SAND2017-8128C

Predictive Surveillance Intelligence

Machine Learning & Deep Learning Conference 2017

Stephen Jackson, 2662

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525

Data Systems Engineering

Any Data • Any Where • Any Time

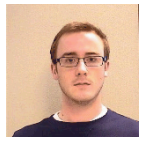
Who We Are

Erin Acquesta

5853

Gary Ashcraft

2662



Michael Bridges

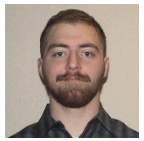
2662

Stephen Jackson

2662

Rosalie Multari

2257



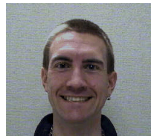
Gunnar Teague

2668

Randy Wells

5852

Managers:



Brian Post

2666

Shawn Kerr

2662

Predictive Surveillance Intelligence (PSI) has been funded for FY17 through Mission Area Alignment Initiative (MAAI) Funds

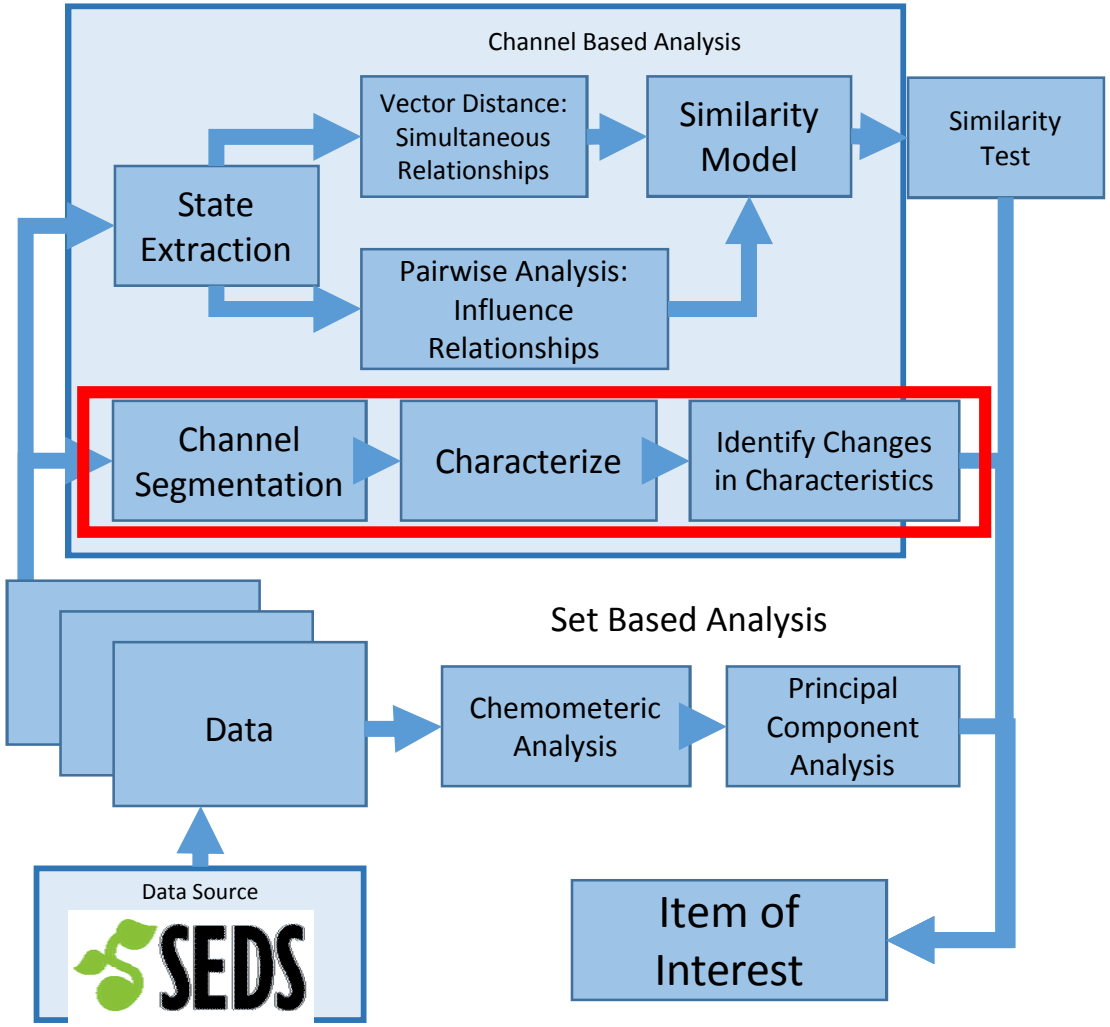
Motivation

- Assessing the viability of the nuclear weapons stockpile is a keystone of nuclear deterrence.
- Weapons in the stockpile are regularly assessed to ensure their functionality.
- Historically, this has been a primarily “human-in-the-loop” process.
- Many weapons in the stockpile are due for modernization
