Development of Data-Driven Methods for Materials Discoveries: Representation, Learning, and Uncertainty Modeling and Management

ID: INV02
Invited Talk
Aug. 23, 11.20 - 11.50

DAM Hieu-Chi,[1,2] NGUYEN Duong-Nguyen,[1] HA Minh-Quyet,[1] VU Tien-Sinh,[1] NGUYEN Viet-Cuong,[3] KINO Hiori,[4] and MIYAKE Takashi[5]

[1] *Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan*

[2] *Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan*

[3] *HPC SYSTEMS Inc., 3-9-15 Kaigan, Minato, Tokyo 108-0022, Japan*

[4] *National Institute for Materials Science, 1-2-1 Sengen, Tsukuba 305-0044, Japan*

[5] *National Institute of Advanced Industrial Science and Technology, 1-1-1 Umezono, Tsukuba 305-8568, Japan*

Email: dam@jaist.ac.jp

Keywords: Machine Learning, Data Mining, Evidence Theory

In this presentation, we explain our efforts in developing data-driven methods for materials studies, utilizing machine learning (ML) and data mining to enhance materials research and design. Our focus is on methods that deepen the understanding of mechanisms governing material properties in both molecular and crystalline forms, derived from computational and experimental data.

Central to our research is the development of advanced ML and data mining models that enhance comprehension of structure-property relationships and predictive capabilities for materials properties. Our work starts with designing descriptors that incorporate structural information, electronic structures, and conventional physicochemical knowledge, facilitating property predictions through ML techniques [1]. Our recent use of interpretable deep learning with attention mechanism has advanced the understanding of structure-property relations in materials, contributing to more transparent and explainable models in materials science. These models not only predict material properties but also elucidate underlying factors, revealing key structure-property relationships [2].

We also focus on managing and quantifying uncertainty in inference and prediction to ensure robust outcomes. Using the Dempster-Shafer theory, we convert material data into evidence on similarities, appropriately modeling and managing uncertainty for data-driven inferences using unbalanced, contradictory, and biased materials data. We have developed evidence-based methods to evaluate material similarity and compositions, providing deeper insights and improving predictive accuracy. This approach has shown superior performance in identifying promising material compositions, validated through experimental studies [3].

The synergy of data-driven and conventional computational approaches in our studies accelerates materials discovery, emphasizing interpretable models and the explainability of results. This comprehensive approach demonstrates the transformative potential of data-driven methods in understanding and discovering novel materials.

References

1. T. L. Pham, H. Kino, K. Terakura, T. Miyake, K. Tsuda, I. Takigawa, and H. C. Dam, Sci. Technol. Adv. **18**, 756-765 (2017).
2. T.-S. Vu, M.-Q. Ha, D.-N. Nguyen, V.-C. Nguyen, Y. Abe, T. Tran, H. Tran, H. Kino, T. Miyake, K. Tsuda, and H.-C. Dam, npj Comput. Mater. **9**, 215 (2023).
3. M.-Q. Ha, D.-N. Nguyen, V.-C. Nguyen, T. Nagata, T. Chikyow, H. Kino, T. Miyake, T. Denœux, V.-N. Huynh, and H.-C. Dam, Nat. Comput. Sci. **1**, 470-478 (2021).