A regression-based feature selection study of the Curie temperature of transition-metal rare-earth compounds: prediction and understanding

Hieu Chi Dam$^{1,2,3}$, Viet Cuong Nguyen$^4$, Tien Lam Pham$^{1,5}$, Anh Tuan Nguyen$^6$, Hori Kino$^{2,5}$, Kiyoyuki Terakura$^{1,2}$, and Takashi Miyake$^{2,5,7}$

$^1$Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan
$^2$CMF$^2$, MaDIS, NIMS, Tsukuba, Ibaraki 305-0047, Japan
$^3$JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama, 332-0012, Japan
$^4$HPC Systems Inc., Japan
$^5$ESICMM, NIMS, Tsukuba 305-0047, Japan
$^6$Hanoi Metropolitan University, 98 Duong Quang Ham, Cau Giay, Hanoi, Vietnam
$^7$CD-FMat, AIST, Tsukuba, Ibaraki 305-8568, Japan

(Dated: May 4, 2017)

The Curie temperature ($T_C$) of binary alloy compounds consisting of 3$d$ transition-metal and 4$f$ rare-earth elements is analyzed by a machine learning technique. We first demonstrate that nonlinear regression can accurately reproduce $T_C$ of the compounds. The prediction accuracy for $T_C$ is maximized when five to ten descriptors are selected, with the rare-earth concentration being the most relevant. We then discuss an attempt to utilize a regression-based model selection technique to learn the relation between the descriptors and the actuation mechanism of the corresponding physical phenomenon, i.e., $T_C$ in the present case.

PACS numbers:

The development of strong permanent magnets is an urgent technological issue as well as a fundamental challenge in materials science. Most strong permanent magnets are rare-earth magnets that mainly consist of transition-metal and rare-earth elements. To date, various rare-earth transition-metal compounds have been synthesized. The strongest magnetic compound is Nd$_2$Fe$_{14}$B developed by Sagawa et al., which is the main phase of neodymium magnets $^1$. Although Sagawa’s intention of raising $T_C$ by adding B was successful ($T_C = 335$ K (Nd$_2$Fe$_{17}$) vs. 586 K (Nd$_2$Fe$_{14}$B)), $T_C$ of Nd$_2$Fe$_{14}$B is still much lower than those of Sm$_2$Co$_{17}$ (1,193 K) and α-Fe (1,043 K). Because of this, dysprosium is added to ensure a sufficient coercivity at high temperatures for technological usage.

It is highly desirable to establish a technique that enables an accurate description of $T_C$ and to clarify the controlling parameters (descriptors) that influence $T_C$ for the development of a new strong magnet. From a theoretical point of view, an accurate description of $T_C$ is a demanding task. The spin magnetic moment at each atomic site and the magnetic exchange coupling between them can be evaluated within the framework of density functional theory $^2$. The Curie temperature can be obtained by solving a derived classical spin model. It is known that this approach works reasonably well for systems where the longitudinal spin fluctuation can be neglected. In the case of rare-earth compounds, however, the 4$f$ electrons bring further complications. Both the electron correlation and spin–orbit interaction are strong in these compounds; hence, an advanced theoretical treatment is needed for a reliable description of the magnetic properties.

The recent rapid development of data-driven approaches in materials research offers another possibility $^3$–$^6$. Nowadays, we are able to obtain $T_C$ data for many compounds from databases or the literature. Machine learning may be utilized to predict $T_C$ of a new compound from the existing data. This approach is particularly suited to materials exploration because of its high efficiency. In the present work, we demonstrate how this idea is achieved by employing the Gaussian kernel ridge regression (GKR) technique for machine learning. We focus on $T_C$ of transition-metal rare-earth bimetals. We present a comparison between the value of $T_C$ predicted by GKR and the observed data and analyze the selection of descriptors for obtaining a model for predicting $T_C$ with a high prediction accuracy. In this way, we obtain information about the important descriptors for the $T_C$ prediction. We may also expect that insights into the actuating mechanisms of physical phenomena can be obtained through the process mentioned above. However, “prediction” and “understanding” are not necessarily achieved simultaneously. This is because the “correlation relation” obtained by the machine learning prediction is generally different from the “causality relation” required for the understanding of the actuating mechanism. In the following, we first discuss how a prediction with a high accuracy can be achieved and then present our attempts to obtain insights into the actuating mechanism.
TABLE I: Descriptors representing the structural and physical properties of binary compounds. (Details can be found in Section A in the Supplemental Materials [10]).

