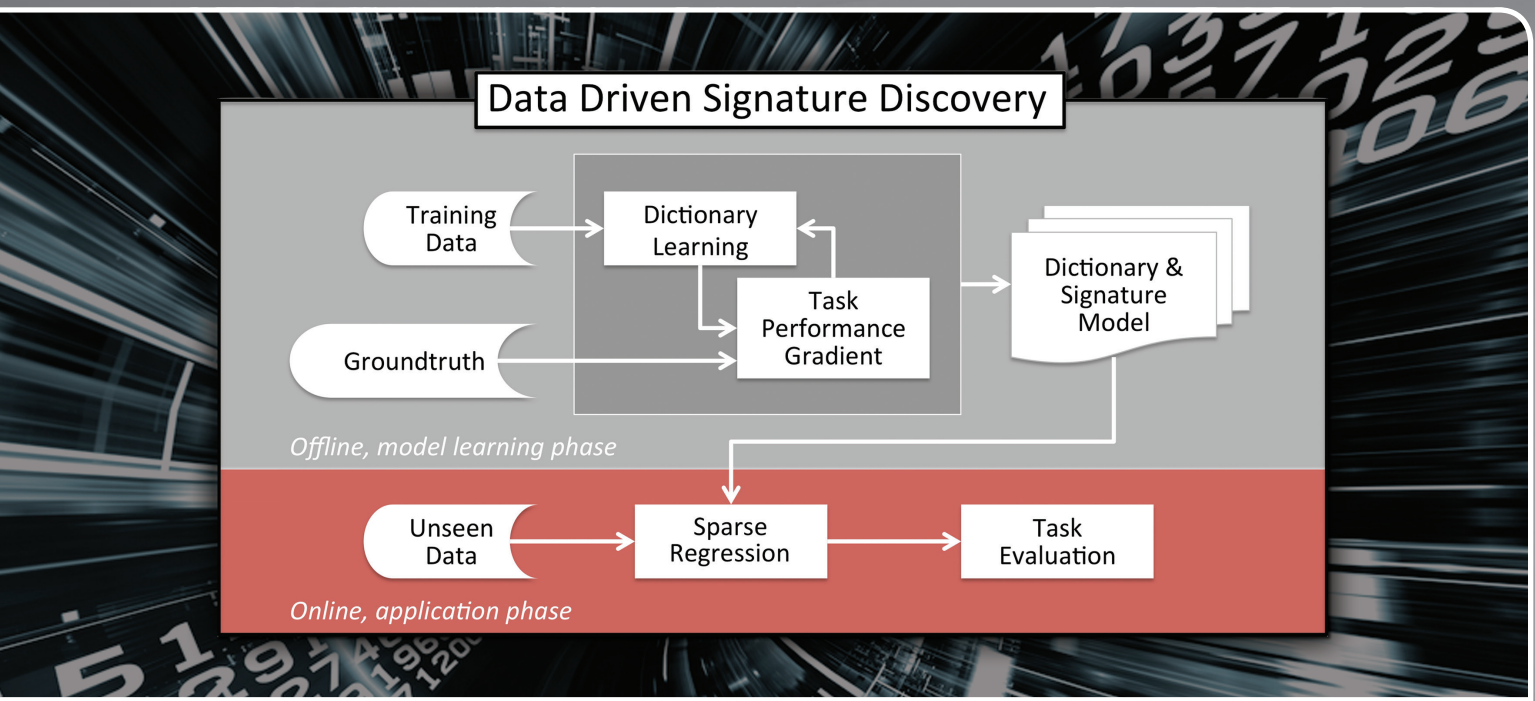


Compressive Sensing for Threat Detection

Data-driven signature learning can discover signatures that are optimal for specific tasks such as detection, multi-class discrimination, and regression. Methods to validate these signatures and quantify parametric uncertainty of the learning algorithm are also developed.

CHALLENGE

Our goal is to build a data-driven and data-agnostic signature discovery toolkit. Given data and some ground truth information, the toolkit will learn the optimal signature model that can subsequently be applied to new, unseen data. Once the model is learned, new data can be tested using simple linear algebra procedures, and testing can take place at or near sensors instead of in a data analysis center. Because our approach is data-driven, it is applicable to several domains and has been most recently applied to hyperspectral imaging and electron microscopy.

