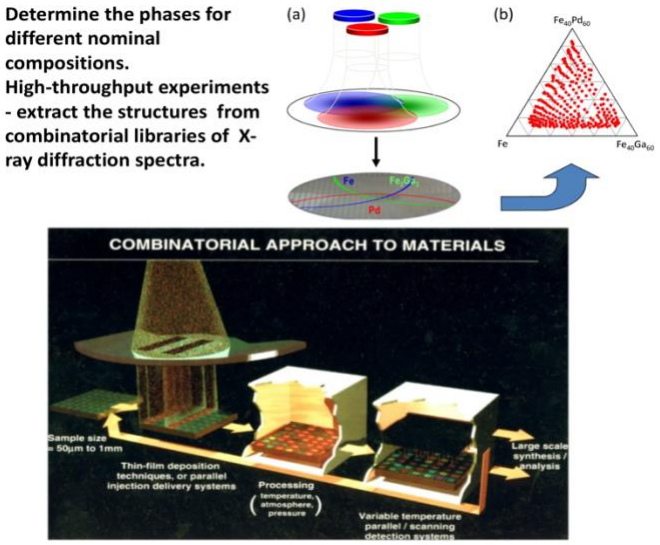
August 2018

Unsupervised phase mapping of X-ray diffraction data by nonnegative matrix factorization integrated with custom clustering

Analyzing X-ray material combinatorial libraries



Analyzing large X-ray diffraction datasets is a key step in high-throughput mapping of the compositional phase diagrams of combinatorial materials libraries. We have developed a new unsupervised machine learning method for pattern analysis and phase extraction of such datasets. The new method is capable of robust determination of the number of basis patterns present in the data, which, in turn, enables straightforward identification of any possible peak-shifted "spurious" patterns. Such peak-shifted patterns arise due to small lattice variations of the same structure. These are caused by chemical alloying and are ubiquitous in X-ray datasets of combinatorial thin-film libraries. Successful exclusion of the peak-shifted patterns permits to quantify and classify the contribution of all distinct structures to each data point, which can be used to determine accurately the compositional phase diagram of the studied system.

The Algorithms and Mathematics

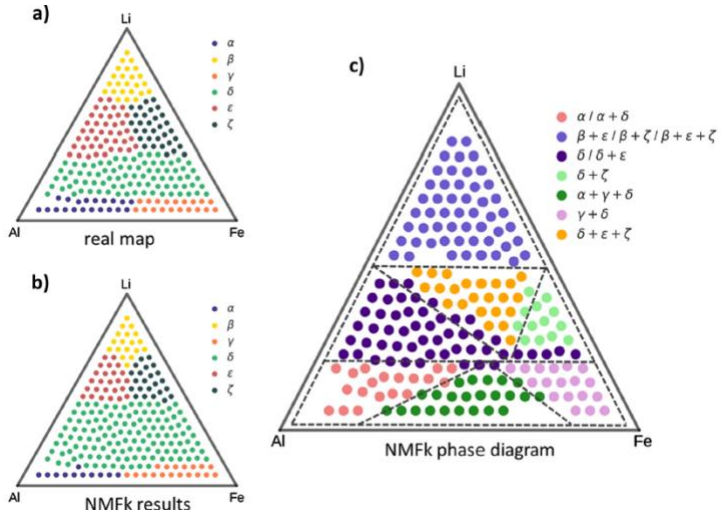
The new unsupervised machine learning method is based on the Nonnegative Matrix Factorization (NMF) algorithms, by organically combining two procedures. The first procedure expands on the conventional NMF by adding a robust protocol for determination of the number of end members in the X-ray set of spectra. The second procedure examines the patterns obtained at the end of the first procedure for peak-shifting. It estimates, via cross-correlation analysis, the unique end members of the investigated dataset, combines the peak-shifted copies, and modifies accordingly the abundancies obtained from the NMF analysis. The thus-extracted unique end members represent the constitutive single phases, and their abundancies can be used for determining the compositional phase diagram of the system.

The Impact

Because peak-shifting is ubiquitous in many characterization datasets, such as Fourier transform infrared spectroscopy, Raman spectroscopy, and X-ray photoelectron spectroscopy, and for each technique peak-shifting takes place for different reasons, the method developed by us is applicable not only to X-ray datasets but could be used for analyzing large variety of materials data.

Summary

Unsupervised Machine Learning (ML) methods extracts sets of hidden (and often previously unknown) features from uncategorized datasets. NMF has shown great promise for the task of analyzing large volumes of X-ray measurements. Its simplicity and ease-of interpretability offer great advantages, and several systems relying on this method were created and tested successfully on large datasets. One solved key problem is determining the number of basis patterns (end members) in the data. Another is peak-shifting of X-ray patterns – a common consequence of lattice changes caused by alloying. Our new method addresses both of these problems. First, it extends the conventional NMF algorithm by leveraging a robust protocol for determining the optimal number of NMF-extracted end members and then a second procedure examines the obtained basis patterns for peak-shifting. By applying our method to both real and synthetic datasets we demonstrated that it can be used for extracting the basis patterns, followed by a quick identification of the peak-shifted ones. We also show that the unique end members and their abundancies can be used for determining the outlines of the compositional phase diagram.



Contact

Boian S. Alexandrov, boian@lanl.gov
Los Alamos National Laboratory

Funding

This research was funded by the LDRD program at Los Alamos National Laboratory, the reserve DR# 20180060, *Robust Deep Unsupervised Machine Learning for Big-Data Analytics*.

Publication

Unsupervised phase mapping of X-ray diffraction data by nonnegative matrix factorization integrated with custom clustering, Valentin Stanev, Velimir V. Vesselinov, A. Gilad Kusne, Graham Antoszewski, Ichiro Takeuchi and Boian S. Alexandrov, npj Computational Materials (2018) 4:43; doi:10.1038/s41524-018-0099-2, https://rdcu.be/35Uc, Springer Nature.

LA-UR-18-20888