Machine Learning Algorithms Review: from the Chemical Compound Space to the Hilbert Space

Bastien Casier
Université de Lorraine and CNRS, LPCT, UMR 7019, F-54000 Nancy, France

The past decade has seen an impressive growth in the development and application of machine-learning techniques in quantum chemistry and computational condensed matter physics [1]. In fact, these new paradigms of studies serve as powerful tool to analyze, classify or predict various properties requiring a large amount of data generated via computationally demanding quantum mechanical calculation. As an instance, machine-learning approaches can be applied to broad range of problems, including, among the other, the potential energy surface fitting, the ab-initio molecular dynamics or the prediction of scalar, vectorial and tensorial quantities (e.g. atomization energies, electronic structure correlation energies, forces, polarizability tensors, etc).

The main idea of machine learning methods in quantum chemistry and material sciences is to train a machine learning model (neural-networks and kernels) on a subset of chemical structures for which simulations at a quantum mechanical level were done. The trained machine is then used to predict the target properties of the rest of the systems. This technique can be applied in the chemical compound space but also in the conformation space. However, the efficiency of these methods strongly depends of the way to represent the molecular or solid-state information. In this aim, a lot of descriptors have been developed like the Coulomb matrices [2] or the smooth overlap of atomic positions [3].

During this seminar, I will illustrate all these points by regression of a potential energy surface associated to proton-transfer reaction of the acetone molecule [4] and by description of the chemical compound space via a new graph kernel representation (Figure) [5]. Finally, I will present the new neural-network quantum state theory [6] where the goal is to describe the quantum Hilbert space. An example will be given with the selection of configuration interaction (CI) determinants.

References:

[1] Noé, F. *et al.*, *Annu. Rev. Phys. Chem*. 71, 361 (**2020**).
[2] Rupp, M. *et al.*, *Phys. Rev. Lett*. 108, 058301 (**2012**).
[3] De, S. *et al.*, *Phys. Chem. Chem. Phys*. 18, 13754 (**2016**).
[2] Casier, B. *et al*., *J. Chem. Phys.*, 152, 234103 (**2021**).
[4] Casier, B. *et al*., *J. Comput. Chem.*, 1-12 (**2021**).
[5] Choo, K. *et al*., *Nat. Commun.* 11:2368 (**2020**).

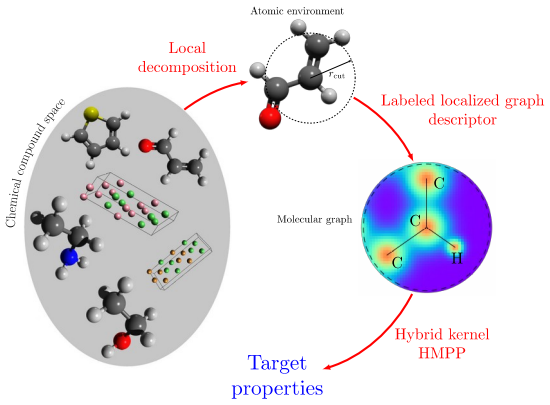

Figure: Graph kernel machine learning method.