

Materials Research Method using Smart Materials Informatics

Atsushi FURUKAWA* Tomohiro IKEDA* Tatsuya OKAYAMA*

ABSTRACT

The achievement of outcomes through the use of informatics techniques employing big data is today being reported. However, because of the lack of opportunities to obtain high volumes of data (upwards of several thousands or several tens of thousands of data), these techniques still present issues in terms of practical applicability. In the case of materials in particular, because it is necessary to evaluate characteristics following the creation of test pieces, obtaining upwards of several thousands or several tens of thousands of data is unrealistic. A Smart Materials Informatics method that would make it possible to make predictions based on even a limited amount of data was therefore defined. Study of the amount of data necessary for predictions showed that predictions could be made using only several tens to hundreds of data. The use of the new method in studying functional materials (catalyst, magnet, and thermoelectric materials) made it possible to identify materials possessing characteristics suited to their respective purposes in a short period of time, demonstrating the practical applicability of the method.

1. Introduction

As the process of economic globalization continues and international competition between automakers intensifies, there is increasing demand for manufacturers to enhance their research and development capabilities in the area of materials⁽¹⁾. Given that they form the components used in automobiles, the characteristics of the materials employed have a tremendous effect on various automotive specifications and design elements. The creation of new high-performance materials is therefore highly significant, producing fundamental changes in automotive performance. The approach to seeking new materials, i.e. the methodology employed in materials research, is an important element in this creation of new materials. Empirical methods (study of the correlation between basic elements such as electronegativity and ion radius and material characteristics, theoretical consideration of the principles by which specific functions are realized, formulation and testing of hypotheses) have up to the present contributed to the creation of large numbers of new materials. However, in the case of functional materials like catalyst, magnet, and thermoelectric materials, elements that are challenging to observe, such as crystal structure and electrons, affect material characteristics in complex ways. As a result, even if the principle by which a characteristic is realized is understood, it may not be possible to easily

propose a method of manifesting this principle. Because of this, the search for functional materials has often involved a repeated process of creation and testing of materials. This method of seeking new materials entails repeated procedures including preparation of starting materials, formulation of materials, manufacture of test pieces, and evaluation of characteristics. This process is inefficient, in addition to involving considerable time and expense.

At the same time, computers have undergone a marked evolution. The evolution has been exponential; over the course of several decades, the power of CPUs has increased approximately 300,000-fold, and the storage capacity of computer memories by approximately one million times. Software technologies such as machine learning have also evolved with these evolutions in computer capacity. Machine-learning technologies are evolving that make it possible to predict unknown data with a high degree of accuracy by learning known data. Against this background, the prediction of material characteristics using materials informatics (MI) methods representing a fusion of materials science and information science is attracting attention as a means of reducing the necessity for the creation of test materials and the evaluation of material characteristics⁽²⁾. Between several hundreds and several tens of thousands of data are normally used as the training data for these methods. In the search for new materials for automotive applications, however, data is initially obtained by

* Automobile R&D Center

formulating materials and evaluating their characteristics. Given this, increasing the number of data is inherently not an easy task. While MI methods may be effective, in cases such as these there is almost no opportunity to put them to practical use. Realizing the ability to make predictions based on a limited number of data is therefore an important focus.

Designating a method of making it possible to predict the performance of materials based on limited data Smart Materials Informatics (SMI), the research discussed in this paper therefore attempted to determine the minimum essential number of data necessary to help enable the making of predictions.

2. Study of Essential Number of Data for Prediction

As the response of target variables to explanatory variables becomes increasingly complex, the number of data necessary for accurate predictions necessarily increases. In studying the essential number of data for prediction, it is therefore important to understand the nature of data concerning functional materials.

Figure 1 shows an example of a response surface for performance P of a functional material in relation to explanatory variables x_1 and x_2 .

