



Contribution ID: 77

Type: **Poster**

Machine learning methods research on predicting protein profile parameters by synchrotron radiation small-angle X-ray scattering

Tuesday 9 April 2024 18:30 (20 minutes)

Synchrotron radiation light sources are widely used in various scientific fields due to their high performance and high energy. In the field of biology, protein function mechanism research is promoted by decoding protein structures and studying the correlation with their functions. Then the function mechanism advances scientific development in many fields, such as drug and new material research and development. The traditional method of X-ray diffraction experiments on crystalline protein samples makes it difficult to visually characterize the true structure of proteins in physiological states. The synchrotron radiation small-angle X-ray scattering experiment is convenient to perform for protein samples in the solution state. By analyzing the measured scattering intensity data, some low-resolution structural information of proteins can be obtained. Traditional software analysis methods usually fit a profile model which takes minutes to hours. We develop a machine learning analysis method to enable fast and accurate prediction of protein profile parameters, including molecular weight, maximum diameter, and radius of gyration, from scattering intensity data. We first create a labeled dataset of protein profile parameters and scattering intensities, including simulation data, experimental data from the literature, and data collected from small-angle X-ray scattering experiments. The dataset is preprocessed and then divided into three categories: training, validation and test. Next, we build the training framework for the supervised machine learning model and design the objective loss function between predicted and labeled profile parameters. Finally, the prediction model is trained and optimized by iteratively minimizing the value of the objective loss function. Thus, it is realized to directly and efficiently predict protein profile parameters from bulk scattering data. The prediction model can achieve second responses and improve the efficiency by a hundred to a thousand times, thus enabling real-time online high-throughput data parsing. The real-time feedback on the dynamics of protein profiles can be monitored during experimental processes with this prediction model. More in-depth research on protein function mechanisms can be promoted and provide the potential to extend scientific applications.

Author: LI, Qingmeng

Presenter: LI, Qingmeng

Session Classification: Posters

Track Classification: MLC