  - Data generated surveilling modernized weapons will exceed current analytical capabilities.
  - Machine learning has high potential to augment the capacity of engineers and provide automated analytical capabilities.

What We're About

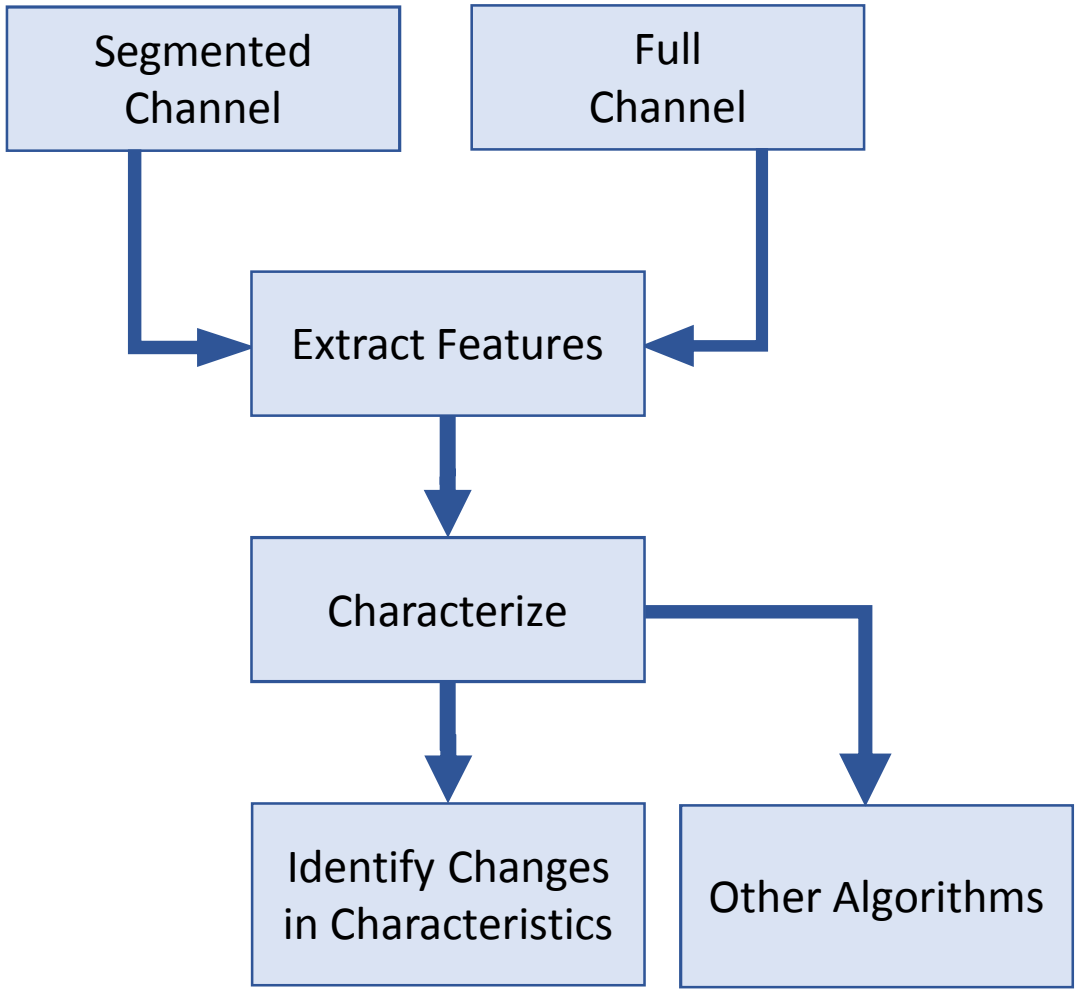
- The PSI project is an effort to explore application of existing machine learning techniques to surveillance data.
  - As a path finding activity, PSI seeks to integrate with other NW data efforts in exploring machine learning applications to the NW enterprise.
- PSI is actively exploring the problem space; all results are preliminary and are in the process of being refined.
  - Goal 1: Demonstrate prototype application of advanced analytics / data science techniques on nuclear weapons test data
  - Goal 2: Back-test analytics against previously analyzed data to gauge efficacy.
  - Goal 3: Collaborate with other efforts:
    - Stockpile Evaluation Data System (SEDS)
    - Telemetry Analysis and Visualization Suite (TAVS)
    - Cognitive Foundry (Citrus)
    - PANTHER

