Recommender system for discovery of inorganic compounds

Hayashi, Hiroyuki; Seko, Atsuto; Tanaka, Isao

npj | computational materials

www.nature.com/npjcompumats

REVIEW ARTICLE OPEN

Recommender system for discovery of inorganic compounds

Hiroyuki Hayashi[1], Atsuto Seko[1] and Isao Tanaka[1,2]✉

A recommender system based on experimental databases is useful for the efficient discovery of inorganic compounds. Here, we review studies on the discovery of as-yet-unknown compounds using recommender systems. The first method used compositional descriptors made up of elemental features. Chemical compositions registered in the inorganic crystal structure database (ICSD) were supplied to machine learning for binary classification. The other method did not use any descriptors, but a tensor decomposition technique was adopted. The predictive performance for currently unknown chemically relevant compositions (CRCs) was determined by examining their presence in other databases. According to the recommendation, synthesis experiments of two pseudo-ternary compounds with currently unknown structures were successful. Finally, a synthesis-condition recommender system was constructed by machine learning of a parallel experimental data-set collected in-house using a polymerized complex method. Recommendation scores for unexperimented conditions were then evaluated. Synthesis experiments under the targeted conditions found two yet-unknown pseudo-binary oxides.

npj Computational Materials (2022)8:217 ; https://doi.org/10.1038/s41524-022-00899-0

INTRODUCTION

Innovation in materials technology often initiates with the discovery of materials. To take an example, the discovery of powerful permanent magnets and lithium battery materials has led to the emergence of modern and mass-produced electric vehicles, making a significant impact on our society. Two scenarios are possible for the materials discovery. The first is the discovery of unknown functions in already known compounds. For this purpose, an experimental database of known compounds is searched using features representing the function. The features are chosen based on physical and/or empirical rules using information on constituent elements and crystal structures of compounds. Systematic first-principles calculations are sometimes performed to obtain features. The second scenario begins with discovering a compound as-yet-unreported by experiments, i.e., an as-yet-unknown compound. This is challenging since the chemical composition space of inorganic compounds with multiple elements and multiple crystal sites is vast. The space cannot be explored efficiently without a good strategy to narrow down the search space. A combination of an experimental database and its data-driven analysis is a powerful approach. In this article, we will focus on the second scenario, i.e., the discovery of as-yet-unknown compounds.

Currently, several inorganic compound databases are available, such as the inorganic crystal structure database (ICSD)[1] in which approximately 250000 compounds are registered. The yearly trend of the number of unique compositions registered in ICSD is shown in Fig. 1. They are only for ternary and quaternary compounds consisting of multiple cations and a single anion having chemical compositions of integer ratios, which can be selected using the ANX formula in ICSD. Anions are taken from groups 15 (pnictogen), 16 (chalcogen), and 17 (halogen) in the periodic table. Cations are from the remaining groups, except for group 18 (noble gas) and hydrogen. These compounds include complex or pseudo-binary (-ternary) pnictogenides, chalcogenides, and halides. According to the rule, carbonates and silicates are included, but nitrates, phosphates, and sulfates are not. The number of ternary compounds composed of two cations and one anion registered to date is 5823. Similarly, a quaternary compound consisting of three cations and one anion counts 4897. Among them, 2574 (44%) for ternary and 3428 (70%) for quaternary are oxides. The predominance of oxides is natural since they are relatively easy to find as natural minerals or to synthesize artificially. The bar chart shows that the annual increase in the number of registered compounds has saturated or declined. The trend suggests that the discovery of ternary compounds is getting more difficult each year if we continue the same traditional approach. At the same time, given the high diversity of elemental combinations, there is a good chance to discover compounds in quaternary compounds, especially for non-oxides.

Chemically relevant compositions (CRCs) means the chemical composition that gives a stable or metastable compound under given thermodynamical conditions. Thermodynamically stable compounds are on the convex hull of formation energies, while metastable compounds show slightly higher formation energies above the convex hull. It is typically not easy for experiments to estimate the convex hull of the formation energy for a given thermodynamic condition. Identifying stable and metastable compounds by experiments is time and labor intensive. On the other hand, the convex hulls at zero temperature can be drawn based on energetics by systematic first principles calculations. Additional phonon and configurational calculations must be performed to incorporate temperature effects, which are possible but rather time-consuming. It should be noted, however, estimation of the formation energies for compounds is quite costly when their crystal structures are unknown, since the structures should be determined prior to the first principles calculations. If the CRC can be estimated prior to experiments or first principles calculations, the information is very useful in narrowing down the chemical composition in the search for compounds.