| Category                          | Descriptors                                                                 |
|-----------------------------------|-----------------------------------------------------------------------------|
| Atomic properties of transition metals ($\mathbf{T}$) | $Z_T, r_T, r_{T}^2, J_{Pt}, \chi_{T}, S_{ad}, L_{3d}, J_{3d}$              |
| Atomic properties of rare-earth metals ($\mathbf{R}$) | $Z_R, r_R, r_{R}^2, J_{Pt}, \chi_{R}, S_{af}, L_{4f}, J_{4f}, g_{j}, J_{j}g_{j}, J_{4f}(1-g_{j})$ |
| Structural information ($\mathbf{S}$) | $C_T, C_R, d_{T-R}, d_{T-R}, d_{R-R}, N_{T-T}, N_{T-R}, N_{R-R}, N_{R-T}$ |

**Preparation for the data analysis**

We collected the experimental data of 108 binary compounds consisting of transition metals and rare-earth metals from the AtomWork database of NIMS Pauling File [Atomwork], including the crystal structure of the compounds and their observed $T_C$. Our task is to learn a model for predicting $T_C$ of a new compound from the training data of known compounds. For this purpose, one of the most important steps is the choice of an appropriate data representation that reflects knowledge of the application domain, i.e., a model of the underlying physics.

To represent the structural and physical properties of each binary compound, we use a combination of 28 descriptors. We divide all 28 descriptors into three categories, as summarized in Table I.

The first and second categories pertain to the descriptors describing the atomic properties of the transition-metal elements ($\mathbf{T}$ descriptors) and rare-earth elements ($\mathbf{R}$ descriptors), respectively. The descriptors related to the magnetic properties are included. It has been well established that information related to the crystal structure is very valuable in understanding the physics of binary compounds with transition metals and rare-earth metals. Therefore, we design the third category with structural descriptors ($\mathbf{S}$ descriptors) whose values are calculated from the crystal structures of the compounds in the literature. Note that the concentration of the transition metal ($C_T$) and that of the rare-earth metal ($C_R$) are both used. If we use the atomic percent for the concentration, the two quantities are not independent. Here, we measure the concentration in units of (atoms/A$^3$), which is more informative than the atomic percent because the former contains information about the constituent atomic size. Then, ($C_R$) and ($C_T$) are not totally dependent.

**Prediction**

To learn a function for predicting the values of $T_C$ of compounds from the data represented by using vectors of the descriptors, we apply the GKR technique [11], which has recently been applied successfully to many materials science issues [12, 13]. For GKR, the predicted property $f(\mathbf{x})$ at the point $\mathbf{x}$ is expressed as the weighted sum of Gaussians:

$$f(\mathbf{x}) = \sum_{i=1}^{N} c_i \exp\left(-\frac{||\mathbf{x}_i - \mathbf{x}||_2^2}{2\sigma^2}\right),$$

where $N$ is the number of training data points, $\sigma^2$ is a parameter corresponding to the variance of the Gaussian kernel function, and $||\mathbf{x}_i - \mathbf{x}||_2^2 = \sum_{i=0}^{N} (x_i^2 - x^2)$ is the squared $L^2$ norm of the difference between the two $N_D$-dimensional descriptor vectors $\mathbf{x}_i$ and $\mathbf{x}$. The coefficients $c_i$ are determined by minimizing

