Methods and Prospects of Machine Learning applied to Challenges in Crystallography

Markus Reischl, Jan Schützke

Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, Markus.Reischl@kit.edu, Germany,

Within the domain of analyzing powder X-ray diffraction scans, manual examination of the recorded data is the most popular method, but it requires expertise, is time-consuming and provides subjective results. Nowadays, numerous papers are published showing advances and benefits of artificial intelligence, especially of deep learning methods, using data-centered approaches to automatically derive mathematical models (e.g. for classification). With this contribution, we give a brief overview about possibilities of modern data-based modelling, about implementations, necessary data-structures and their application to crystallography. Coming back to X-ray diffractometries, we show how to design an automated phase-analysis model based on a Convolutional Neural Network (CNN). Therefore, a framework for the efficient generation of simulated diffraction scans is developed, since real measured and labeled scans are hardly available, and deep learning approaches require an extensive dataset to learn a general representation. Using this synthetic database, the CNN is parameterized, trained and compared against the manual analysis. As a supportive approach, a second network, a denoising autoencoder is presented, which can be used to eliminate background and other disturbing effects from the signal [1].

[1] Schuetzke J, Benedix A, Mikut R, Reischl M. Enhancing deep-learning training for phase identification in powder X-ray diffractograms. 2021. IUCrJ, 8 (3), https://doi:10.1107/S2052252521002402