In the last decade, large databases of first principles calculations of inorganic compounds have been constructed and made available for many users[2–7]. Combining machine learning models

[1]Department of Materials Science and Engineering, Kyoto University, Kyoto, Japan. [2]Nano Research Laboratory, Japan Fine Ceramics Center (JFCC), Nagoya, Japan. ✉email: tanaka@cms.MTL.kyoto-u.ac.jp

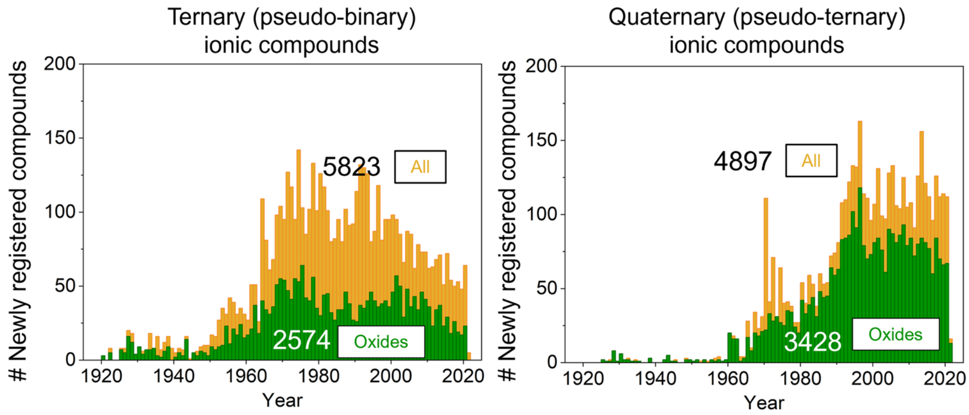

Fig. 1 The yearly trend of the number of unique compositions registered in ICSD (2021 Ver.2) for ternary and quaternary ionic compounds (orange bars). Only compounds reported to be experimentally synthesized, satisfying the charge-neutral condition, and having no partially occupied sites were adopted. Oxides are shown separately by green bars. Data-extraction procedures from ICSD to construct these figures are given in the Supplementary Information.

and first principles data, attempts to find CRCs have been reported[8–12]. In ref. [8], a procedure was given to estimate the probability as CRC using compositional similarity. In ref. [9] using a database of the first principles formation energies, a machine-learning model was constructed only with chemical compositions to predict yet-unknown CRCs. The present authors used a list of compounds registered in ICSD as training data, and adopted methods to establish recommender systems for the discovery of CRCs[13,14]. Recommender system[15–17] is a type of information filtering system, which is increasingly popular in a variety of fields, for example, E-commerce and social networking services. It attempts to estimate personalized recommendation scores of items to users based on their history of purchase or ratings. When this method is used for material discovery, the purchase history corresponds to the experimental database of compounds. The recommendation score is then related to the probability of finding a CRC. In our studies[13,14], two types of algorithms were used to estimate the recommendation scores. One is a descriptor-based recommender system with features specific to chemical elements. The other is a tensor-based recommender system. They will be explained in the following chapters, together with some successful examples to synthesize as-yet-unreported compounds. In the last chapter, we describe the construction of a recommender system for experimental processing conditions for compounds based on a parallel experimental data-set collected in-house. Synthesis condition data was put into a tensor-based recommender system to evaluate recommendation scores for unexperimented conditions.

Compositional descriptor-based recommender system

Firstly recommender system of CRC was constructed using compositional descriptors[13]. They are made up of 22 elemental features, such as the atomic number and Pauling electronegativity, which can be classified into (1) intrinsic quantities of elements, (2) heuristic quantities of elements, and (3) physical properties of elemental substances. The compositional space was made from the means, standard deviations, and covariances of these 22 elemental features weighted by the concentration of the constituent chemical element. Here, the compositional space was restricted to ionic compounds with integer-valency cation and anion. Grid points were placed on the compositional space at integer composition-ratios. The points corresponding to compounds registered in the ICSD were designated as 'entries'. The rest of the grid points were treated as 'no-entries'. The data were then supplied to the machine learning for the binary classification in which responses have two distinct values of $y = 1$ and 0. A score of $y = 1$ was given to 'entries', and $y = 0$ for 'no-entries'. Although the composition of $y = 1$ can be regarded as CRC, the composition of $y = 0$ does not necessarily mean that the composition is not CRC. There may be insufficient synthesis experiments at that chemical composition of 'no-entry'. There is also a possibility that the composition is a CRC, but the corresponding compound is difficult to synthesize experimentally.

After the machine learning using classifiers, a recommendation score, $\hat{y}$, was estimated at approximately 1.3 million pseudo-binary and approximately 3.8 million pseudo-ternary compositions that were not registered in ICSD. The recommendation scores were then arranged in descending order. To verify whether the chemical compositions with high recommendation scores correspond to currently unknown CRCs, we examined if they were listed in another database, ICDD-PDF[18]. As there was a large overlap between registered compositions in ICSD and ICDD-PDF, the data-set that were not included in ICSD were extracted from ICDD-PDF. We then examined whether chemical compositions with high recommendation scores were included in ICDD-PDF. Figure 2a shows the cumulative numbers of verified CRCs for pseudo-binary compositions with the ranking of the recommendation scores. Results by three classifiers, i.e., random forest, gradient boosting, and logistic regression, are much better than that of the random sampling in all cases, indicating that the approach is helpful for discovering the currently unknown CRCs that are not present in the training database. Among the three classifiers, the random forest method performed the best. The histogram of the number of verified CRCs by the random forest method is shown in Fig. 2b. The discovery rate defined by the numbers of verified CRCs in the candidate CRCs is 18% for the top 1000, and 15% for the top 3000 candidates. The discovery rate for the top 1000 is 60 times greater than that by the random sampling, 0.29%. It should be noted, however, that the discovery rate evaluated in this way is only a lower limit, since unknown compounds not registered in the ICDD-PDF cannot be counted. First principles calculations can be used to examine if the candidate CRCs are on the convex hull of formation energies. This will be discussed in the next chapter with Fig. 6.