In the case of the nonlinear response surface shown in Fig. 1, when the explanatory variable x_1 is small, the increase of x_2 in relation to P is positive, as shown by the curve L1; when the explanatory variable x_1 is large, the increase of x_2 in relation to P is negative, as shown by the curve L2. In the case of systems in which target variables respond nonlinearly in relation to explanatory variables in this way, it is not easy to establish guidelines for the process of seeking new materials. In addition, the level of challenge involved increases when the number of explanatory variables increases. Catalyst, magnet, and thermoelectric functional materials, in which nonlinearity is comparatively strong, were therefore selected to provide case studies of the use of SMI methods in seeking new materials. It would

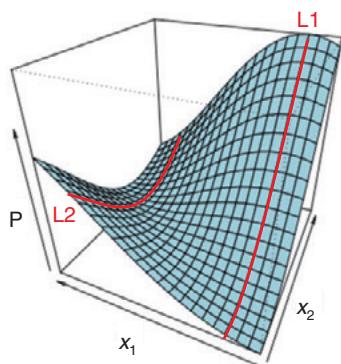


Fig. 1 Example of response surface

therefore be important to utilize data correlating with the nonlinearity displayed by these functional materials in studying the necessary number of data for prediction.

2.1. Relationship between Number of Data and Prediction Error

A specific number of data would be necessary in studying the practical applicability of the SMI method. It would not be practical to actually manufacture materials and evaluate their characteristics in order to acquire these data for the purposes of the study. Formulas making it possible to calculate target variables from arbitrary explanatory variables were therefore employed to obtain large numbers of data, and the non-experimental data generated by these formulae were utilized in studying the practical applicability of the SMI method. A non-parametric machine learning method able to comparatively freely generate regression equations in response to data was selected as the prediction model. The hyper-parameters of the selected machine learning method were tuned to help ensure a high level of predictive accuracy even when a limited number of data were employed. In order to further increase predictive accuracy, the data were subjected to preprocessing using principal component analysis to help ensure the orthogonality of the explanatory variables. In addition, functions compatible with nonlinearity were selected as the functions for use in the machine-learning program.

Internal validation was employed as an indicator of the predictive accuracy of the prediction model, making it possible to determine predictive accuracy for data other than training data, even when there were a limited number of data. Leave-one-out cross-validation (LOOCV) was selected as the form of internal validation. In addition, in order to make it possible to determine the relationship between the number of data and predictive error, a six-dimensional data set composed of five explanatory variables and target variables was formulated using the formulas for the generation of non-experimental data mentioned above. Jackknife sampling, a random sampling method that does not allow duplication, was applied to this data set for the various numbers of data, and average values for LOOCV when multiple repetitions were conducted were calculated. The number of jackknife samples taken was restricted to 100 from considerations of computation time and accuracy. Figure 2 shows the results.

As Fig. 2 indicates, the trend of the results is for LOOCV to become lower as the number of data increases. This is considered to be due to the fact that as the number of data increases, the volume of information also increases, increasing the accuracy of the prediction model. When the number of data increases over 50, LOOCV becomes less prone to reduction, and the amplitude of the decline in LOOCV narrows. This is considered to be because the representational ability of the model is reaching its limit, and the effect of the increase in the number of data in relation to LOOCV is becoming reduced. The relationship

of LOOCV to the number of data shown in Fig. 2 will be termed the LOOCV profile below.

Mathematically speaking, the LOOCV profile is known to be inversely proportional to the number of data. However, while the LOOCV profile shown in Fig. 2 is basically inversely proportional to the number of data, during the initial period when the number of data is low, the LOOCV fluctuates wildly, and its behavior in this domain is not theoretically explained. It was therefore considered that clarifying the causes of this fluctuation would be important in helping to enable the use of a valid SMI method.

