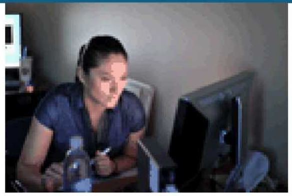
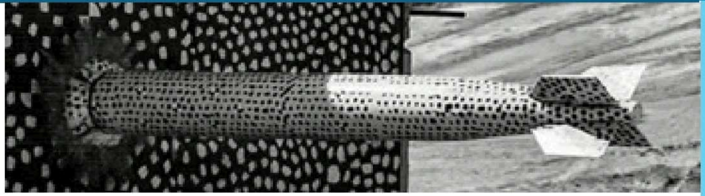
SAND2019-9217C

RF Feature Extraction using Wireless Protocol Informatics, Pattern Matching, and Natural Language Processing





PRESENTED BY

Nicholas Jacobs

Team: Chris Howerter (PI), Jenny Galasso, Nicholas Jacobs,
Tim Draelos, Christine Lai, Amanda Gonzales, Deepu Jose

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & 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.

Outline

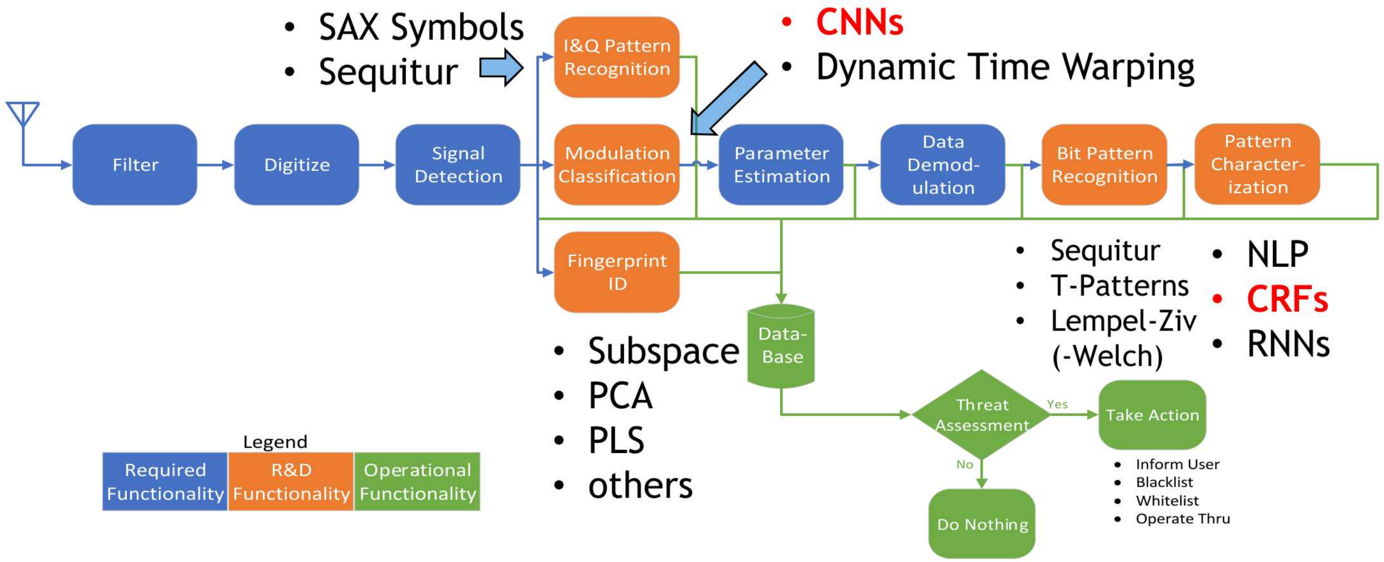
- Problem Description
- Modulation Classification
  - Convolutional Neural Network (CNN)
  - Data Generation and Format
  - Results
- Pattern & Behavior Characterization
  - Data Format
  - Conditional Random Fields (CRFs)

Problem you are trying to solve

With the proliferation of nations using space to extend their footprint of influence, the US Government has a need to monitor space assets to understand the related communications behavior. The focus of this work is to characterize and identify signals from various emitters that are either space or ground-based.

