ARTICLE OPEN

Automated crystal structure analysis based on blackbox optimisation

Yoshihiko Ozaki^{1,2}, Yuta Suzuki^{3,4}, Takafumi Hawai³, Kotaro Saito^{3,5,6}, Masaki Onishi¹ and Kanta Ono^{3,4}✉

In the present study, we show that time-consuming manual tuning of parameters in the Rietveld method, one of the most frequently used crystal structure analysis methods in materials science, can be automated by considering the entire trial-and-error process as a blackbox optimisation problem. The automation is successfully achieved using Bayesian optimisation, which outperforms both a human expert and an expert-system type automation despite the absence of expertise. This approach stabilises the analysis quality by eliminating human-origin variance and bias, and can be applied to various analysis methods in other areas which also suffer from similar tiresome and unsystematic manual tuning of extrinsic parameters and human-origin variance and bias.

npj Computational Materials (2020)6:75; <https://doi.org/10.1038/s41524-020-0330-9>

INTRODUCTION

The physical properties of materials are governed by their crystal structure; thus, crystal structure analysis is an indispensable element in materials science research^{1,2}. Compared to the drastic improvement in material fabrication and measurement throughput, the throughput of crystal structure analysis has not been improved because the analysis heavily relies on manual time-consuming trial-and-error methods^{3–8}. Rietveld refinement or the Rietveld method^{9,10}, one of the most widely used crystal structure analysis methods for powder diffraction data, such as X-ray diffraction (XRD) and neutron diffraction, faces this problem as well.

Rietveld refinement involves complex curve fitting using an experimental result and a simulation based on a physical model. The physical model comprises intrinsic parameters for governing crystal structure such as lattice constants and atomic positions and extrinsic parameters such as background signals and peak shape defined by instrumental resolution. The method is designed to minimise the difference between the observed and simulated results by updating the parameters in a given physical model until the difference reaches a threshold. All parameters need initial values, including categorical ones such as background and peak shape functions. For each numerical parameter, a binary parameter (fix or refine) to control the scope of the refinement needs to be set. The number of intrinsic parameters depends on the complexity of the crystal structure and ranges from 10 to 100, while that of extrinsic parameters is typically less than 20.

It is commonly known that refining all parameters at once often leads to physically unreasonable results. The basic strategy to avoid such an issue is to gradually increase the number of parameters to be refined; that is, beginning with a few to finally incorporate all in order to keep them physically acceptable. Consequently, one needs to update the scope of the refinement at each step of the refinement process and, eventually, to determine the sequence of scopes. To guide researchers among a massive number of possible combinations and sequences of different refinement scopes with several tens of refinable parameters,

empirical guidelines which significantly limit the size of "scope space" are often employed. However, even with the limited scope space, one has to search the right path to a reasonable solution in the search space by trial and error; furthermore, it is not guaranteed whether the guidelines will lead researchers to the optimal crystal structure in the whole scope space. Considering the wide use of Rietveld refinement in various materials science communities, this situation, that only proficient experts can exploit Rietveld refinement properly, should be improved.

Such complexity in parameter tuning would have consequences on analysis quality in terms of both variance and bias. For example, the consistency in experts' results shown by round robin tests in refs. ^{11–13} would have declined due to a wider distribution of individual skills to deal with the complexity if more researchers had participated in the tests. On the contrary, the consistency itself indicates the success of the refinement guidelines shared by experts. While such guidelines contribute to the consistency as expected, they can also cause a systematic bias in refinement results. In this study, based on blackbox optimisation (BBO)¹⁴, we offer a general automation framework for Rietveld refinement to alleviate these problems and allow practitioners to spend more time examining candidate structures proposed by the framework, and not searching for them in a vast parameter space.