2.2. Study of Minimum Necessary Number of Data

The causes of the extreme fluctuation in LOOCV were studied to help enable the use of a valid SMI method. In order to realize a visual understanding of the behavior of the prediction model in the low-data-number domain, the number of dimensions of the formulas for the generation of non-experimental data mentioned in Section 2.1. was reduced, and a three-dimensional non-experimental data set consisting of target variables (TV) in relation to two explanatory variables (x_1 and x_2) was formulated. The lengths of the maximum and minimum values of the explanatory variables x_1 and x_2 in this non-experimental data set were each divided into 100 uniform intervals, and TV were calculated in relation to 100,000 grid points representing all the combinations of x_1 and x_2 , creating a 100,000-point data set. Figure 3 shows a three-dimensional response surface generated for this data set.

The response surface shown in Fig. 3 is a theoretical surface produced by the formulas for the generation of non-experimental data. The response surface shown in Fig. 3 is calculated when the error of the machine learning prediction model is zero.

Next, the lengths of the maximum and minimum values of x_1 and x_2 were divided into three, four, five, and six uniform intervals, and non-experimental data consisting of explanatory variables and target variables for 9, 16,

25, and 36 points, representing all the combinations of x_1 and x_2 , were learned, and prediction models for each data set were formulated. The explanatory variables from the 100,000-point data set were input to the prediction models, and predicted values for target variables were calculated for each model. Figures 4(a), (b), (c), and (d) show the resulting response surfaces. As these results show, as the number of data increases, the predicted surfaces more closely approach the theoretical surface shown in Fig. 3. The change in the response surfaces with the increase in the number of data is particularly great in the low-data-number domain. It is considered that because the number of training data in this domain is low, the prediction model is unable to represent the entire shape, and a slight increase in the number of data results in a major change in the response surface. It was therefore considered that, despite the fact that LOOCV values are calculated low in the domain in which there is a major change in the shape of the predicted response surface, there is a strong possibility of randomness that negatively affects the reliability of the values, and they should not be employed as LOOCV values.

By contrast, it was considered that, because there was no major change in the shape of the response surfaces in the domain in which the number of training data is high despite the increase in the number of training data, and the prediction model was able to represent the entire surface, the reliability of the calculated LOOCV values was high.

Based on the above, the cause of the extreme fluctuations in LOOCV in the initial domain in which the number of data was low in the results of calculations of the LOOCV profile was the fact that the basic shape of the prediction model was not fixed in this domain. It was considered that if the number of data were increased up to the domain in which the fluctuations in LOOCV became low, the prediction model would capture the basic shape of the response surface, and the LOOCV values would be valid.

In addition, within the range of the non-experimental data employed, the number of data at which it becomes possible to apply the prediction method is between several

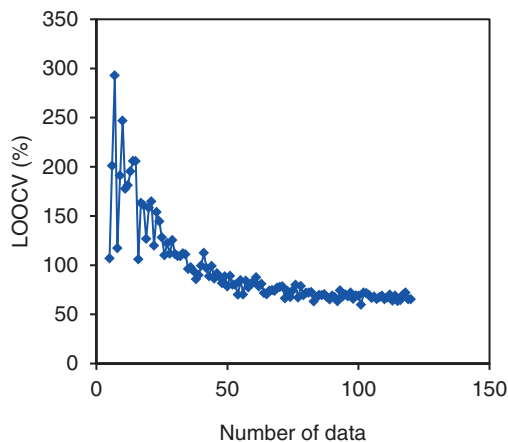


Fig. 2 Relationship between number of data and LOOCV

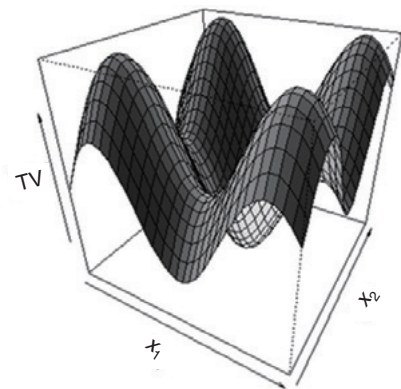
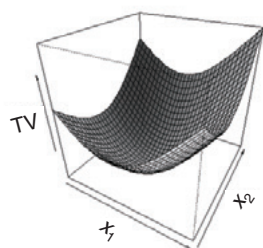


