Sparse feature learning for condition monitoring

Sergio Martin-del-Campo[1,2] & Fredrik Sandin[2,1]

Objectives

Learn a dictionary of statistically independent features from a signal using sparse coding [1] in order to automatically characterize physical phenomena, including emerging faults.

- Probabilistic feature learning
- Data-driven approach, few assumptions
- Vibration and acoustic emission signals
- Detection of new and abnormal features

Introduction

Condition monitoring of machine elements is used to detect faults and reduce machine downtime. Early detection and characterization of abnormal operational conditions and emerging faults is an important and challenging problem. We study the use of feature-learning methods for automatic structuring and characterization of condition-monitoring signals from individual industrial machines.

Methods

The signal, $x(t)$, is modeled as a linear superposition of noise and features with compact support

$$x(t) = \epsilon(t) + \sum_i a_i \phi_{m(i)}(t - \tau_i). \tag{1}$$

The functions $\phi_m(t) \in \Phi$ are morphological features called *atoms*. The weights, a_i, and temporal offsets, τ_i, of atoms are calculated with matching pursuit [2].
The dictionary of atoms, Φ, is random at start and is *learned* by maximization of the expectation of the log data probability

$$\Phi = \arg\max_\Phi \langle \log [p(x \mid \Phi)] \rangle. \tag{2}$$

Equation (2) is solved with gradient ascent using a prior that promotes sparse coding and statistical independence of atoms [3, 4]

$$\frac{\Delta \phi_m}{\eta} = \frac{\partial \log [p(x \mid \Phi)]}{\partial \phi_m} = \frac{1}{\sigma_\epsilon^2} \sum_i a_i (x - \hat{x})_{\tau_i}. \tag{3}$$

The factors $(x - \hat{x})_{\tau_i}$ are residuals after matches with atoms $\phi_{m(i)}$ of amplitudes a_i at offsets τ_i. A learning rate parameter, $\eta \ll 1$, is introduced in the gradient ascent so that atoms converge. The number of atoms in the dictionary is set at start.

Results 1

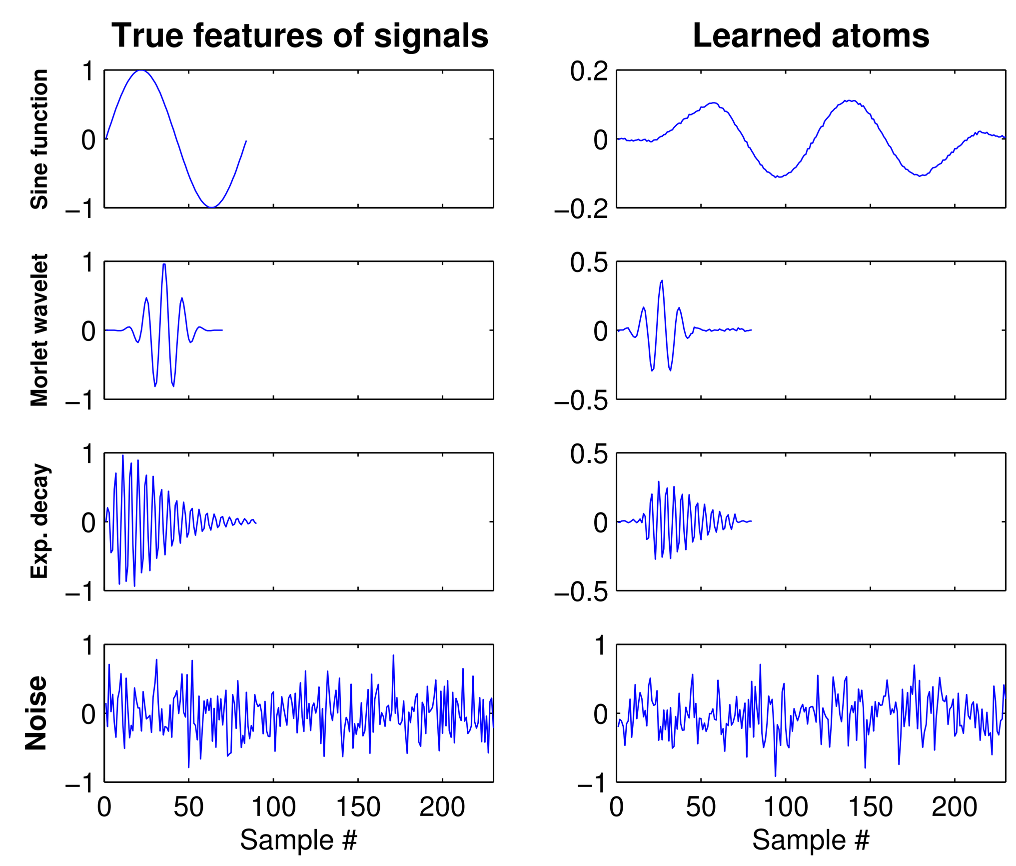

Figure 1: Basic blind-source separation problem.