BBO refers to a problem setup, in which the objective function and/or constraint function are given by blackboxes, i.e., the analytic forms of the function are unknown. BBO has recently emerged as a popular concept in the machine learning community because it performs an important task of tuning hyperparameters in machine learning models automatically, which is termed as hyperparameter optimisation (HPO)¹⁵. HPO plays a vital role in machine learning; (1) it reduces the human effort necessary for applying machine learning, (2) it improves the performance of machine learning algorithms, and (3) it facilitates fair comparisons among the results of machine learning studies by providing the same level of tuning¹⁵. HPO is formalised as the maximisation of machine learning model performance in the hyperparameter configuration space. For this case, the relationship between a hyperparameter configuration and the machine learning model

¹Artificial Intelligence Research Center (AIRC), National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan. ²GREE, Inc., Tokyo, Japan. ³Institute of Materials Structure Science (IMSS), High Energy Accelerator Research Organization (KEK), Ibaraki, Japan. ⁴SOKENDAI (The Graduate University for Advanced Studies), Ibaraki, Japan. ⁵Paul Scherrer Institute, Villigen, Switzerland. ⁶Present address: Medley, Inc, Tokyo, Japan. ✉email: kanta.ono@kek.jp

performance is considered as a blackbox. At present, Bayesian optimisation¹⁶, a global search method for BBO, is the most successful in this HPO problem^{17–21}. This topic is closely related to our focus, i.e., to stabilise the analysis quality and bypassing unsystematic (and often unexciting) tasks in scientific studies. Herein, we realised that the concept of HPO can be applied to Rietveld refinement.

Whereas numerous studies have been conducted on Rietveld refinement^{9,10,22–33} and software^{34–39}, up to now, the number of studies that aim to address aforementioned problems is very limited. Powder 3D Parametric⁴⁰ is a semi-automated software toolkit for Rietveld refinement based on user-provided configuration. This software is particularly designed for sequential analysis of a large number of data such as time-resolved measurement. SrRietveld^{41,42}, an automated software toolkit for Rietveld refinement, provides several pre-written scripts for automating the common tasks in the method. AutoFP⁴³ uses an expert system algorithm to simulate the manual process performed by human experts. The expert system consists of simple if-then rules, based on a collection of empirical strategies and simple ideas. Recently, PowderBot⁴⁴, which applies reinforcement learning⁴⁵ to automate Rietveld refinement, has been proposed. Reinforcement learning is suitable for solving problems with stochastic state transitions which are formulated as a Markov decision process. This method is famous for its great success in playing games like Go. While the reinforcement learning-based approach is conceptually interesting, it seems that another simpler approach such as BBO is sufficient for our purpose because Rietveld refinement is deterministic. In single-crystal diffraction, SCAR⁴⁶, machine-learning-based automated crystal structure analysis software has been proposed very recently. It supports precise analysis and is applicable to various single-crystal diffraction data, and is of interest in the field of crystal structure analysis. However, in powder X-ray diffraction, there is still room to apply ideas from machine learning. To the best of our knowledge, no attempt has been made to apply the BBO approach to Rietveld refinement for automating the time-consuming trial-and-error process and search beyond the boundaries defined by the common sense of human experts.

RESULTS AND DISCUSSION

Formalisation

Here, we introduce our "BBO-Rietveld" approach that automates the tuning of the refinement scope in the Rietveld method in the same manner as previously done in HPO. Fig. 1 illustrates the overview of the BBO-Rietveld approach described below, and Table 1 shows the list of optimised parameters. Most of them are

for defining the scope and an extrinsic categorical parameter for the background function. Note that we did not perform numerical initial parameter search contrary to the possible expectation that some of the readers would have.

In Rietveld refinement, the weighted profile residual factor (R_{wp}), which is the difference between calculated and observed diffraction intensity, is widely used as a metric to evaluate the Rietveld refinement result²⁹. Let $\mathbf{x}$ be a fixed parameter configuration of Rietveld refinement including the scope of the refinement, y_i^{Obs} be the i -th observed diffraction intensity, $y_{\mathbf{x},i}^{Calc}$ be the diffraction intensity calculated using the physical model with the configuration $\mathbf{x}$, and w_i be the weight corresponding to the inverse of the uncertainty of i -th intensity. Then, R_{wp} is defined as follows:

$$R_{wp}(\mathbf{x}) = \left(\frac{\sum_i w_i (y_i^{Obs} - y_{\mathbf{x},i}^{Calc})^2}{\sum_i w_i (y_i^{Obs})^2} \right)^{1/2}. \quad (1)$$

The crystal structure with the minimum R_{wp} is the most probable structure based on the given experimental data.