$$\sum_{i=1}^{N} y_i^2 + \lambda\sum_{i=1}^{N} ||c_i||_2^2,$$

where $y_i$ is the observed data value. The regularization parameters $\lambda$ and $\sigma$ are chosen with the help of cross-validation, i.e., by excluding some of the materials as a test set during the training process and measuring the coefficient of determination $R^2$ defined by

$$R^2 = 1 - \frac{\sum_{j=1}^{N_{test}} (y_j - f(\mathbf{x}_j))^2}{\sum_{j=1}^{N_{test}} (y_j - \bar{y})^2},$$

where $N_{test}$ is the number of test points, and $\bar{y}$ is the average of the test set for testing how the predicted values for the excluded materials agree with the actually observed values. In this study, we use $R^2$ as a measure of the prediction accuracy. To obtain a good estimate of the prediction accuracy, we carried out GKR with 100 times 10-fold cross-validation using the collected data, of which $T_C$ is the target quantity.

To find the most appropriate set of descriptors for the prediction of $T_C$, we train the GKR models for all combinations of the descriptors. With each combination, we search for the regularization parameters $\lambda$ and $\sigma$ to maximize the prediction accuracy. Then, we obtain new data for all possible sets of descriptors and the corresponding best prediction accuracy that the GKR model can achieve. By analyzing these new data, we aim (1) to find the set of descriptors that yields the best prediction accuracy for the prediction of $T_C$ and (2) to find the descriptors with strong relevance for the prediction of $T_C$.

We define the prediction ability $PA(\mathbf{S})$ of descriptors by the maximum prediction accuracy that the GKR model can achieve by using the descriptors in a subset $\mathbf{s}$ of a set $\mathbf{S}$ of descriptors as follows:

$$PA(\mathbf{S}) = \max_{\mathbf{s} \in \mathbf{S}} R^2_s,$$

where $R^2_s$ is the value of the coefficient of determination $R^2$ achieved by GKR using a descriptor set $\mathbf{s}$.

Once a large amount of new data (data of the GKR models) is obtained ($2^{N_D} - 1$ with $N_D = 28$), we are ready to analyze these new data to find the most appropriate model for the prediction of $T_C$. We first examine how $T_C$ can be predicted by using the designed descriptors of the compounds. Figure I shows the dependence of the best prediction accuracy on the number of descriptors recruited in the GKR model. The results indicate that it is
possible to accurately predict $T_C$ of rare-earth transition-metal bimetal alloys by using GKR with the designed descriptors. The prediction accuracy reaches a maximum with five to ten descriptors (see Table S1 in the Supplemental Materials [10]). Note that for each fixed number of descriptors recruited in GKR, a couple of different sets of descriptors also attain a high prediction accuracy that is almost comparable to the best one because the descriptors are not independent. These sets are listed in Table S1 in the Supplemental Materials. By using a set of eight descriptors ($C_R$, $C_T$, $Z_R$, $Z_T$, $I$, $I_T$, $S$, and $L$), we can obtain an excellent prediction accuracy (as seen in Fig. 2) with $R^2$ and the mean absolute error (MAE) being approximately 0.96 and 41 K, respectively. It is clear that the prediction accuracy gradually decreases when the number of recruited descriptors increases. This result originates from the fact that the overuse of many weakly relevant descriptors weakens the correlation between the similarity of the compounds, which is measured using the Gaussian kernel of the descriptors and the differences in their values of $T_C$.