PSI Roadmap



- Multi-effort approach to data analysis: channel based and set based analysis
- Presentation will focus on efforts in channel segmentation and characterization realm
- Analysis is performed in black-box manner: no specialized knowledge of weapon systems or tests

Algorithmic Approach



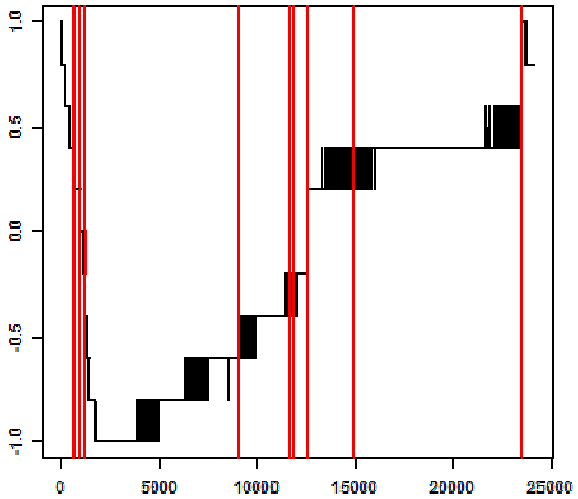
- Run clustering to determine classifications of signals
- 40% percent of clustered signals are used for classifier training and verified against remaining 60%
- Classification creates a model used across multiple tests
- Example for identifying item of interest: signal usually is classified as X, Y classification is an item of interest

Dataset Used

- NW Flight Data
  - 15 tests from one system
    - Tests vary in length
    - Tests are not time aligned (T_0 for test A ≠ T_0 for test B)
  - ~170 channels per test
    - Channels vary in bitrate and signal type (analog vs digital)
    - Channels share a T0 across the test.
    - Channels are assigned a number consistent across tests.
  - Efforts are “black-box”: No engineering units, known signal information, etc.

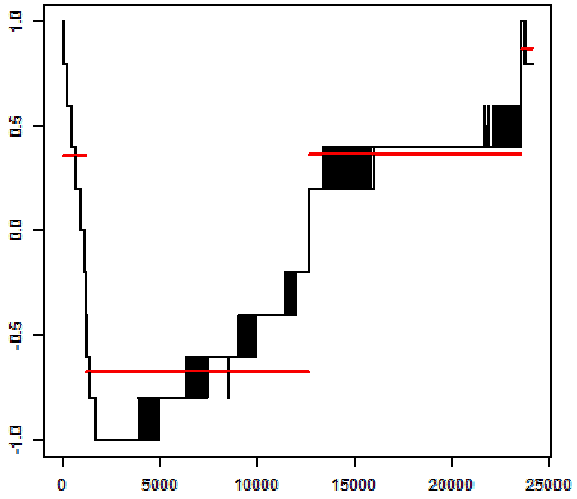
Change Point Detection

- Segmentation accomplished via change point detection.
  - Black lines represent the signal
  - Red lines represent the segmentation based on each particular method