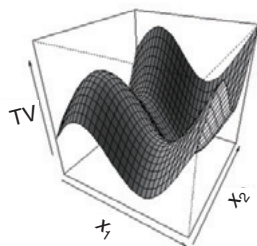
Fig. 3 Theoretical surface

tens and hundreds of data. With this quantity of data, it becomes possible to actually produce a material and evaluate its characteristics.

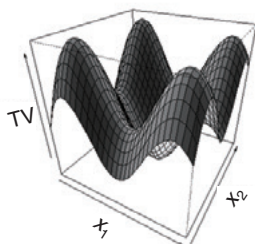
The studies discussed above verified that it would be possible to employ SMI methods to search for new materials. In order to determine whether SMI methods would be practically applicable, searches for new materials were conducted using the method in relation to the catalyst, magnet, and thermoelectric materials selected as case studies.



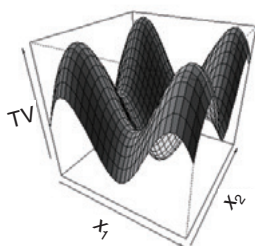
(a) Predicted surface: 9 training data



(b) Predicted surface: 16 training data



(c) Predicted surface: 25 training data



(d) Predicted surface: 36 training data

Fig. 4 Predicted surfaces for different numbers of training data

3. Study of Practical Applicability of SMI Method through Creation of New Materials

3.1. Catalyst Material

Exhaust gas conversion performance at the low temperatures at engine start is a particular demand on automotive three-way catalysts⁽³⁾. Three-way catalysts are made up of precious metals and a material that supports them. These precious metals initially exist in the form of particles of around several nm, but the particles aggregate in the high-temperature automotive exhaust gas environment, and these aggregates grow to between several tens and several hundreds of nm.

The support material was focused on in an attempt to control this aggregation of precious metal particles in high-temperature environments. With element 1 termed A, element 2 termed B, and oxygen termed O, a comprehensive search for the elements A and B to be employed in a composite oxide ABO was attempted. Only a 1:1 ratio of A to B was considered. Figure 5 shows the prediction procedure of this catalyst material.

Element-specific property data for elements A and B when they exist separately, including primary ionization energy and melting point, were considered to be explanatory variables, and the specific surface area of palladium (Pd) particles when supported by the material composed of elements A and B was the target variable. Given that it was necessary that elements A and B could be employed in an automotive catalyst, elements that are gaseous at ordinary temperatures and radioactive elements were excluded. When elements that could not be employed were excluded, the number of potential elements was 64 for both elements A and B. This produced 2080 possible combinations of A and B, including the combination A=B. A and B were varied to produce 52 ABO support materials. The specific surface area of Pd particles supported on these materials was measured, generating 52 training data. LOOCV was calculated for this number of data to determine whether it fell within the scope for application of the prediction model using the method discussed in Chapter 2. Figure 6 shows the results.

Based on Fig. 6, it was judged that at 45 data and above, the fluctuations in LOOCV began to abate, and at 52 data the fluctuations tapered off. The number of data was

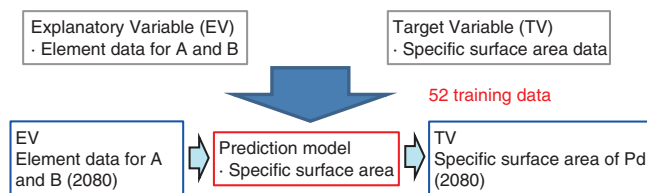


Fig. 5 Procedure for prediction of specific surface area of Pd

therefore in the domain in which LOOCV in the prediction model was reliable, and it was judged that the prediction results could be utilized in selecting a catalyst support material. The explanatory variables for the abovementioned 2080 combinations of elements A and B were input to the prediction model formulated using these 52 data, and the specific surface area of Pd particles supported by materials composed of elements A and B was predicted. Combinations of elements A and B for which high specific surface areas were predicted were given priority for selection. Catalysts were created using these combinations, and the specific surface area of Pd particles supported by each support material was measured using CO gas absorption measurements. Table 1 shows the measurement results.