Our task can be formalised as the following BBO problem:

minimise $R_{wp}(\mathbf{x})$,

subject to $\mathbf{c}(\mathbf{x}) \geq 0$,

$\mathbf{x} \in \mathbf{X}$, (2)

where $\mathbf{X}$ is the configuration space and $\mathbf{c}(\mathbf{x})$ is the set of constraint functions. In the problem, a configuration $\mathbf{x}$ is considered as a variable and it affects R_{wp} through $y_{\mathbf{x},i}^{Calc}$. The constraints are set to ensure that the refined parameters are physically valid, or sometimes to limit the search space and reduce manual trial-and-error attempts with a plausible explanation. In this study, we limit ourselves to introduce the latter type of constraints and we only introduce an essential constraint to keep human-origin biases to a minimum (see Methods section for more details).

By solving the problem, we obtain the best configuration $\mathbf{x}^* = \text{argmin}_{\mathbf{x}} R_{wp}(\mathbf{x})$, which achieves the best fit without manual process.

Additionally, goodness-of-fit (GOF) is used as a metric to evaluate the optimised Rietveld refinement result. The definition of GOF is given as follows:

$$\text{GOF} = \frac{\sum_i w_i (y_i^{Obs} - y_{\mathbf{x},i}^{Calc})^2}{N_{obs} - p}, \quad (3)$$

where N_{obs} is the number of data points of observed XRD pattern,

Fig. 1 Overview of the BBO-Rietveld approach. An optimiser iteratively samples a new promising configuration and then runs a Rietveld refinement software with the sampled configuration. A refined crystal structure that achieves the best fit is obtained as a final output.

Table 1. List of optimised parameters. The parameter type "conditional" means that the parameter search space depended on the other parameters.

Parameter	Parameter type	Search space
Initial 2θ bound refinement	Categorical	{True, False}
Initial 2θ lower bound (deg.)	Real	[15, 130] for Y_2O_3 , [15, 80] for DSMO
Initial 2θ upper bound (deg.)	Real/conditional	[$l + 20$, 150] for Y_2O_3 , [$l + 20$, 100] for DSMO (l is the initial 2θ lower bound)
Background function	Categorical	{Chebyshev, cosine, Q^2 powerseries, Q^{-2} powerseries, linearinterpolate, inverseinterpolate, loginterpolate}
Degree of background function	Integral	[1, 15]
Background refinement	Categorical	{True, False}
Instrument zero position refinement	Categorical	{True, False}
Sample shift X refinement	Categorical	{True, False}
Sample shift Y refinement	Categorical	{True, False}
Scale factor refinement	Categorical	{True, False}
Peak shape refinement (Gaussian: U, V, W)	Categorical	{True, False}
Peak shape refinement (Lorentzian: X, Y)	Categorical	{True, False}
Peak shape refinement (SH/L)	Categorical	{True, False}

Fig. 2 Histograms of R_{wp} with optimised configurations. **a** Histograms of R_{wp} for Y_2O_3 . **b** Histograms of R_{wp} for DSMO. The values for AutoFP are taken from ref. ⁴³. The result of AutoFP for DSMO, $R_{wp} = 12.8\%$, is omitted.

and p is the number of variables in the fitting model. This metric is called as GOF in GSAS-II³⁹ and χ^2 in FullProf³⁵.

It should be noted that small R_{wp} or GOF does not guarantee that the corresponding refined crystal structure would be appropriate for the obtained experimental data. The examination of the analysis result is still an important task for human experts²⁹.

Experiments

To evaluate the proposed BBO-Rietveld approach, we optimised parameter configurations of Rietveld refinement for XRD patterns of Y_2O_3 , $\text{Dy}_{0.5}\text{Sr}_{0.5}\text{MnO}_3$ (DSMO) and LiCoO_2 which are chosen as benchmark materials. The tree-structured Parzen estimator (TPE)^{17,20}, one of the variants of Bayesian optimisation, is used as the optimiser. The details of our optimisation system are provided in the Methods section. Here we discuss the results of Y_2O_3 and DSMO. Readers who are interested in the result of LiCoO_2 should refer to the supplementary resources provided in the BBO-Rietveld repository (see Code availability). For each optimisation run, the optimiser iteratively evaluated 200 configurations including the initial 20 random configurations and returned the best one with the minimum R_{wp} . The search spaces for each parameter are shown in Table 1. The initial crystal structure models were taken from crystallographic information files (CIFs) on the GitHub repository of AutoFP⁴⁷, presumably those used in ref. ⁴³. To evaluate statistical property and reproducibility, optimisations have been performed 100 times with different random seeds each for Y_2O_3 and DSMO. Histograms of 100 R_{wp} values for both materials are shown in Fig. 2. Reference values for

a human expert and AutoFP are also shown for comparison. The latter ones are taken from the previous study⁴³.

