Supplementary Materials

Efficient autonomous material search method combining *ab initio* calculations, autoencoder, and multi-objective Bayesian optimization

Yuma Iwasaki[a,b*], Hwang Jaekyun[a], Yuya Sakuraba[c], Masato Kotsugi[d], and Yasuhiko Igarashi[e]

[a] *Research and Services Division of Materials Data and Integrated System, National Institute for Materials Science (NIMS), Tsukuba, Japan;*

[b] *Institute for AI and Beyond, The University of Tokyo, Tokyo, Japan;*

[c] *Research Center for Magnetic and Spintronic Materials, National Institute for Materials Science (NIMS), Tsukuba, Japan*

[d] *Department of Materials Science and Technology, Tokyo University of Science, Tokyo, Japan;*

[e] *Graduate School of System and Information Engineering, University of Tsukuba, Tsukuba, Ibaraki, Japan*

*E-mail: IWASAKI.Yuma@nims.go.jp

S1. The KKR-CPA method

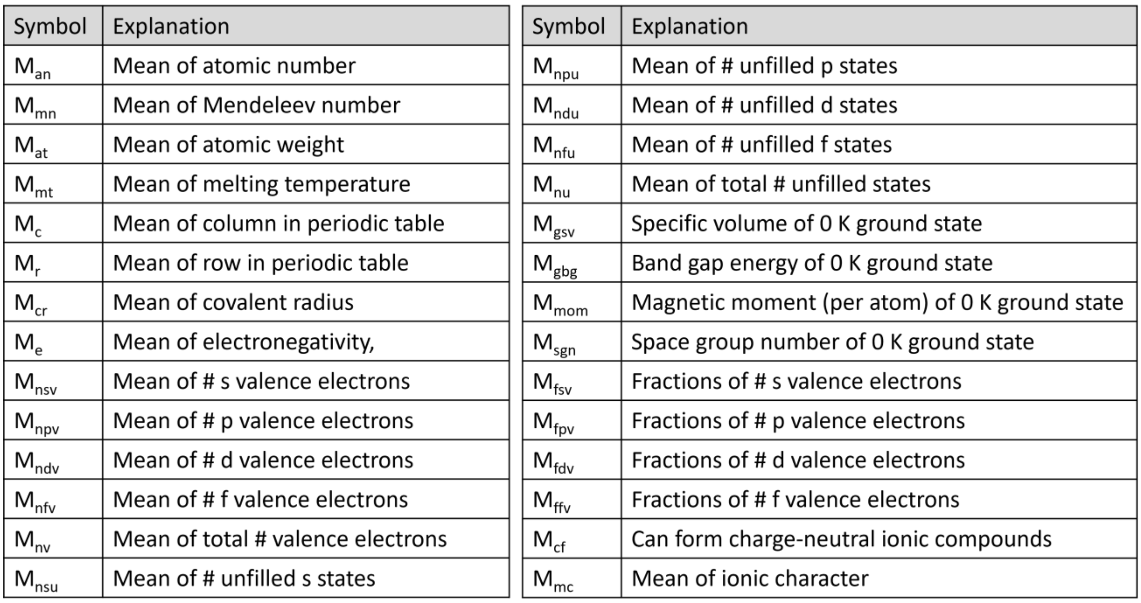
In the *ab initio* calculations, Green's function-based density functional theory (DFT) calculations were conducted by the Korringa Kohn Rostoker coherent potential approximation (KKR-CPA) method using the AkaiKKR software [1]. Multi-elemental disordered phases can be calculated using CPA, which can simulate them with high accuracy, especially in alloy systems [2–4].

