



## Important Descriptors and Descriptor Groups of Curie Temperatures of Rare-earth Transition-metal Binary Alloys

Hieu Chi Dam<sup>1,2,3</sup>, Viet Cuong Nguyen<sup>4</sup>, Tien Lam Pham<sup>1,5</sup>, Anh Tuan Nguyen<sup>6</sup>,  
Kiyoyuki Terakura<sup>1,2</sup>, Takashi Miyake<sup>2,5,7</sup>, and Hiori Kino<sup>2,5</sup>

<sup>1</sup>Japan Advanced Institute of Science and Technology, Nomi, Ishikawa 923-1292, Japan

<sup>2</sup>CMF<sup>2</sup>, MaDIS, NIMS, Tsukuba, Ibaraki 305-0047, Japan

<sup>3</sup>JST, PRESTO, Kawaguchi, Saitama 332-0012, Japan

<sup>4</sup>HPC Systems Inc., Minato, Tokyo 108-0022, Japan

<sup>5</sup>ESICMM, NIMS, Tsukuba, Ibaraki 305-0047, Japan

<sup>6</sup>Hanoi Metropolitan University, 98 Duong Quang Ham, Cau Giay, Hanoi, Vietnam

<sup>7</sup>CD-FMat, AIST, Tsukuba, Ibaraki 305-8568, Japan

(Received September 8, 2018; accepted October 2, 2018; published online October 29, 2018)

We analyze the Curie temperatures of rare-earth transition metal binary alloys using machine learning. In order to select important descriptors and descriptor groups, we introduce a newly developed subgroup relevance analysis and adopt hierarchical clustering in the representation. We execute exhaustive search and demonstrate that our approach results in the successful selection of important descriptors and descriptor groups. It helps us to choose the combination of descriptors and to understand the meaning of the selected combination of descriptors.

Magnets are now widely used and play an important role in energy savings.<sup>1,2)</sup> One of the most important applications of magnets is electric motors, whose performance significantly depends on the performance of magnets. Nd–Fe–B based rare-earth magnets are the strongest among the existing permanent magnets, and are almost the only type of permanent magnets that meets the stringent performance requirements of the recent electric motors. However one of the problems with Nd–Fe–B magnets is the relatively low Curie temperature compared to the operation temperatures of the motors. Therefore, many researchers have carried out studies to overcome this drawback, including the exploration of new magnets.

The Curie temperature ( $T_C$ ) is one of the most important physical quantities of magnets, but unfortunately, it is one of the most difficult physical quantities to predict correctly. There are several theory-driven methods for evaluating the  $T_C$  of magnetic materials.<sup>3)</sup> One of the basic approaches is to solve an (extended) Hubbard model by using various low-energy solvers. In principle, this method is expected to be accurate. However Anisimov et al. showed that the results are sensitive to the effective parameters and details of the low energy solver.<sup>4–6)</sup> Therefore, this approach is still at the level of testing the formalism for simple systems like pure transition-metal magnets.

Atomistic spin model is the most common choice for practical application to more complex systems.<sup>3)</sup> The spin model is constructed from the magnetic moment at each atomic site and the intersite magnetic exchange-couplings based on the assumption of fixed magnitude of spin moments. The parameters are evaluated using the first-principles calculations.<sup>3)</sup> This method can be applied to rare-earth magnets. Usually, the model is simplified further, and is restricted to the TM-3d and RE-4f spins. Then,  $T_C$ s is evaluated, usually in the mean field approximation. The mean field approximation, however, usually overestimates  $T_C$ s. Thus, there exist many sources of error in the  $T_C$  evaluation using the atomistic spin model. The development of theoretical methods for the estimation of the  $T_C$  is still underway.

