicates the set of parameters that define experimental details of the measurement and sample, such as \( Q_{\text{max}}, r \)-range of the PDF, and so on (Proffen & Billinge, 1999), and listed in table 1. However, in general, different parameter values that do not conform to the values used in the training might be relevant in a scientist’s measured PDFs. We denote these as \( G'(r, \Omega') \), where the prime on the \( \Omega \) indicates the set contains an experimental value that was not used in the training. A natural and important question is whether the trained model with \( G(r, \Omega) \) could provide reasonable predictions on materials associated with \( G'(r, \Omega') \). If the accuracy of the model predictions on \( G'(r, \Omega') \) is close to its performance on the PDFs, \( G(r, \Omega) \), that the model learned from, we believe that the model is robust. In this paper we assess how well the model performs when it is tested on PDFs that were calculated using experimental parameters different from those for the training set (blue section in Figure 1). We conclude that overall it performs quite well with respect to \( Q_{\text{max}}, r_{\text{max}} \) and atomic displacement factor (ADP), or \( U_{\text{iso}} \), of the measurement, hence providing evidence to the robustness of the CNN machine learning model developed in (Liu et al., 2019).
Fig. 1. $G(r, \Omega)$ are the calculated PDFs w.r.t. the original experimental parameters used to train the CNN model, while $G'(r, \Omega')$ are PDFs with varying parameters. $p$ and $p'$ are the corresponding model outputs from which we could measure the accuracy of the model predictions. The red (blue) section depicts the training (testing) process respectively.

2. Method

2.1. Data and Model

Our main objective is to test the robustness of the originally trained model. However, we are not able to identity the exact datasets that constituted the training set in the original training. In order to avoid testing robustness by inadvertently using a dataset that might be part of the original model’s training data, our first step is to rebuild the model again.

The input PDF data are calculated from 98,830 structures in the 45 most heavily represented space groups in the ICSD (Belsky et al., 2002) structural database. The PDFs are calculated from crystallographic information framework (CIF) (Hall et al., 1991) files obtained from ICSD using the diffpy-cmi (Juhás et al., 2015) package with parameters $\Omega$ defined in Table 1. The parameters are the same as the ones used in (Liu IUCr macros version 2.1.10: 2016/01/28
et al., 2019), except that the grid size is $\pi/40$ in our experiment (the paper used $\pi/23$), so that we could calculate the PDF with higher $Q_{\text{max}}$’s. 80% of the data is considered as training samples, and the rest is treated as test samples. The choice of $r_{\text{min}}$, $r_{\text{max}}$, and $r_{\text{grid}}$ in Table 1 discretizes the input PDFs to 1D signal sequences of dimension $209 \times 1 \times 1$. We further normalize the PDF input, $G(r, \Omega)$, to ensure that it lies between 0 and 1 for each entry.

Table 1. Experimental parameters used to calculate PDFs used for training, validation and testing of the model.

| Parameter         | Value |
|-------------------|-------|
| $r_{\text{min}}$ (Å) | 1.5   |
| $r_{\text{max}}$ (Å) | 30.0  |
| $r_{\text{grid}}$ (Å) | $\pi/40$ |
| $Q_{\text{min}}$ (Å$^{-1}$) | 0.5   |
| $Q_{\text{max}}$ (Å$^{-1}$) | 23.0  |
| $U_{\text{iso}}$ (Å$^2$)     | 0.008 |
| $Q_{\text{damp}}$ (Å$^{-1}$) | 0.04  |
| $Q_{\text{broad}}$ (Å$^{-1}$) | 0.01  |

To rebuild the model, we use the architecture based on the convolutional neural network (CNN) used in Liu’s paper (Liu et al., 2019). The output, $p$, of the model is a $45 \times 1$ vector, which represents the probability of the input PDF being in each of the 45 space groups considered in our study. We use weighted categorical cross entropy loss,

$$\text{Loss} = -\sum_{i=1}^{45} w_i \cdot p_{\text{true}_i} \cdot \log p_i,$$

(1)

to mitigate the effects of unbalanced data, where the weight $w_i$ is defined as the number of structures in the training set over the number of structures of each space group in the training dataset. Adaptive moment estimation (Adam) with a mini-batch size of 64 is used to train the model. Furthermore, we modify the learning rate as an exponential decay, $l = 5 \times 10^{-4} e^{-0.025 \times \text{epoch}}$. The model is trained using Keras on a single Nvidia Tesla P100 GPU.