Lattice constants were determined to minimize the total energy. In the lattice constant optimization calculations, the spin–orbit interactions and relativistic effects were not considered. Therefore, the *reltyp* parameter was set to *nrl*. The imaginary part at Fermi level (*edelt*) was set to 0.001. The *bzqlty* parameter, which determines the quality of the Brillouin zone mesh, was set to 4. The maximum considered angular momentum (*xml*) was 3. The exchange-correlation potential (*sdftyp*) was set to the local density approximation (*mjw*). The maximum number of iteration loops (*maxitr*) was set to 300. The density of states (DOS) was then calculated using the optimized lattice constant. In this calculation, the spin–orbit interaction and relativistic effects were considered. Therefore, the *reltyp* parameter was set to *srals*. The *edelt*, *bzqlty*, *xml* and *maxitr* were set to 0.001, 6, 3 and 500, respectively. The width of the energy contour (*ewidth*) was automatically selected from {1.0, 1.5, 2.0, 2.5}.

S2. List of the Magpie descriptors used in this study

Ward et al. developed a Magpie software [5]. This software creates a set of descriptors for each material, including elemental property statistics (i.e., the mean and standard deviation) of different elemental properties (e.g., period/group on the periodic table, atomic numbers, atomic weight, and melting temperatures) and electronic structure attributes such as the average fraction of electrons from the s, p, d, and f valence shells of all the present elements [6]. In this study, 28 Magpie descriptors were manually selected and used. The details are given in a previous work [5].

Table S1. List of the Magpie descriptors used in this study



Symbol	Explanation	Symbol	Explanation
M_{an}	Mean of atomic number	M_{npu}	Mean of # unfilled p states
M_{mn}	Mean of Mendeleev number	M_{ndu}	Mean of # unfilled d states
M_{at}	Mean of atomic weight	M_{nfu}	Mean of # unfilled f states
M_{mt}	Mean of melting temperature	M_{nu}	Mean of total # unfilled states
M_{c}	Mean of column in periodic table	M_{gsv}	Specific volume of 0 K ground state
M_{r}	Mean of row in periodic table	M_{gbg}	Band gap energy of 0 K ground state
M_{cr}	Mean of covalent radius	M_{mom}	Magnetic moment (per atom) of 0 K ground state
M_{e}	Mean of electronegativity,	M_{sgn}	Space group number of 0 K ground state
M_{nsv}	Mean of # s valence electrons	M_{fsv}	Fractions of # s valence electrons
M_{npv}	Mean of # p valence electrons	M_{fpv}	Fractions of # p valence electrons
M_{ndv}	Mean of # d valence electrons	M_{fdv}	Fractions of # d valence electrons
M_{nfv}	Mean of # f valence electrons	M_{ffv}	Fractions of # f valence electrons
M_{nv}	Mean of total # valence electrons	M_{cf}	Can form charge-neutral ionic compounds
M_{nsu}	Mean of # unfilled s states	M_{mc}	Mean of ionic character

S3. The autoencoder

To define a material space for efficient autonomous search, an autoencoder was used [7], which is one of the dimensionality reduction methods. The information on the composition vector ***C*** and the Magpie descriptor vector ***M*** is compressed in the middle layer of a neural network with the same input and output layers. Here, a 10-dimensional latent variable ***Z*** was created. ***Z*** contains information on both ***C*** and ***M***. The importance of the crystal structure information in the autonomous material search can be adjusted by changing the number of latent variables created by the autoencoder. In other words, to perform an autonomous material search in which the crystal structure information is more important, the number of latent variables should be reduced and vice versa. Here, the '*h2o*' package in the R software was used [8,9]. The activation function (*activation*) was set to *Tanh*. The iteration time (*epochs*) was set to 300. The hidden layer size (*hidden*) was set to 10. Default settings were used for the other parameters. Thus, the mean square error (MSE) and root mean square error (RMSE) were 0.03098466 and 0.1760246, respectively.