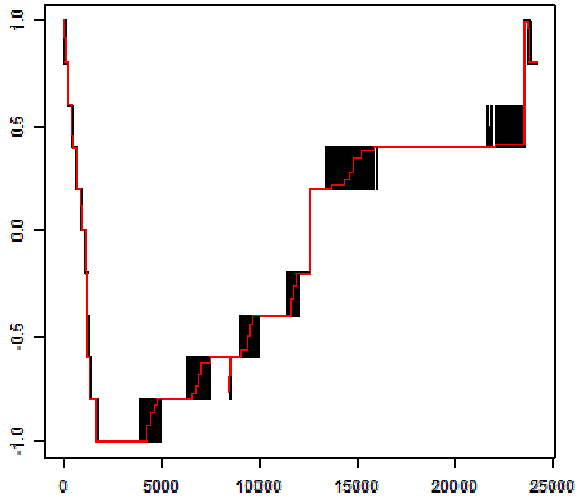
Variance Method

Red line indicates areas in the signal in which the variance is similar



Mean Method

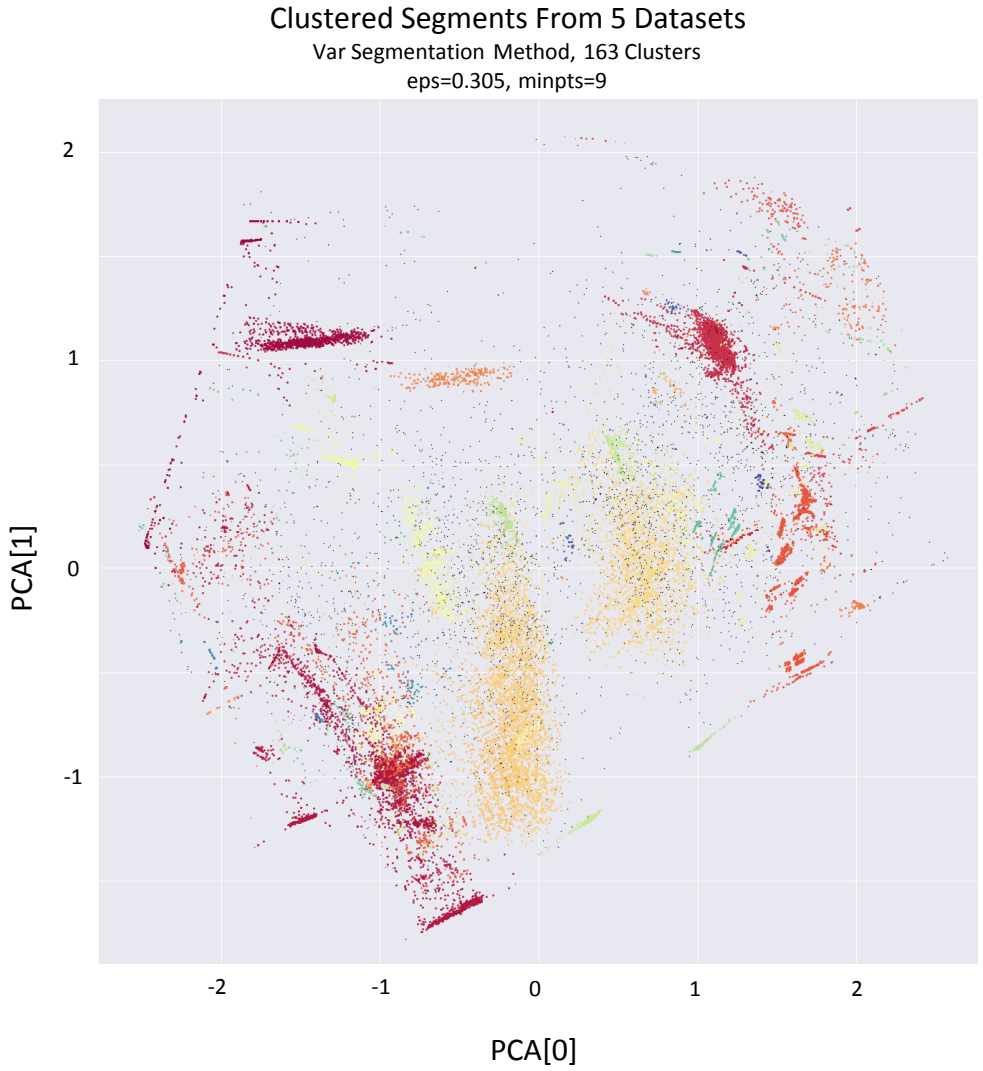
Red line in indicates areas in the signal in which the mean is similar