Next, we evaluate the relevance of each descriptor for the prediction of $T_C$. We compare $PA(S)$ of the full set of descriptors $S$ ($N_S = 28$) and $PA(S - \{d_i\})$ for of the descriptors $d_i$. We call this a leave-one-out test. We find that most of the descriptors are weakly relevant (see the Supplemental Materials for the definitions of strong and weak relevance and irrelevance [10]), and the prediction accuracy does not significantly change, except in one case for $C_R$. Figure 2 shows some examples of the leave-one-out test when $\{d_i\}$ is either $C_R$, $Z_R$, $C_T$, or $Z_T$. It is clearly seen that the absence of $C_R$ in the GKR model results in a dramatic decrease in the accuracy: $PA(S) > PA(S - \{C_R\})$; therefore, $C_R$ is surely assigned as a strongly relevant descriptor in terms of the prediction of $T_C$. On the other hand, for the other three descriptors $Z_R$, $C_T$, and $Z_T$, only a marginal reduction in the average of the prediction accuracy can be observed, particularly for the number of descriptors ranging from five to ten. We can conclude that the information embedded in each of $Z_R$, $C_T$, and $Z_T$ can be compensated by other descriptors, whereas the information embedded in $C_R$ cannot be compensated by other descriptors. The essentiality of $C_R$ can be confirmed easily in Fig. 3, where the upper limit of $T_C$ linearly depends on $C_R$. Note, however, that the $C_R$ dependence of $T_C$ is even qualitatively different among transition-metal counterparts. For Mn and Co as transition metals, $T_C$ tends to decrease with $C_R$; however, it tends to increase for Fe. For Ni, $T_C$ is rather insensitive to $C_R$. It is important to note that the GKR model can reproduce the situation quite well.

**Understanding**

Since the embedded information in a weakly relevant descriptor can be recovered by other descriptors, it is important to compare the degrees of relevance of descriptors within each category for the prediction of $T_C$ to obtain an insight into the mechanism determining $T_C$. This can be achieved with an add-one-in test, in which we first remove all of the descriptors belonging to one of the three
categories of descriptors and evaluate $R^2$ using only the descriptors of the remaining two categories. Then, we estimate $R^2$ by recovering each of the removed descriptors. This test enables us to know which descriptor is the most relevant in each category. At the same time, we choose the minimum set of descriptors among the removed descriptors to recover nearly the same value of the highest $R^2$. The results of these tests are described below for each of the three categories of descriptor sets in Table III.

1) Structural descriptors ($\mathbf{S}$ descriptors): We first remove all of the descriptors describing the crystal structures ($\mathbf{S}$ descriptors). Then, as the rest of descriptors cannot distinguish different materials properly, $R^2$ becomes 0.0. The results of the add-one-in test in the order of high $R^2$ are presented in the first two columns of Table III for each of the $\mathbf{S}$ descriptors. These $R^2$ values were obtained with two different methods using the descriptors in the other two categories: 1) the $R^2$ values obtained for the optimal choice of the $\mathbf{R}$ and $\mathbf{T}$ descriptors are shown in parentheses in Table III and 2) those obtained by excluding $Z_R$, $\chi_R$ (electronegativity), and $Z_T$ are shown without parentheses. The reasoning behind the second method will be discussed below. $R^2$ only for $C_R$ suggests that the $\mathbf{S}$ descriptors can be well represented solely by $C_R$. This is consistent with the above assignment of $C_R$ as a descriptor with strong relevance. Therefore, in the add-one-in tests for the $\mathbf{R}$ and $\mathbf{T}$ descriptors, we use only $C_R$ for the $\mathbf{S}$ descriptors.

2) Atomic descriptors for rare-earth elements ($\mathbf{R}$ descriptors): The third and fourth columns of Table III present the results of the add-one-in test for the $\mathbf{R}$ descriptors. With only $Z_R$, $R^2$ reaches 0.956 (0.957); any possible combinations of $\mathbf{R}$ descriptors without $Z_R$ can only give an $R^2$ of 0.947 (0.950). Therefore, $Z_R$ can represent the entire set of $\mathbf{R}$ descriptors for prediction.