As a result, it was possible to create four support materials (I, II, III, and IV) on which the specific surface area of supported Pd particles was higher than the specific surface area when a mass-production support material was utilized. A supported Pd catalyst was created using support material I, the material producing the highest specific surface area in supported Pd particles, and its conversion characteristic following aging treatment was evaluated using model gases. Figure 7 shows the results.

As Fig. 7 indicates, the catalyst employing the developed support material displayed a superior conversion characteristic to a catalyst employing the mass-production support material in relation to CO, HC, and NOx. This was considered to be because the supported Pd particles were small and dispersed and a high specific surface area had been maintained, even following aging. Pd support material I was therefore analyzed using a scanning electron microscope (SEM), with the results shown in Fig. 8.

The Pd particles supported by support material I were around several nm. Energy-dispersive X-ray spectroscopy showed that the particles were elemental Pd. These results indicate that a catalyst material had been developed that supported Pd particles with a high specific surface area

Table 1 Summary of catalysts with higher specific surface area of Pd particles than conventional support material

Support material	Element A	Element B	Specific surface area of Pd particles
I	$\alpha 2$	$\beta 2$	1.39
II	$\alpha 1$	$\beta 1$	0.21
III	$\alpha 2$	$\beta 3$	0.13
IV	$\alpha 3$	$\beta 3$	0.34
Reference (Conventional support material)	Al	Al	0.11

even following aging, and that the SMI method employed in predicting the specific surface area of the Pd particles was valid. Support material I was selected based on these results, making it possible to create a supported Pd catalyst displaying superior performance to a catalyst using the mass-production support material.

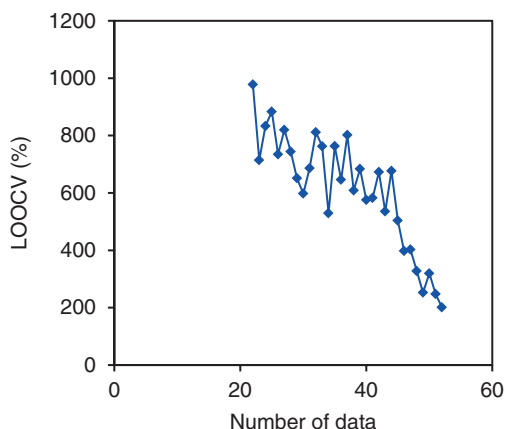


Fig. 6 Relationship between number of data and LOOCV when studying support materials