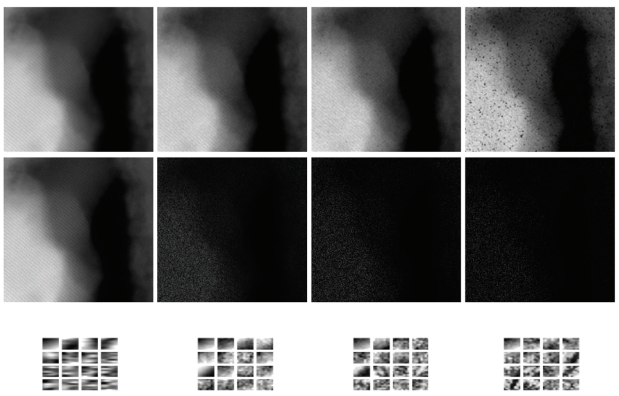
APPROACH

Our approach applies dictionary learning to find a measurement basis that represents each data sample as a linear combination of a small number of basis elements. In the past, it has been common to use principle components, wavelets, or sinusoids (e.g., JPEG) as the measurement basis. Instead of choosing the basis a priori, our approach discovers a basis that is optimal for a specific task. These tasks could be simply data reconstruction and denoising, or multi-class discriminative analysis such as hyperspectral unmixing. In hypserspectral imagery, using a random two percent of a noisy data cube, it is possible to give a high fidelity reconstruction and automatically label the spectra in their respective classes.

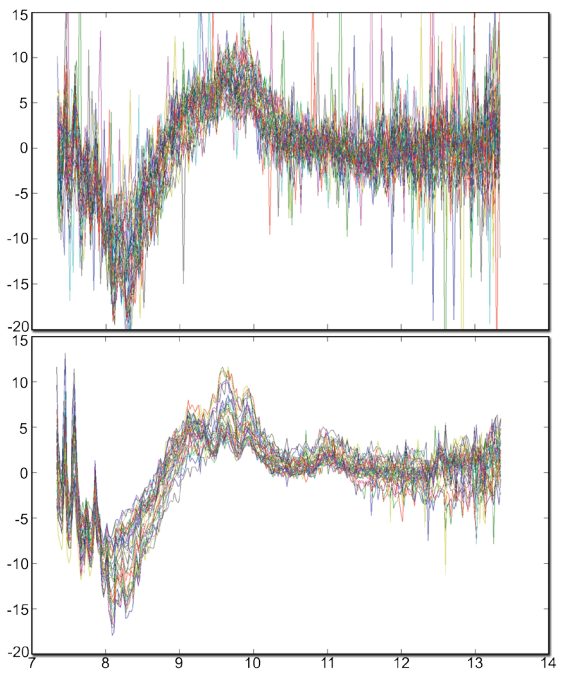
We employ both the Bayesian and optimization approaches for dictionary learning. There are trade-offs associated with each approach. The Bayesian approach automatically tunes model parameters, which results in a more accurate model, but usually requires a higher computational cost. Conversely, the optimization approach is faster, but requires parameters to be guessed or determined in a meta-learning process.

IMPACT

With this approach, we can reduce the number of samples (or sample with lower resolution hardware), and still obtain high fidelity reconstruction. In the case of electron



From Top to Bottom: Reconstructed Electron Microscopy Image (Zeolite), Data Used for Reconstruction, Dictionary Learned. From left to right: 100%, 20%, 10%, and 5% samples. Dictionary learning recovers the image from a fraction of the data.



Top: Original Spectra (Steel); Bottom: Spectra Recovered Using Two Percent of the Original Data Cube. The recovered spectra have significantly less noise and are easier to use for spectral unmixing.

microscopy, where taking measurements damages the specimen, we have shown that this approach allows for complete reconstruction with less than 10 percent of the standard measurments. Moreover, we can map sensor data directly to class labels without the need to reconstruct the data in the final visualizable form. Applying this methodology directly to sensors effectively elicits the capability to sense signatures instead of raw data.

ACCOMPLISHMENTS/OUTCOMES

1. We create a proof of concept application to scanning transmission electron microscopy. We have shown that high quality images can be reconstructed from less than 10 percent of the data normally used. For 3D tomographic reconstruction, we expect that the sampling ratio can be decreased even more.
2. We found an application to hyperspectral imaging. We have validated results from our collaborators at Duke University, reconstructing a hyperspectral cube from two percent of the original data. We have also begun to study the effects of weathering on signature drift in hyperspectral unmixing.
3. We developed methods for algorithm parameter tuning and parametric uncertainty quantification. We submitted a paper to IEEE evolutionary computing journal.
4. We developed a method for pre-conditioning non-negative matrix factorization (a special case of dictionary learning). The method allows state of the art NMF algorithms to work on problems an order of magnitude larger in the same running time; our methods was published at the International Machine Learning Conference Workshop on Spectral Learning.

Andrew Stevens
Pacific Northwest National Laboratory
andrew.stevens@pnnl.gov | (509) 375-2879

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle *Since 1965*

U.S. DEPARTMENT OF ENERGY