In contrast to the deductive approaches described so far, there is now a movement toward utilizing inductive approaches, i.e., data-driven methods for estimating  $T_C$ , and there have been many reports of successful prediction of the physical quantities using such methods.<sup>7–12)</sup> The data-driven approach accumulates data, prepares descriptors, makes a model with the descriptors, and finally predicts the values of physical quantities of new materials. One of the key points to be considered for successful prediction is the choice of descriptors. A typical example of descriptor selection can be seen in the work by Ghiringhelli et al., where a regression model is used to predict the energy difference between zinc blende or wurtzite and rocksalt structures.<sup>13)</sup> They used a linear regression model, and first prepared basic descriptors. However, a linear regression model with only the basic descriptors has low description power. Then, they performed various operations on the basic descriptors and produced a number of nonlinear combinations among the basic descriptors. This resulted in an increase in the prediction power. They shrank the number of descriptors using LASSO and finally employed exhaustive search to find the best linear regression model. Their work shows that the combination of descriptors is important for increasing the accuracy of the regression model.

Usually, we select the best regression model and discard all the others (performance-optimized model). However we know that there exist many regression models, where the combination of the descriptors is different from the one that has the best score, but the score of which is as good as the best one indicated by the exhaustive search method. (The best score means, for example, the largest  $R^2$  value in the regression model.) There exists another strategy where we choose the regression model the score of which is not the best, but is high. For example, we can choose low cost descriptors, where “low cost” means easy or literally low cost to evaluate through experiments or calculations. This model is usually referred to an operation-optimized model. Okada et al. devoted considerable effort to the latter problem. They showed the scores of regression models as the density of states to understand the overall structure in one way, and



plotted the best scores as a function of the combinations in another way, such as the indicator diagram, to select the best combinations depending on the purpose of the analysis.<sup>14–16</sup>

Yet, it is not easy to understand the relationship and structures among descriptors from a huge list of scores and descriptors. Informatics treatment usually ignore the importance of the meaning of the descriptors, though they are physical parameters that physicists regard as important. However we hope that we can extract more information from the huge data. In the present work, we introduce a well-defined subgroup concept to clarify the relationship among descriptors. Our method can also elucidate how to choose combination of descriptors systematically as well as how to understand the meaning of descriptors.

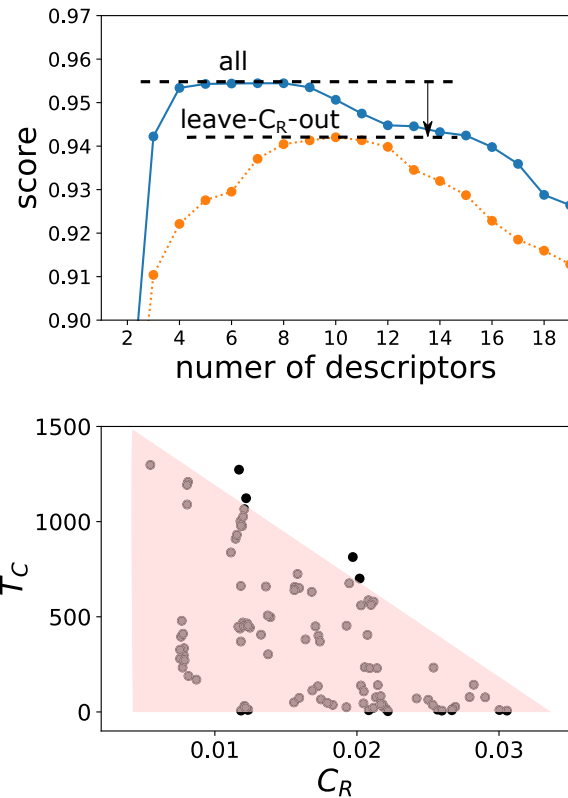
Our target variable is the experimental  $T_C$  of the rare-earth transition-metal binary stoichiometry alloys considered in this study.<sup>17</sup> We select the descriptors from the element dependent categories (R for rare-earth elements and T for transition metal elements), and utilize the knowledge of the conventional theory-driven method. The key parameters of the effective theory-driven models are related to the properties of the constituent elements and/or structural parameters. For example, the orbital energy level increases (becomes deeper) as the atomic number  $Z$  increases. The electron interaction becomes stronger as the atomic orbital becomes more localized. The magnetic exchange-couplings are associated with the strength of the electron interaction and transfer integrals. The coupling strength between TM-3d and RE-4f (through RE-5d) is crucial for discussing the RE dependence of magnetism. This strength is proportional to the 3d-4f effective exchange coupling and the 4f total spin projected onto the 4f total angular momentum  $J_{4f}$ . The latter quantity is given by  $J_{4f}(1 - g_J)$ , with  $g_J$  being the Landé  $g$ -factor. We also add the descriptors from the structure-related category (S) to describe the ratio of the elements as well as the real volume or spatial dependent simple variables to distinguish, e.g.,  $\text{Th}_2\text{Zn}_{17}$  and  $\text{Th}_2\text{Ni}_{17}$  polytypes. We list the descriptors in Table I, and give their detailed explanations in the supporting information.<sup>18</sup>