Experimental efforts were carried out in collaboration with synthetic experts to synthesize unknown compounds with high recommendation scores[19]. Figure 3 shows Li_2O-GeO_2-P_2O_5 pseudo-ternary system with chemical compositions registered in three databases, i.e., ICSD, ICDD-PDF, and Springer Materials (SpMat)[20]. Chemical compositions of CRCs with high recommendation scores but not registered in any database are numbered according to their recommendation scores. Synthesis experiments were performed at target compositions by firing the mixed

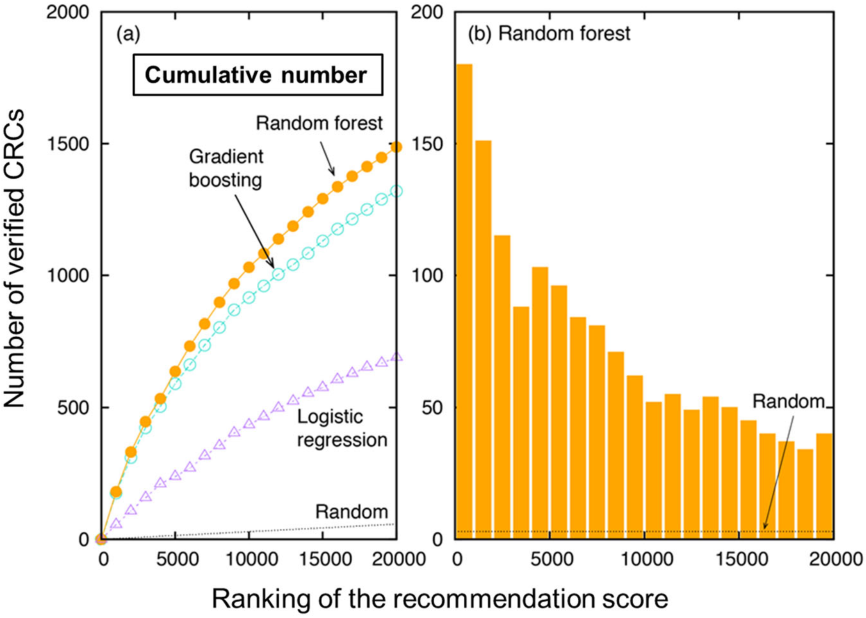

Fig. 2 The numbers of verified CRCs for pseudo-binary compositions with the ranking of the recommendation scores. a The cumulative numbers of verified CRCs by three classifiers, i.e., random forest, gradient boosting, and logistic regression, are compared with that by random sampling. **b** The histogram by the random forest method, i.e., the differential form of **a** for the random forest method.

starting powders in air. The products were supplied to powder x-ray diffraction experiments. At the composition of 6 in Fig. 3, $Li_6Ge_2P_4O_{17}$, the diffraction patterns were not able to be assigned to any known compound. After optimizing synthesis conditions and detailed characterization, a phase having the composition $Li_6Ge_2P_4O_{17}$ was identified. The discovered phase showed a crystal structure different from any known compounds in the three databases.

Another set of synthesis experiments was carried out for AlN-Si_3N_4-LaN pseudo-ternary system[21]. Fifteen compositions with high recommendation scores were selected as candidates for CRCs. Synthesis experiments were performed at target compositions by firing the mixed starting powders at 1900 °C under 1.0 MPa N_2. A pseudo-ternary nitride, $La_4Si_3AlN_9$, forming a crystal structure different from any known compounds was successfully identified. An as-yet-unknown variant (isomorphous substituent) of a known compound was also discovered at the composition of $La_7Si_6N_{15}$.

Tensor-based recommender system

Different from the case in the previous chapter, the recommender system in this chapter does not use any descriptors. The CRCs registered in the ICSD database were used as training data. They were stored in a tensor, which was decomposed assuming a low-rank structure of the tensor. The recommendation scores for unknown data were then evaluated. A simplified scheme of the matrix-based recommender system often used in E-commerce is shown in Fig. 4a. The vertical axis corresponds to a customers' list. The history of each customer is stored on the horizontal axis as purchased records of items. Low-rank structure of the matrix means that customers with similar preferences are interested in purchasing similar items. The matrix in E-commerce contains an enormous number of data, but is typically sparse. Combined with an appropriate decomposition technique, this type of recommender system is known to be very helpful for both customers and E-shops.

In the work reported in ref. [14], the compositional space was restricted to ionic compounds composed of two, three, and four cations {A, B, C, D} and one anion {X} having integer valency.

