

# Machine learning for materials design and discovery: how do we learn from little?

<sup>1</sup>\*Ryo Yoshida

<sup>1</sup> *The Institute of Statistical Mathematics, Research Organization of Information and Systems,*  
10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan

\* yoshidar@ism.ac.jp

The ability of machine learning (ML) models trained on massive amounts of data has reached or even outperformed humans in intellectually demanding tasks across various fields. As such, ML has received much attention as a key driver to the next frontier of materials science, which can contribute to substantial savings, in terms of both time and costs in the development of new materials. In this talk, I will describe some key technologies of ML to achieve this goal.

The first topic is ML for computational materials design. In general, the material spaces are considerably huge. For instance, the chemical space of small organic molecules is known to contain as many as  $10^{60}$  candidates. In the past several years, we have developed a series of inverse material design algorithms based on state-of-the-art ML technologies coupled with quantum chemistry calculation. The objective is to computationally create promising materials that exhibit desired properties. It is expected that the emergence of such ML technologies will accelerate the pace of expanding the frontier in the vast universe of various materials.

The second topic is a subject of data scarcity. In recent years, a broad array of materials property databases has begun to emerge towards digital transformation of materials science. However, volume and diversity of the current database remain far from fully enjoying recent advances made in ML. A ML framework called transfer learning has the great potential to break the barrier of limited amounts of materials data in which various kinds of properties are physically interrelated. For a given target property to be predicted from a limited supply of training data, models on related proxy properties are pre-trained using an enough amount of data, which capture common features relevant to the target task. Repurposing such machine-acquired features on the target task brings an outstanding achievement in the prediction performance even with exceedingly small data as if highly experienced human experts can perform rational inferences even on considerably less experienced tasks. To facilitate the widespread use of transfer learning, we have developed a pre-trained model library, called XenonPy.MDL. In this first release, the library constitutes more than 100,000 pre-trained models for various properties of small molecules, polymers, and inorganic solid-state materials. Along with this library, I will demonstrate outstanding successes of transfer learning.

## References

1. Ikebata et al. (2017). Bayesian molecular design with a chemical language model, *Journal of Computer-Aided Molecular Design*, 31(4), 379–391.
2. Wu et al. (2019). Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm.
3. Yamada et al. (2019). Predicting materials properties with little data using shotgun transfer learning.
4. XenonPy: <http://xenonpy.readthedocs.io/en/latest>
5. iQSPR: <https://github.com/yoshida-lab/iqspr>