As a regression model, we employ kernel ridge regression with the radial basis function kernel. Kernel ridge regression can include the non-linear effects of the descriptors and has much stronger power to fit the target functions with the descriptors, though there exist a demerit of taking much more time to fit/predict the regression models than the linear regression does. We used Python scripts with mpi4py, scipy and scikit-learn.<sup>19–21</sup> Our scores in the regression models are the  $R^2$  values, which we evaluate in the leave-one-out cross validation.

First, we analyze the descriptors. We take Pearson's correlation coefficient between the descriptors. For the T category, the absolute values of Pearson's correlation coefficient among the three descriptors,  $Z_T$ ,  $r_T$ , and  $S_{3d}$ , are the same, namely 1, which means that their contributions are the same in the regression model after the normalization procedure. Therefore, the number of independent descriptors is reduced from 27 to 25. Then, we perform exhaustive search for  $2^{25} - 1 = 3.3 \times 10^7$  regression models where the combinations of descriptors are different, and evaluate their accuracy values (scores).

**Table I.** Transition metal, rare-earth, and structural descriptors. See also the supporting information.<sup>18)</sup>

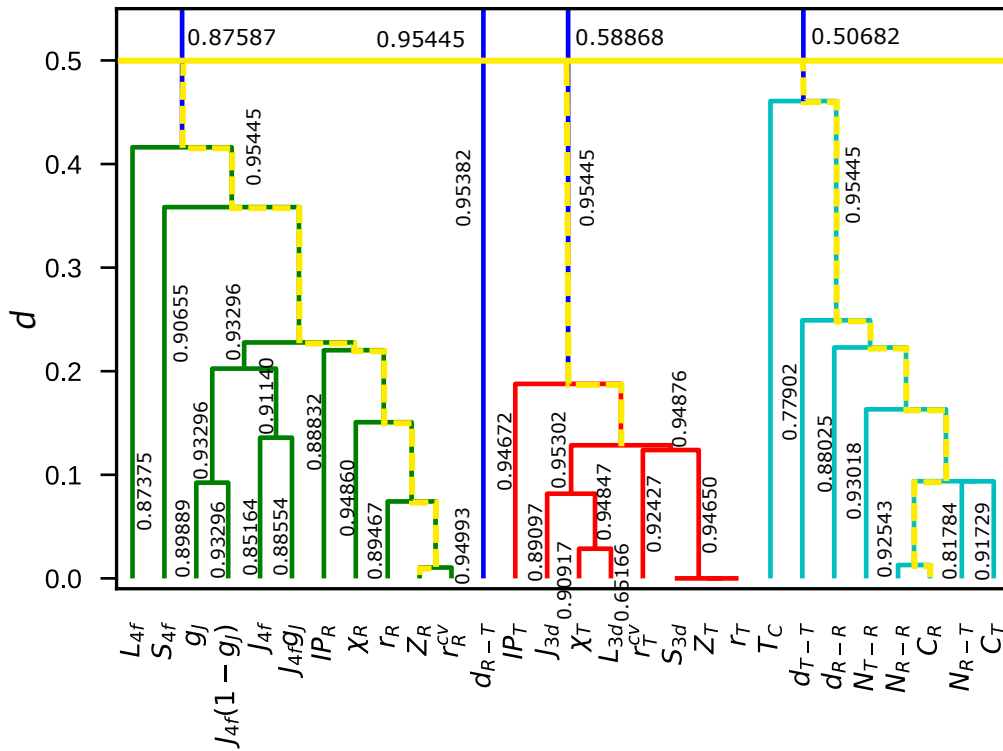
Category	Descriptors
Atomic properties of transition metals (T)	$Z_T, r_T, r_T^{cv}, IP_T, \chi_T, S_{3d}, L_{3d}, J_{3d}$
Atomic properties of rare-earth metals (R)	$Z_R, r_R, r_R^{cv}, IP_R, \chi_R, S_{4f}, L_{4f}, J_{4f}, g_J, J_{4f}g_J, J_{4f}(1 - g_J)$
Structural information (S)	$C_T, C_R, d_{T-T}, d_{T-R}, d_{R-R}, N_{T-R}, N_{R-R}, N_{R-T}$



