Theoretical and Physical Chemistry Institute

National Hellenic Research Foundation

Vass. Constantinou 48, Athens

LECTURE

"Coupling of Machine Learning with Electronic Structure Theory for Chemical Applications"

Dr. Konstantinos D. Vogiatzis

Department of Chemistry, University of Tennessee,

Knoxville, Tennessee, United States

Thursday, June 6, 2019, 12:00

Seminar room, ground floor, NHRF

Coupling of Machine Learning with Electronic Structure Theory for Chemical Applications

Konstantinos D. Vogiatzis

Department of Chemistry, University of Tennessee, Knoxville, Tennessee 37996, United States

Our recent efforts on the development of new computational methods that couple electronic structure theory with machine learning which target *an artificial intelligence environment which can help humans solve problems, associated with the fundamental chemical questions of the fourth industrial revolution* [1] will be discussed.

First, we have developed a novel molecular fingerprinting method based on persistent homology that can encode the geometrical and electronic structure of molecules for chemical applications. We have demonstrated its applicability on studies on non-covalent interactions between functional groups of materials and small gas molecules for environmental applications. [2] Quantum chemical calculations were performed on a small number of molecules (100) for the generation of meaningful data. We have used these data in order to train a statistical model that includes the new fingerprinting method and machine learning algorithms. The trained models have been used for high-throughput virtual screening by predicting the properties of larger molecular databases (more than 133,000 entries) where quantum chemical data are not available.

Second, a novel approach based on machine-learning algorithms and data that accelerates the convergence of the coupled-cluster (CC) methods, and it is demonstrated for the singles-and-doubles (CCSD) level. [3] For solving the CCSD equations, an initial guess of the t_2 amplitudes is needed, and usually the MP2 amplitudes (t_{MP2}) are chosen as the starting point. As an alternative, we suggest a methodology that "learns" the relation between the t_{MP2} and t_2 amplitudes of CCSD and provides a more accurate guess for the CCSD iterative solver. Features are collected from quantum chemical data, such as orbital energies, one-electron Hamiltonian, Coulomb and exchange terms. The data-driven CCSD (DDCCSD) is not an alchemical method since the actual iterative coupled-cluster equations are solved. Our preliminary data show a remarkable speed-up for the convergence of the CCSD equations. We are currently applying our *Data-driven acceleration of iterative eigensolvers* algorithm on other electronic structure theory methods.

If time permits, the massively parallel implementation of the complete active space self-consistent field (CASSCF) method will be discussed. Our implementation allowed record calculations with large active spaces beyond the limits of conventional CASSCF. [4]

[1] A. Aspuru-Guzik, R. Lindh, M. Reiher *ACS Cent. Sci.* **2018**, *4*, 144.
[2] J. Townsend, C. M. Micucci, J. Hymel, V. Maroulas, K. D. Vogiatzis *In Preparation*.
[3] J. Townsend, K. D. Vogiatzis *Submitted*.
[4] K. D. Vogiatzis, D. Ma, J. Olsen, L. Gagliardi, W. de Jong *J. Chem. Phys.* **2017**, *147*, 184111.