In the present problem, as $Z_R$ covers only the $4f$ series and $T_C$ mostly varies in systematic ways across the rare-earth-element series in the experimental data, rare-earth elements can be suitably identified with $Z_R$. If $Z_R$ covers more of the $f$ series with different principal quantum numbers, $Z_R$ will not be a suitable descriptor for measuring the similarity among elements. Moreover, according to our domain knowledge, $Z_R$ itself will not directly appear in the physical model for $T_C$. From a causal analysis, $Z_R$ gives information about the electronic configuration based on which magnetic properties are evaluated. Table III also indicates that $\chi_R$ serves as a similarly relevant $\mathbf{R}$ descriptor. Note, however, that this is simply due to its nearly linear relationship with $Z_R$. On the basis of such a consideration, we remove $Z_R$ and $\chi_R$ from the $\mathbf{R}$ descriptors for understanding and look for alternative physical quantities as relevant $\mathbf{R}$ descriptors. This is the reason of why we adopt the second method for choosing descriptors described above. We see from Table III that $J_{4f}(g_{j} - 1)$ gives nearly the same $R^2$ as that for $Z_R$ and that it is more relevant than $S_{4f}$ and $J_{4f}$. Moreover, the last line of the third and fourth columns of Table III suggests that $J_{4f}(g_{j} - 1)$ can only represent the $\mathbf{R}$ descriptors without $Z_R$. In the $4f$ state of the rare-earth element, the spin–orbit interaction is strong, the total angular momentum $J_{4f}$ rather than each of $S_{4f}$ and $L_{4f}$ is conserved, and $J_{4f}(g_{j} - 1)$ is the projection of $S_{4f}$ onto $J_{4f}$. The Curie temperature $T_C$ will be controlled by the effective $4f$–$3d$ exchange interaction through the RE–5d states in which $J_{4f}(g_{j} - 1)$ plays a role. With this physical picture, the above result of the add-one-in test is physically meaningful.

3) Atomic descriptors for transition-metal elements ($\mathbf{T}$ descriptors): The results of the add-one-in test for the $\mathbf{T}$ descriptors are summarized in the last two columns of Table III. Similar to the $\mathbf{R}$ descriptor case, although the atomic number $Z_T$ is the best single choice of the $\mathbf{T}$ descriptors for prediction in the present problem, we search for a more physically meaningful choice of $\mathbf{T}$ descriptors. The results in Table III clearly show that $S_{3d}$ is much more relevant than $J_{3d}$ for $3d$ transition-metal elements. This is again physically meaningful because of the weak spin–orbit interaction and significant $3d$ band width. From Table III we can also see that the combination of $S_{3d}$ and $IP_T$ gives an even higher $R^2$ than that for $Z_T$. This is also physically meaningful because $T_C$ can be controlled not only by $S_{3d}$ but also by the $3d$–$3d$ exchange interaction, which may be affected by $IP_T$ (related to the $3d$ energy level).

Summarizing the above results, we select the following four descriptors compatible with the physical model of $T_C$: $C_R$, $J_{4f}(g_{j} - 1)$, $S_{3d}$, and $IP_T$. This set of descriptors leads to $R^2 = 0.940$, which is sufficiently high.

Conclusion

In this study, we analyze the Curie temperatures of binary alloys consisting of $3d$ transition-metal and $4f$ rare-earth elements by a machine learning technique. Non-
TABLE II: Dependence of the prediction accuracy of the GKR model on the descriptors in the add-one-in test.

| $S$ descriptors | $R$ descriptors | $T$ descriptors |
|-----------------|-----------------|-----------------|
| descriptors in use | $R^2$ | descriptors in use | $R^2$ | descriptors in use | $R^2$ |
| none | 0.0 | none | 0.0 | none | 0.0 |
| $C_R$ | 0.940 (0.957) | $Z_R$ | 0.956 (0.957) | $Z_T$ | 0.945 |
| $C_T$ | 0.909 (0.911) | $\chi_R$ | 0.945 (0.950) | $S_{3d}$ | 0.937 |
| $d_{R-R}$ | 0.876 (0.882) | $J_{4f}(gJ - 1)$ | 0.940 (0.950) | $IP_{R}$ | 0.894 |
| $d_{T-T}$ | 0.621 (0.632) | $S_{4f}$ | 0.919 (0.920) | $\chi_T$ | 0.883 |
| $d_{T-R}$ | 0.518 (0.518) | $g_4$ | 0.909 (0.911) | $J_{3d}$ | 0.702 |
| $C_R, C_T$ | 0.934 (0.960) | $J_{4f}g_4$ | 0.905 (0.907) | $L_{3d}$ | 0.426 |
| none | 0.0 | $IP_R$ | 0.892 (0.894) | $IP_{R}, S_{3d}$ | 0.950 |
| any without $Z_R$ | 0.947 (0.950) | $J_{3d}$ | 0.871 (0.870) | $IP_{R}, S_{3d}, J_{3d}$ | 0.956 |
| none | 0.0 | any $T$ descriptor | 0.957 |