**Fig. 1.** (Color online) Top panel: The blue line shows the best score for each number of descriptors. The orange dotted line shows the score when  $C_R$  is removed. Bottom panel:  $C_R$  ( $\text{\AA}^{-3}$ ) vs  $T_C$  ( $^{\circ}\text{C}$ ).

Usually, we evaluate the score of the regression model; however, we want to evaluate the importance of the descriptors. Therefore, we change the viewpoint from the regression model to the descriptor in order to discuss the importance of the latter. We use relevance analysis,<sup>22,23</sup> which roughly corresponds to the linear response theory with respect to the descriptors. (We explain the scores and relevance analysis in the supporting information.<sup>18</sup>) It originally utilizes the change in values when we remove/add a descriptor. The former corresponds to the leave-one-out experiment, while the latter corresponds to the add-one-in experiment. The descriptor is strongly or weakly relevant when its accuracy score changes meaningfully in the leave-one-out or the add-one-in experiment, respectively.

Our first relevance analysis is based on strong relevance. We found that only the descriptor,  $C_R$ , is strongly relevant. We can verify the importance of  $C_R$  when we plot  $C_R$  vs  $T_C$ . Almost all the points are placed in the bottom-left side of the right panel of Fig. 1. Thus, it is clear that  $C_R$  has a considerable influence on the  $T_C$ . It should be noted that we



**Fig. 2.** (Color online)  $R^2$  scores of the subgroup relevance analysis on the hierarchical clustering of the descriptors. We include  $T_C$  in the dendrogram. The group R (green) is from  $L_{4f}$  to  $r_R^{cv}$ . The group T (red) is from  $IP_T$  to  $r_T$ . The group S (cyan) is from  $d_{TR}$  to  $C_T$ . The group  $d_{TR}$  is made of the descriptor  $d_{TR}$ . The horizontal values are strong relevance values and the tilted values are weak relevance values. The vertical axis shows the distance,  $d$ , and the values are one minus the absolute values of Pearson’s correlation coefficient. The paths of the highest value (0.95445) are colored in yellow dashed lines. See details in the main body also.

will not able to find such a relationship if we simply execute the regressions.

We notice that relevance analysis can be done not only for a descriptor, but also for a subgroup of descriptors. We define groups and subgroups in this paragraph. The second relevance analysis is based on weak relevance, where, in the original prescription, we add another *descriptor* to the set of descriptors, which we must define. We define the groups and subgroups here, and make use of them in the relevance analysis. We utilize hierarchal clustering analysis, where the distance between descriptors is one minus the absolute values of Pearson’s correlation coefficient. We can define the groups or subgroups of descriptors that are clustered based on the criteria of them being within distance,  $d$ , of each other. For example, we can define four groups at  $d = 0.5$ . Two of them have the same descriptors as those of the T and R categories, while the other two have that of the original S category. (We call the original cluster as *category* and the cluster by the hierarchical analysis as *group*.) The  $d_{TR}$  constitutes a group, while the other S category descriptors constitute the other. It is not surprising that the grouping at  $d = 0.5$  is almost the same as the categories defined a priori as T, R, and S when we remember the definition of the descriptors of the materials. Here, we successfully defined the groups and subgroups, where the groups are almost the same as the original category but are clustered from the data themselves. (We redefine the group S as a result of this clustering. The group S that does not include  $d_{TR}$  is different from the category S.)

