Machine-Learning-Accelerated Data Analysis for X-Ray Diffraction

Identification of crystal structures is one of the core objectives of materials science. Mathematical optimization and machine-learning techniques have allowed rapid estimations of crystal structures from powder X-ray diffraction data. It has been suggested that these techniques can help concretize the rules of thumb for measured data analyses based on large amounts of materials data and obtain candidate crystal structures that are unaffected by bias of the analyst. These techniques also help researchers solve more complex problems more quickly.

Rapid advances in materials informatics, the application of machine learning (ML) and information processing techniques to materials science are changing the field of materials science. In contrast to physical measurements, which have long been regarded as targets for efficiency improvement, data analysis has largely remained dependent on trial and error by humans. Consequently, data analysis has become a time bottleneck in the development of materials. Herein, we introduce some recent studies on ML-accelerated data analysis methods for quantum beam experiments.

Powder X-ray diffraction (PXRD) is one of the essential tools used in materials analysis. Rietveld analysis, a method to obtain a precise crystal structure model with curve fitting to the measured PXRD patterns, is a popular method for PXRD pattern analysis.

Although Rietveld analysis has a long history and is a reliable method, it requires setting numerous parameters independent of the properties of the material, such as the choice of a background function and the setting of peak shape. Because these settings significantly impact the results, trial and error and expertise are required to obtain appropriate analysis results. We realized that this situation is similar to hyperparameter optimization for ML models and formulated the trial-and-error process of Rietveld analysis as a black box optimization (BBO) problem to find a setting that minimizes the weighted profile residual factor (R_{wp}) [1]. As the optimization target by BBO, we chose 13 parameters (e.g., background function and variables to be fitted) and solved this optimization problem using the tree-structured Parzen estimator (TPE), an algorithm based on Bayesian optimization.

The optimization history of Y_2O_3 is shown in Fig. 1. The value of R_{wp} decreased with the progress of optimization by the TPE, and, on average, our proposed method exceeded human expertise at ~100 evaluations. In each experiment. The calculation time required for 100 evaluations was ~30 min on a general computer. Moreover, in 100 experiments with different random seeds to run TPE, we confirmed that $90R_{wp}$ was less than the result obtained by the expert. These results demonstrate that the proposed method based on BBO has the potential to realize automatic Rietveld analysis with the same or better performance than that of experts.

Similarly, for $Dy_{0.5}Sr_{0.5}MnO_3$, the experiment was repeated 100 times with different random seeds used to run the TPE. Using multidimensional scaling, we visualized the similarity between the crystal structures obtained using BBO (Fig. 2). Each point corresponds to one crystal structure, and the distance in the figure corresponds to the similarity of the structures. In the lower right of the figure, crystal structures similar to the results obtained by the expert formed a loose cluster. This result indicates that structures comparable to the expert's results could be obtained automatically. In addition, there is an outlier structure in the upper left corner of Fig. 2, where R_{wp} converges well enough. The difference from the other structures, that is, the positional shift corresponding to 5% of the lattice parameter, is considered sufficient to affect the physical properties. Although conventional criteria would reject this structure because isotropic atomic displacement factors (U_{iso}) of atoms with similar masses are not comparable, the ability to propose quite different structure candidates is a major feature of the BBO approach. We contend that this approach removes the human-origin bias and may lead to new discoveries that have been overlooked in the past.

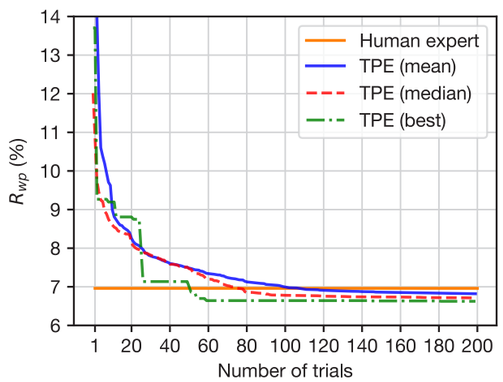

Figure 1: Optimization history for Y_2O_3. The curves show the optimization history of R_{wp} for Y_2O_3. Our proposed method exceeded the human expert at ~100 evaluations [1].

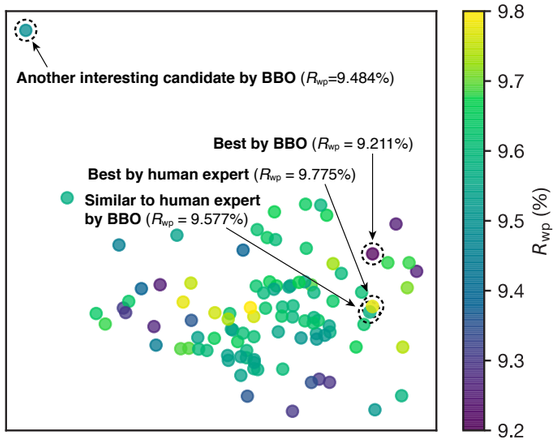

Figure: 2 Multidimensional scaling visualization. Each point in this figure represents a crystal structure refined with the 100 configurations optimized by BBO (i.e., 100 crystal structures with the best R_{wp} among 200 configurations from each of 100 runs) or the crystal structure refined with the best configuration by a human expert [1].

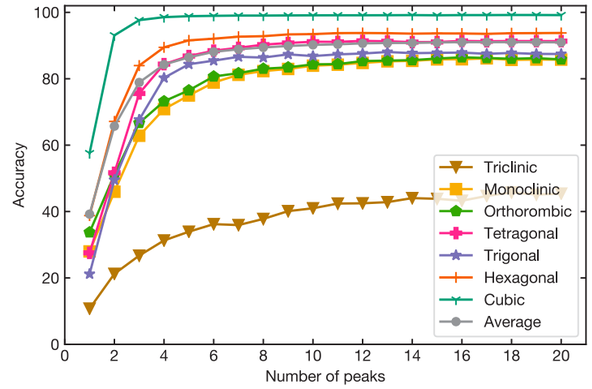

Figure: 3 Classification accuracy of crystal system versus number of peak positions used for prediction [2].

As a different approach, we proposed a method to predict crystal structure features directly from PXRD patterns using ML without using Rietveld analysis [2]. In this study, we developed a technique that can predict the crystal system and space group, which are essential features of the crystal structure, with an accuracy of >90% and >80%, respectively, and in <1 ms. This performance is sufficient for screening a large amount of data.

We not only developed a prediction model but also gained data-driven insight, such as concerning the effect of the number of diffraction peaks used for prediction of the prediction accuracy (Fig. 3) and the importance of data features. Furthermore, from the learned ML model analysis, we quantified some of the tacit knowledge used by expert users to predict the approximate crystal system when they look at the PXRD pattern. Such knowledge may help in designing a minimum but effective measurement setup and an experimental plan for a specific purpose.

ML-accelerated data analysis in quantum beam experiments is developing rapidly and will help researchers focus on more creative work and solve more complex problems.

REFERENCES

[1] Y. Ozaki, Y. Suzuki, T. Hawai, K. Saito, M. Onishi and K. Ono, *npj Comput. Mater.*, **6**, 75 (2020)
[2] Y. Suzuki, H. Hino, T. Hawai, K. Saito, M. Kotsugi and K. Ono, *Sci. Rep.*, **10**, 21790 (2020).

Y. Suzuki[1], Y. Ozaki[2, 3], T. Hawai[4], K. Saito[4, 5], K. Ono[4, 6], M. Onishi[2] ([1]SOKENDAI, [2]AIST-AIRC, [3]GREE, Inc., [4]KEK-IMSS, [5]Randeft Inc., [6]Osaka Univ.)