In this presentation, we will discuss two components to this research:

- Modulation Classification using CNN's
- Sequence Learning / Behavior Characterization using CRF's

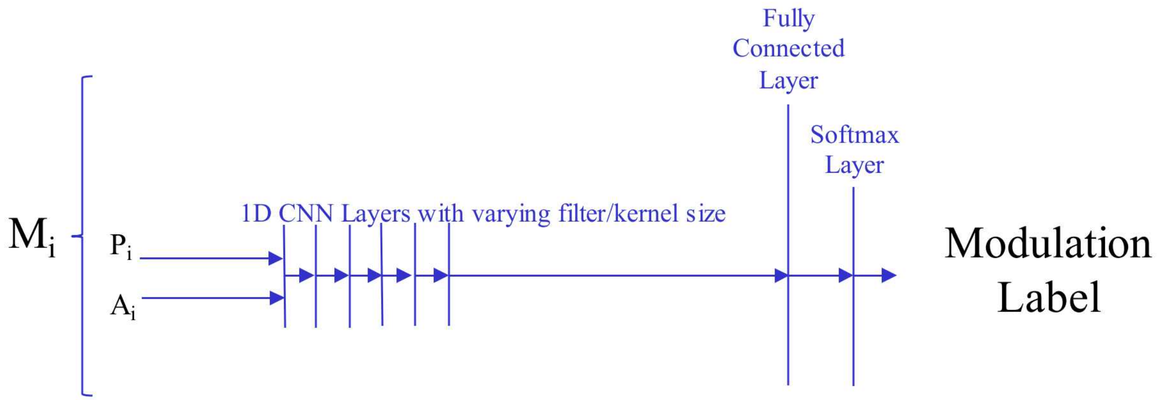


Modulation Classification

- Upon detecting a signal burst, the first step to adequately determining and characterizing signal type and move closer to finding any patterns in data is to determine the modulation scheme
  - Frequency Shift Keying (FSK)
  - Phase Shift Keying (PSK)
  - Quadrature Amplitude Modulation (QAM)
- With successful classification of modulation type, we can attempt to demodulate any data being transmitted and continue with identifying and characterizing any patterns in the data itself
- This builds on work completed in the Blind Signal Characterization (BSC) LDRD (2015)

Modulation Classification

- Semi-supervised ML with regularization via Mean-Teacher
- 6-layer CNN (7/2-kernel)
  - Conv/BN/MaxPool/Relu
  - Filters 7x1x32 -> 2x1x64 -> 2x1x128 -> 2x1x256 -> 2x1x512 -> 2x1x1024
  - AvgPool(8) -> FC (1024->15)

