

Quantitative Prediction of Properties of Organic Molecules from ELNES via Artificial Neural Network

Kakeru Kikumasa¹, Shin Kiyohara², Kiyou Shibata¹ and Teruyasu Mizoguchi¹

¹The University of Tokyo, Meguro, Tokyo, Japan, ²Tokyo Institute of Technology, Yokohama, Kanagawa, Japan

Introduction

Electron energy loss near edge structure (ELNES) and x-ray adsorption near edge structure (XANES), so-called core-loss spectroscopy, correspond to absorption spectrum originating from an electronic transition from a core orbital to conduction bands, and reflect partial density of states of the unoccupied orbitals at the excited state. As ELNES and XANES commonly provide information on local atomic structures and electronic states, they have been widely used for nanoscale analysis of wide variety of functional materials.

Although ELNES and XANES are powerful spectroscopic technique, the relationships between their spectral features and material properties are ambiguous. ELNES/XANES spectra have been interpreted with a fingerprinting approach combined with a theoretical calculation using high expertise and many computations, hence this approach requires lots of computational time and efforts. On the other hand, recent improvements of experimental equipment enable time-resolved and/or space resolved ELNES/XANES observation, which provides up to tens of thousands of spectra in an experiment. For this reason, individual interpretation of such massive spectrum data via the conventional approach is becoming unrealistic.

In recent years, machine learning (ML) approaches, which can overcome the limitation of analysis using multidimensional big-data, have been utilized in spectroscopy. Several studies have applied ML to ELNES/XANES spectra and succeeded in extracting information on the local chemical environments such as average bond length and Mulliken charge [1]. Nevertheless, there has been little study done concerning carbon-K edge ELNES/XANES spectra of organic molecules, which are often used for investigating local chemical environment of various organic functional molecules.

In this research, we applied ML to carbon-K edge spectra dataset generated by *ab initio* calculations. We constructed a neural network model to extracted hidden information about molecular properties quantitatively. Moreover, we included some data about molecules in addition to spectrum data to improve the accuracy of properties prediction.

Methodology

A feedforward neural network was used for constructing prediction models in this study. Input data was composed of 240 dimensional vector data of spectral intensities, which were referred to range of 24 eV in 0.1 eV steps. For output data, 12 properties of the molecules were selected. Hyperparameters, including the number of hidden layers, the number of neurons and the weight decay parameter, were tuned through cross validation.

To construct the carbon-K edge spectra dataset, 7,941 organic molecules are selected from the QM9 database [2,3]. Most of the molecules have multiple carbon site, hence 38,442 spectra were calculated. These spectra were calculated by a first principle pseudopotential method using CASTEP code [4]. The cut-off energy was set at 500 eV, and 15 Å supercells were used. All the calculated spectra were aligned with their threshold and broadened by Gaussian function with the FWHM at 0.6 eV.

Results and discussion

Several organic molecules and their corresponding calculated carbon-K edge spectra are shown in Fig. 1. Their spectral features are remarkably different reflecting the difference of chemical bonding and/or atoms in the molecules despite these molecular structures are similar. Figure 2 shows the result of the regression analysis on HOMO-LUMO gap and internal energy by the constructed prediction model. HOMO-LUMO gap was successfully predicted, whereas the regression accuracy of internal energy remained at a low level. Careful investigation suggests that the accurate prediction of extensive properties is difficult using only by the spectral data as the input data. By further investigations, we have succeeded in improving the prediction accuracy of internal energy.

In summary, we confirmed that the carbon-K edge ELNES/XANES spectra contain sufficient information for predicting molecular properties, especially intensive properties, with the aid of ML. Our model provides a new way to investigate time and space resolved materials properties.