Learning of features in an artificially constructed signal, which is a superposition of:

- Continuous sine function
- Morlet wavelets at random offsets
- Impulse functions, with exponential rise and decay, at random offsets
- Noise at 10 dB SNR

The two columns in the figure shows the true features of the signal and the learned atoms.

Results 2

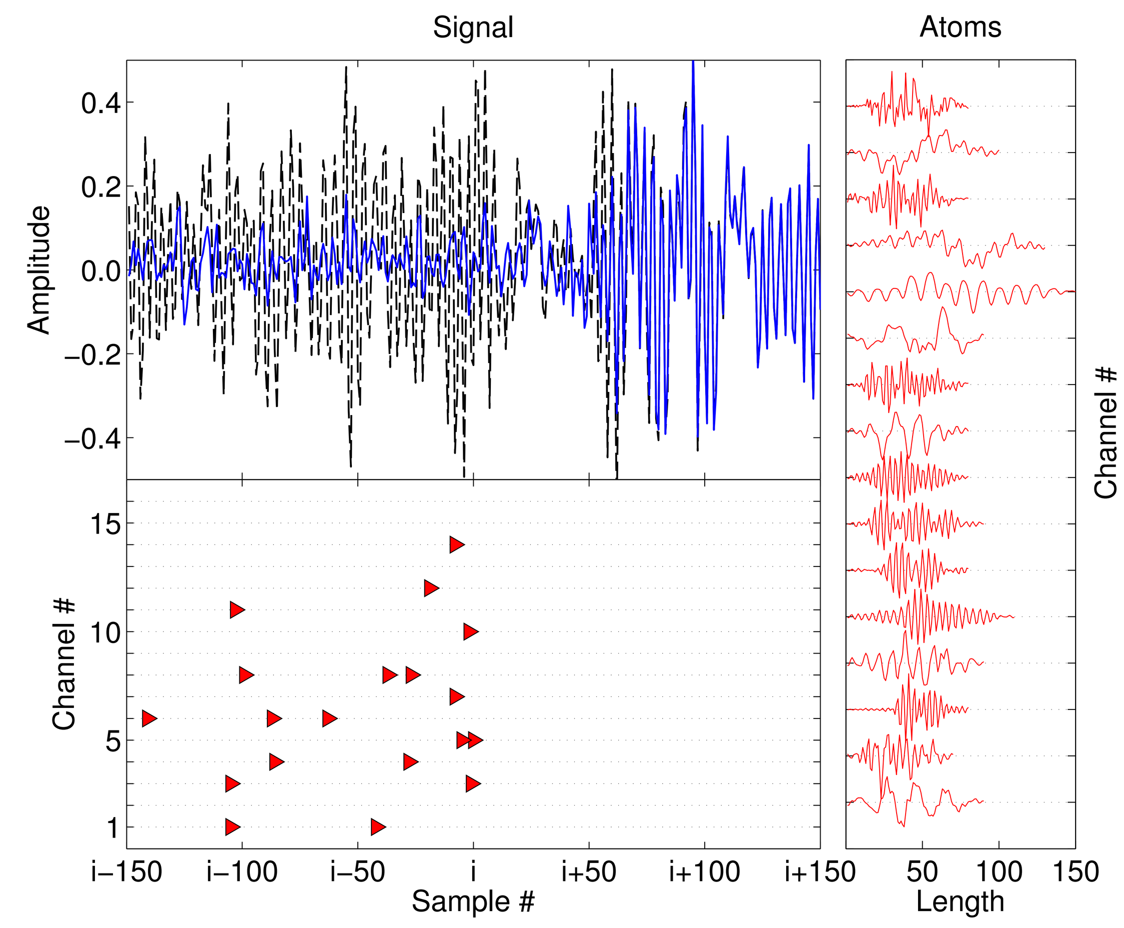

Figure 2: Event representation of ball-bearing vibration data using a dictionary of 16 learned atoms.

The signal (dashed line) is decomposed in events (triangles) representing activated atoms, which are substracted from the original signal resulting in a residual (solid line). The learned dictionary contains impulse-like atoms because of a defect in the bearing. On average 150 samples are represented by about 21 events at 10 dB SNR, i.e., the code is sparse.