We can make further advances in this grouping. We notice that the definition of the value of  $d$  is unnecessary, but we

only have to define the vertical line of the decomposition tree to define the subgroups because the child nodes below the vertical line is the same. (See also Fig. 2. The vertical axis corresponds to  $d$ .) Thus, we are able to define many subgroups of the descriptors as sets of the child nodes of the *dendrogram*.

We apply the relevance analysis not to a descriptor but to a subgroup/group. We call this method *subgroup relevance analysis*. We plotted the result in Fig. 2. The horizontal score is evaluated in the leave-one-out experiment and is related to the strong relevance, while the vertical scores are evaluated in the add-one-in experiment and is related to the weak relevance. Note that the score of a subgroup belonging to a group is evaluated under the condition that we must use at least one descriptor in the subgroup, and any descriptors belonging to the other groups can be added in the weak relevance analysis.

In Fig. 2, the weak relevance values, or add-one-in values, are written as vertical values. The subgroup containing only  $r_R$  has the score, 0.89467, which is the highest score in the condition that we must take the subgroup  $r_R$  in the group R and we can take any descriptors in the other groups. (A subgroup which has a descriptor is also a subgroup.) The subgroup containing  $r_R$ ,  $Z_R$ , and  $r_R^{cv}$  has the score, 0.95445, which is the highest score in the condition that we must take at least one descriptor in the subgroup  $r_R$ ,  $Z_R$ , and  $r_R^{cv}$  of the group R and we can take any descriptors in the other groups as explained in the previous paragraph.

The sole descriptor  $Z_R$  in the group R has the highest score (0.95445). It means that  $Z_R$  can solely represent the group R. This is also the case for the  $C_R$  subgroup in the group S.

However the structure of the group T is different from those of the groups R and S. The subgroup made of  $J_{3d}, \chi_T, r_T^{cv}, Z_T$  (and  $r_T$  and  $S_{3d}$ ) has the highest score (0.94876), but its child subgroup descriptors have smaller scores (0.92427 and 0.94650). It means that there exists no single descriptor that can represent the overall nature of the group T. When we examine all the combinations made of  $J_{3d}, \chi_T, r_T^{cv}, Z_T$ , we find that  $Z_T$  takes the best score (0.95450) if we choose only one of the descriptors among them, a set of  $Z_T$  and  $J_{3d}$  is the best (0.95339) for two descriptors, and a set of  $Z_T, J_{3d}$ , and  $L_{3d}$  is the best (0.95445) for three descriptors. We note that the descriptor  $Z_T$  has the same effect as  $S_{3d}$ . We discuss interpretation of the result later.

We can also obtain the importance of the groups from the horizontal values above the yellow solid line in Fig. 2. They are the strong relevance values, or leave-one-out values of the groups T, R, and S. For example, the group R has the value, 0.87587, which is the best score when we remove all the descriptors of the group R. The better the score is, the less important the group is. The value, 0.50682, is the smallest among them, which means that the group S is the most important among the groups. On the other hand, the least important group is R, the value of which is 0.87587. It means that the score still holds a high value even if we exclude all the descriptors in the group R. Therefore, the importance of group R is the lowest among T, S, and R.

We have added additional explanation in Fig. 2. The descriptor  $J_{4f}(1 - g_j)$  can represent the subgroup containing  $g_j, \dots, J_{4f}g_j$ , but the score is 0.93296, which is lower than the score 0.95445 of  $Z_R$ . We have also added a comment on the group of  $d_{TR}$ . The strong relevance value is 0.95445 and the weak relevance value is 0.95382. The facts that their difference is small and that the weak relevance value is smaller than the strong relevance value mean that the existence of the group  $d_{TR}$  makes the regression model worse.