Surprisingly, 90% and 99% of our optimisation runs for both test datasets, Y_2O_3 and DSMO, achieve better R_{wp} than the human expert and AutoFP, although no additional expertise in Rietveld refinement is used other than the conventional refinement procedure (see Methods section for details). Fig. 3 shows the refinement results of the best configuration derived by BBO for each benchmark material. We cannot find apparent flaws in the results refined with the best configuration. The changes of R_{wp} and goodness-of-fit (GOF) are shown in Fig. 4 (Y_2O_3) and 5 (DSMO). Both R_{wp} and GOF were improved as optimisation progresses, and in both metrics, our proposed method exceeded human expertise at around 100 evaluations. R_{wp} and GOF will be further improved if we continue the optimisation.

To visualise the refined structure features (x , y , z positions and U_{iso} of each atom) in two-dimensional space, we utilise the multidimensional scaling (MDS)^{48,49}, one of the dimensionality reduction algorithms. MDS can represent high-dimensional data in a low-dimensional space by approximating the distance in the original space. The number of original structure feature dimension is 16 for Y_2O_3 and 20 for DSMO. The MDS visualisation for DSMO is shown in Fig. 6. Each point in this figure represents a refined crystal structure, and the distance between points indicates their similarity. The majority of the refined structures form a loose cluster at the lower-right side of the figure, which means they are similar to each other. In the cluster, there are three special points; the best one from BBO, the one from the human expert, and the one from BBO and very close to the human expert result as shown

Fig. 3 Rietveld plots. **a** Rietveld plot of Y_2O_3 with the best configuration. **b** Rietveld plot of DMSO with the best configuration. The refinements with the best configurations achieve good fits for Y_2O_3 ($R_{\text{wp}} = 6.610\%$, $\text{GOF} = 1.748$) and DMSO ($R_{\text{wp}} = 9.211\%$, $\text{GOF} = 2.130$).

Fig. 4 Optimisation history for Y_2O_3 . **a** Optimisation history of R_{wp} . **b** Optimisation history of GOF. The curves show the optimisation history of R_{wp} and the optimisation history of GOF for Y_2O_3 . In both metrics, our proposed method exceeded the human expert at around 100 evaluations.

Fig. 5 Optimisation history for DMSO. **(a)** Optimisation history of R_{wp} . **(b)** Optimisation history of GOF. The curves show the optimisation history of R_{wp} and the optimisation history of GOF for DMSO. The result of AutoFP in **(b)**, $\text{GOF} = 3.56$, is omitted. In both metrics, our proposed method exceeded the human expert at around 100 evaluations as same as the case of Y_2O_3 .

in Fig. 6. While the best structure obtained by BBO (Table 2) scores smaller R_{wp} than that of the human expert result (Table 2), U_{iso} of O1 atom in the BBO result is noticeably smaller than that of O2 atom, which violates a conventional criterion that U_{iso} of atoms with similar mass should be comparable. On the other hand, the third special point, obtained by BBO and close to the human expert result, passes the criterion (Table 2). This demonstrates that BBO efficiently obtains candidates with small R_{wp} values but additional examinations are still required to choose the most preferable one.

Next, we discuss the interesting outlier at the upper left of Fig. 6. The outlier (Table 2) has good converged R_{wp} , but the x position of

O1 atom in the refined structure is considerably different from others. The difference exceeds the uncertainty calculated by the software, and, from the structural point of view, the positional shift corresponding to 5% of the lattice parameter would be enough to affect physical properties. This implies that the outlier corresponds to a local minimum not belonging to the cluster discussed above. In such a situation, multiple crystal structure candidates can explain the experimental data sufficiently with a similar level, which may affect the conclusion of a study or evoke a new discussion. Despite that it meets the constraint in the optimisation (positive U_{iso}), the outlier can be rejected due to the violation of the conventional criterion for U_{iso} as well as the best result by BBO.