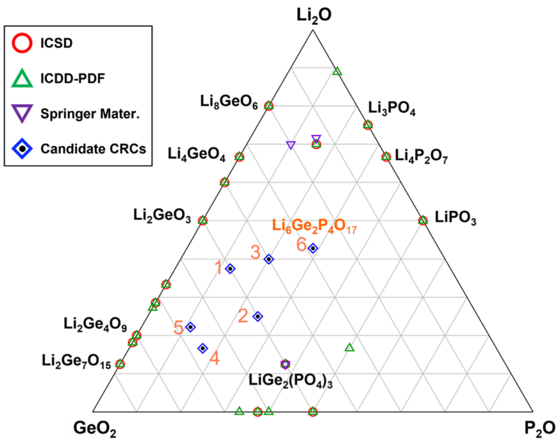

Fig. 3 Candidate CRCs on the Li_2O-GeO_2-P_2O_5 pseudo-ternary system with chemical compositions registered in three databases, i.e., ICSD, ICDD-PDF, and SpMat. Adopted from ref. [19] with small modifications.

The number of candidates was approximately 7.4 million for ternary $A_aB_bX_x$ with max(*a, b, x*) = 8, approximately 1.2 billion for quaternary $A_aB_bC_cX_x$ with max(*a, b, c, x*) = 20 and approximately 23 billion for quinary $A_aB_bC_cD_dX_x$ with max(*a, b, c, d, x*) = 20. The number of the training data in ICSD was 9313, 7742, and 1321 for ternary, quaternary and quinary, respectively. Figure 4b shows an example of a 3rd-order tensor expressing binary compounds. Three axes are cation type, anion type, and integer set showing the chemical composition. Using the Tucker decomposition method[22], the 3rd-order tensor can be approximated by a product of a core tensor and three matrices, as shown in Fig. 4c. For verification, the data-set unregistered in ICSD but included in two other databases, ICDD-PDF and SpMat, were used. Figure 5 shows the cumulative numbers of verified CRCs with the ranking of the recommendation scores for ternary, quaternary and quinary systems. The discovery rate was 59%, 52%, and 15% for the top 100 candidates for ternary, quaternary and quinary systems,

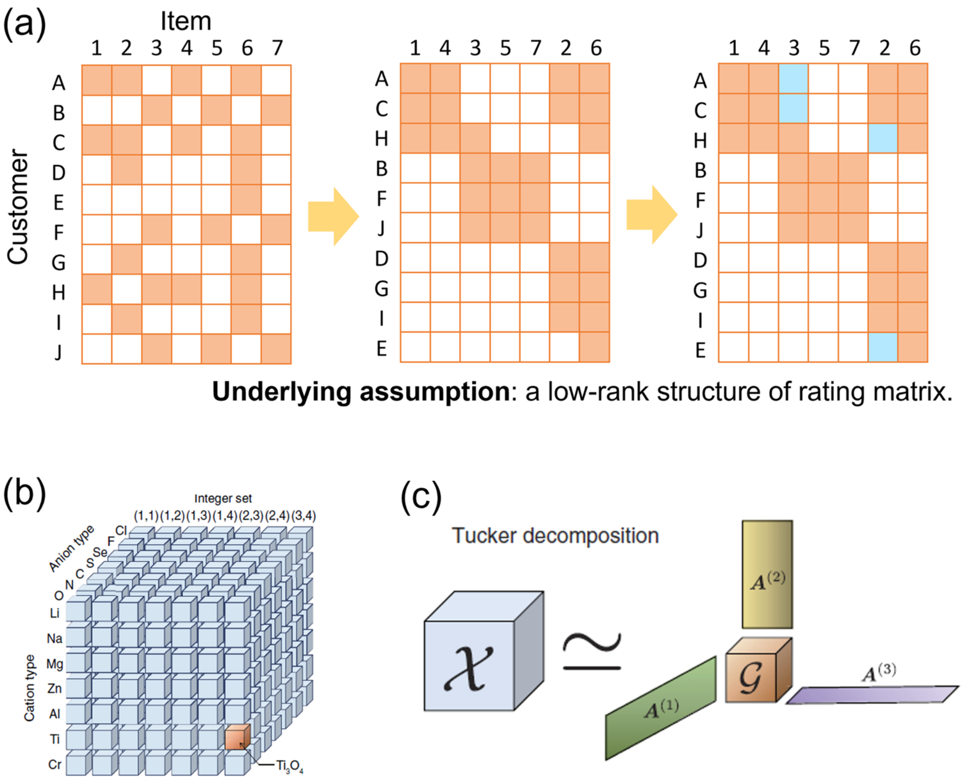

Fig. 4 Schematic illustration of matrix- and tensor-based recommender systems. a A simplified scheme of the matrix-based recommender system used in E-commerce. **b** An example of a 3rd order tensor expressing binary compounds. **c** Using the Tucker decomposition method, a large tensor can be approximated by a product of a small core tensor and three matrices. Adopted from ref. [14] with small modifications.