linear regression with a Gaussian kernel accurately reproduces the values $T_C$ of 108 compounds. A regression-based model selection technique is utilized for learning the relation between the descriptors and $T_C$. The prediction accuracy of $T_C$ is maximized when eight descriptors are selected, with the rare-earth concentration being the most relevant. We demonstrate how the regression-based model selection technique can be utilized for mining the scientific connection between the descriptor and the actuation mechanisms of a physical phenomenon.

Acknowledgments

This work was partly supported by PRESTO and by the “Materials Research by Information Integration” Initiative (MI²I) 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.

[1] M. Sagawa, S. Fujimura, N. Togawa, H. Yamamoto, and Y. Matsuura, J. Appl. Phys. 55, 2083 (1984).
[2] Computational Materials Science 50, 841 (2011).
[3] S. Yousef, G. Da, N. Thanh, B. Scotty, C. J. R., and A. Wanda, Phys. Rev. B 85, 104104 (2012).
[4] S. Yang, M. Lach-hab, I. I. Vaisman, and E. Blaisten-Barojas, Phys. Chem. C 113, 21721 (2009).
[5] G. Hautier, C. C. Fischer, A. Jain, T. Mueller, and G. Ceder, Chem. Mater 22, 3762 (2010).
[6] S. J. C., R. Matthias, H. Katja, K.-R. Müller, and B. Kieron, Phys. Rev. Lett 108, 253002 (2012).
[7] O. Isayev, D. Fourches, E. N. Murato, C. Oses, K. Rasch, A. Tropsha, and S. Curtarolo, Chem. Mater 27, 735 (2015).
[8] Pauling File, http://paulingfile.com/.
[9] Atomwork, http://crystdb.nims.go.jp/.
[10] See supplemental material at [URL will be inserted by AIP] for details of the linear representation of atomic energies and statistical analysis of the GMM results.
[11] K. P. Murphy, ed. (MIT Press, 2012).
[12] R. Matthias, International Journal of Quantum Chemistry 115, 1058 (2015).
[13] V. Botu and R. Ramprasad, Int. J. Quant. Chem. 115, 1074 (2014).
[14] G. Pilania, C. Wang, X. Jiang, S. Rajasekaran, and R. Ramprasad, Scientific Reports 3, 2810 EP (2013).
[15] L. Yu and H. Liu, J. Mach. Learn. Res. 5, 1205 (2004).
[16] S. Visalakshi and V. Radha, in 2014 IEEE International Conference on Computational Intelligence and Computing Research (2014), pp. 1–6.
I. SUPPLEMENTAL MATERIALS

A. Data collection and data representation

We collected the experimental data of 108 binary compounds consisting of transition metals and rare-earth metals from the Atomwork database of NIMS [Pauling File, Atomwork], including the crystal structure of the compounds and their observed $T_C$. To represent the structural and physical properties of each binary compound, we use a combination of 28 descriptors. We divide all 28 descriptors into three categories.

