Hybrid localized graph kernel for machine learning energy-related properties of molecules and solids

Bastien Casier[1*], Mauricio Chagas Silva[1], Michael Badawi[1], Fabien Pascale[1], Tomáš Bučko[2,3], Sébastien Lebègue[1] and Dario Rocca[1]

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

2 Department of Physical and Theoretical Chemistry, Faculty of Natural Sciences, Comenius University in Bratislava, Mlynská Dolina, Ilkovičova 6, SK-84215 Bratislava, Slovakia

3 Institute of Inorganic Chemistry, Slovak Academy of Sciences, Dúbravská cesta 9, SK-84236 Bratislava, Slovakia

*bastien.casier@hotmail.fr

ABSTRACT

Nowadays, the coupling of electronic structure and machine learning techniques serves as a powerful tool to predict chemical and physical properties of a broad range of systems [1-3]. With the aim of improving the accuracy of predictions, a large number of representations for molecules and solids for machine learning applications has been developed [4-5]. In this work we propose a novel descriptor based on the notion of molecular graph. While graphs are largely employed in classification problems in cheminformatics or bioinformatics [6], they are not often used in regression problem, especially of energy-related properties. Our method is based on a local decomposition of atomic environments and on the hybridization of two kernel functions: a graph kernel contribution that describes the chemical pattern and a Coulomb label contribution that encodes finer details of the local geometry. The accuracy of this new kernel method in energy predictions of molecular and condensed phase systems is demonstrated by considering the popular QM7 and BA10 datasets [5,7]. These examples show that the hybrid localized graph kernel outperforms traditional approaches such as, for example, the smooth overlap of atomic positions (SOAP) [4] and the Coulomb matrices [5].

References

[1] Noé, F. *et al.*, *Annu. Rev. Phys. Chem.* **2020**, 71, 361

[2] Dral, P. O., *Phys. Chem. Lett.* **2020**, 11, 2336

[3] Casier, B. *et al.*, *J. Chem. Phys.* **2020**, 152, 234103.

[4] De, S. *et al.*, *Phys. Chem. Chem. Phys.* **2016**, 18, 13754.

[5] Rupp, M. *et al.*, *Phys. Rev. Lett.* **2012**, 108, 058301.

[6] Lavecchia, A., *Drug Discov. Today* **2015**, 20, 318.

[7] Nyshadham, C. *et al.*, *Npj Comput. Mater* **2019**, 5.