Efficient materials exploration based on systematic density-functional calculations and machine learning techniques

Isao Tanaka[1,2] , Atsuto Seko[1], Yukinori Koyama[1] and Atsushi Togo[1]

1 Department of Materials Science and Engineering, Kyoto University, Kyoto, 606-8501, Japan
2 Nanostructures Research Laboratory, Japan Fine Ceramics Center, Nagoya, 456-8587, Japan

Recently, challenges for accelerated discovery of materials with the aid of information technology have been demonstrated. One of the approaches uses high-throughput screening of materials database that is generated by first principles density functional theory (DFT) calculations. Thanks to recent progress of computational power and technique, a large number of DFT calculations can be made with the accuracy comparable to experiments, which can be used for the efficient materials exploration. Another approach is based upon state-of-the-art machine-learning algorithm to search the optimum. A combined approach of them should also be useful.

Two examples are shown in my talk. Firstly, we apply a combination of DFT calculations and regression techniques to the estimation of an approximated function describing experimental melting temperatures for single and binary component solids[1]. We adopt four kinds of regression techniques for training melting temperatures of binary compounds, i.e. ordinary least-squares regression (OLSR), partial least-squares regression (PLSR), support vector regression (SVR) and Gaussian process regression (GPR). Results by four regression methods are compared. The GPR is then used for efficient searching of the compound with the highest melting temperature called kriging.

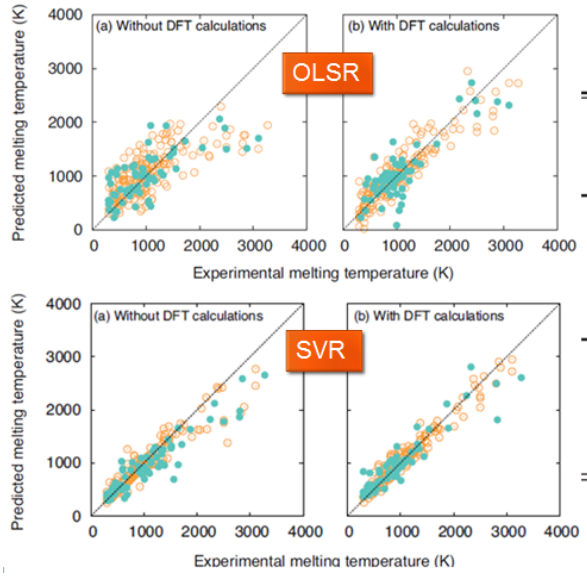
Secondly, Li-ion conductivities in lithium superionic conductors (LISICONs) are examined [2]. A systematic set of DFT calculations based on the cluster expansion method, as well as first-principles molecular dynamics (FPMD) simulations are carried out for a diverse range of chemical compositions. A machine-learning technique is used to combine theoretical and experimental datasets to predict the conductivity of each composition at 373 K. The insights in the present study imply that an iterative combination of first-principles calculations and focused experiments can greatly accelerate materials-design process through efficient screening of a wide compositional and structural phase space.

This study was supported by the Grant-in-Aid for Scientific Research on Innovative Areas "Nano Informatics" (grant number 25106005) from JSPS.

References

[1] A. Seko, T. Maekawa, K. Tsuda and I. Tanaka, *Phys. Rev*. B 89 054303 (2014).

[2] K. Fujimura, A. Seko, I. Tanaka et al., *Adv. Energy. Mater.* 3 980 (2013).



	CV score (K)	RMS error for test data (K)
(a) Without DFT calculations		
OLSR	473	472
SVR	376	364
(b) With DFT calculations		
OLSR	293	306
SVR	265	262

FIG.1. Melting temperature of 248 compounds predicted by the OLSR performed with (a) a predictor set without DFT data and (b) that with DFT data. Training and test data are shown by open and closed circles, respectively. CV scores and RMS errors for test data are shown together. [1]

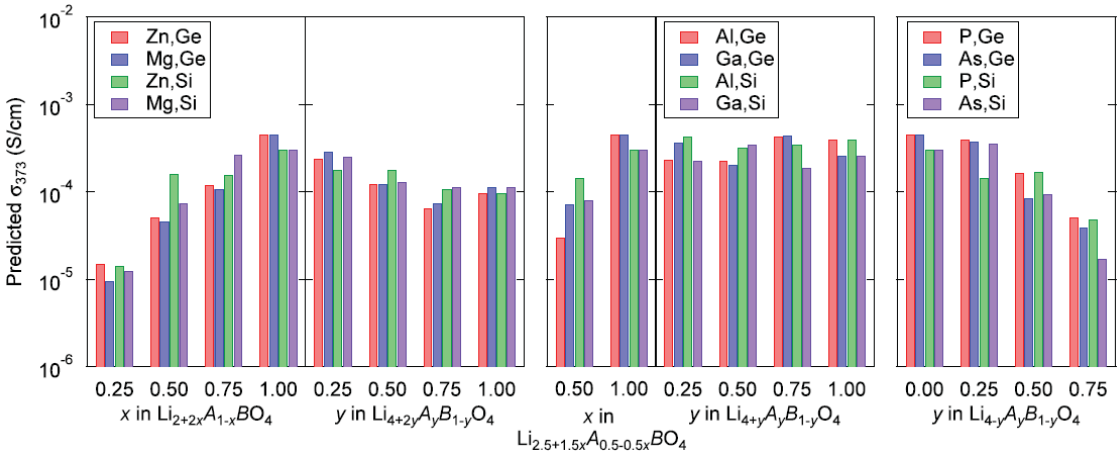


FIG.2. Predicted ionic conductivities at 373K for 72 compositions in the LISICON system. [2]