Materials informatics for discovery of ion conductive ceramics for batteries

Masanobu NAKAYAMA

1Department of Advanced Ceramics and Frontier Research Institute for Materials Science (FRIMS), Nagoya Institute of Technology, Gokiso, Showa, Nagoya 466–8555, Japan
2Unit of Elements Strategy Initiative for Catalysts & Batteries (ESICB), Kyoto University, Katsura, Saikyo-ku, Kyoto 615–8520, Japan

In this review, we introduce our work in the field of materials informatics for the prediction of ionic conduction properties in inorganic crystalline solids. Rational material development based on information-derived prediction of the ionic conductivity for the materials listed in the crystal structure database is attractive to reduce processing time and labor costs. For this purpose, the development of general descriptors and a sufficient volume of ionic conductivity datasets are required. As an example, herein we describe machine learning regression and Bayes optimization schemes and their results by using histogram descriptors and a bond valence-based force field approach.

©2021 The Ceramic Society of Japan. All rights reserved.

Key-words: Battery materials, Solid electrolytes, Materials informatics, Machine learning, Bayes optimization

[Received February 17, 2021; Accepted March 24, 2021]

1. Introduction

The development of all-solid-state lithium (Li)-ion batteries (ASSLIB) is vital to achieving the safe and stable charge-discharge and long-term (>10 years) performance of electric vehicles. In particular, finding solid electrolytes (i.e., Li-ion conductive materials) is crucial to produce ASSLIBs with fast Li-ion conduction and wide electrochemical window.1,2) Until now, several well-known structural skeletons for solid electrolytes, such as Li3N-, NASICON-, LISICON-, perovskite-, argyrodite-, thio-LISICON-, Garnet-, LGPS-, and inverse Perovskite-type structures, and their related compounds have been discovered.3) Significant research efforts have been devoted to optimizing the structure, composition, and/or processing conditions of the compounds described above. The optimization of the structure, composition, and processing conditions improves the material performance. However, the extent of the improvement is often limited. An alternative approach is to discover a new structural framework for solid electrolytes, which may lead to a dramatic improvement of the material performance. This approach is attractive and risky because one needs to consider several candidates without any guarantee of success. In this approach, researchers rationally choose the material structure and composition by considering the governing factors affecting Li-ion migration, based on their knowledge and experience. Subsequently, experimental validation, including synthesis and evaluation of the selected materials, is carried out. The results obtained from experimental validation are fed back to the next selection scheme. This scheme is known as the “trial-and-error” approach, and it may reduce the number of experiments in contrast to exhaustive search from numerous candidates. Nevertheless, it relies entirely on human resources. Materials informatics has been gaining attention because it introduces a new way of accelerating the development and discovery of materials, using computational and artificial intelligence-derived techniques.4)–7) In this paper, some of the authors’ own works on materials informatics for inorganic solid-state ionic conductors are introduced.

2. Eligibility of the materials informatics approach

Informatics has historically proved useful in various research fields, such as gene analysis and organic chemistry. To evaluate the applicability of the materials informatics approach for the prediction of ionic conductivity of inorganic crystalline materials, and since the olivine-type LiFePO4 has been thoroughly investigated for its migration mechanism,8,9) n-olivine-type LiMnO4 compounds were chosen for validation purposes. Where M and X correspond to cations with $M^{2+}$X$^{5+}$ and $M^{3+}$X$^{4+}$ combinations. A total of 66 MX cation pairs were selected (including virtual compositions), and their structure information

©2021 The Ceramic Society of Japan

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by-nd/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
and migration energies were computed by first-principle density functional theory (DFT) calculations. We set 42 structural descriptors and ion migration energies as objective values for machine learning (ML) regression schemes, such as partial least square (PLS) and artificial neural network regressions. Note that the migration energy calculation is computationally demanding. Figure 1(a) shows the diagnostic plots for the Li migration energies between DFT and prediction function using 15 training datasets. (b) Predicted activation energies as a function of M ionic radius for 66 olivine-type compounds. (c-1) and (c-2) represent coefficients and value importance parameter of the 42 composition/structure descriptors of the prediction function, respectively. (Figures are reused with the permission of respective publishers.)

and migration energies were computed by first-principle density functional theory (DFT) calculations. We set 42 structural descriptors and ion migration energies as objective values for machine learning (ML) regression schemes, such as partial least square (PLS) and artificial neural network regressions. Note that the migration energy calculation is computationally demanding. Figure 1(a) shows the diagnostic plots for the Li migration energies between DFT and prediction function using 15 training datasets. (b) Predicted activation energies as a function of M ionic radius for 66 olivine-type compounds. (c-1) and (c-2) represent coefficients and value importance parameter of the 42 composition/structure descriptors of the prediction function, respectively. (Figures are reused with the permission of respective publishers.)