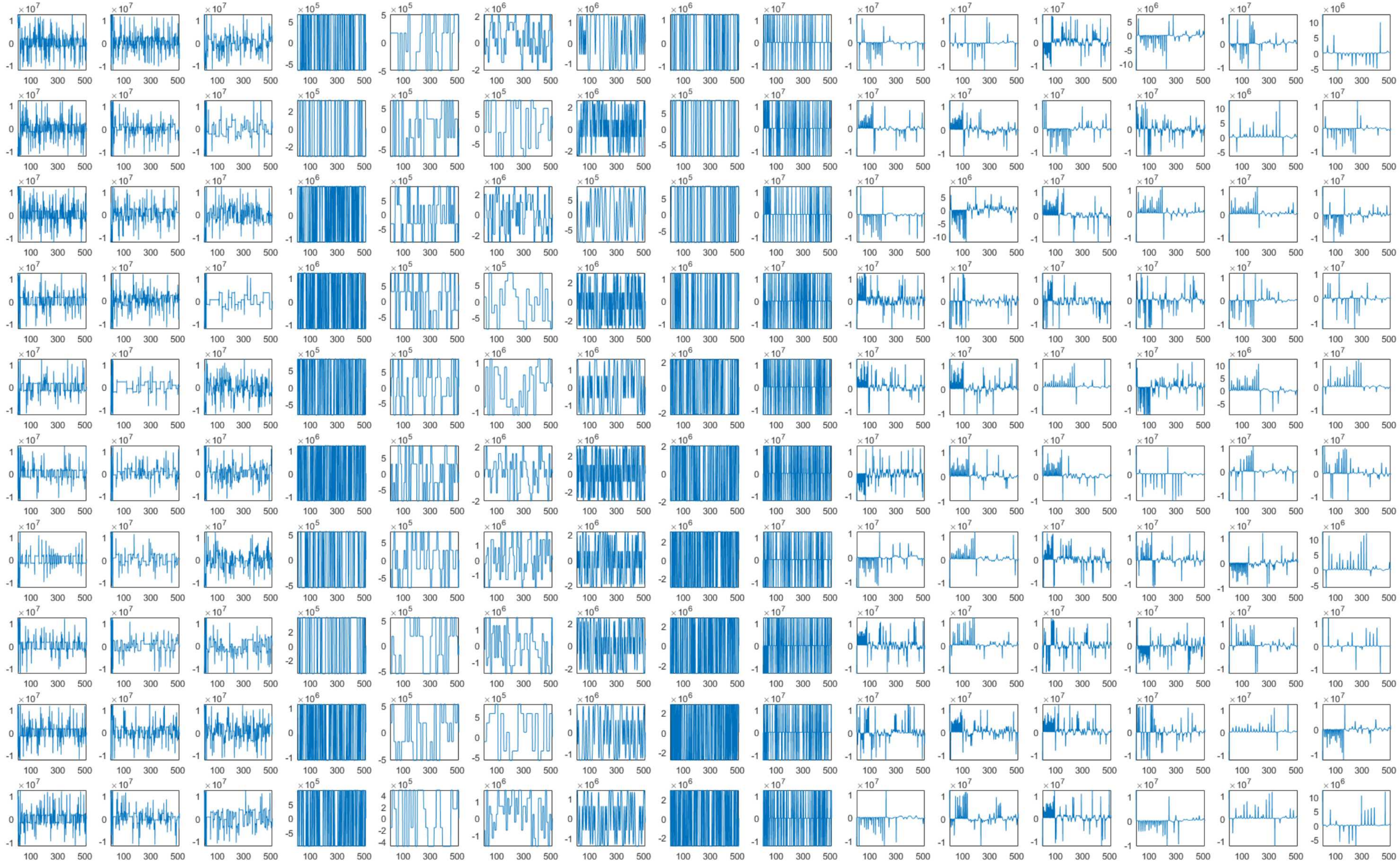


Network to classify a modulation example, M_i
Pwelch signal input, P_i
Angle signal input, A_i

Modulation Classification

- 14 modulation types, each with constant baud rate, frequency, and SNR split into FSK, PSK, and QAM
  - Modulation breakdown for 3-cat initial classifier
    - Frequency-Shift Keying (FSK), Phase-Shift Keying (PSK), Quadrature Amplitude Modulation (QAM)
  - Modulation breakdown for 2nd level training
    - FSK – MSKs vs FSK vs CPM/2CPFSK
    - PSK – BPSK vs QPSK vs 8PSK vs OQPSK
    - QAM - 16QAM vs 64QAM vs 256QAM
- Power, Angle, Frequency, and magnitude values used for input to CNN
- 20% held out test set and validation set each