Wild Binary Search Method

The red line is a fitted piecewise constant function equal to the mean between change-points

Clustering & Classification

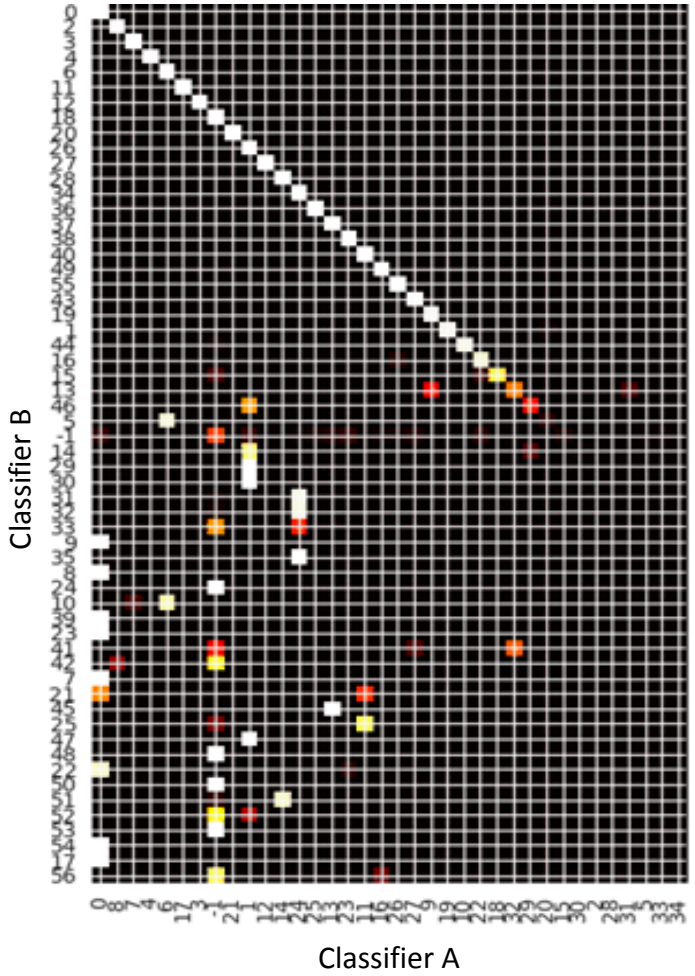


- Identify features of signals to describe the shape of the signal
- Principal component analysis to reduce dimensionality.
- Cluster & classify signals based on identified properties
- Representative signal set has features extracted and feature sets are clustered using DBSCAN
- Voting classifier determines best classifier (SVC, K-neighbors, or Decision Tree) for dataset to maximize classifier accuracy