The first category pertains to the descriptors describing the atomic properties of the transition-metal constituent, including the (1) atomic number ($Z_T$), (2) atomic radius ($r_T$), (3) covalent radius ($r_{co}^T$), (4) ionization potential ($IP_T$), (5) electronegativity ($\chi_T$), (6) spin angular moment ($S_{3d}$), (7) orbital angular moment ($L_{3d}$), and (8) total angular moment ($J_{3d}$) of the 3d electrons. The selection of these descriptors originates from the physical consideration that the intrinsic electronic and magnetic properties will determine the 3d orbital splitting at transition-metal sites.

In the same manner, we design the second category pertaining to the descriptors for describing the properties of the rare-earth metal constituent, including the (9) atomic number ($Z_R$), (10) atomic radius ($r_R$), (11) covalent radius ($r_{co}^R$), (12) ionization potential ($IP_R$), (13) electronegativity ($\chi_R$), (14) spin angular moment ($S_{4f}$), (15) orbital angular moment ($L_{4f}$), and (16) total angular moment ($J_{4f}$) of the 4f electrons. To capture the effect of the 4f electrons better, we add three additional descriptors for describing the properties of the constituent rare-earth metal ions, including (17) the Landé factor ($g_f$), (18) the projection of the total magnetic moment onto the total angular moment ($J_{4f}g_f$), and (19) the projection of the spin magnetic moment onto the total angular moment ($J_{4f}g_f$) of the 4f electrons. The selection of these features originates from the physical consideration that the magnitude of the magnetic moment will determine $T_C$.

It has been well established that information related to the crystal structure is very valuable in relation to understanding the physics of binary compounds with transition metals and rare-earth metals. Therefore, we design the third category with structural descriptors that roughly represent the structural information at the transition metal and rare-earth metal sites, which are (20) the concentration of the transition metal ($C_T$), (21) the concentration of the rare-earth metal ions ($C_R$), (22) the average distance between a transition-metal site and the nearest transition-metal site ($d_{T-T}$), (23) the average distance between a transition-metal site and the nearest rare-earth-metal site ($d_{T-R}$), (24) the average distance between a rare-earth-metal site and the nearest rare-earth-metal site ($d_{R-R}$), (25) the average number of nearest transition-metal sites surrounding a transition-metal site ($N_{T-T}$), (26) the average number of nearest rare-earth-metal sites surrounding a transition-metal site ($N_{T-R}$), (27) the average number of nearest rare-earth-metal sites surrounding a transition-metal site ($N_{R-R}$), and (28) the average number of nearest transition-metal sites surrounding a rare-earth-metal site ($N_{R-T}$). The values of these descriptors are calculated from the crystal structures of the compounds from the literature.

B. Prediction of the Curie temperature

To learn a function for predicting the values of $T_C$ of compounds from the data represented by using vectors of the descriptors, we apply the Gaussian kernel ridge regression (GKR) technique [11]. The results indicate that it is possible to accurately predict $T_C$ of rare-earth transition-metal bimetal alloys by using GKR with the designed descriptors. The prediction accuracy reaches a maximum with five to ten descriptors (see Table S1 in the Supplemental Materials). Note that for each fixed number of descriptors recruited in the GKR, a couple of different sets of descriptors also attain a high prediction accuracy almost comparable to the best one because the descriptors are not independent. By using a set of eight descriptors ($C_R$, $C_T$, $Z_R$, $Z_T$, $IP_T$, $S_{3d}$, $J_{3d}$, and $L_{3d}$), we can obtain an excellent prediction accuracy (as seen in Fig. 2) with $R^2$ and the mean absolute error (MAE) being approximately 0.96 and 41 K, respectively.