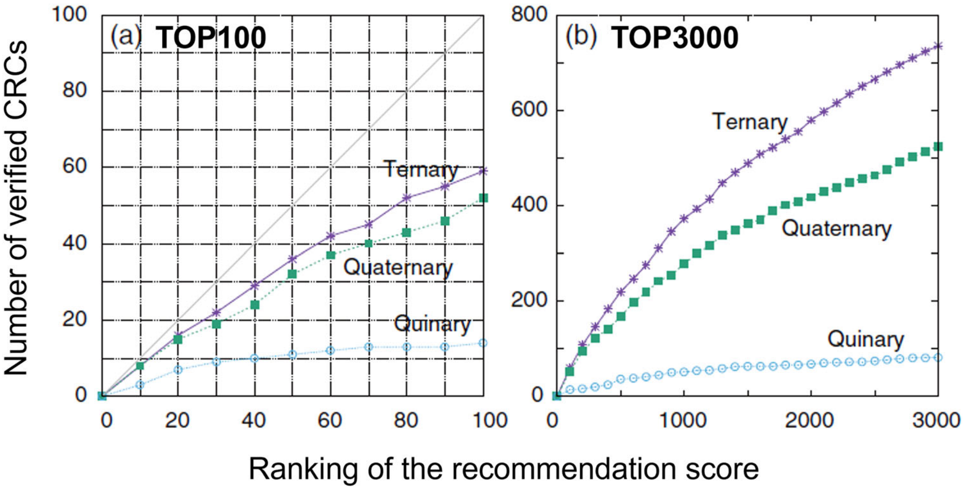

Fig. 5 The cumulative numbers of verified CRCs with the ranking of the recommendation scores for ternary, quaternary and quinary systems. a The top 100 candidates. **b** The top 3000 candidates. Adopted from ref. [14] with small modifications.

respectively. The lower discovery rate for the quinary system can be ascribed to the smaller number of training data than the ternary and quaternary systems. The high discovery rate for the present tensor-based recommender system, which does not use any descriptors, was well confirmed.

A set of first principles calculations was made to examine if the candidate CRCs are on the convex hull of formation energies. Pseudo-binary systems that contain candidate CRCs with the top 27 recommendation scores were selected. First principles calculations were performed using the plane-wave basis projector augmented wave (PAW) method[23,24] as implemented in the VASP code[25,26]. Since crystal structures were scarcely known a priori, calculations were exhaustively made, adopting all possible prototype structures registered in ICSD. Lowest energy structures were then used to draw the convex hull. A part of the results for pseudo-binary oxides is shown in Fig. 6 together with discovered CRCs and their recommendation scores in parentheses. Known CRCs registered in three databases are also plotted. As described in ref. [14], among 27 candidate CRCs, 23 compositions (85%) were found on the convex hull. Recalling that the 23 CRCs are not registered in any of the three databases, this result demonstrates the high performance of the present recommender system.

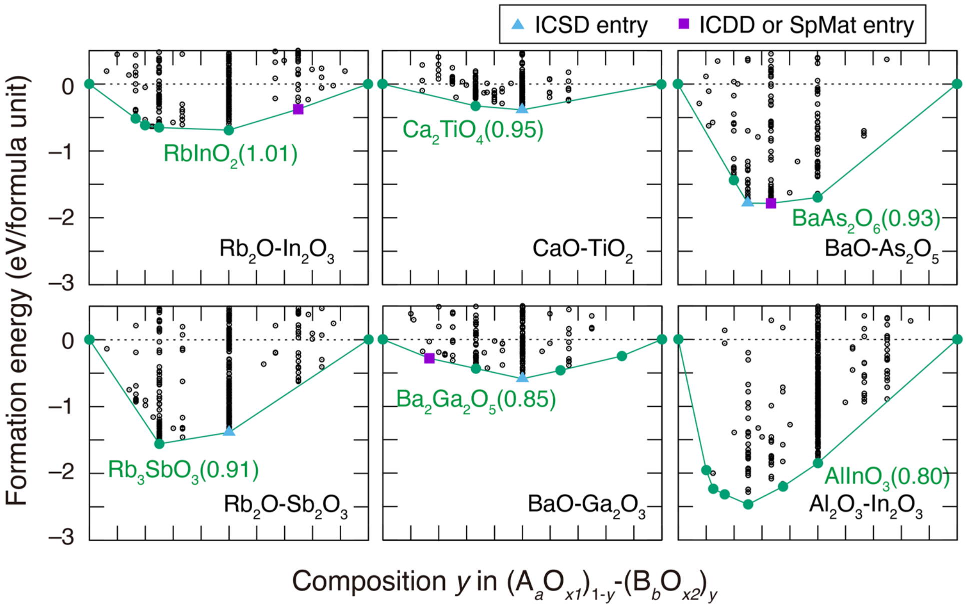

Fig. 6 The convex hull of the formation energy by the DFT calculations for pseudo-binary-oxide systems containing candidate CRCs. Closed circles (green) denote compounds on the convex hull. Closed triangles (blue) and squares (violet) denote CRCs registered in ICSD and ICDD-PDF + SpMat, respectively. Candidate compositions are given with recommendation scores in parentheses. Adopted from ref. [14] with modifications.

Synthesis condition recommender system

Methods to estimate recommendation scores for unknown CRCs have already been described in previous chapters of this article. It is true that some compounds were experimentally discovered at the proposed CRC based on the recommendation. However, we also experienced that synthesis experiments were often unsuccessful at the proposed CRCs. Since the predictive performance was well confirmed as described in the previous chapters, the failure is likely attributed to the lack of knowledge to find successful synthesis conditions. It is natural that a yet-undiscovered compound is difficult to synthesize. Experts in experimental chemistry attempt to synthesize compounds based on their experiences and knowledge of similar compounds. If there is a database of various synthesis conditions for diverse compounds, a computer may indicate the synthesis conditions efficiently instead of a human expert through machine learning. Synthesis conditions can be collected through text mining of scientific literature[27–30]. Such databases have been constructed recently, which may be useful for finding successful synthesis conditions. While such databases provide valuable information, there is a major problem when applied to machine learning. The data-set obtained from literature is strongly biased toward successful synthesis results. But, a good combination of successful and unsuccessful synthesis results is preferred for reliable machine learning. For this purpose, it is desirable to develop equipment that can automatically perform a large number of synthesis experiments in parallel without human bias, which is called combinatorial or parallel synthesis equipment.

