Grammar-guided Feature Extraction from Signals and Images

Damian R. Eads[1,2], Simon J. Perkins[1], James Theiler[1], Karen Glocer[1,2], Ethan L. Miller[2]

[1] Intelligent Searching of Images and Signals (ISIS) Team

Space and Remote Sensing Sciences Group, ISR-2, LANL

[2] Department of Computer Science, UCSC

Image and signal sensors collect data at increasingly greater rates, making it a great challenge to search and organize data in archives. Work is underway to engage the sensor, data management, and machine learning expertise of LANL and UCSC to tackle adaptive, content-based search in large remote sensing archives. We demonstrate the utility of a new method for extracting features from imagery and signals to aid the archival search problem.

MOTIVATION

• Advances in remote sensing technology have greatly enhanced our ability to observe the Earth at a distance. At the same time, data rates have greatly increased, thereby demanding the development of scalable algorithms for search, identification, and analysis of data.

Example: A single 60 cm GSD, pan-sharpened, 4-band, 11-bit CB, 10 km x 10 km QuickBird scene is about 2.2 GB and 277 MegaPixels!

• Manual development of highly specialized algorithms is time-consuming, error-prone, and expensive. Manual analysis of all data is not practical.

EXAMPLE TASKS

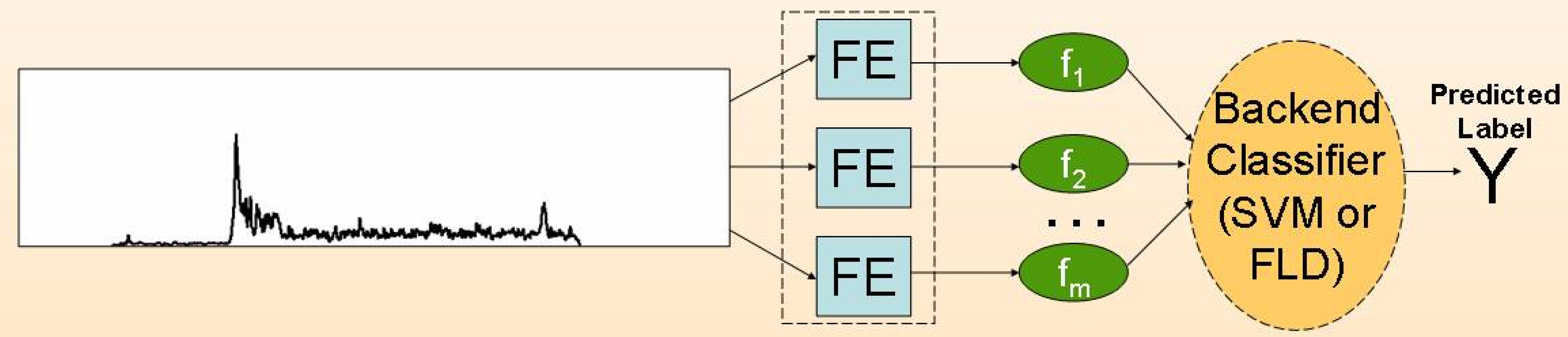
Images: We wish to teach the system to classify pixels (e.g. healthy vs. sick wheat crop) The system classifies pixels. Generate land-cover maps from imagery.

Signals: We wish to classify events recorded in time series (e.g. lightning).

OUR APPROACH

The feature extraction process is automated using grammar. Grammars enable a large space of intelligent feature extraction architectures to be expressed in a compact syntax. We employ simple hill climbing to build a suite of feature extraction algorithms. We use FLDs and SVMs for final pixel classification.