Fig. 6 MDS visualisation. Each point in this figure represents a crystal structure refined with the 100 configurations optimised 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 the human expert.

We believe that imposing constraints other than universal physical requirements in BBO can result in biased optimisation of results toward human expectations. Experts can examine the list of candidates using conventional criteria and other knowledge, and can eliminate inappropriate solutions at any time. This ability to propose multiple candidates without human effort and bias is a great advantage of automation and may lead to the discovery of hitherto unnoticed new knowledge beyond the standard practices of conventional manual tuning and the expert system simulating the practices. The use of empirical constraints in a manual analysis by experts also has the purpose of bounding the search space to reach a valid solution as soon as possible. However, Rietveld refinement using BBO can evaluate a large number of parameter combinations and suggest candidate structures much more efficiently than manual approach, that is, the constraints imposed to save time are no longer essential.

The results demonstrate that the proposed approach is reliable, allowing more practitioners without much expertise to utilise Rietveld refinement. In addition, the automation of the trial-and-error processes not only gives better results, but also stabilises the quality of results by eliminating human-origin variance and bias accumulated in manual tuning processes as HPO does in machine learning.

Table 2. Refined results of DSMO with different methods.

(a)					
Atom	x	y	z	U_{iso}	Site
Mn1	0.0000	0.5000	0.0000	0.0059(4)	4b
Dy1	0.0037(7)	0.0253(1)	0.2500	0.0148(3)	4c
Sr1	0.0037(7)	0.0253(1)	0.2500	0.0148(3)	4c
O1	0.5623(29)	−0.0070(11)	0.2500	0.0058(50)	4c
O2	0.2738(22)	0.2840(19)	0.0315(15)	0.0208(31)	8d
(b)					
Atom	x	y	z	U_{iso}	Site
Mn1	0.0000	0.5000	0.0000	0.0066(5)	4b
Dy1	0.0040(8)	0.0253(2)	0.2500	0.0154(18)	4c
Sr1	0.0040(8)	0.0253(2)	0.2500	0.0154(18)	4c
O1	0.5669(30)	−0.0084(12)	0.2500	0.0125(50)	4c
O2	0.2738(23)	0.2833(19)	0.0288(16)	0.0176(31)	8d
(c)					
Atom	x	y	z	U_{iso}	Site
Mn1	0.0000	0.5000	0.0000	0.0067(4)	4b
Dy1	0.0036(9)	0.0254(2)	0.2500	0.0153(3)	4c
Sr1	0.0036(9)	0.0254(2)	0.2500	0.0153(3)	4c
O1	0.5733(40)	−0.0067(13)	0.2500	0.0129(50)	4c
O2	0.2741(24)	0.2833(19)	0.0272(16)	0.0172(30)	8d
(d)					
Atom	x	y	z	U_{iso}	Site
Mn1	0.0000	0.5000	0.0000	0.0065(4)	4b
Dy1	−0.0064(7)	0.0251(1)	0.2500	0.0156(3)	4c
Sr1	−0.0064(7)	0.0251(1)	0.2500	0.0156(3)	4c
O1	0.5176(80)	−0.0075(10)	0.2500	0.0058(50)	4c
O2	0.2760(18)	0.2723(19)	0.0446(11)	0.0287(30)	8d

(a) Best by BBO. (b) Best by human expert. (c) Similar to human expert by BBO. (d) Another interesting candidate by BBO. The R_{wp} values for (a)–(d) are 9.211%, 9.775%, 9.577%, and 9.484% respectively. The GOF values for (a)–(d) are 1.748, 1.855, 1.817, and 1.799 respectively. The atom positions are indicated in fractional coordinates. The numbers in parentheses are standard uncertainties obtained from Rietveld refinement.

The efficiency against execution (wall-clock) time is also important in practice. In this study, each optimisation run (200 trials) takes less than an hour using a general workstation (2.3 GHz 18 core CPU, 256 GB RAM), and it can be performed within a practical time limit using a general laptop. The execution time of the optimiser is improved by parallelisation. For the same number of trials, the performance of parallel optimisation is typically slightly worse than in sequential optimisation, but the execution time will be about $1/n$ when the number of workers is n . Increasing the number of experts is difficult, but increasing computational resources is relatively easy. It is also an advantage of automation over manual work.