Automated experimental equipment to construct such a database has been reported recently[31–35]. The present authors reported parallel synthesis experiments to prepare precursor powders of various inorganic oxides in four different ways, i.e., solid-state reaction, polymerized complex, cyclic ether sol–gel, and spray coprecipitation[31]. In the work of ref. [32], pseudo-binary inorganic oxide compositions were targeted and parallel synthesis experiments were made by a polymerized complex method. There were ${}_{28}C_2 \times 27 = 10206$ combinations of two cations from 28 elements and 27 compositional ratios, as shown in Fig. 7a. Among them, 1139 compositions were known to be CRCs and registered in ICDD-PDF. The remaining 9067 compositions were unknown if they were CRCs. Some of them may be unstable. Others may be difficult to synthesize experimentally and require special conditions for synthesis. Since the synthesis of pseudo-binary inorganic oxides has a long history of in-depth investigation, the chance of discovering yet-to-be-found oxides may be quite low. Therefore, to discover compounds, it is necessary to employ a much more efficient method than random trials on chemical compositions and synthesis conditions.

Here, the synthesis condition space was composed of 66150 conditions. At each target composition, a maximum of five different synthesis temperatures, ranging from 873 to 1273 K, was adopted. Three starting materials were used for V and Mo; for the rest, one starting material was used for each cation. In order to obtain training data for the machine learning, synthesis experiments were performed under 1542 conditions in total, which included 600 conditions at compositions where the presence of CRC was unknown, and 942 conditions at known CRC. Both of them were randomly selected. As shown in Fig. 7a, at the known CRC, the target compound was successfully synthesized under 499 of 942 conditions. On the other hand, at the unknown CRC, not a single condition out of 600 was successful. Results of the synthesis experiments were put into a fourth-order tensor with four axes, namely, 'starting material #1', 'starting material #2', 'cation mixing ratio', and 'firing temperature', as shown in Fig. 7b. Then the tensor was subjected to the Tucker decomposition and recommendation scores for unexperimented conditions were estimated. In order to verify the predictive performance of the recommender system, additional synthesis experiments were conducted at the top 300 synthesis conditions of unexperimented compositions. A histogram in Fig. 7c displays the number of successful and unsuccessful results as a function of the recommendation score. The fractions of the successful synthesis conditions, i.e., success rate, for each bin of the recommendation score are shown in Fig. 7d. Although the success rate was about 20% when the recommendation score was 0.2, it increased proportionally with the recommendation score. It became about 50% when the recommendation score was 0.5. In this way, the usefulness of the

