From materials data to evidence for understanding materials properties

(¹*School of Knowledge Science, Japan Advanced Institute of Science and Technology*, ²*International Center for Synchrotron Radiation Innovation Smart, Tohoku University*) ◯Hieu-Chi DAM[1,2]

Keywords: *Materials informatics, Data-driven approach, Machine learning, Data mining, Evidence theory*

Generally, materials science studies use data analysis to study already synthesized materials or calculated hypothetical materials to generate ideas for the next experiment or design the material they want to develop. The results of material data analysis usually include similarities between materials and correlations between the structure or composition of materials and their physical properties. However, it is not easy to manually and properly evaluate similarities between materials and correlations between multiple factors to gain insights and reveal their physicochemical mechanisms.

With the recent development of data-driven AI, it is expected that data-driven approaches will accelerate materials science research. However, applying the developed data-driven AI techniques to materials science is not easy, even if they have been successful in other fields. The problem is that the effectiveness of a data-driven approach depends on the compatibility of the applied data-driven algorithms with the materials science problem to be solved and with the representation of the collected materials data.

Another problem is that material data is still scarce and usually includes results that are difficult to interpret consistently. In addition, the reported materials data is often biased toward successful examples. Therefore, although learning similarities between materials and extracting knowledge from data is intuitive and straightforward and is the preferred application of data-driven approaches in materials science, still, it is not easy to properly evaluate similarities and dissimilarities between materials even when using data-driven techniques to analyze materials data.

This talk will introduce our efforts to develop an evidence-based data-driven approach[1,2] that can overcome the difficulties currently encountered in applying data-driven to materials science to handle multiple materials data, including insufficient, inconsistent, biased data, in a unified manner.

1) N. N. T. Ton, M. Q. Ha, T. Ikenaga, A. Thakur, H. C. Dam, T. Taniike, *2D Mat.* **2020**, *8*, 015019.
2) M. Q. Ha, N. D. Nguyen, V. C. Nguyen, T. Nagata, T. Chikyow, H. Kino, T. Miyake, T. Deneoux, V. N. Huynh, H. C. Dam, *Nat. Comp. Sci.* **2021**, *1*, 470–478.
.