The problems of human-origin variance and bias and low stability of the quality of model fitting results are very common in other research areas. We consider that the framework of automation as a blackbox optimisation in this study is powerful and our approach can be easily applied to similar problems in other fields.

METHODS

Details of implementation

To solve the problem (Eq. (2)), we connect GSAS-II³⁹, a crystallography data analysis software, which provides Rietveld refinement to Optuna⁵⁰, an automatic HPO software framework. Extrinsic parameters listed in Table 1 are optimised using the TPE. Other extrinsic parameters related to Rietveld refinement are initialised with GSAS-II default values, and the intrinsic parameters for a crystal structure are obtained from ref.⁴³. The TPE is easy to use and supports real, integral, categorical, and conditional types of parameters, making it suitable for application in our case. It iteratively samples a new promising configuration based on the past observations (i.e. the pairs of the configurations and the corresponding R_{wp} values) and then runs GSAS-II to perform Rietveld refinement with the sampled configuration to find the best one. The refinement procedure described herein consists of the following conventional steps: (1) set initial 2θ range, (2) refine the background, (3) refine the 2θ zero correction and the unit cell parameters, (4) refine sample displacement parameters and the histogram scale factor, (5) refine atom parameters, and (6) refine the above parameters without 2θ bounds in step 1. During the optimisation, the constraint that all refined U_{iso} must be non-negative is imposed to avoid unreasonable structure models by setting R_{wp} to a large constant ($=10^9$) for penalty when the constraint is violated. The TPE does not mimic human experts but simply optimises the objective function. This is completely different from a conventional expert system that aims to reproduce the empirical strategies of human experts.

Visualisation of results

To visualise the crystal structure features with MDS in Fig. 6, feature standardisation, which makes the values of each feature in the data have zero-mean and unit-variance, was performed before applying MDS. We used the MDS implementation in scikit-learn⁵¹, a machine learning toolkit for Python.

DATA AVAILABILITY

The dataset of diffraction patterns and initial crystal structures are available in the GitHub repository of AutoFP⁴⁷ (Y_2O_3 , DSMO) and PowBase⁵² ($LiCoO_2$).

CODE AVAILABILITY

The codes for automated crystal structure analysis, documents, and supplementary resources that support the findings of this study are available at <https://doi.org/10.5281/zenodo.3747147>. These are also available at the BBO-Rietveld repository on GitHub (<https://github.com/quantumbeam/BBO-Rietveld>). As the supplementary resources, we provide not only the experimental results of Y_2O_3 , DSMO that we discussed in this paper but also the result of $LiCoO_2$ that are also comparable to the result of human expert.

Received: 4 October 2019; Accepted: 20 April 2020;

