Artificial Intelligence Methods in Analytical Data Interpretation: Bayesian Approach and Fisher Information

Alex Ulyanenkov[1], Alexander Mikhalychev[2], Svetlana Vlasenko[2], Nikita Lappo[2], Ilya Perapechka[2]

[1]Atomicus LLC, 4730 University Way NE Suite 104-32882, Seattle, WA 98105, USA

[2]Atomicus OOO, Mogilevskaya Str. 39a-215, 220007 Minsk, Belarus

info@atomicus.de

Computer-based automated data analysis procedures are not only much faster and less demanding for expert knowledge of operators than manual techniques, but also provide reliability and reproducibility of the analysis results due to much lesser vulnerability to accidental errors. Generally, analysis can be based on the previously gained experience (human expertise for manual approaches or machine learning for computer systems) or rely on fundamental principles of physics, mathematics, statistics, etc., when the analysis algorithms are derived instead of being "guessing". In this contribution, we advocate the latter approach: namely, we demonstrate the capabilities of two efficient mathematical tools (Bayes' formula and Fisher information) suitable for automatic analysis of X-ray data, but often overlooked by physicists.

The Bayes' theorem quantitatively connects our *a posteriori* knowledge about the investigated sample with *a priori* knowledge and the measurement results and is applicable to ranking of models in the pattern recognition problem. We apply a Bayesian approach to quantitative analysis of X-ray powder diffraction data [1] (identification of phases from a measured X-ray spectrum), and to peak indexing in high-resolution X-ray diffraction (HRXRD) spectra (finding the correspondence between each measured peak with an expected Bragg peak, a layer thickness oscillation fringe or a superlattice fringe) – Figure 1. The initial information about the investigated sample (e.g. absence of some elements in a powder sample) is encoded in prior probabilities for the models. The expected inaccuracies of the peak parameters determination due to statistical and systematic noise in the data are characterized by the likelihood function. Then, models of the spectrum decomposition (peak identification) are ranked according to the estimated posterior probability. The optimal peak labeling is provided by the most likely model.

Fisher information [2] quantifies the ability to reconstruct model parameters from measured data with a statistical noise. Cramer-Rao bound [3] connects the so called Fisher matrix with the expected variances and covariances of the estimated parameters, thus giving the information about the reliability of the obtained results and the possible correlations between the model features. First, we show that this tool is suitable for optimizing the planned HRXRD measurements by choosing the most informative reflections and measurement geometries depending on the sample parameters of interest. The proposed approach gives a way to rate all possible sets of scans on the base of the amount of their information on the investigated sample. Also, banded diagonal structure of the inverse Fisher matrix implies that the reconstruction problem can be solved locally, by iterative consideration of only a subset of parameters to be reconstructed at each step. We apply this technique to the problem of decomposing an X-ray spectrum, consisting of a large number of strongly overlapping peaks, into contributions of individual peaks.
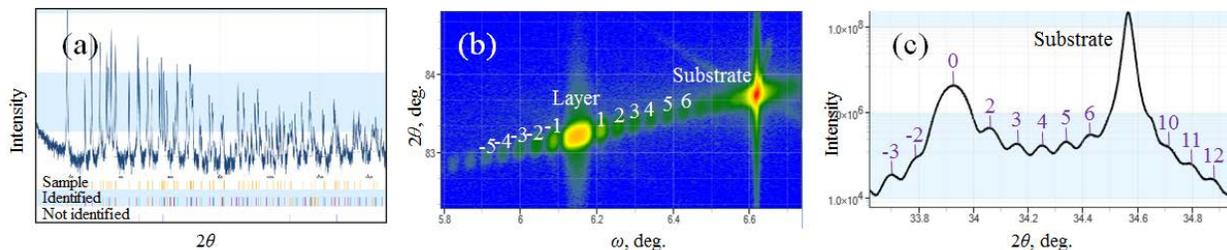

Figure 1. Application of Bayesian approach to HRXRD and powder diffraction spectra: automatic peak identification results for powder diffraction data for a mixture of zincite, fluorite, brucite, and corundum (a), for a reciprocal space map for $In_{0.06}Ga_{0.94}As$ layer at GaAs substrate (b), and for a scan for GaN – $In_{0.15}Ga_{0.85}N$ superlattice (c).

1. A. Mikhalychev, A. Ulyanenkov, *Bayesian approach to powder phase identification*, J. Appl. Cryst., 2016; **50**: 776-786.
2. A.Ly et al, *A Tutorial on Fisher information*, Journal of Mathematical Psychology, 2017; **80**:40–55.
3. H. Cramer, *Mathematical Methods of Statistics* (Princeton, Princeton University Press, 1946).