Here, we compare the result of the subgroup relevance analysis shown in Fig. 2 with the best score having  $n$  descriptors without the subgroup relevance analysis, which is shown in Table II. The set of  $C_R, Z_R$ , and  $Z_T$  has the best score (0.94222) for  $n = 3$ . The set of  $C_R, Z_R, Z_T$ , and  $J_R$  has the best score (0.95339) for  $n = 4$ . The set of  $C_R, Z_R, Z_T, J_R$ , and  $L_{3d}$  has the best score (0.95429) for  $n = 5$ . The descriptor sets are made of the most important descriptors in group R ( $Z_R$ ), group S ( $C_R$ ), and group T ( $Z_T$  when we choose a descriptor;  $J_{3d}$  and  $Z_T$  when we choose two descriptors; and  $J_{3d}, L_{3d}$ , and  $Z_T$  when we choose three descriptors.) These combinations are the same as the analysis in the previous paragraph. Thus, the subgroup relevance analysis successfully illustrates the structure among the descriptors and their importance.

One may think that the difference in the scores are quite tiny. For example, 99.0% value of the global best score is 0.944, which roughly corresponds to the best score with 12 descriptors (see also Table I in the supporting information).<sup>18)</sup> However the predicting ability changes drastically. We plot the ‘‘RMSE’’ between the best models with  $n$  descriptors in Fig. 2 in the supporting information.<sup>18)</sup> It can be clearly seen that the prediction abilities for  $n = 3$  to 8 is qualitatively different from those for  $n \geq 9$ , but the difference of the score of the best model with 9 (10) descriptors to the

**Table II.** The best  $R^2$  score and descriptors as a function of the number of descriptors  $n$ .

$n$	Score	Descriptor(s)
2	0.87015	$C_R, Z_T$
3	0.94222	$C_R, Z_R, Z_T$
4	0.95339	$J_{3d}, C_R, Z_R, Z_T$
5	0.95429	$L_{3d}, J_{3d}, C_R, Z_R, Z_T$
6	0.95439	$L_{3d}, J_{3d}, \chi_T, C_R, Z_R, Z_T$
7	0.95445	$L_{3d}, J_{3d}, \chi_T, C_R, Z_R, Z_T, r_T^{cv}$
8	0.95445	$L_{3d}, J_{3d}, \chi_T, IP_T, C_R, Z_R, Z_T, r_T^{cv}$

global best model is only 0.1% (0.4%). The difference in the score looks tiny at a glance, but is meaningful in this data and regression model. (One must also discuss the total density of state of the scores to discuss the meaningful difference of the scores, but it is beyond the scope of this study.<sup>14–16)</sup>

The ordering of the scores of the models (combinations of descriptors) can be changed according to the details of the regression scheme and noise in the data, because the differences in the scores are quite small (Table II in the main body and Table I in the supporting information).<sup>18)</sup> Thus, just showing the best models with  $n$  descriptors may give us wrong information. However the relevance analysis can give us more significant differences. The dendrogram, or grouping, does not depend on the scores of the models because it is made only of the distances between the descriptors. Even if there exists noise in the data, which may affect the scores of the model, we can expect that similar descriptors will give similar scores. The subgroup relevance analysis can illustrate how the distances, or the similarities, between the descriptors affect to the models.

Here, we further explain the advantage of the expression with the dendrogram. For example, we can easily choose  $r_R^{cv}$  if we do not want to use  $Z_R$  if the importance is expressed as in Fig. 2. It enables us to find the next best route, that is, to go upward and try a new branch downward in the tree structure. We believe that this expression is much better than simply providing a list, and it is much easier to find out the operation-optimized regression models.