Published online: 05 June 2020

REFERENCES

- Callister, W. D. & Rethwisch, D. G. *Materials Science and Engineering* (John Wiley and Sons, 2010).
- De Graef, M. & McHenry, M. E. *Structure of Materials. An Introduction to Crystallography, Diffraction and Symmetry* (Cambridge University Press, 2012).
- Koinuma, H. & Takeuchi, I. Combinatorial solid-state chemistry of inorganic materials. *Nat. Mater.* **3**, 429–438 (2004).
- Hofmann, D. W. M. & Kuleshova, L. N. *Data Mining in Crystallography*, vol. 134 of *Structure and Bonding* (Springer Science & Business Media, Berlin, Heidelberg, 2009).
- Potyrailo, R. et al. Combinatorial and high-throughput screening of materials libraries: Review of state of the art. *ACS Comb. Sci.* **13**, 579–633 (2011).
- Gregoire, J. M. et al. High-throughput synchrotron X-ray diffraction for combinatorial phase mapping. *J. Synchrotron Rad.* **21**, 1262–1268 (2014).
- Lookman, T., Alexander, F. J. & Rajan, K. *Information Science for Materials Discovery and Design* (Springer-Verlag, 2015).
- Alberi, K. et al. The 2019 materials by design roadmap. *J. Phys. D* **52**, 013001 (2018).
- Rietveld, H. M. Line profiles of neutron powder-diffraction peaks for structure refinement. *Acta Crystallogr.* **22**, 151–152 (1967).
- Rietveld, H. M. A profile refinement method for nuclear and magnetic structures. *J. Appl. Crystallogr.* **2**, 65–71 (1969).
- Hill, R. Rietveld refinement round robin. i. analysis of standard x-ray and neutron data for pbso4. *J. Appl. Crystallogr.* **25**, 589–610 (1992).
- Hill, R. & Cranswick, L. International union of crystallography. commission on powder diffraction. rietveld refinement round robin. ii. analysis of monoclinic zro2. *J. Appl. Crystallogr.* **27**, 802–844 (1994).
- León-Reina, L. et al. Round robin on rietveld quantitative phase analysis of portland cements. *J. Appl. Crystallogr.* **42**, 906–916 (2009).
- Audet, C. & Hare, W. Introduction: tools and challenges in derivative-free and blackbox optimization. In *Derivative-Free and Blackbox Optimization*, 3–14 (Springer, 2017).
- Feurer, M. & Hutter, F. Hyperparameter optimization. In *Automated Machine Learning*, 3–33 (Springer, 2019).
- Mockus, J., Tiesis, V. & Zilinskas, A. The application of Bayesian methods for seeking the extremum. *Towards Glob. Optim.* **2**, 2 (1978).
- Bergstra, J., Bardenet, R., Bengio, Y. & Bengio, Y. Algorithms for hyper-parameter optimization. In *Advances in Neural Information Processing Systems*, 2546–2554 (2011).
- Hutter, F., Hoos, H. H. & Leyton-Brown, K. Sequential model-based optimization for general algorithm configuration. In *International Conference on Learning and Intelligent Optimization*, 507–523 (Springer, 2011).
- Snoek, J., Larochelle, H. & Adams, R. P. Practical Bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems*, 2951–2959 (2012).
- Bergstra, J., Yamins, D. & Cox, D. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In *International Conference on Machine Learning*, 115–123 (2013).
- Eggenberger, K. et al. Towards an empirical foundation for assessing Bayesian optimization of hyperparameters. In *NIPS workshop on Bayesian Optimization in Theory and Practice* (2013).
- Young, R. A., Prince, E. & Sparks, R. A. Suggested guidelines for the publication of Rietveld analyses and pattern decomposition studies. *J. Appl. Crystallogr.* **15**, 357–359 (1982).
- LeBail, A., Duroy, H. & Fourquet, J. L. Ab-initio structure determination of $LiSbWO_6$ by X-ray powder diffraction. *Mater. Res. Bull.* **23**, 452 (1988).
- Hill, R. J. Rietveld refinement round robin. I. Analysis of standard X-ray and neutron data for $PbSO_4$. *J. Appl. Crystallogr.* **25**, 589–610 (1992).
- Harris, K. D. M., Tremayne, M., Lightfoot, P. & Bruce, P. G. Crystal structure determination from powder diffraction data by Monte Carlo methods. *J. Am. Chem. Soc.* **116**, 3543–3547 (1994).
- Young, R. A. *The Rietveld Method*. IUCr monographs on crystallography (Oxford University Press, 1995).
- Shankland, K., David, W. I. F. & Csoka, T. Crystal structure determination from powder diffraction data by the application of a genetic algorithm. *Z. Kristallog.* **212**, 550–552 (1997).
- Harris, K. D. M., Johnston, R. L. & Kariuki, B. M. The genetic algorithm: foundations and applications in structure solution from powder diffraction data. *Acta Crystallogr. Sect. A* **54**, 632–645 (1998).