Example Modulation plots



Modulation Classification Results

3-category top level classifier

- 99.7% accuracy
  - FSK vs. PSK vs. QAM

Level 2 FSK Classifier

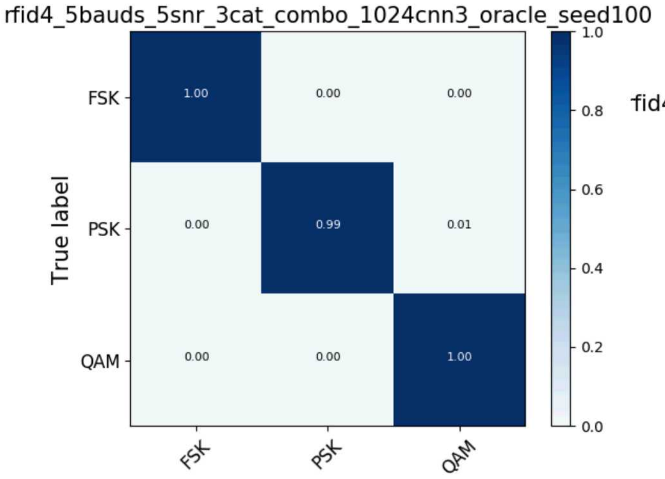
- 96.9% accuracy

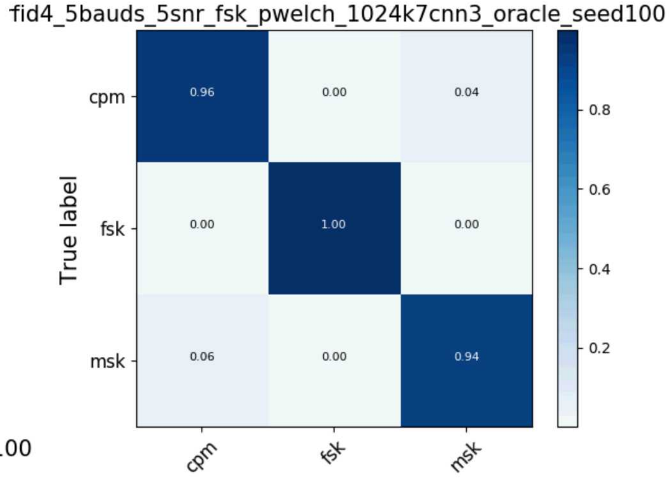
Level 2 PSK Classifier

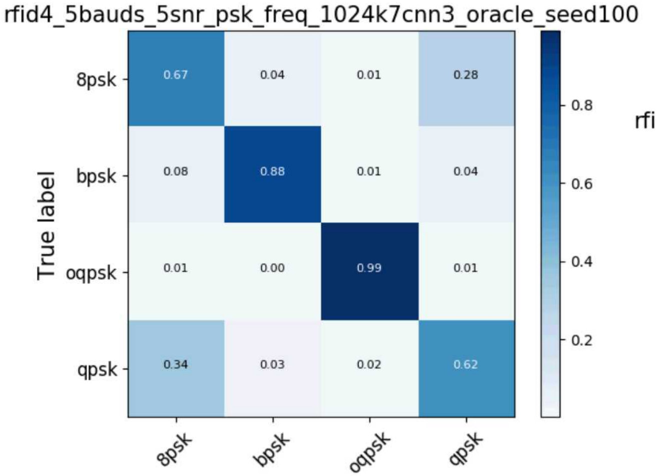
- 78.9% accuracy

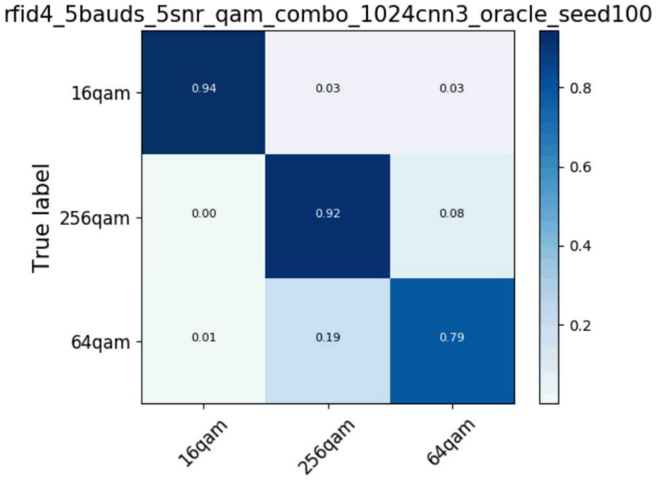
Level 2 QAM Classifier

- 88.5% accuracy





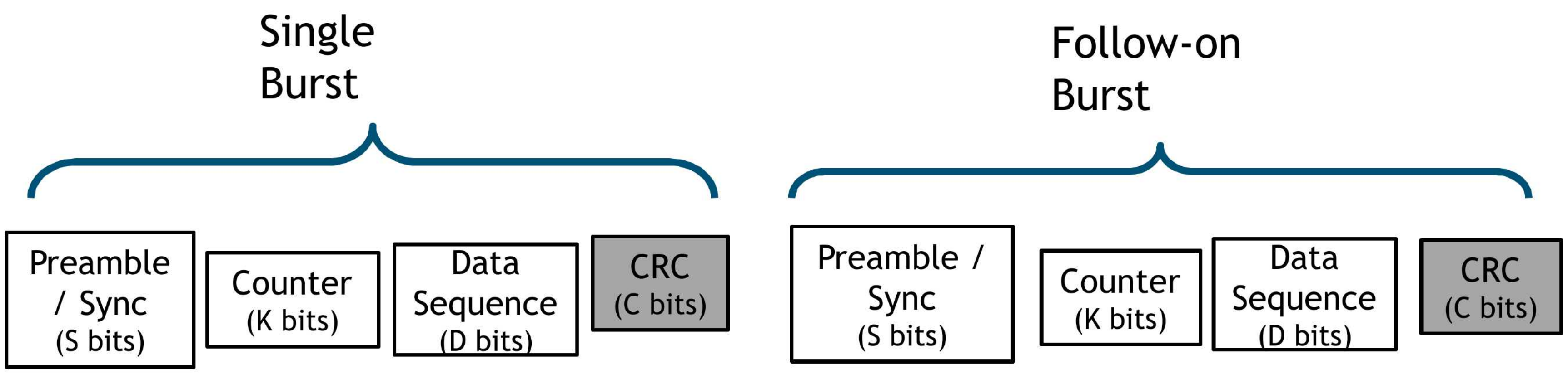




Pattern / Behavior Characterization

Data used: Synthetic Data

- Simple synthetic data developed for MLSPPI ExEx LDRD (completed last year) that generates signal bursts representative of RF signal traffic and protocols
- Contains several known patterns, such as pre-determined preambles, sync patterns, etc… that we can detect, find, and characterize in the symbol sequence of the decoded traffic