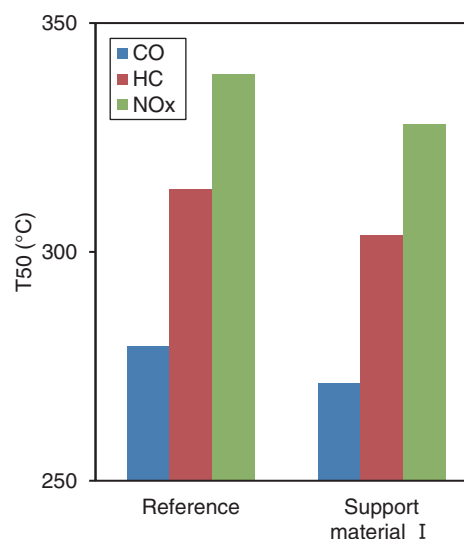


Fig. 7 Temperature for 50% conversion of individual gases

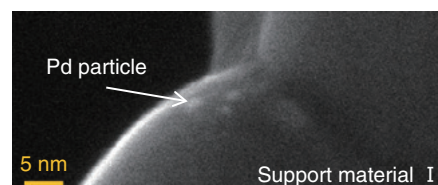


Fig. 8 SEM image of support material- I-with Pd particle

3.2. Magnet Material

Today, as automotive electrification proceeds, a higher level of performance is demanded from the magnets employed in drive motors. It is necessary for magnets employed in automotive use environments to maintain coercivity at high temperatures, and the realization of a fine structure in the magnet material is an important factor in achieving this goal⁽⁴⁾. Past studies have indicated that the two-stage low- and high-temperature processing undergone by hot-deformed magnets refines the structure of the material and renders it uniform⁽⁵⁾. The SMI method was therefore applied in relation to the parameters of the hot deformation process shown in Fig. 9, and a magnet material displaying a high level of performance was sought.

Figure 10 shows the procedure employed in applying the abovementioned SMI method.

The temperature (T1), amount of pressing (AP1), and velocity of pressing (VP1) employed in the first procedure and the temperature (T2) and amount of pressing (AP2) employed in the second procedure of the hot deformation process were considered explanatory variables, and the products of the magnetic flux density, Br, and coercivity, iHc, of magnets created while varying these explanatory variables were considered target variables. Twenty-eight types of training data were obtained. LOOCV profiles were calculated for these 28 types of training data, and it was verified that the extreme fluctuations in LOOCV had tapered off, and the data were within the domain in which results for LOOCV were reliable. The 28 types of training data were learned using a machine learning method compatible with nonlinearity, and a prediction model was formulated.

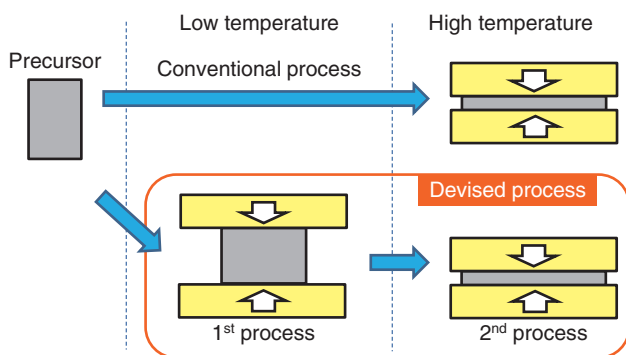


Fig. 9 Schematic image of hot deformation process

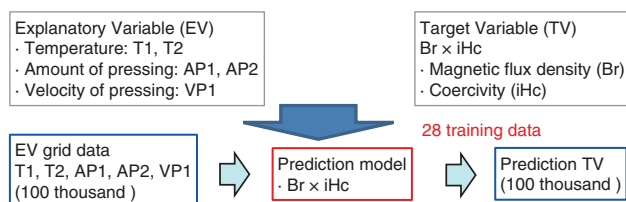


Fig. 10 Prediction procedure for magnet material

Table 2 Prediction result yielding highest level of magnet performance

TV (Br x iHc)	T1 (°C)	AP1 (%)	VP1 (-)	T2 (°C)	AP2 (%)
252	494	27.8	2.0E-05	700	20

100,000 hypothetical explanatory variables, representing all possible combinations when each explanatory variable was divided into 10 increments, were input into the prediction model, and 100,000 target variables were predicted. Table 2 shows the prediction results that yielded the highest level of magnet performance.

A magnet was created using the hot deformation process conditions shown in Table 2, and its performance was evaluated. Figure 11 shows the results, together with past results.

“Base” in Fig. 11 is the material employed when research on magnets produced using the hot deformation process commenced. “Y” is the magnet material obtained following approximately two years of research. When material Y was obtained, 28 data concerning the performance of magnets produced using different hot deformation conditions were obtained. A prediction model was formulated based on these 28 data, and the optimum conditions shown in Table 2 were predicted. The calculations for this prediction took around five minutes. A magnet was created using the conditions shown in Table 2, and magnet material X, which displayed superior performance to magnet material Y, was obtained.