S4. Multi-objective Bayesian optimization

A 12-dimensional material space is defined by combining the latent variables ***Z*** (Z_1, Z_2,..., Z_{10}) created by the autoencoder and the one-hot vector ***S*** (S_F, S_H) representing the crystal structure. This materials space is explored by the KKR-CPA combined with the multi-objective Bayesian optimization. Here, the following Gaussian process regression models were constructed for the spin polarization, *P*, and the half-metallic gap, *G*.

$$P = f(Z_1, Z_2, Z_3, \dots Z_{10}, S_F, S_H)$$

$$G = f(Z_1, Z_2, Z_3, \dots Z_{10}, S_F, S_H)$$

where *P*, *G*, *Z* and *S* are the spin polarization, half-metallic gap, latent variables created by the autoencoder and one-hot-vector of crystal structure, respectively. The *gausspr* function of the *kernelab* package in the R software was used for the Gaussian process regression [10,11]. The radial basis kernel function (*rbfdot*) was used as the kernel function (*kernel*). The hyper parameter (*sigma*) was determined by the heuristics (*sigest*) every time (the *kpar* parameter was set to '*automatic*'). Both the initial noise variance (*var*) and the tolerance of termination criterion (*tol*) were set to 0.001. Default settings were used for the other parameters. The upper confidential bound (UCB) for each material is calculated as an acquisition function from the Gaussian process regression models [12].

$$UCB = \sigma + C\mu$$

The expected uncertainty, σ, and expected value, μ, are calculated using the Gaussian process regression. The exploration weight, *C*, is used to tune the trade-off between exploration and exploitation. Here, C was set to the following 5 patterns.

$$C = \{0, 1{,}5{,}20{,}50\}$$

This means that five different target materials were derived for the next KKR-CPA calculation at different ratios of exploration to exploitation. Therefore, five KKR-CPA calculations were performed per each Gaussian process regression model in the autonomous materials search. The candidate materials with the largest Pareto hypervolume were determined based on the UCB value and the material data (training data), in which *P* and *G* are already observed. The Pareto hypervolume was calculated using the *ecr* package in the R software [13]. These are then used as the target for the next KKR-CPA calculations. Using this method, it is possible to autonomously search for materials with high *P* and *G*.

References

[1] Akai H. Electronic structure Ni-Pd alloys calculated by the self-consistent KKR-CPA method. J Phys Soc Jpn. 1982;51:468–474.

[2] Khan SN, Staunton JB, Stocks GM. Statistical physics of multicomponent alloys using KKR-CPA. Phys Rev B. 2016;93:054206.

[3] Yang L, Liu B, Luo H, et al. Investigation of the site preference in Mn_2RuSn using KKR-CPA-LDA calculation. J Magn Mater. 2015;382(15):247–251.

[4] Jin K, Sales BC, Stocks GM, et al. Tailoring the physical properties of Ni-based single-phase equiatomic alloys by modifying the chemical complexity. Sci Rep. 2016;6:20159.

[5] Ward L, Agrawal A, Choudhary A, et al. A general-purpose machine learning framework for predicting properties of inorganic materials. npj Comput Mater. 2016;2:16028.

[6] Stanev V, Oses C, Kusne AG, et al. Machine learning modeling of superconducting critical temperature. Npj Comput Mater. 2018;4:29.

[7] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science 2006;313:504–507.

[8] Aiello S, Eckstrand E, Fu A, et al. Machine learning with R and H2O. http://h2o.ai/resources/

[9] Arora A, Andel A, Lanford J, et al. Deep Learning with H2O. http://docs.h2o.ai/h2o/latest-stable/h2o-docs/booklets/DeepLearningBooklet.pdf.

[10] Karatzoglou A, Smola A, Hornik K. kernlab: Kernel-Based Machine Learning Lab. R package version 0.9-31. https://CRAN.R-project.org/package=kernlab

[11] Karatzoglou A, Smola A, Hornik K, et al. kernlab – An S4 Package for Kernel Methods in R. J Stat Softw. 2004;11(9):1–20.

[12] Auer P. Using confidence bounds for exploitation-exploration trade-off. J Mach Learn Res. 2002;3:397–422.

[13] Bossek J. Ecr 2.0: a modular framework for evolutionary computation in R. In Proceedings of the Genetic and Evolutionary Computation Conference, 2017:1187–1193.