The Paper

Charting the low-loss region in electron energy loss spectroscopy with machine learning
DOI: `10.1016/j.ultramic.2021.113202`
In this paper the authors use machine learning algorithms originally developed for particle physics to subtract the zero-loss peak from the useful low-loss region in an electron energy-loss spectroscopy spectrum. This new method allows the authors to have an assumption- and bias-free method for unmasking the low-loss regime with known uncertainty propagation.

Background

In the paper the authors aim to study the low-loss region of a tungsten-di-sulfide nanostructure, the same 2H/3R poly type nanoflower as seen in the Z-contrast imaging segment of the lecture on TEM sensors, by employing their own developed machine learning method. The employ the machine learning algorithm to subtract the intensity profile of the zero-loss peak from the overall electron energy-loss spectroscopy spectra, (EELS)-spectra. Electron energy-loss spectroscopy is a transmission electron microscopy technique by which the electronic properties of an electron transparent material are probed by measuring the energy-loss of scattered electrons from the illuminating beam of the microscope. The impinging electrons either lose their energy inelastically by interacting with the electrons in the specimen lattice or pass through the sample without interacting or interacting elastically with the atoms cores and thus contributing to the zero-loss peak (ZLP). The energy-loss is probed by using a magnetic prism positioned post-column, this prism spreads the electrons based on their velocity which in turn is determined by the beam voltage and electron-loss. After the prisms a selecting slit allows only electrons with a certain range of energy-losses through which are then imaged by an imaging device.

Neural network architecture

The goal of the neural network is the subtraction of the ZLP from the overall EELS spectrum, to this end the ZLP needs to provide an estimate of the intensity of the inelastically scattered electrons. With designing this neural network the authors employed a Neural Network Parton Distribution Functions (NNPDF) approach [1] which was previously pioneered for use in the field of high-energy particle physics. This approach is based on a combination of three main components, the use of neural networks as unbiased interpolants, the Monte Carlo method to estimate and propagate uncertainties, and, genetic algorithms minimizations for the training of the neural networks to avoid over-fitting.

The neural network employed was designed to estimate the intensity of inelastic scattering based on the machine parameters at the time of the experiment. To this end the first layer of neurons serve as inputs of these parameters to the neural network. The intensity of the zero-loss peak (ZLP) depends on parameters, such as beam energy, exposure time, energy loss, etc, therefore these parameters are supplied to the first layer. The neural network is a feed forward network with three more layers after the first and a single end neuron outputs an intensity level for the given input parameters. The second and third layer a made up of non-linear sigmoid activation neurons with ten and fifteen neurons for the second and third layer respectively, the fourth layer consists of five ReLU neurons. The non-linear activations of the second and third layer allow the neural network to learn the

---

[1]J. Rojo, Machine learning tools for global PDF fits, in: 13th Conference on Quark Confinement and the Hadron Spectrum, Vol. 9, 2018, http://arxiv.org/abs/1809.04392

non-linearities in the dependence of the ZLP intensity on machine parameters while the ReLU neurons in the final layer ensure that the system is positive-definite.

Uncertainty propagation

On of the key aspects with the approach undertaken in the paper is that unlike other methods of removing the zero-loss peak (ZLP) the uncertainties of the intensities of the ZLP that are subtracted are known, allowing the estimation of error in the subtraction is a more transparent and trustworthy approach. To estimate the uncertainties and propagate them the authors use the Monte Carlo replica method. The idea behind the method is that for gathered testing data there is still underlying information, such as, central values, uncertainties, and, correlations. These are used to construct a sampling of the probability distribution in the space of possible data. This data can then be fed into the trained neural network to map the space of possible data to a space of possible ZLP models.

Training strategy

Due to the incomparability of the two scenarios the model is trained differently for electron energy-loss spectroscopy (EELS) spectra taken of a vacuum or material sample. Both spectra are however only trained on the low-loss region of the spectra to ensure the model is trained on the data points were the elastic scattering of the zero-loss peak is most prevalent. For *vacuum spectra* the model is trained on a set of Monte Carlo replications of the actual vacuum EELS spectrum. The training of the model is sped up by scaling all in- and outputs to lie in the region of most variability in the sigmoid function. Furthermore, the model is constraint with a Lagrange multiplier by adding pseudo-data points to the trailing edge of the ZLP to make it tend to zero for large energy-loss without effecting the training dataset. For *sample spectra* the EELS spectrum is split into three regions, the low-loss region where the ZLP dominates, the intermediate region where contributions from the sample and the ZLP are equal, and, the high energy-loss region where the ZLP tends to zero. Since there is high confidence in the behaviour of the ZLP in both the first and third region the model is only trained on these regions. Then for both vacuum spectra and sample spectra the learning process is started with non-deterministic values being assigned to the nodes and weights. These are adjusted and optimised by stochastic gradient descent combined with back propagation. Since a small deviation is always applied after the normal gradient descent it is less likely for the neural network to become stuck in a local minimum of the cost function.

Possible pitfalls

In the paper the authors use replications of experimental data to enlarge the training data set. The newly generated data should therefore not be to correlated to any one measurement in the original data set, this is circumvented by enriching the training data with Monte Carlo replications of the original data by drawing from a distribution such that the central values and errors of the enriched training data do not deviate.