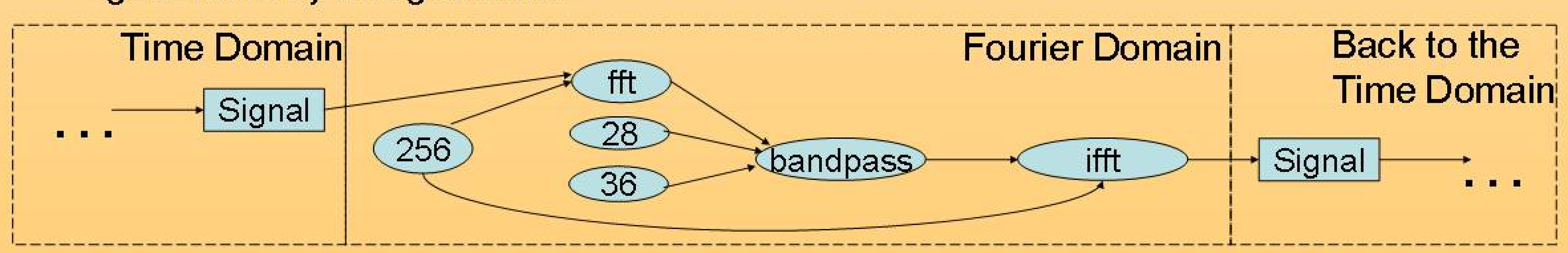
GRAMMAR-GUIDED METHOD

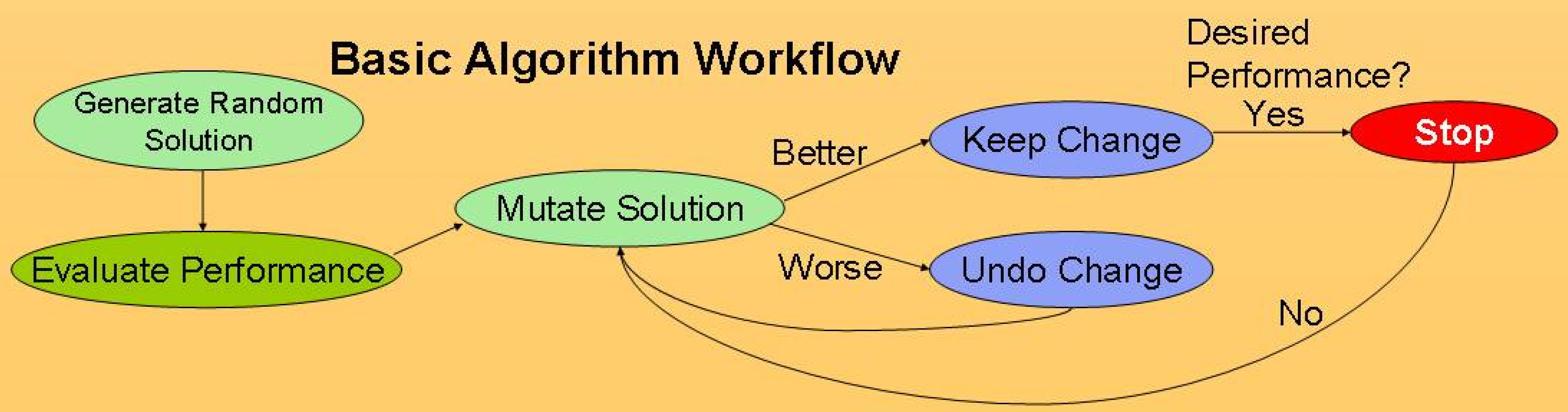


Time series classifier solution: is a collection of feature extraction algorithms (the dashed box) combined with a backend classifier (the dashed circle). When applied to time series, the time series is fed into each individual feature extractor (labeled FE), and each produces at least one numerical descriptor of the input signal. These features are then fed into a back-end classifier and a predicted label (Y) results.

```
TimeProc(X)        ::= TimeProc(Fourier(X)) | autocorr(X) | . . .;
RandNum            ::= 0:10^-12:1;
RandFFTPoint       ::= *{64, 128, 256, 512};
Fourier(X)         ::= ifft(FourierProc(fft(X, PT = RandFFTPoint()), PT));
FourierProc(X, Y)  ::= lowpass(X, *(1:Y/2)), | highpass(X, *(1:Y/2))
                      | bandpass(X, *(1:Y/2), *(1:Y/2));
```

Grammar Snippet: ensures feature extraction algorithms operate in Fourier space in a sensible way. The Fourier production forbids any time domain operators from working on frequency-domain data. Shown below is an example of a program piece generated by this grammar.





RESULTS

Method	FORTE 2	FORTE 6	FORTE 7
GGFE+FLD	13.22%	21.49%	25.87%
GGFE+LSVM	20.66%	28.1%	28.67%
FLD, no FE	40.5%	50.41%	54.54%
LSVM, no FE	30.58%	40.50%	34.97%
SVM+RBF, no FE	28.93%	38.02%	37.06%

10-fold Cross Validation Error for Grammar-guided FE vs. no Feature Extraction on FORTE 2, 6, and FORTE 7 data sets.

CONCLUSIONS

• Demonstrated on FORTE lightning data set as well as other more standard data sets.
• Attained better accuracy than baseline.
• More extensive experimental comparisons are underway with better baselines.
• The grammar-guided approach has been extended to pixel classification of imagery. Integrated into Genie Pro imagery tool. Genie Pro placed first in a DoD competition of other competitors.
• MATLAB and C++ implementations are available.

ACKNOWLEDGMENTS

This work was supported by grants from the Department of Energy, Department of Defense, and various other government agencies. We would like to thank the ISIS team for helpful discussions and feedback.