Additionally, structural factors affecting ionic conductivity were assessed quantitatively [Fig. 1(c)], which supported the rational material design. MO6 octahedra related parameters are largely affected by the ion conduction property, which corresponds to the dependence on M cation size as mentioned above. Although this study deals only with a single structure framework and 66 composition pairs, it was confirmed that the materials informatics approach was useful for the prediction of ion conduction properties in inorganic crystalline solids. Good fitting results were observed even by PLS simple linear regression, and despite the non-continuous oxidation state pairs (M\textsuperscript{2+}X\textsuperscript{5+} and M\textsuperscript{3+}X\textsuperscript{4+}) were included in the dataset. Ion migration properties were largely affected by Coulombic interactions. However, satisfactory fitting results were observed without previous data classification.

3. Descriptors for inorganic crystalline materials

There is a high demand for novel ionic conductors materials, using information-driven prediction of ionic conductivities not only for specific structures (as mentioned in section 2) but for arbitrary structures. Therefore, the development of general descriptors for crystalline solids representing crystal structure information in a more comprehensive way is required. We suggested descriptors using histograms from compositional and structural information (histogram descriptor). For example, the electronegativity for the chemical composition of Li\textsubscript{7}PS\textsubscript{6} is summarized in Fig. 2, where we directly bin the real electronegativity values and the counting numbers of the corresponding atom, giving general vector-form descriptors. Likewise, other histogram descriptors, such as atomic radius, bond distance vs. number of bonds (radial distribution function), atom centered Voronoi real-feature values from the Voronoi partitioning of crystalline solids, are simply connected into vector form representation.

4. High-throughput computations to evaluate ionic conduction performance

As mentioned above, the development of general descriptors for crystalline solids is an important requirement for the high-throughput computations of migration energies.
to apply materials informatics to arbitrary compositions and structures. Additionally, since informatics is purely a posteriori approach, a sufficient amount of ion conduction data, which is the objective value, is needed to build a prediction function by ML regression. Although several experimental data are available in the literature, measurement conditions are somewhat scattered. Moreover, the reported data may be biased because authors and/or reviewers are prone to publish high-conductivity rather than low-conductivity data. A highly accurate yet computationally demanding alternative approach to assess ion conductivity performance is first-principle DFT-driven molecular dynamics. On the contrary, the force field (FF) approach is much faster than DFT-derived techniques. Adams et al. suggested the use of universal FF potentials and bond valence-based FF (BVFF) to estimate the conductivity of ionic bonds. Using BVFF parameter sets, we developed an automatic scheme to evaluate the migration energies of ions, described as follows.

1) First, the oxidation states of the cations were estimated by the BV sum approach [44], and the FF parameter sets were assigned from the literature.
2) Then, all mobile ions were removed to avoid blocking the ion migration pathway by other ions.
3) After that, the given lattice was divided into appropriate voxel sizes, and the potential energy of Li\(^{+}\) in each voxel was calculated using the BVFF.
4) Finally, the migration pathway was investigated using a percolation algorithm.

By applying this technique, migration energies were evaluated within a few minutes when the number of atoms in the cell was <100. Therefore, exhaustive calculations for the data listed in databases can be performed within a reasonable time. Figure 3 shows the calculated migration energies for Li\(^{+}\), Na\(^{+}\), and Mg\(^{2+}\) in oxides extracted from the Inorganic Crystal Structure Database as a function of cell volume per number of oxygen atoms. The computational results show that the ratio of fast ion conductors, which satisfies its migration energy (<0.5 eV), is only 3, 4, and <1 % for Li, Na, and Mg oxides, respectively. Thus, finding new fast ion conductors by a simple random search scheme is a difficult task.