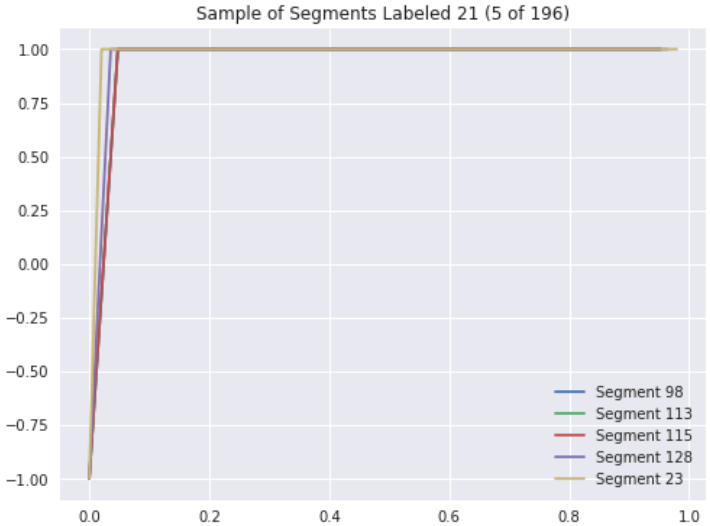
Results and Tuning

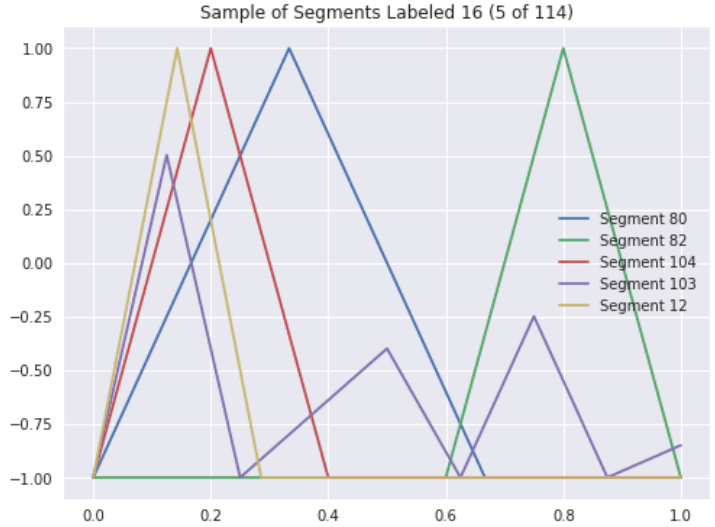
- Currently comparing how clusters relate across training datasets
- Right: comparing labels applied by two different classifiers to dataset
  - White is more similar.
  - Read as “segments labeled 32 by classifier A are labeled 13 and 41 by classifier B”
- Several parameters control output of PCA & DBSCAN
  - Number of PCA components used for clustering
  - Minimum Points, Distance parameter (ϵ)
- Refining parameters improves clustering efficacy and better understanding of clustering results

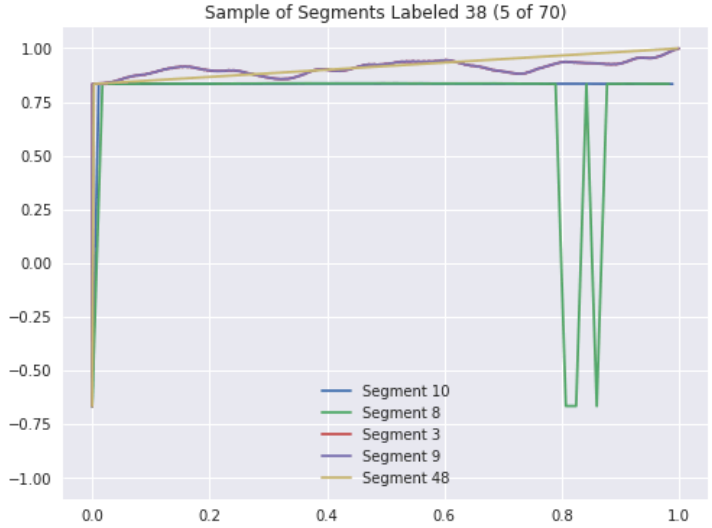
Mapped Comparison Between Two Classifiers