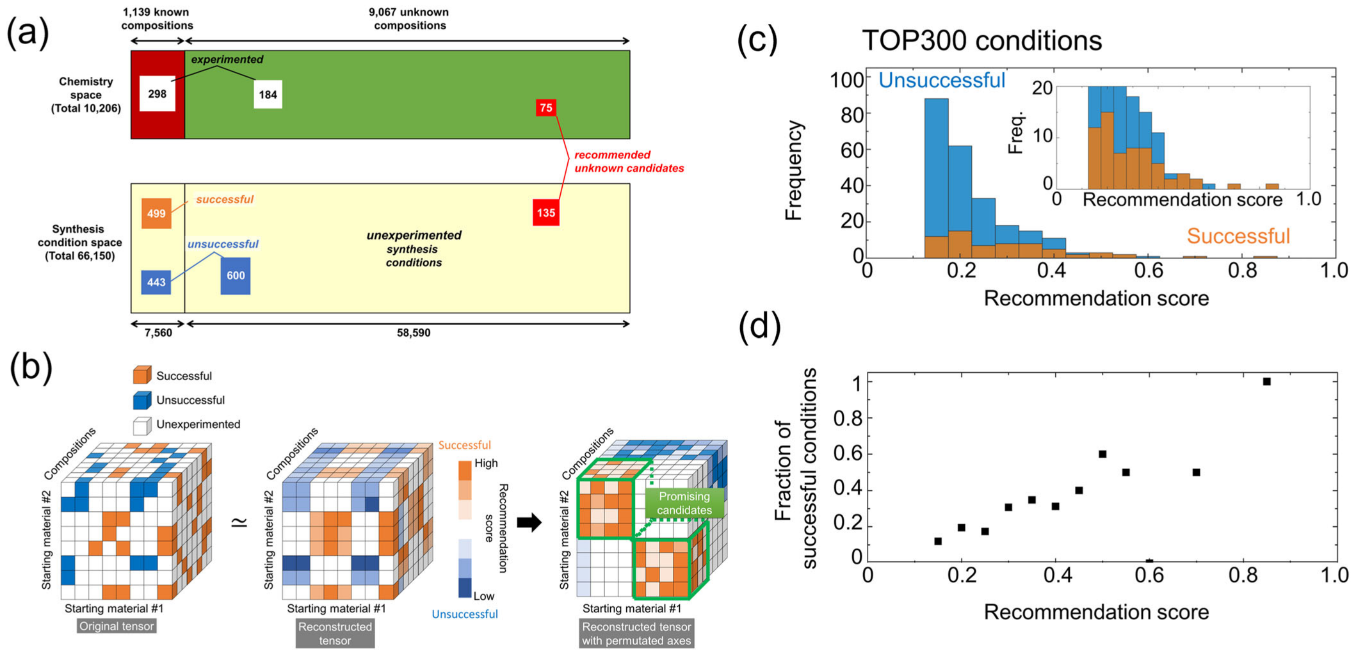

Fig. 7 A synthesis-condition recommender system. a The chemistry space and the synthesis condition space. **b** A schematic of the Tucker decomposition of the synthesis condition tensor. **c** Results of additional synthesis experiments for the top 300 synthesis conditions. The number of successful (orange) and unsuccessful (blue) results were shown as a function of the recommendation score. **d** The fractions of the successful synthesis conditions, i.e., success rate, for each bin of the recommendation score in **c**. Adopted from ref. [32] with small modifications.

recommendation score to estimate the success rate of synthesis conditions was demonstrated.

The top 300 synthesis conditions included 135 conditions for 75 unknown compositions. Synthesis experiments under the targeted conditions successfully found two as-yet-unknown pseudo-binary oxides: $La_4V_2O_{11}$ and $La_7Sb_3O_{18}$. Their powder X-ray diffraction profiles were analyzed by the Rietveld method using the RIETAN-FP program[36] after the crystal structure determination using the EXPO2014 code[37] to identify their crystal structures. $La_4V_2O_{11}$ and $La_7Sb_3O_{18}$ were found to be isostructural to known compounds, γ-$Bi_4V_2O_{11}$ and $La_7Ru_3O_{18}$, respectively. Although the discovery of inorganic pseudo-binary oxides was thought to be difficult, two as-yet-unreported compounds were successfully synthesized using the recommender system of the process conditions.

Conclusion and outlook

The recommender system is increasingly popular in a variety of fields in our society, such as E-commerce and social networking services. Based on a database, it attempts to suggest to an individual user what products to buy, what movies to watch, and so on. The method can be applied to materials discovery using an experimental database. The recommendation score can be related to the probability of finding the most pertinent chemical composition, synthesis conditions, etc. In this article, we described such studies on recommender systems for materials discovery.

Firstly, studies on the discovery of as-yet-unknown compounds using the recommender system were reviewed. A training dataset was obtained from those registered in ICSD. Two kinds of techniques were used to estimate recommendation scores. One method used compositional descriptors made up of elemental features. The other method used a tensor decomposition technique. The predictive performance for currently unknown CRCs was determined by examining their presence in other databases (ICDD-PDF and SpMat) in which overlapped data with ICSD was omitted. According to the recommendation, synthesis experiments were made. Two pseudo-ternary compounds, $Li_6Ge_2P_4O_{17}$ and $La_4Si_3AlN_9$ with currently unknown structures were successfully discovered.

Next, a synthesis-condition recommender system was constructed by machine learning of a parallel experimental data-set collected in-house using a polymerized complex method. Recommendation scores for unexperimented conditions were then evaluated. Additional synthesis experiments were conducted at the top 300 synthesis conditions of unexperimented compositions to verify the predictive performance of the recommender system. Although inorganic pseudo-binary oxides have historically been the subject of much research and discovering compounds was thought to be difficult, two as-yet-unknown pseudo-binary oxides, $La_4V_2O_{11}$ and $La_7Sb_3O_{18}$ were successfully synthesized.

High performance of the recommender system for the discovery of CRC and synthesis conditions was well demonstrated in these works. It may be interesting to know the advantages between the tensor-based and descriptor-based approaches. In general, they are dependent on the quality and quantity of the problems and datasets. When many data are uniformly distributed in the search space, the tensor-based approach should be preferred. Otherwise, the descriptor-based approach helps avoid so-called cold-start problems, which occur when few known CRCs are available. Especially when the descriptors representing the target property (formation energy, synthesis condition, etc.) are clearly identified, the descriptor-based approach should be worthwhile to adopt.

As for the synthesis condition recommender system, the data acquisition speed is rate-controlling. A breakthrough is expected to occur when the recommender system is combined with a high-speed and automated synthesis robot to improve the quality of the recommendation iteratively.

The use of recommender systems is still in infancy, it would be important to consider its application to a variety of problems and data in materials science and technology.

DATA AVAILABILITY

The database for the tensor-based recommender system in this study is available at https://github.com/sekocha/recommender. Other data supporting the findings of this study are available from the corresponding author on reasonable request.