29. Toby, B. H. R factors in Rietveld analysis: how good is good enough? *Powder Diff.* **21**, 67–70 (2006).
30. Tsubota, M. & Jiro, K. A necessary criterion for obtaining accurate lattice parameters by Rietveld method. *Sci. Rep.* **7**, 15381 (2017).
31. Fancher, C. M. et al. Use of Bayesian inference in crystallographic structure refinement via full diffraction profile analysis. *Sci. Rep.* **6**, 31625 (2016).
32. Oishi-Tomiyasu, R. Application of convex optimization to identification of atomic species from diffraction patterns. *Powder Diff.* **32**, S179–S185 (2017).
33. Tomiyasu, K., Oishi-Tomiyasu, R., Matsuda, M. & Matsuhira, K. A new mathematical approach to finding global solutions of the magnetic structure determination problem. *Sci. Rep.* **8**, 16228 (2018).
34. Rodríguez-Carvajal, J. Recent advances in magnetic structure determination by neutron powder diffraction. *Phys. B: Condens. Matter* **192**, 55–69 (1993).
35. Rodríguez-Carvajal, J. Recent developments of the program FULLPROF. *IUCR* **26**, 12–19 (2001).
36. Toby, B. H. EXPGUI, a graphical user interface for GSAS. *J. Appl. Crystallogr.* **34**, 210–213 (2001).
37. Larson, A. C. & VonDreele, R. B. General structure analysis system (gsas). Tech. Rep. LAUR-86-748, Los Alamos National Laboratory, New Mexico (2004).
38. Izumi, F. & Momma, K. Three-dimensional visualization in powder diffraction. *Solid State Phenom.* **130**, 15–20 (2007).
39. Toby, B. H. & VonDreele, R. B. GSAS-II: The genesis of a modern open-source all-purpose crystallography software package. *J. Appl. Crystallogr.* **46**, 544–549 (2013).
40. Rajiv, P., Dinnebier, R. E. & Jansen, M. “Powder 3D Parametric” a program for automated sequential and parametric Rietveld refinement using topas. *Mater. Sci. Forum* **651**, 97–104 (2010).
41. Tian, P. & Billinge, S. Testing different methods for estimating uncertainties on Rietveld refined parameters using srrienveld. *Z. Kristallogr.* **226**, 898–904 (2011).
42. Tian, P. et al. SrRietveld: a program for automating Rietveld refinements for high-throughput powder diffraction studies. *J. Appl. Crystallogr.* **46**, 255–258 (2013).
43. Cui, X. et al. AutoFP: a GUI for highly automated Rietveld refinement using an expert system algorithm based on FullProf. *J. Appl. Crystallogr.* **48**, 1581–1586 (2015).
44. Feng, Z. et al. Method of artificial intelligence algorithm to improve the automation level of Rietveld refinement. *Computational Mater. Sci.* **156**, 310–314 (2019).
45. Sutton, R. S. & Barto, A. G. Reinforcement learning: an introduction. In *Adaptive Computation and Machine Learning Series* (MIT Press, 2018).
46. Viswanathan, G. et al. Single-crystal automated refinement (SCAR): a data-driven method for determining inorganic structures. *Inorg. Chem.* **58**, 9004–9015 (2019).
47. AutoFP GitHub repository. <https://github.com/xpcclove/autofp/>.
48. Torgerson, W.S. *Theory and Methods of Scaling* (Wiley, 1958).
49. Borg, I. & Groenen, P. Modern multidimensional scaling: theory and applications. *J. Educ. Meas.* **40**, 277–280 (2003).
50. Akiba, T., Sano, S., Yanase, T., Ohta, T. & Koyama, M. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623–2631 (ACM, 2019).
51. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
52. Powbase. <http://www.cristal.org/powbase/>.

ACKNOWLEDGEMENTS

This work is partly supported by JST-Mirai Program, Grant Number JPMJM19G1. K.O. and T.H. are partly supported by the Elements Strategy Initiative Center for Magnetic Materials (ESICMM), Grant Number 12016013, through the Ministry of Education, Culture, Sports, Science and Technology (MEXT). Y.O. and M.O. are supported by the New Energy and Industrial Technology Development Organization (NEDO). Y.S. is supported by the Japan Science and Technology Agency (JST), ACT-I, grant number JPMJPR18UE. K.S. has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement No. 701647. K.O. and K.S. gratefully acknowledge the financial support by Toyota Motor Corporation.

AUTHOR CONTRIBUTIONS

Y.O., Y.S. and K.O. conceived the idea for the present work. Y.O. and Y.S. carried out the numerical experiments. T.H. and K.S. examined the Rietveld refinement results as experts in materials science. All authors discussed the results and wrote the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to K.O.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020