The use of the SMI method made it possible to formulate a prediction model based on a limited number of data and determine the optimum conditions for the manufacture of a magnet material in a short period of time.

3.3. Thermoelectric Material

Half or more of the chemical energy of gasoline is exhausted as heat energy in automobiles. There are

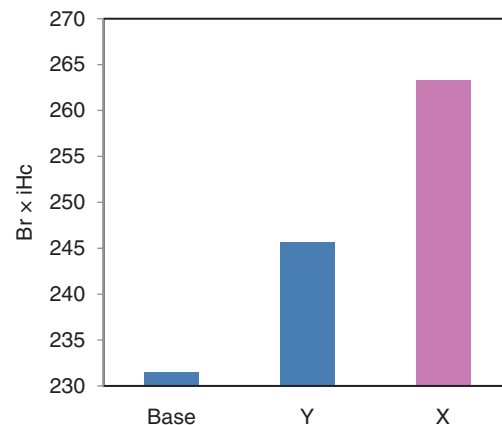


Fig. 11 Results of evaluation of magnet performance together with past data

expectations on thermoelectric materials to function as elements positioned between high- and low-temperature sections, able to recover heat energy by statically converting it into electrical energy⁽⁶⁾. Theoretically, the efficiency of thermoelectric materials in recovering heat energy increases in proportion to the temperature difference, but the temperature resistance of the materials represents an issue. The realization of increased temperature resistance through modification of the elemental composition of a thermoelectric material was therefore investigated. Figure 12 shows the crystal structure of skutterudite, the thermoelectric material selected for study.

The thermoelectric material shown in Fig. 12 is made up of the elements lanthanum (La), iron (Fe), cobalt (Co), and antimony (Sb). The thermoelectric characteristic, η , and the level of temperature resistance of the material are known to change with changes in these constituent elements and the concentration of the carrier.

A prediction model was formulated using the elemental composition shown above, and a study was conducted to seek a thermoelectric material displaying heat resistance while maintaining the same thermoelectric characteristic. Figure 13 shows the prediction procedure.

The quantities of La, Fe, Co, and Sb in the thermoelectric material and the carrier concentration were considered to be explanatory variables, and the thermoelectric characteristic, η , and the level of temperature resistance of the material were considered to be the target variables. The thermoelectric characteristic, η , and the level of temperature resistance of the material were measured when the explanatory variables, the material's elemental composition and carrier concentration, were changed as considered appropriate, and 19 types of training data were

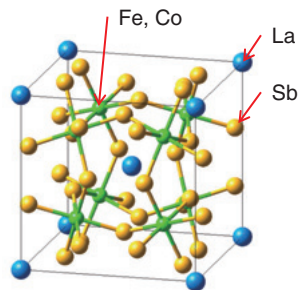


Fig. 12 Illustration of Skutterudite crystal structure

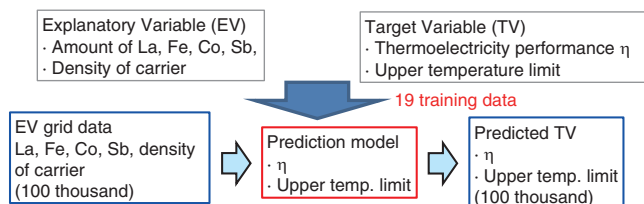


Fig. 13 Prediction procedure for thermoelectric material

obtained. The LOOCV profiles were calculated, and it was verified that the data were within the domain in which the LOOCV values could be considered reliable. The training data were employed in the formulation of two prediction models, for the material's thermoelectric characteristic, η , and its level of temperature resistance. 100,000 grid data, representing all possible combinations when each explanatory variable for each of the two prediction models was divided into 10 increments, were input into the prediction models, and 100,000 target variables were predicted. Figure 14 shows a Pareto diagram of the results.

