OPTICAL PROPERTIES FOR ALL SYNTHESIZABLE MOLECULES FROM QUANTUM CHEMISTRY-BASED MACHINE LEARNING

CHENXI SUN, *Department of Chemistry, University of Massachusetts- Amherst, Amherst , MA, USA*; YILI SHEN, CHENGWEI JU, ZHOU LIN, *Department of Chemistry, University of Massachusetts, Amherst, MA, USA*.

Luminescent organic molecules have been widely applied in optoelectronics and biological research. First-principles time-dependent density functional theory (TDDFT) and machine learning methods have demonstrated great success in the predictions of optical properties of large organic molecules. However, the systematic error, large time cost, and narrow range of predictable properties of TDDFT hinder its applications in high-throughput screening of real-life systems. On the other side, statistics-based methodologies have the advantages of high accuracy and low costs. While the generalizability of the models and synthesizability of the molecules still pose challenges.

Herein, we developed a ML model that implemented semi-empirical quantum chemical properties to accurately predict the absorption frequencies, emission frequencies, and photoluminescence quantum yield (PLQY) of organic molecules. Based on the evaluation on chromophore families and chromophore-solvent pairs, we illustrated that our extension of the semi-empirical quantum chemical properties remarkably improved the accuracy and generalizability of the model with only a margin increase of computational costs. Meanwhile, tree-based algorithms outperformed neural networks and managed to reach mean absolute errors (MAEs) as low as 0.061 eV for absorption frequencies, 0.065 eV for emission frequencies, and 0.10 for PLQY. Tested on another database containing 96 million compounds with semi-empirical calculations, our model exhibited great success in the predictions of optical properties for all synthesizable molecules at very low computational costs, and thus substantially promote the discovery of potential optical materials at a large scale.