Using unsupervised machine learning for "better" TEM analysis

Antonius T. J. van Helvoort

Department of Physics, Norwegian University of Science and Technology (NTNU), N-7491 Trondheim, Norway

Corresponding author: a.helvoort@ntnu.no

Machine learning is an intriguing topic that has made disruptive transformations in our society. In science and technology, it has led to breakthroughs in many disciplines. In this contribution I will present, through some practical examples, how unsupervised machine learning can be a valuable practical tool for materials characterization based on transmission electron microscopy (TEM). Machine learning can be applied on image and for spectrum image analysis. Here we will focus on scanning electron diffraction (i.e. 4D scanning TEM) data. Further, only unsupervised machine learning will be addressed here, and to be specific mostly non-negative matrix factorization (NMF) [1]. This contribution will demonstrate the basic principles, added value and limitations of unsupervised machine learning in the field of electron microscopy.

The first example analyzes the crystal structure of III-V nanowires. These structures are interesting for future optoelectronic devices. They are ideal test structures to show the principles of NMF applied to scanning electron diffraction data. The code and similar data are freely available for who is interested [2]. To help the unsupervised machine learning process, dynamical diffraction effects are suppressed by collecting scanning precession electron diffraction [3,4]. The resulting components obtained resemble the electron diffraction pattern from the expected phases and orientations. Deconvolution in a higher number of components can for example extract thickness effects. Finding the correct number for deconvolution is in general a challenge for unsupervised methods. The basic approach is applied to thick (~400 nm) nanowires and a 10 μm long focused ion beam cross-section of a nanowires. Data set consisting over 1.10^6 patterns have been analyzed. Bending of the lamella is limiting the area from which data can be deconvoluted into only a few straightforward interpretable components.

The second example will demonstrate that unsupervised machine learning makes statistical analysis by TEM possible, over large area and various samples, while preserving nm-scale spatial resolution. Thereby unsupervised machine learn can tackle one of the main shortcomings of conventional TEM. In this study the evolution and number density of five precipitates in a 6xxxx alloy during a series of heat treatments has been deduced [5]. To obtain such new results a tailored data treatment prior to the NMF step is required. Verification of the predicted phases is achieved by other techniques such as lattice imaging for some randomly selected precipitates. Possible overlap between different phases in the volume illuminated by the electron probe should be taken into consideration. It will be shown that the NMF can even with overlap present deconvolute data into separate phases. This aspect of NMF approach is particularly relevant for the analysis of polycrystalline areas and nanomaterials [4, 6].

There are adapted forms of NMF and other unsupervised routines such as cluster analysis that have been successfully applied to TEM data analysis as well. These tools are or will become also freely available through different open-source packages. Furthermore, new routines are continuously developed. With the ever-growing data sizes unsupervised machine learning tools will gain a more essential role in data exploration. One of the main take-home messages from this contribution is however that verification of unsupervised machine learning algorithms is an essential step in developing and applying them to TEM data and to achieve “better” final results [7].

[1] Lee & Seung, Nature **401**, 788 (1999).
[2] D. Johnstone, pyxem 0.13.0: Demo 1. Available at 10.5281/zenodo.4436723.
[3] R. Vincent & P.A. Midgley, Ultramicroscopy **53**, 271 (1994).
[4] B.H. Martineau, Adv. Struct. Chem. Imaging **5**, 1 (2020).
[5] J.K Sunde, Materials Characterization **142**, 458 (2018).
[6] T. Bergh, J. Microscopy. **279**, 158 (2020).
[7] Acknowledgement: I am depth to students working with me on open-source TEM data analysis since 2013. Regarding unsupervised machine learning I want to specially name, V.T. Fauske, J.S. Nilsen, S. Høgås, T. Bergh, H.W. Ånes and D.N Johnstone.