Results 3

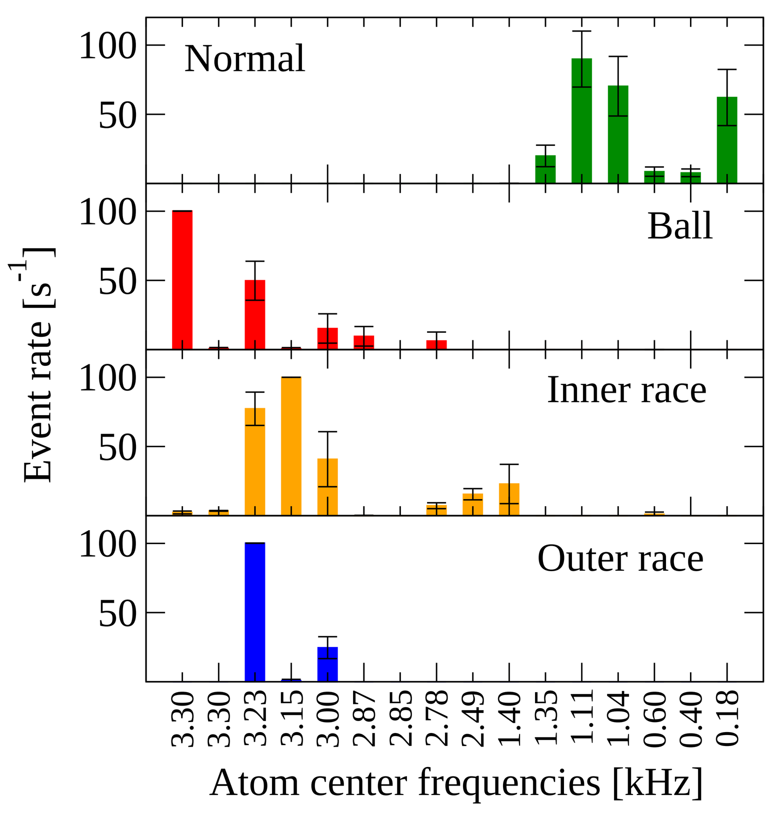

Figure 3: Histogram of event rates for each of the sixteen atoms in the learned dictionary.

Analysis of event rates is sufficient to distinguish normal bearings from faulty bearings, indicating that the learned features are related to different physical conditions.

- Faulty bearings are divided in three categories depending on the location of the faults (ball, inner race or outer race)
- The type of defect is characterized by the channel event rates, even thought the rpm, load and size of faults are varying
- Atoms learned from normal bearings have low center frequencies.
- Atoms learned from faulty bearings have higher center frequencies.

Conclusion

Sparse coding with dictionary learning is an interesting probabilistic method for feature extraction in condition monitoring applications. We find that the learned dictionary is useful in a basic classification task involving bearings with different faults. In addition, it is possible to reach about one order of magnitude reduction in data rate with little loss of useful information. A prototype implementation is presented in [5, 6, 7], including a discussion of a high-speed online processing system based on an FPGA.

Future Work

- Further development of the feature learning method and analysis of more realistic failure modes
- Processing of resulting events (pattern recognition) and waveforms (hybrid model?)
- Develop test-rig for acoustic emission experiments
- Further development of the FPGA prototype implementation [5, 6, 7]

Contact information

[1]SKF University Technology Center, Luleå University of Technology, Sweden.

[2]EISLAB, Luleå University of Technology, Sweden.

sergio.martindelcampo@ltu.se (Ph.D. student)
fredrik.sandin@ltu.se

References

[1] M. Elad. Sparse and redundant representation modeling – what next? *Signal Processing Letters, IEEE*, 19(12):922–928, 2012.
[2] S.G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Processing*, 41(12):3397 –3415, 1993.
[3] E. C. Smith and M. S. Lewicki. Efficient auditory coding. *Nature*, 439(7079):978–982, 02 2006.
[4] B.A. Olshausen and D.J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.
[5] S. Martin del Campo, K. Albertsson, J. Nilsson, J. Eliasson, and F. Sandin. FPGA prototype of machine learning analog-to-feature converter for event-based succinct representation of signals. In *Machine Learning for Signal Processing (MLSP), 2013 IEEE International Workshop on*, pages 1–6, 2013.
[6] K. Albertsson. Towards self-learning sensors: Matching pursuit with dictionary learning on FPGA. Master's thesis, Luleå University of Technology, 2013.
[7] J. Nilsson. Towards self-learning sensors: FPGA-Based ADC Front End. Master's thesis, Luleå University of Technology, 2013.