We can conclude that the descriptor  $C_R$  is strongly relevant when we define the subgroups at  $d \sim 0$  and execute the leave-one-out experiment. The original relevance analysis is the special case of the subgroup relevance analysis. Therefore, the subgroup relevance analysis is a natural extension of the original relevance analysis.

Here, we note the possible interpretation of the regression model in the context of condensed matter physics, where we know that physics should depend not on  $J_{4f}$  but on  $J_{4f}(1 - g_j)$  in the effective model Hamiltonian. We, however, found more important descriptors, e.g.,  $Z_R$  and  $r_R^{cv}$  in the group R and  $J_{3d}$  in the group T. It is more plausible that the regression model found a relationship similar to the generalized Slater–Pauling curve for Curie temperature as a function of  $C_R$  and  $Z_T$  and  $Z_R$ , and that the other effects are only marginal.<sup>24)</sup> We introduced many descriptors that cannot appear in the atomic-scale effective model Hamiltonian, and the regression model simply selected the *inter-scale* regression model including the macro scale parameter  $C_R$  first and  $Z_T$  and  $Z_R$  next, which do not directly appear in the effective model Hamiltonian because their relationships are more



apparent. It should be noted that the number of data, only about a hundred, is too few to discuss the details because it can easily change the prediction accuracy as discussed in the supporting information.<sup>18)</sup>

We cannot avoid errors in  $T_{CS}$  because of experimental errors and human errors. The latter is mainly because AtomWork does not allow web scraping. We examine the possibility of outlier detection using machine learning. We show a plot of experimental  $T_{CS}$  versus predicted ones in the supporting information.<sup>18)</sup> The overall coincidence is good from 0 K to  $\sim 1300$  K, but there exist a few outliers. We mainly check the outliers of  $T_{CS}$  and fix the errors again and again if there are any. We found three major errors and a minor error. After fixing these errors, we evaluated the cross-validation test scores again for the best  $n$  descriptors of the original regression model. The best  $R^2$  was 0.96688. By using machine learning, it may be able possible to find data errors efficiently; however, it cannot detect data prediction of which appears consistent with the experimental values accidentally.

We employed Pearson's correlation coefficient to define the distance in this study. However, there exist many choices for the distance. It depends on the problem whose representation is the most appropriate in the unsupervised learning part. We use the similarity, or distance, between materials to find the regression model, but usually discard the similarity between descriptors to make the regression model. We, however, utilized the latter similarity, and therefore took full advantage of the similarity of the data in this prescription.

We showed that the distances between the descriptors are useful to illustrate the importance of descriptors and descriptor groups. This result is not strange when the descriptors have some physical meaning. There exists, however, minor discrepancies in the subgroup containing  $Z_R$ ,  $J_{3d}$ , and  $L_{3d}$  in the dendrogram. This is a limitation of this theory; however, it is possible to overcome this difficulty. We used the distance between the descriptors to explain the scores of the relevance analysis, but its inverse problem is also possible. We can set the value of distances between the descriptors, or the structures of the dendrogram, to be more consistent with the scores of the relevance analysis.

We can consider many variants of the subgroup relevance analysis. We took the best *descriptor* from the subgroup shown in yellow in Fig. 2. Thus, we were able to show the best descriptors in the subgroup. Another method is to take the best *subgroup* in the downstream to a specified subgroup. Then, we will be able to understand the relationship among subgroups, and we can easily change them depending on the purpose.

Note that the Monte-Carlo tree search also utilizes the same nature of tree structures. There may be a route to find out the almost best regression model by utilizing subgroup decomposition without performing expensive exhaustive search.

In summary, we studied the data-driven approach on the Curie temperature of rare-earth transition metal stoichiomet-

ric alloys. We successfully made regression models that achieved high scores from our descriptors. We developed subgroup relevance analysis and successfully illustrated the importance, relationship, and structures among the descriptors from a huge list of exhaustive search. In addition, it should be noted that our method makes full use of the similarity of the given data.