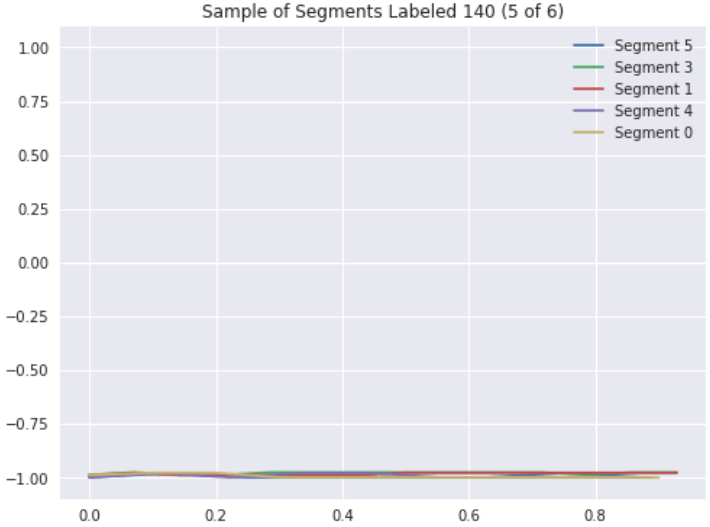


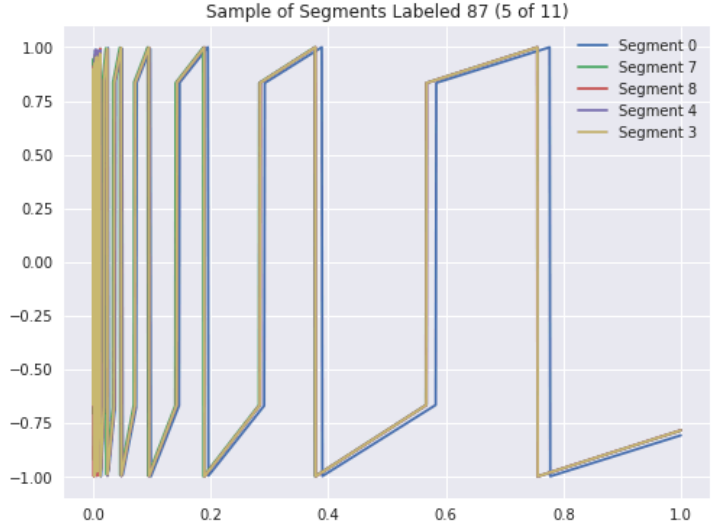
Classification Output Samples

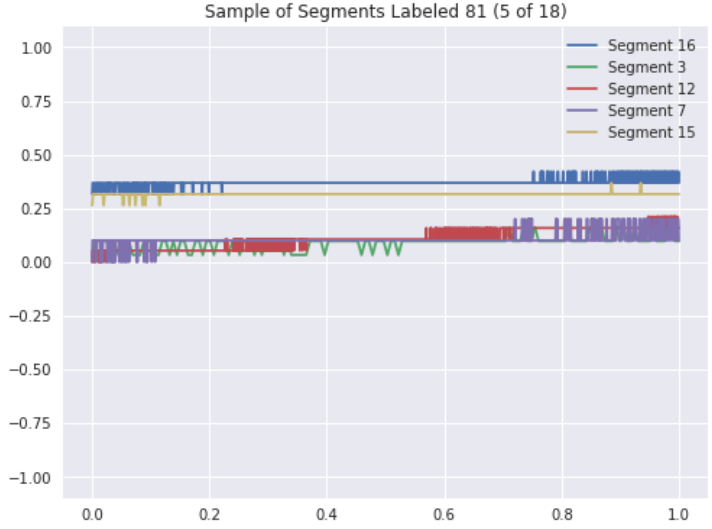












Future Objectives

- More & Different Types Of Data
  - Laboratory
  - Development
  - Different Systems
- Integrating Channel Relationships
  - Identify immediate and time delayed relationships between channels.
  - Develop models of expected channel interactions
- Integration of project as full pipeline to integrate it into surveillance business practices
  - Upload -> Automated Analysis -> Flagged Items of Interest

Thank You

- PSI is an exploratory effort to improve the efficacy of NW surveillance using black-box machine learning techniques.
- Multiple approaches to identify items of interest in test data.
- Segmentation and classification of segments of signals

- We are very open to recommendations and ideas.