An accuracy of 67.7% from top-1 prediction and 90.2% from top-6 predictions is achieved. The performance of our reconstructed model is similar to the one shown in
the original paper, which was 70.0% top-1 accuracy and 91.9% top-6 accuracy. The model rebuilt here is used, without any further retraining, in subsequent robustness tests on datasets involving PDFs having different parameter values, as illustrated schematically in Fig. 1.

2.2. Robustness Test

In order to test the robustness, we consider the three experimental parameters that are used to calculate the PDFs from the structural CIFs, which are $Q_{\text{max}}$, $r_{\text{max}}$, and $U_{\text{iso}}$. We randomly choose structures from the testing set (10% of the testing samples are chosen), and compute their PDFs while varying each of these parameters separately between limits that are experimentally reasonable. These calculated PDFs are then given to the trained model, without model retraining despite the changes in parameter values of the input PDFs, to predict the space group, and the model accuracy is computed as a function of the experimental parameter value.

First we consider the robustness against a variation in $r_{\text{max}}$. The model was trained with an $r_{\text{max}}$ of 30 Å and we want to test its performance when given PDFs computed (or measured) over a narrower $r$-range, from 10 Å to 30 Å every 2 Å. Variations in $r_{\text{max}}$ will change the length of the PDF vector, which is not allowed in our model. Since we are only considering $r$-ranges that are shorter than 30 Å, to keep the dimension of all input PDFs consistent, the data are padded with zero’s up to the value of $r_{\text{max}} = 30$ Å before being interpolated on to the $209 \times 1 \times 1$ grid using quadratic interpolation.

To test the $Q_{\text{max}}$ sensitivity, computed PDFs in the range of $12 \leq Q_{\text{max}} \leq 30$ Å$^{-1}$ in steps of 3 Å$^{-1}$ were tested against the trained model, and for the ADP, $U_{\text{iso}}$, from $0.005 \leq U_{\text{iso}} \leq 0.01$ Å$^2$ in steps of 0.001 Å$^2$, where the model was trained on values $Q_{\text{max}} = 23$ Å$^{-1}$ and $U_{\text{iso}} = 0.008$ Å$^2$, respectively.
3. Results

3.1. Robustness with respect to $r_{\text{max}}$

Figure 2 shows the top-6 accuracy against a variation in $r_{\text{max}}$ from 10 Å to 30 Å. When $r_{\text{max}}$ is larger than 20 Å, top-6 accuracy is always above 87.1%, which is close to the optimal value of 90.2%. It is recommended to give the model a PDF with a $r_{\text{max}} \geq 30$ Å. However, the robustness test shows that if the signal is from data over a narrower range, such as a nanoparticle whose signal dies on a shorter length-scale, the model can still be categorized into space group with reasonably good accuracy, though the performance drops off more quickly below an $r_{\text{max}}$ of 20 Å or so.

![Fig. 2. Top-6 accuracy as $r_{\text{max}}$ is varied. The value of $r_{\text{max}}$ used to train the model is shown as a vertical green dashed line.](image)

3.2. Robustness with respect to $Q_{\text{max}}$

Next we consider the robustness of the model when PDFs are generated using different $Q_{\text{max}}$ values. As shown in Figure 3, when $Q_{\text{max}}$ is larger than 18 Å$^{-1}$, top-6 accuracy is above 81.1%. The bump around 23 Å$^{-1}$ makes sense, as the model favors...
the $Q_{\text{max}}$ value that it is trained on. But the performance with $Q_{\text{max}}$ value deviated from 23 Å$^{-1}$ is still fairly good over the entire range of values considered, the accuracy never falls below 77.7%, and so the model is quite robust against variations in $Q_{\text{max}}$.

![Graph](image.png)

Fig. 3. Top-6 accuracy as $Q_{\text{max}}$ is varied. The value of $Q_{\text{max}}$ used to train the model is shown as a vertical green dashed line.

3.3. Robustness with respect to Atomic Displacement Parameter, $U_{\text{iso}}$

Finally (Fig. 4), we consider robustness against variations in $U_{\text{iso}}$. The results are even less sensitive to the choice of ADP. When $U_{\text{iso}}$ of the PDFs were in the range 0.005 Å$^2$ to 0.01 Å$^2$, the top-6 accuracy is always above 87.3%.
Fig. 4. Top-6 accuracy as the ADP, $U_{iso}$, is varied. The value of $U_{iso}$ used to train the model is shown as a vertical green dashed line.

The numbers from all the robustness tests are reproduced in the supplementary information associated with this paper.

