M6

Data mining study of physical properties of transition-metal rare-earth compounds: prediction and understanding

HC. Dam[1,2], VC. Nguyen[3], TL. Pham[1,2], H. Kino[2], K. Terakura[2], and T. Miyake[2,4]
[1]Japan Advanced Institute of Science and Technology, Ishikawa 923-1292, Japan
[2]National Institute for Materials Science, Ibaraki 305-0047, Japan
[3]HPC Systems Inc., Japan
[4]National Institute of Advanced Industrial Science and Technology, Ibaraki 305-8568, Japan

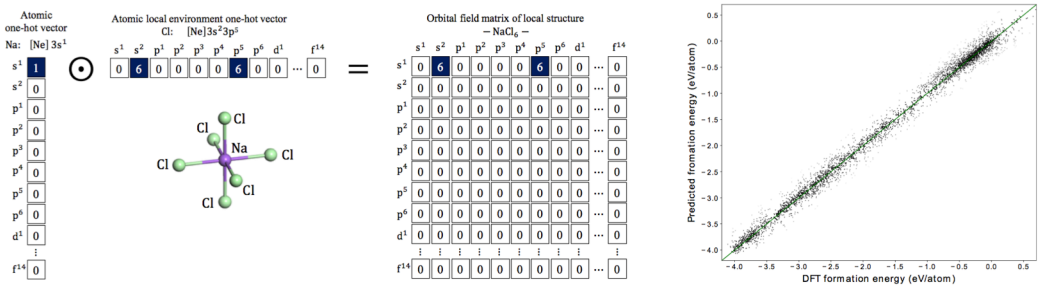
Recently, the increasing volume of available experimental and quantum-computational material data, along with the development of machine-learning techniques, has provided a new opportunity to develop methods for accelerating discoveries of new materials and physical and chemical phenomena. By using machine-learning algorithms, hidden information on materials, including patterns, features, chemical rules, and physical laws, can be automatically discovered from both first-principles-calculated data and experimental data [1-8]. To render data-driven approaches meaningful and useful for materials science studies, it is necessary to design material representations with which the physical properties can be predicted with high accuracy and the results derived using machine-learning methods can be interpreted in the language of physical chemistry.

Information on the structure of a material is usually described using a set of atoms with their coordinates and periodic unit cell vectors, which are required for crystalline systems. The material data using this primitive representation can be categorized as unstructured data, and the mathematical operations performed on such material data involve the algebra of sets only. Therefore, advanced quantitative machine-learning algorithms cannot be applied directly and effectively to conventional material data, owing to the limitation of the algebraic operations of the primitive data representation. In order to apply well-established machine-learning methods, including predictive learning and descriptive learning, it is necessary to convert the primitive representation into fixed-dimensional vectors or matrices, such that the comparison and calculations using the new representation reflect the nature of the materials and the actuating mechanisms of the chemical and physical phenomena. In response to this request, various methods for encoding materials have been developed in the field of materials informatics [9, 10].

In this paper, we introduce our effort in designing a material descriptor, with emphasis on the interpretability of the derived learning results, that (1) utilizes information on the local structure, (2) incorporates the valence atomic configuration, and (3) accepts algebraic operations to construct global descriptors from local descriptors. We utilize the intuition from the ligand field and crystal field theories, in classifying or categorizing local atomic environments by using the number of valence orbitals (electrons) coordinating the valence orbital of the center atom, to implement the above-mentioned concept of developing a novel representation (Fig 1a). We name this type of descriptor the “orbital field matrix (OFM)."

To verify the applicability of the proposed material representation, we focus on magnetic materials based on rare earth – transition metal (RT) alloys and RT alloys including a light element X, which may be B, C, N, or O (RTX). We first examine the

decision trees for predicting the magnetic moments of Mn, Fe, Co, and Ni in RT alloys. The decision trees learned from the RT alloy data show that the coordination numbers of the occupied *d* orbitals of the transition metals and the occupied *f* orbitals of the rare-earth metals play important roles in determining the local magnetic moments of the transition metal sites. In addition, kernel ridge regression (KRR) analyses using standard techniques and similarity measures are implemented in learning prediction models to quantitatively predict the local magnetic moments of transition metal sites in RT alloys, formation energies for RTX materials, and atomization energies for organic molecules. Our computational experiments show that the proposed material representation can accurately reproduce the DFT-calculated local magnetic moments of transition metal sites in RT alloys, formation energies of crystalline systems, and atomization energies of molecular systems (Fig. 1b).



(**a**) (**b**)

Fig. 1 (**a**) OFM representation for an Na atom in a regular octahedral site surrounded by 6 Cl atoms: atomic one-hot vector for Na (middle), representation for the 6 Cl atoms surrounding the Na atom (left), and representation for the Na atom surrounded by 6 Cl atoms (right). (**b**) Comparison of formation energies calculated using DFT and those predicted through machine learning (ML-predicted), using OFM.

References

[1] S. Yousef, G. Da, N. Thanh, et al., Phys. Rev. B 85, 104104 (2012).
[2] S. Yang, M. Lach-hab, et al., Phys. Chem. C 113, 21721 (2009).
[3] G. Hautier G, CC. Fischer, A. Jain, et al., Chem. Mater. 22, 3761 (2010).
[4] R. Matthias, H. Katja, et al., Phys. Rev. Lett. 108, 253002 (2012).
[5] Isayev, D. Fourches, EN. Muratov, et al., Chem Mater. 27, 735 (2015).
[6] LM. Ghiringhelli, J. Vybiral, et al., Phys. Rev. Lett. 114, 105503 (2015).
[7] HC. Dam, TL. Pham, TB. Ho, et al., J. Chem. Phys. 140, 044101 (2014)..
[8] TL. Pham, H. Kino, K. Terakura, et al., J. Chem. Phys. 145, 154103 (2016).
[9] J. Behler, M. Parrinello, Phys. Rev. Lett. 98, 146401 (2007).
[10] M. Rupp, A. Tkatchenko, KR. Muller, et al., Phys. Rev. Lett. 108, 058301 (2012).