Application of High-Throughput Framework and Machine Learning on the Prediction of Material Properties

Wen-Jay Lee* [1,2], Chia-Yung Jui[1]
[1]*National Center for High-performance Computing, No.22, Keyuan Rd., Central Taiwan Science Park, Taichung City, Taiwan (R.O.C.)*
[2]*Department of Physics, National Chung Hsing University, 145 Xingda Rd., South Dist., Taichung City, Taiwan(R.O.C.)*

Material genome initiative (MGI) aims to develop materials innovation infrastructure for accelerating the finding of new and emerging materials, and further realize its application in product. It emphasizes on the integration of tools, theories, models, and data from basic scientific research with the processing, manufacturing, and deployment of materials. The infrastructure enable this integration by providing access to digital resources that contain the property data of known materials as well as the computational and experimental tools to predict these characteristics. Until today, several platforms for database and high throughput engine have been proposed. AFlow(organic & inorganic material)[1], Material Project (organic & inorganic material)[2], Khazana (polymer genome)[3], and NanoMine (polymer nanocomposites)[4] provides the computational results of material properties by density function theory and machine learning. Pymatgen[5] and AiiDA[6] are robust, open-source Python library for materials analysis using VASP and Quantum Expesso, respectively, which provide the simulation workflow and high-throughput engine. Recently, we have developed a transferable and extendable framework of high-throughput and dependency (HT&D) flow for material simulations. Compared to the well-known platform (Pymatgen and AiiDA), our framework is easy and friendly to implement a series of programs together (such VASP, LAMMPS, GAUSSIAN, Material Studio, and homemade programs), so that one can create personal working flow and high-throughput simulation including preprocess, kernel simulation, and post-process programs to realize the virtual screening purpose. In addition, we have employed machine learning (ML) with database of Material Project to predict the material properties of semiconductor. Both HT&D flow and ML will play important roles for accelerating the material screening and finding in the future.

References:
[1] http://aflowlib.org/
[2] https://www.materialsproject.org/
[3] http://khazana.uconn.edu/index.php?m=1
[4] http://nanomine.northwestern.edu:8000/
[5] http://pymatgen.org/
[6] http://www.aiida.net/docs/

Email: wjlee@nchc.narl.org.tw