The validity of this high-throughput algorithm was assessed by comparing it with first-principle DFT calculations. The diagnostic plot between BVFF and DFT calculations for olivine-type LiM0\(_x\)O\(_4\) datasets showed a positive correlation, and the BVFF approach tended to overestimate the migration energy. This deviation may stem from the removal of migrating ions from the structure, which reduces Coulombic repulsive interactions, and disregards local structure relaxation effects during ion jumps. We also evaluated the experimental ionic conductivities for promising samples from exhaustive search results. Samples of Li\(_2\)FeCl\(_4\), LiAlCl\(_4\), Na\(_2\)V\(_2\)O\(_7\), and Mg\(_{0.6}\)Al\(_{1.2}\)Si\(_{1.8}\)O\(_6\) were selected as fast ion conductors and showed relatively good ionic conductivity in the experiment. However, discrepancies between the BVFF calculation and the experimental measurement (incl. contribution of grain boundary resistance) were observed (~0.4 and 1.3 eV, respectively) for the activation energy of Mg ion conduction in Mg\(_{0.6}\)Al\(_{1.2}\)Si\(_{1.8}\)O\(_6\). Therefore, further validation and improvement of the computational scheme is needed. Currently, we consider the DFT-derived FF approach using a metaheuristic fitting procedure as reported previously.

5. Machine learning regression and Bayes optimization (BO)

The two requisites to obtain a reasonable prediction function by materials informatics and crystal structure database are building general descriptors and sufficient numbers of objective values. The former and latter can be satisfied by histogram descriptors and BVFF-derived migration energy datasets. To demonstrate this, various ML regression techniques were performed for the Li ion migration energies for 711 Li-ion oxides, containing selected metal oxides from the Materials Project datasets. The detailed selection criteria are as follows: first, the compounds should contain both Li and O atom; and second, the inclusion of the following atoms was allowed in the compounds Mg, Ca, Sr, Ba, Sc, Y, La, Ti, Zr, Hf, Nb, Ta, Zn, B, Al, Ga, In, C, Si, Ge, Sn, P, As, Sb, and S atom(s).

We performed eight ML regression algorithms, PLS, ridge, least absolute shrinkage and selection operator, elastic-net, linear support vector regression (SVR), kernel SVR, random forest, and gradient boosting regression (GBR). In this study, 80 % of the dataset was randomly selected for training, test predictions were performed on the remaining 20 %, and hyperparameter tuning was performed with a 10-partition cross-validation method. Figure 4(a) shows the root mean square errors (RMSE) and the corresponding diagnostic plot of the prediction functions derived from the eight ML regression algorithms above for the 711-compound dataset. The GBR algorithm showed the lowest RMSE [Fig. 4(a)]. The diagnostic plots indicate a strong correlation between the calculated and predicted values, and were largely scattered. Indeed, the RMSE value (~0.4 eV) was too large, since the researchers searched
for fast Li-ion conductors with migration energies below \( \sim 0.3 \) eV.

As mentioned before, the quality of the ML regression predictions was unsatisfactory. An alternative way to improve the quality of the predictions is by introducing a slight change in the informatics task itself. The original task was to build a prediction function, which assessed the ionic conductivity using descriptor inputs. Here, the redefined task was to distinguish the optimal sample, i.e., identify the sample with the lowest migration energy. For this purpose we applied the BO technique.\(^{23}\) BO is more efficient sampling than random sampling, accelerating the search for the best compounds. A Gaussian process model was embedded in the present BO method, and the expected improvement strategy was chosen to obtain the acquisition function, using the software library COMBO.\(^{23}\) In the first step, five samples were randomly chosen, followed by a BO-derived selection. Note that this was a proof-of-concept study, since all objective values were calculated. Figure 5 shows the discovery rate of the optimal compound as a function of sampling number, averaged over 1000 optimization trials. Results show that the best compound was discovered within \( \sim 250 \) samples (discovery rate \( > 99.9 \% \)), which reduced the total computational cost by 65\%. Thus, the study above demonstrates and validates that efficient materials search for the ionic conductor is possible using histogram descriptors and BO. Further improvement may be possible by including additional descriptors related to ionic conductivities. Additionally, this study used BVFF-derived migration energies calculations, which can be evaluated within a few minutes in most cases.

Hence, the materials informatics approach is not necessarily required, unless uncalculated compounds are numerous. However, the informatics scheme described above is useful for highly accurate computationally demanding calculations such as DFT-derived migration energy datasets.

6. Materials informatics for experiments

Most of the reports on materials informatics have considered computational datasets, since it is easy to prepare computational structure and property datasets from various open resources, such as The Materials Project,\(^ {21}\) OQMD,\(^ {24}\) AFLOW,\(^ {25}\) among others. Additionally, computational data is uniform and often independent of the process parameters, which is advantageous for the informatics approach. On the other hand, experimental data includes various aspects except for composition and/or crystal structure. For example, ionic conductivity includes micro-structure-derived grain boundary resistance, which is not often considered in atomistic simulations.

