COMPUTATIONAL RESEARCH in BOSTON and BEYOND SEMINAR

Which parts matter? Interpretable random forest models for X-Ray absorption spectra

DR. STEVEN TORRISI

Department of Physics, Harvard University
Research Scientist, Toyota Research Institute
(Starting August 2021)

ABSTRACT:

X-ray absorption spectroscopy (XAS) produces a wealth of information about the local structure of materials, but interpretation of spectra often relies on easily accessible trends and prior assumptions about the structure. Recently, researchers have demonstrated that machine learning models can automate this process to model the environments of absorbing atoms from their XAS spectra. However, machine learning models are often difficult to interpret, making it challenging to determine when they are valid and whether they are consistent with physical theories. In this work, we present three main advances to the data-driven analysis of XAS spectra: we demonstrate the efficacy of random forests in solving two new property determination tasks (predicting Bader charge and mean nearest neighbor distance), we address how choices in data representation affect model interpretability and accuracy, and we show that multiscale featurization can elucidate the regions and trends in spectra that encode various local properties. The multiscale featurization transforms the spectrum into a vector of polynomial-fit features, and is contrasted with the commonly-used "pointwise" featurization that directly uses the entire spectrum as input. We find that across thousands of transition metal oxide spectra, the relative importance of features describing the curvature of the spectrum can be localized to individual energy ranges, and we can separate the importance of constant, linear, quadratic, and cubic trends, as well as the white line energy.

This work has the potential to assist rigorous theoretical interpretations, expedite experimental data collection, and automate analysis of XAS spectra, thus accelerating the discovery of new functional materials. We expect that this featurization strategy could be useful for broad domains of application, such as one-dimensional time-series analysis or other forms of spectroscopy.

Paper: https://www.nature.com/articles/s41524-020-00376-6

FRIDAY, JULY 23, 2021
12:00 PM – 1:00 PM

https://math.mit.edu/sites/crib/

ZOOM MEETING info:

https://mit.zoom.us/j/96155042770

Meeting ID: 961 5504 2770

Massachusetts Institute of Technology