Two plots for which the temperature resistance of the material was high at a specific thermoelectric characteristic, η , or above were selected from the prediction results, and a thermoelectric material was manufactured under conditions that made it possible to reproduce the selected elemental composition and carrier concentration. The thermoelectric characteristic, η , and level of temperature resistance of the material were then measured. Results for thermoelectric characteristic, η , and level of temperature resistance matched within the error bar, indicating that it had been possible to create a thermoelectric material possessing a higher level of temperature resistance than conventional materials.

These results indicate that the use of the SMI method in relation to characteristic values that are in a trade-off relationship, such as the thermoelectric characteristic, η , and the level of temperature resistance in this study, makes it possible to develop materials that respond to specific needs with pinpoint accuracy.

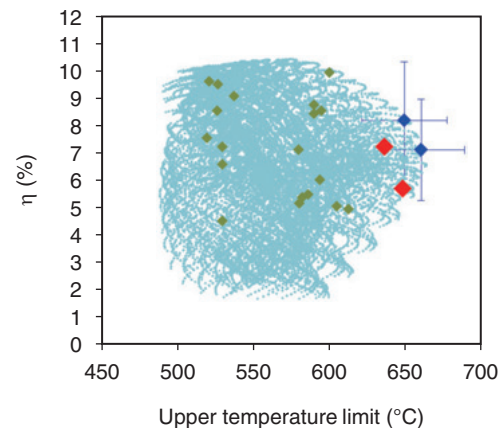


Fig. 14 Pareto diagram of η and upper temperature limit

4. Conclusion

A study of the behavior of LOOCV in relation to the number of data when applying the SMI method to material data demonstrated that the number of data that would be

effective in making predictions was between several tens of data and 100 data, indicating that the method could be employed in seeking new materials. The application of the SMI method to catalyst, magnetic, and thermoelectric materials demonstrated that the method made it possible to develop new materials in line with the aim for the material, for example boosting characteristics or increasing temperature resistance, within a short period.

While predictions were made only for functional materials in this study, the scope of applicability of the method is not restricted to these materials. If training data made up of target variables and explanatory variables that adequately explain the target variables is available, the method makes it possible to predict the results of tests not yet performed.

References

- (1) Materials Genome Initiative for Global Competitiveness, Executive Office of the President National Science and Technology Council Washington, D.C. 20502, p.4-16, June 24, (2011)
- (2) Nanotechnology/Materials Unit, Center for Research and Development Strategy, Japan Science and Technology Agency, 195p., CRDS-FY2013-WR-03
- (3) Inose, S., Okayama, T.: Research for High Concentration of Precious Metal and Thin Layer of Catalyst Top Layer, Honda R&D Technical Review, Vol. 17, No. 2, p. 129-135
- (4) Shimizu, H., Kato, R., Nakazawa, Y., Hattori, A., Hioki, K., Oikawa, T.: Heavy Rare Earth-free Hot-deformed Magnet for Traction Motor, Honda R&D Technical Review, Vol. 28, No. 2, p. 87-92
- (5) Mouri, T., Kumano, M., Yasuda, H.Y., Nagase, T., Kato, R., Nakazawa, Y., Shimizu, H.: Effect of two-stage deformation on magnetic properties of hot-deformed Nd-Fe-B permanent magnets, Scripta Materialia, Vol. 78-79, p. 37-40, (2014)
- (6) Mori, M., Yamagami, T., Oda, N., Hattori, M., Haraguchi, T., Sorazawa, M.: Heat Recovery Technology of Automobile Using Thermoelectric Element and Its Effect on Fuel Economy, Honda R&D Technical Review, Vol. 21, No. 2, p. 73-79

■ Author ■



Atsushi FURUKAWA



Tomohiro IKEDA



Tatsuya OKAYAMA