**Acknowledgments** This work was partly supported by PRESTO and by the "Materials Research by Information Integration" Initiative (MI<sup>2</sup>I) project of the Support Program for Starting Up Innovation Hub, both from the Japan Science and Technology Agency (JST), Japan; by the Elements Strategy Initiative Project under the auspices of MEXT; and also by MEXT as a social and scientific priority issue (Creation of New Functional Devices and High-Performance Materials to Support Next-Generation Industries; CDMSI) to be tackled by using a post-K computer. The calculations were partly carried out on Numerical Materials Simulator at NIMS.

- 1) S. Sugimoto, *J. Phys. D* **44**, 064001 (2011).
- 2) S. Hirotsawa, M. Nishino, and S. Miyashita, *Adv. Nat. Sci.: Nanosci. Nanotechnol.* **8**, 013002 (2017).
- 3) T. Miyake and H. Akai, *J. Phys. Soc. Jpn.* **87**, 041009 (2018), and references therein.
- 4) A. S. Belozеров, I. Leonov, and V. I. Anisimov, *Phys. Rev. B* **87**, 125138 (2013).
- 5) A. S. Belozеров and V. I. Anisimov, *J. Phys.: Condens. Matter* **26**, 375601 (2014).
- 6) A. A. Katanin, A. S. Belozеров, and V. I. Anisimov, *Phys. Rev. B* **94**, 161117 (2016).
- 7) R. Potyrailo, K. Rajan, K. Stoewe, I. Takeuchi, B. Chisholm, and H. Lam, *ACS Comb. Sci.* **13**, 579 (2011).
- 8) K. Rajan, *Informatics for Materials Science and Engineering: Data-driven Discovery for Accelerated Experimentation and Application* (Butterworth, Oxford, U.K., 2013).
- 9) A. Agrawal and A. Choudhary, *APL Mater.* **4**, 053208 (2016).
- 10) A. Jain, G. Hautier, S. P. Ong, and K. Persson, *J. Mater. Res.* **31**, 977 (2016).
- 11) Y. Liu, T. Zhao, W. Ju, and S. Shi, *J. Materiomics* **3**, 159 (2017).
- 12) W. Lu, R. Xiao, J. Yang, H. Li, and W. Zhang, *J. Materiomics* **3**, 191 (2017).
- 13) L. M. Ghiringhelli, J. Vybiral, S. V. Levchenko, C. Drax, and M. Scheffler, *Phys. Rev. Lett.* **114**, 105503 (2015).
- 14) K. Nagata, J. Kitazono, S. Nakajima, S. Eifuku, R. Tamura, and M. Okada, *IPJS Trans. Math. Model. Appl.* **8**, 23 (2015).
- 15) T. Kuwatani, K. Nagata, M. Okada, T. Watanabe, Y. Ogawa, T. Komai, and N. Tsuchiya, *Sci. Rep.* **4**, 7077 (2014).
- 16) H. Ichikawa, J. Kitazono, K. Nagata, A. Manda, K. Shimamura, R. Sakuta, M. Okada, M. K. Yamaguchi, S. Kanazawa, and R. Kakigi, *Front. Hum. Neurosci.* **8**, 480 (2014).
- 17) The values of experimental  $T_C$  are taken from the AtomWork database, <http://crystdb.nims.go.jp/>.
- 18) (Supplemental Material) More detailed explanations are available online.
- 19) F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, *J. Mach. Learn. Res.* **12**, 2825 (2011).
- 20) T. E. Oliphant, *Comput. Sci. Eng.* **9**, 10 (2007).
- 21) K. J. Millman and M. Aivazis, *Comput. Sci. Eng.*, **13**, 9 (2011).
- 22) L. Yu and H. Liu, *J. Mach. Learn. Res.* **5**, 1205 (2004).
- 23) S. Visalakshi and V. Radha, *IEEE Int. Conf. Computational Intelligence and Computing Research*, 2014, p. 1.
- 24) For example, C. Takahashi, M. Ogura, and H. Akai, *J. Phys.: Condens. Matter* **19**, 365233 (2007).