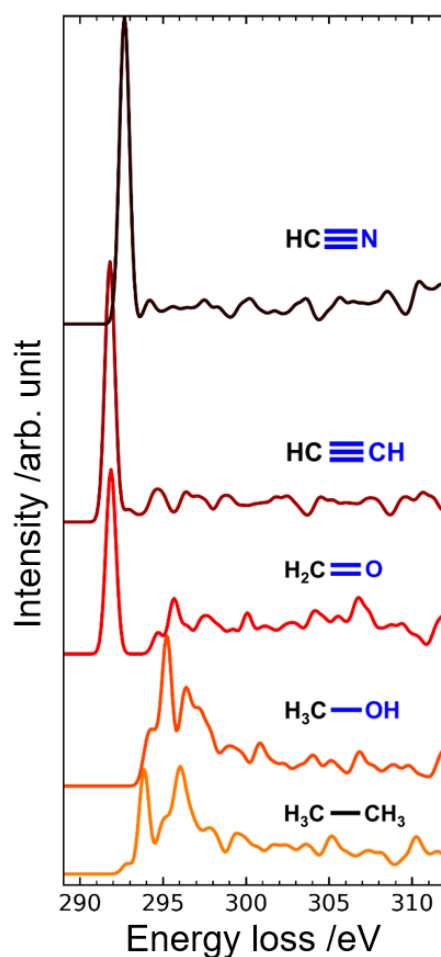


Figure 1. Some of the calculated carbon-K edge spectra of organic molecules in this study.

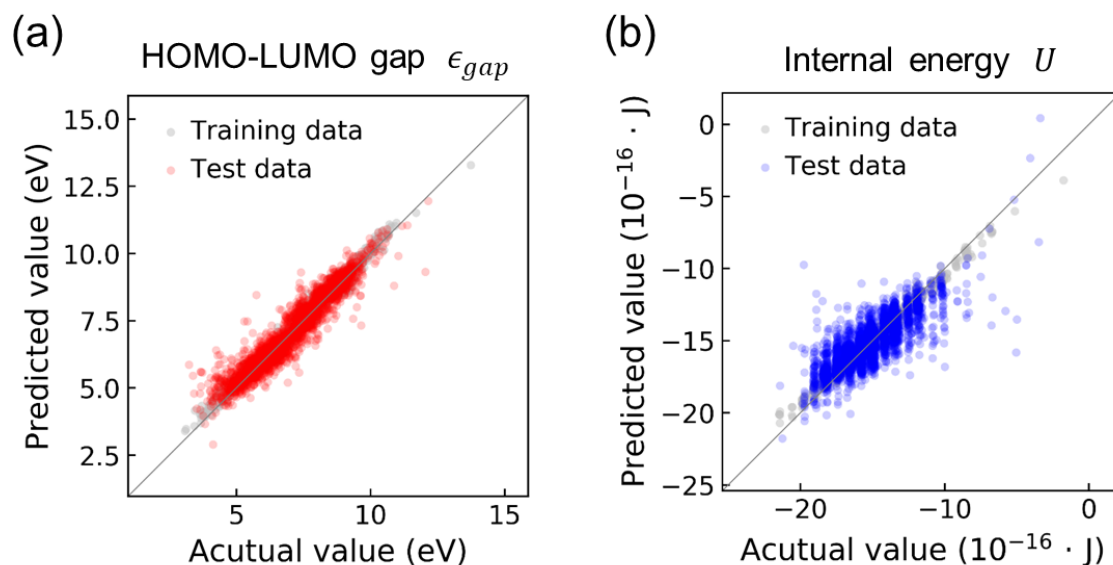


Figure 2. Actual values (colored) and predicted values (grey) of material properties. (a) HOMO-LUMO gap ($\epsilon_{LUMO}-\epsilon_{HOMO}$), (b) Internal Energy. Grey diagonal lines indicate points of equal predicted values and actual values.

References

- [1] S. Kiyohara *et al.*, *J. Phys. Mater.*, **2** (2019) 024003
- [2] L. Ruddigkeit *et al.*, *J. Chem. Inf. Model.*, **52** (2012), 2864-2875
- [3] R. Ramakrishnan *et al.*, *Scientific Data*, **1** (2014), 140022.
- [4] M. D. Segall *et al.*, *J. Phys. Condens. Matter.*, **24** (2002), 2717-2744