Pattern / Behavior Characterization

- Leverages prior work from the Wireless Protocol Informatics (WPI) LDRD (2014), the MLSPPI ExEx LDRD (2018), and from several other related projects to identify raw patterns in the demodulated data stream (Sequitur, T-Patterns, Word2Vec)
  - Sequitur is a compression algorithm, used to find recurring patterns. T-Patterns are patterns with a statistically significant temporal dependence, and Word2Vec groups patterns with similar contexts
  - These can now be used as features for sequence learning and behavior characterization

- This research is working on taking this a step further and learning underlying latent structure in the communications using Conditional Random Fields (CRFs)

Pattern / Behavior Characterization

- The latent structure in a CRF can be used to learn categories for identified patterns (preambles vs sync vs crc, etc…), as well as more abstract higher-level behaviors like activities
- CRFs are discriminative models that use conditional probabilities for activations and can utilize a general graphical structure for latent variables.

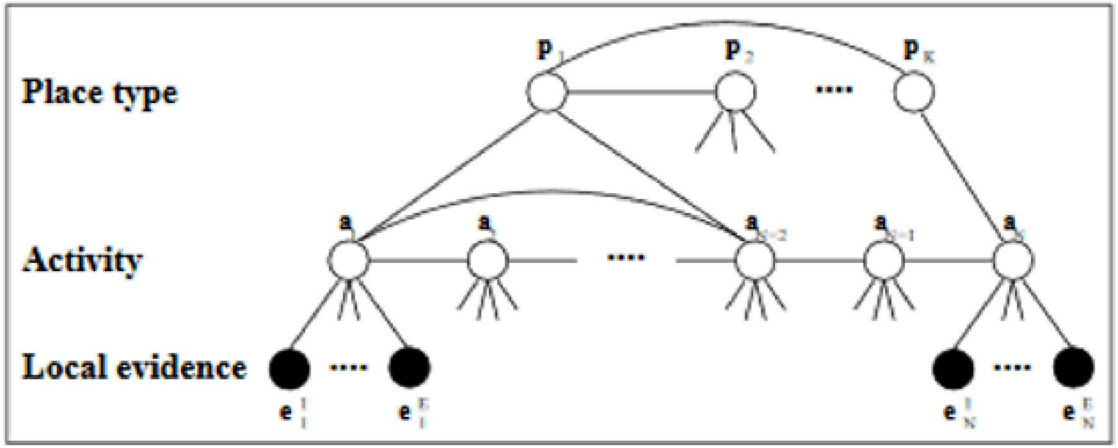
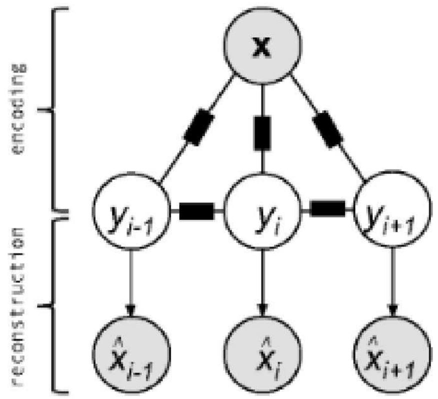


Image source: http://robots.stanford.edu/isrr-papers/final/final-42.pdf

CRF Autoencoders

- Existing literature has looked at using CRF Autoencoders for unsupervised learning of structure (Ammar, NIPS 2014)
- This takes the CRF and adds a reconstruction layer to determine an estimate of observation using conditional probabilities dependent on latent activations

Utilizes gradient-based learning (for encoding) and Expectation-Maximization (for reconstruction), switching between the two every 1-2 steps. See image for a simple version of this structure shown as a factor graph, taken from (Ammar, NIPS 2014)



- This work is still ongoing

Source: http://www.cs.cmu.edu/~wammar/pubs/nips2014.pdf

Conclusion

- Every communication system, unidirectional or bidirectional, uses a predefined protocol to describe the signaling scheme, message construction, and encoding.
- These common protocols and modulation schemes create patterns that can be used to classify and characterize which protocols are used, as well as behaviors of interest. Doing this successfully helps in identifying and characterizing transmitting participants in a digital conversation, and allows sufficient information to be captured for further data demodulation and study while massively reducing data requirements.

Thank You!

Contact Info:

Nicholas Jacobs

njacobs@sandia.gov