Contribution ID: **11** Type: **not specified**

Prediction and Taking Insight of Molecular Quantum Property by Random Forest Regression

Friday, 8 November 2019 15:55 (15 minutes)

Fluorescent molecules are widely used for bio-imaging. They are attached to specific cell organelles or proteins, enabling observation of detailed structure and dynamics in the cell. Efficient fluorescent molecules must have a high quantum yield for effective bio-imaging. Diverse effective fluorescent molecules whose color are distinctive need to get more information of cell or protein. First step of discover novel molecules using computational approach is prediction of compound. Here, we use random forest regression to predict excitation energy and oscillator strength of a molecule. Random forest algorithm is white box. It is easy to extract feature importance. We could get insight from it.
A statistical machine that predicts excitation energies and associated oscillator strengths, the probability of absorption or emission of light in transitions between different energy states, of a molecule were trained using the random forest algorithm. The PubchemQC database was used as a training set. It has over 3 million known compounds. We picked up 0.5 million molecules randomly. 90% of them were in training set and the others were test set. The ECFP4(extended connectivity fingerprints 2) of molecules were used as input features. We found some fragments which can decide molecules' quantum property by feature importance analysis.

Primary author: Mr KANG, Beomchang (Seoul National University)

Co-authors: Prof. SEOK, Chaok (Seoul National University); Prof. LEE, Juyong (Kangwon National University)

Presenter: Mr KANG, Beomchang (Seoul National University)

Session Classification: Contributed talks 2