Received: 10 June 2022; Accepted: 26 September 2022;
Published online: 13 October 2022

REFERENCES

1. Bergerhoff, G. & Brown, I. D. In Crystallographic Databases, edited by F. H. Allen et al. (International Union of Crystallography, Chester, 1987).
2. Materials Project (materialsproject.org).
3. AFLOW (aflowlib.org).
4. OQMD (oqmd.org).
5. NOMAD (www.nomad-coe.eu).
6. Materials Cloud (www.materialscloud.org).
7. AtomWork-Adv (atomwork-adv.nims.go.jp).
8. Hautier, G. et al. Finding nature's missing ternary oxide compounds using machine learning and density functional theory. *Chem. Mater.* **22**, 3762–3767 (2010).
9. Meredig, B. et al. Combinatorial screening for new materials in unconstrained composition space with machine learning. *Phys. Rev. B* **89**, 094104 (2014).
10. Ward, L. et al. Matminer: an open source toolkit for materials data mining. *Comput. Mater. Sci.* **152**, 60–69 (2018).
11. Gossett, E. et al. AFLOW-ML: a RESTful API for machine-learning predictions of materials properties. *Comput. Mater. Sci.* **152**, 134–145 (2018).
12. Huang, B. & von Lilienfeld, O. A. Ab initio machine learning in chemical compound space. *Chem. Rev.* **121**, 10001–10036 (2021).
13. Seko, A., Hayashi, H. & Tanaka, I. Compositional descriptor-based recommender system for the materials discovery. *J. Chem. Phys.* **148**, 241719 (2018).
14. Seko, A., Hayashi, H., Kashima, H. & Tanaka, I. Matrix- and tensor-based recommender systems for the discovery of currently unknown inorganic compounds. *Phys. Rev. Mater.* **2**, 013805 (2018).
15. Resnick, P. & Varian, H. R. Recommender systems. *Commun. ACM* **40**, 56–58 (1997).
16. Aggarwal, C. C. Recommender Systems (Springer, International Publishing, New York, 2016).
17. Symeonidis, P. & Zioupos, A. Matrix and Tensor Factorization Techniques for Recommender Systems (Springer International Publishing, New York, 2016).
18. ICDD-PDF4+ (www.icdd.com/pdf-4/).
19. Suzuki, K. et al. Fast material search of lithium ion conducting oxides using a recommender system. *J. Mater. Chem. A* **8**, 11582–11588 (2020).
20. Springer Materials, http://materials.springer.com.
21. Koyama, Y., Seko, A., Tanaka, I., Funahashi, S. & Hirosaki, N. Combination of recommender system and single-particle diagnosis for accelerated discovery of novel nitrides. *J. Chem. Phys.* **154**, 224117 (2021).
22. Tucker, L. R. Some mathematical notes on three-mode factor analysis. *Psychometrika* **31**, 279–311 (1966).
23. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
24. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).
25. Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* **47**, 558–561 (1993).
26. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
27. Kim, E. et al. Materials synthesis insights from scientific literature via text extraction and machine learning. *Chem. Mater.* **29**, 9436–9444 (2017).
28. Kononova, O. et al. Text-mined dataset of inorganic materials synthesis recipes. *Sci. Data* **6**, 293 (2019).
29. Hong, Z. et al. Challenges and advances in information extraction from scientific literature: a review. *JOM* **73**, 3383–3400 (2021).
30. Makino, K. et al. Extracting and analyzing inorganic material synthesis procedures in the literature. *IEEE Access* **10**, 31524–31537 (2022).
31. Hayashi, H., Hayashi, K., Kouzai, K., Seko, A. & Tanaka, I. Recommender system of successful processing conditions for new compounds based on a parallel experimental data set. *Chem. Mater.* **31**, 9984–9992 (2019).
32. Hayashi, H. et al. Synthesis-condition recommender system discovers novel inorganic oxides. *J. Am. Ceram. Soc.* **105**, 853–861 (2021).
33. Yang, L. et al. Discovery of complex oxides via automated experiments and data science. *Proc. Natl Acad. Sci. USA* **118**, e2106042118 (2021).
34. Szymanski, N. J. et al. Toward autonomous design and synthesis of novel inorganic materials. *Mater. Horiz.* **8**, 2169–2198 (2021).
35. Ziatdinov, M. A. et al. Hypothesis learning in automated experiment: application to combinatorial materials libraries. *Adv. Mater.* **34**, 2201345 (2022).
36. Izumi, F. & Momma, K. Three-dimensional visualization in powder diffraction. *Solid State Phenom.* **130**, 15–20 (2007).
37. Altomare, A. et al. EXPO2013: a kit of tools for phasing crystal structures from powder data. *J. Appl. Crystallogr.* **46**, 1231–1235 (2013).

ACKNOWLEDGEMENTS

H.H. was supported by PRESTO, JST, and JSPS KAKENHI (Grant No. JP20H02423). A.S. was supported by PRESTO, JST, and JSPS KAKENHI (Grant Nos. JP18K18942 and JP19H05787). I.T. was supported by JSPS KAKENHI (Grant No. JP21H04621).

AUTHOR CONTRIBUTIONS

A.S. conceived the idea to use a recommender system for materials discovery and constructed the DFT data-set. H.H. developed parallel experimental equipment and made the experimental condition data-set. I.T. directed the project. All the authors participated in analyzing the results and writing the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-022-00899-0.

Correspondence and requests for materials should be addressed to Isao Tanaka.

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.