4. Conclusions

The use of deep learning to do complex classifications from data is a potentially useful approach that is becoming more widespread in materials science, crystallography and diffraction. Inherent in the process is that the model was trained on a particular set of data and its applicability to do the classification on data that is, in some way, different, for example, measured with different resolutions or over different ranges, might limit its ability to make accurate predictions. In general, the model may be retrained on a wider set of data that incorporates cases of different ranges, resolutions and so on. However, here, for the case of the spacegroupMining model that is deployed on pdfitc.org, we
simply explored its robustness in making accurate predictions on different range and resolution data without retraining the model. The main result is that the model is quite robust and performs well without having to be retrained in most cases. Modest reductions in prediction accuracy were observed, but it still performed well given a rather wide, but reasonable, range of resolution and range parameters, suggesting that it is not of great urgency to retrain it. We note that retraining it with a more diverse set of training data, whilst increasing accuracy for parameter values away from the original training values, it may decrease the prediction accuracy for PDFs with the original set of parameter values, where those values were chosen as being somewhat representative of values in many rapid acquisition x-ray PDF studies. Through this work, it has been shown that, without additional retraining, the spacegroupMining@pdfitc model still performs with reasonable accuracy for a relatively wide range of experimental parameters, and can thus be used as a robust computational tool.

Funding information

This work in the Billinge group was supported by the U.S. National Science Foundation through grant DMREF-1922234 and CCF-1704833.

References

Belsky, A., Hellenbrandt, M., Karen, V. L. & Luksch, P. (2002). Acta Crystallographica Section B Structural Science, 58(3), 364–369.

Billinge, S. J. L. (2014). In International Tables of Crystallography, edited by C. Gilmore et al., vol. H. Buffalo, NY: International Union of Crystallography.

Egami, T. & Billinge, S. J. L. (2012). Underneath the Bragg Peaks: Structural Analysis of Complex Materials. No. 16 in Pergamon Materials Series. Amsterdam: Elsevier, 2nd ed.

Hall, S. R., Allen, F. H. & Brown, I. D. (1991). Acta Crystallographica Section A: Foundations of Crystallography, 47(6), 655–685.

Juhás, P., Farrow, C., Yang, X., Knox, K. & Billinge, S. (2015). Acta Crystallographica Section A: Foundations and Advances, 71(6), 562–568.

Liu, C.-H., Tao, Y., Hsu, D., Du, Q. & Billinge, S. J. L. (2019). Acta Crystallographica Section A: Foundations and Advances, 75(4), 633–643.

Proffen, T. & Billinge, S. J. L. (1999). J. Appl. Crystallogr. 32, 572–575.
Supplemental Materials

Table 2. Top-6 accuracy and top-1 accuracy when $r_{\text{max}}$ is chosen from 10 Å to 30 Å.

| $r_{\text{max}}$ (Å) | 10  | 12  | 14  | 16  | 18  | 20  | 22  | 24  | 26  | 28  | 30  |
|---------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Top-6 accuracy      | 0.648 | 0.717 | 0.747 | 0.778 | 0.820 | 0.871 | 0.882 | 0.890 | 0.901 | 0.901 | **0.902** |
| Top-1 accuracy      | 0.285 | 0.367 | 0.433 | 0.449 | 0.511 | 0.552 | 0.600 | 0.617 | 0.652 | 0.671 | **0.677** |

Table 3. Top-6 accuracy and top-1 accuracy when $Q_{\text{max}}$ is chosen from 12 Å$^{-1}$ to 30 Å$^{-1}$.

| $Q_{\text{max}}$ (Å$^{-1}$) | 12  | 15  | 18  | 21  | 23  | 24  | 27  | 30  |
|-----------------------------|-----|-----|-----|-----|-----|-----|-----|-----|
| Top-6 accuracy              | 0.777 | 0.795 | 0.811 | 0.822 | **0.902** | 0.900 | 0.805 | 0.84  |
| Top-1 accuracy              | 0.516 | 0.591 | 0.597 | 0.604 | **0.677** | 0.663 | 0.598 | 0.610 |

Table 4. Top-6 accuracy and top-1 accuracy when ADP is chosen from 0.005 Å$^2$ to 0.01 Å$^2$.

| adp (Å$^2$) | 0.005 | 0.006 | 0.007 | **0.008** | 0.009 | 0.01 |
|-------------|-------|-------|-------|-----------|-------|------|
| Top-6 accuracy | 0.873 | 0.895 | 0.904 | **0.902** | 0.900 | 0.900 |
| Top-1 accuracy | 0.618 | 0.649 | 0.664 | **0.677** | 0.666 | 0.641 |