We attempted to apply an informatics approach to samples with experimentally measured ion conductivity, including both bulk and grain boundary resistances. NASICON-type LiZr\(_2\)(PO\(_4\))\(_3\) and its compositional derivatives have attracted attention as Li-ion conductors for all-solid-state Li metal batteries, because the compound is stable against metal Li.\(^ {26,27}\) Further improvement of Li ion conductivity is required for practical applications. For demonstration purposes, a double doping system, \( \text{Li}_{1+2x+y}\text{Ca}_x\text{Y}_y\text{Zr}_{2-2x-y}(\text{PO}_4)_3 \), is considered.\(^ {26}\) Where Zr ions are replaced by Ca and Y ions are accompanying interstitial Li formation. Doping is a commonly used technique for improving materials functionality, however, it is time and labor consuming. All the samples were synthesized, sintered, and the ionic conductivity was measured under the same experimental conditions for 47 compositions. Figure 6 shows the compositional dependence of the ratio of \( \alpha \) (hexagonal) or \( \beta \) (monoclinic) phases, concentration of the impurity phase, relative density (sintering degree), and ionic conductivity of Li at 30 °C.
including bulk and grain boundary contributions. At a glance look, the relationship between the observed properties was rather complicated. For example, the relative density was significantly dependent on Ca content, while the amount of Y ions controlled the formation of the $\alpha$ or $\beta$ phase. Thus, the determination of the optimized compositions was difficult only because of limited experimental data. However, as shown in Fig. 7, we confirmed that the BO approach reduced the sampling number significantly only by using compositional descriptors ($x$ and $y$). Additionally, multi-objective BO, which deals with multiple optimization target techniques, seems to be applicable when the amount data is sufficient, as described in our report. Accordingly, the informatics approach is also useful for large experimental datasets. In the future, efficient optimization is possible, including processing parameters, such as sintering temperature and gas atmosphere, by developing appropriate descriptors.

7. Conclusions

In general, finding new ionic conductors is a difficult task according to our exhaustive calculations using the BVFF approach, the success rate is very low. By developing appropriate descriptors for inorganic crystalline solids, the materials informatics approach enables rational and efficient material search, which accelerates research and development in the manufacturing industry. The development of general descriptors that can be used for materials with any composition and structure is crucial to access existing crystal structure databases, which is the result of more than 100 years of research. Our demonstration using histogram descriptors and BVFF-derived ionic conductivity datasets shows that the BO approach can reduce by $\sim65\%$ the computational cost for the optimization of Li-ion conduction from 711 candidate oxides. Additionally, the BO approach showed improved efficiency for the composition optimization of experimental datasets. Shortly, these techniques will lead to changes in our daily research and development activities, making them more efficient and rational.

Acknowledgements The author would like to acknowledge Prof. Ichiro Takeuchi, Prof. Masayuki Karasuyama, Prof. Ryo Kobayashi, Dr. Randy Jalem, Dr. Kunihiro Nobuhara, Dr. Hisatsugu Yamasaki, and laboratory staff and students for fruitful discussions and their support on materials informatics research. This work was partially supported by a Grant-in-Aid for Scientific Research under Grants 18K19129, 18H01707, 19H05815, 20H02436, and 21H01625 from the Ministry of Education Culture, Sports, Science and Technology, Japan (MEXT); the Elements Strategy Initiative to Form Core Research Center of MEXT under grant JPMXP0112101003; Advanced Low Carbon Technology Research and Develop-
Battery & Fuel Cell Project) from MEXT under grant JPMXP1020200301. English language editing was performed by Editage (www.editage.com).

