Application of Machine Learning to XRD Phase Identification

Brian Jones[1], Nathan Henderson[1], S. Nick Rodesney[1]
Jan Schützke[2,3],Markus Reischl[2], Alexander Benedix[3], Karsten Knorr[3]

[1] Bruker AXS – 5465 East Cheryl Pkwy – Madison, WI - US
[2]Institute for Automation and Applied Informatics, KIT, Karlsruhe, Germany
[3]Bruker AXS GmbH, Karlsruhe, Germany

Identification of crystalline phases in mixtures is a frequently performed task in powder XRD. It mostly involves software for searching databases of known compounds, and matching lists of d-spacings and related intensities to the measured (or reduced) data. Figures-of-merit are usually taken as numerical indicators for the probability of the individual phase assignments. However, some expertise of the skilled user is still required for a "manual" validation of the results. This is time-consuming and error prone. Most automated search/match procedures that apply some iterative procedure of the above aim at making the validation step redundant but have failed to be generally reliable to this point.

In recent years, deep learning models established their status as a state-of-the-art approach for automated image analysis, such as detecting cars and pedestrians in a street scene. In analogy, deep learning models were applied here for automated phase identification from one-dimensional XRD data. We used phases and mixtures from the Bruker AXS iron ore and cement application packages that are in wide commercial use, and contain 28 and 76 phases, respectively.

Several models have been tested, which learn the peculiarities of XRD data to support the automated phase identification process. A framework for the efficient generation of hundreds of thousands of simulated scans has been developed, since real measured and labeled scans are only scarcely available, and deep learning approaches require an extensive dataset to learn a general representation. This learning step considered not only varying phase presence and concentrations but noise, background variation, peak broadening as well as variance in lattice parameters.

The trained networks achieve an accuracy of close to 100% for synthetic mixtures of both application packages while analyzing hundreds of scans in under a second, thus outperforming the experts in speed without sacrificing an accurate prediction. Additionally, the models have been tested on measured XRD-scans to confirm the results.