Data assimilation method for materials exploration based on Bayesian optimization

Yosuke Harashima, Nara Institute of Science and Technology

Data assimilation is a concept to construct a prediction model by integrating experimental data and simulated data. This is known to be useful in the weather forecast, for example, and is expected to be the same in the materials science. The data obtained from an experiment require high cost and the number of data tends to be less, while simulation data can be obtained more easily than experiments, however, the simulation data include some approximations and assumptions, and thus, the prediction accuracy is limited. The data assimilation provides a synergistic effect between experiment and simulation, and it enhances efficiency of material exploration. In this study, a new data assimilation method is developed for a multi-dimensional material feature space. This assimilation model can be trained using missing data, either experimental or simulated data. Furthermore, this method can be combined with Bayesian optimization to improve the efficiency of materials search. This data assimilation and Bayesian optimization is demonstrated to find optimal values in a multidimensional feature space and compared with simple Bayesian optimization without assimilation. The result indicates that the performance improves rapidly once a prediction model for simulated data is established.