References
1) K. Takada, Acta Mater., 61, 759–770 (2013).
2) Y. Noda, K. Nakano, H. Takeda, M. Kotobuki, L. Lu and M. Nakayama, Chem. Mater., 29, 8983–8991 (2017).
3) F. Zheng, M. Kotobuki, S. Song, M. O. Lai and L. Lu, J. Power Sources, 389, 198–213 (2018).
4) K. Rajan, Mater. Today, 8, 38–45 (2005).
5) M. Nakayama, H. Taki, T. Nakamura, S. Tokuda, R. Jalem and T. Kasuga, J. Phys. Chem. C, 118, 27245–27251 (2014).
6) M. Harada, H. Takeda, S. Suzuki, K. Nakano, N. Tanibata, M. Nakayama, M. Karasuyama and I. Takeuchi, J. Mater. Chem. A, 8, 15103–15109 (2020).
7) C. Chen, Y. Zuo, W. Ye, X. Li, Z. Deng and S. P. Ong, Adv. Energy Mater., 10, 1903242 (2020).
8) T. Rajkumar, M. Nakayama and M. Nogami, Solid State Commun., 150, 693–696 (2010).
9) W.-J. Zhang, J. Power Sources, 196, 2962–2970 (2011).
10) R. Jalem, T. Aoyama, M. Nakayama and M. Nogami, Chem. Mater., 24, 1357–1364 (2012).
11) R. Jalem, M. Nakayama and T. Kasuga, J. Mater. Chem. A, 2, 720–734 (2014).
12) L. Ward, A. Dunn, A. Faghaninia, N. E. R. Zimmermann, S. Bajaj, Q. Wang, J. Montoya, J. Chen, K. Bystrom, M. Dylla, K. Chard, M. Asta, K. A. Persson, G. J. Snyder, I. Foster and A. Jain, Comput. Mater. Sci., 152, 60–69 (2018).
13) R. Jalem, M. Nakayama, Y. Noda, T. Le, I. Takeuchi, Y. Tateyama and H. Yamazaki, Sci. Technol. Adv. Mat., doi:10.1080/14686996.2018.1439253.
14) S. Adams and R. P. Rao, Phys. Status Solidi A, 208, 1746–1753 (2011).
15) M. Nakayama, M. Kimura, R. Jalem and T. Kasuga, Jpn. J. Appl. Phys., doi:10.7567/JJAP.55.01AH05.
16) N. Tanibata, M. Kato, S. Takimoto, H. Takeda, M. Nakayama and H. Sumi, Adv. Energy Sustain. Res., 1, 2000025 (2020).
17) N. Tanibata, S. Takimoto, K. Nakano, H. Takeda, M. Nakayama and H. Sumi, ACS Mater. Lett., doi:10.1021/acs_materialslett.0c00127.
18) N. Tanibata, Y. Kondo, S. Yamada, M. Maeda, H. Takeda, M. Nakayama, T. Asaka, A. Kitajou and S. Okada, Sci. Rep., 8, 17199 (2018).
19) H. Takeda, K. Nakano, N. Tanibata and M. Nakayama, Sci. Technol. Adv. Mat., 21, 131–138 (2020).
20) R. Kobayashi, Y. Miyaji, K. Nakano and M. Nakayama, APL Mater., doi:10.1063/5.0015373.
21) A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, APL Mater., 1, 11002 (2013).
22) J. Schmidt, M. R. G. Marques, S. Botti and M. A. L. Marques, npj Comput. Mater., 5, 83 (2019).
23) T. Ueno, T. D. Rhone, Z. Hou, T. Mizoguchi and K. Tsuda, Mater. Discov., 4, 18–21 (2016).
24) J. E. Saal, S. Kirklin, M. Aykol, B. Meredig and C. Wolverton, JOM, 65, 1501–1509 (2013).
25) C. E. Calderon, J. J. Plata, C. Toher, C. Oses, O. Levy, M. Fornari, A. Natan, M. J. Mehl, G. Hart, M. Buongiorno Nardelli and S. Curtarolo, Comput. Mater. Sci., 108, 233–238 (2015).
26) Y. Li, W. Zhou, X. Chen, X. Liu, Z. Cui, S. Xin, L. Xue, Q. Jia and J. B. Goodenough, Proc. Natl. Acad. Sci., 113, 13131–13137 (2016).
27) Y. Noda, K. Nakano, M. Otake, R. Kobayashi, M. Kotobuki, L. Lu and M. Nakayama, APL Mater., doi:10.1063/1.5033460.
28) Z. Yang, S. Suzuki, N. Tanibata, H. Takeda, M. Nakayama, M. Karasuyama and I. Takeuchi, J. Phys. Chem. C, 125, 152–160 (2021).

Masanobu Nakayama is professor in the Department of Advanced Ceramics & Frontier Research Institute for Materials Science (FRIMS), Nagoya Institute of Technology and project professor in the ESICB, Kyoto University. His research interest mainly focused on ceramics for batteries and materials informatics. He received his Ph.D from Tokyo Institute of Technology in 2004, and he joined Nagoya Institute of Technology and The Ceramic Society of Japan from 2009.