C. Strong relevance and weak relevance

On the basis of Eq. [14], we can evaluate the relevance of a descriptor for the prediction of $T_C$ by using the expected reduction in the prediction ability caused by removing this descriptor from the full set of descriptors. Let $D$ be a full set of descriptors, $d_i$ a descriptor, and $D_i = D \setminus \{d_i\}$ the full set of descriptors after removing the descriptor $d_i$. The degree of relevance of the descriptors can be formalized as follows:

1. Strong relevance: a descriptor is strongly relevant
if and only if
\[ PA(D) - PA(D_i) = \max_{s \in D} R_s^2 - \max_{s \in D_i} R_s^2 > 0. \]  
(5)

Among the strongly relevant descriptors, a descriptor that causes a larger reduction in the prediction ability when it is removed can be considered as a strong one. The degree of relevance of a strongly relevant descriptor can be computationally estimated by using the leave-one-out approach, i.e., by leaving out a descriptor in the currently considered descriptor set for the GKR analysis and testing how much the prediction accuracy is impaired.

2. Weak relevance: a descriptor is weakly relevant if and only if
\[ PA(D) - PA(D_i) = \max_{s \in D} R_s^2 - \max_{s \in D_i} R_s^2 = 0 \text{ and } \exists D_i' \subset D_i \text{ such that } PA\{\{d_i, D_i'\}\} - PA(D_i') > 0. \]  
(6)

It is clearly seen from Eq. (6) that estimation of the degree of relevance for the weakly relevant descriptors cannot be carried out in a straightforward manner as for the case of the strongly relevant descriptors. Weakly relevant descriptors are descriptors that are relevant for prediction, but they can be substituted by other descriptors. We can only estimate the degree of relevance for this type of descriptor in specified contexts. For example, in terms of the prediction of \( T_C \), the relevance of a descriptor for an atomic property of transition metal can be examined in the context that all of the descriptors for the atomic properties of rare-earth metals are included in the descriptor set. We define the following additional rule for comparing two weakly relevant descriptors:

3. Comparison between weakly relevant descriptors: A weakly relevant descriptor \( d_i \) is said to be more relevant than the descriptor \( d_j \) in the context of having a set of descriptors \( M(d_i, d_j \notin M) \) if and only if
\[ PA\{\{d_i, M\}\} > PA\{\{d_j, M\}\}. \]  
(7)

A comparison of two weakly relevant descriptors can be computationally carried out by using the add-one-in approach, i.e., by exclusively adding the two descriptors to the currently considered descriptor set for the GKR analysis and testing how much the prediction accuracy is improved.

D. Descriptors associated with transition-metal properties

The evaluation of weakly relevant \( T \) descriptors measured by the improvement in the prediction ability (\( \Delta R^2 \)) is summarized in Table II. The largest improvement in the prediction accuracy of \( T_C \) can be obtained by adding the \( S_{3d} \) descriptor. The addition of the \( J_{3d} \) descriptor can also yield a large improvement in the prediction accuracy of \( T_C \), but the addition of the \( L_{3d} \) descriptor yields a far lower improvement in the prediction accuracy. These results are consistent with the understanding so far that the values of \( T_C \) of binary alloys consisting of 3d transition-metal (T) and 4f rare-earth elements (R) are mainly determined by the magnetic interaction in the transition-metal sublattice. Indeed, in R–T compounds, there are three types of interactions including the magnetic interaction between T atoms in the T sublattices (T–T interaction), the magnetic interaction between R atoms and the T sublattices (R–T interaction), and the magnetic interaction between R atoms in the R sublattices (R–R interaction). The R–R interaction is very weak in comparison with the T–T and R–T interactions because the 4f electrons reside far into the interior of the rare-earth atoms and the spatial extent of the 4f electron wave function is rather small compared with the lattice separation. The R–T interaction is also weak in comparison to the T–T interaction; however, the R–T interaction plays an important role in determining the magnetic structure of R–T compounds. The T–T interaction dominates in R–T compounds because the delocalization and spatial extent of the 3d electron wave functions of T atoms are much more pronounced than those of 4f electrons. The understanding so far regarding the dominance of the magnetic interaction in the transition-metal sublattice is also consistent with our analysis result that the improvement in the prediction ability for the descriptors describing the magnetic properties of the constituent transition metals is obviously much larger than that for the descriptors describing the magnetic properties of the constituent rare-earth metals.