SAND2019-12604PE

Multiple Instance Learning: Application to Seismic Event Detection and Open Problems

Sandia National Laboratories
Livermore, CA

Fulton Wang (fulwang@sandia.gov)
Ali Pinar (apinar@sandia.gov)
Jaideep Ray (jairay@sandia.gov)

Unclassified

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.

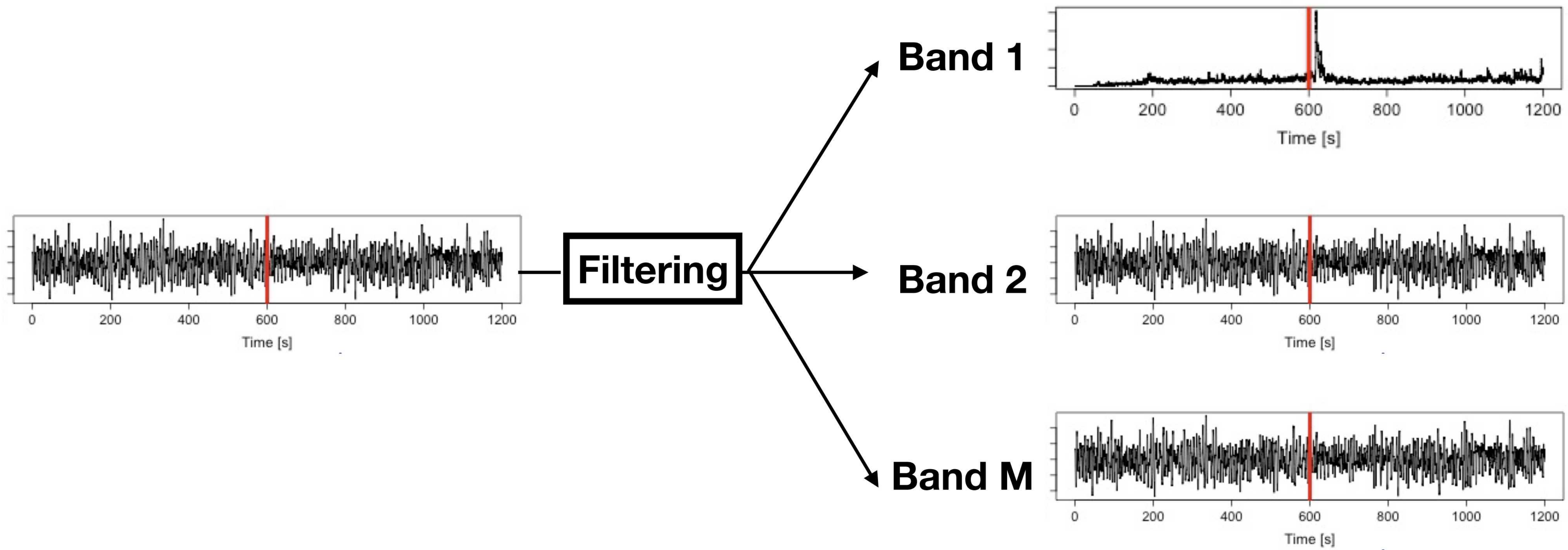
Background

- There exists a network of 120-odd seismic stations to detect nuclear blasts.
- Called International Monitoring System
- Earthquakes, mine blasts, and nuclear blasts give rise to seismic waves
- These travel through the earth and are detected at the IMS stations.
- Most of the detections are earthquakes, followed by mining blasts. ~100 a day per station.
- Being able to detect these seismic wave arrivals quickly with minimum manual intervention is critical.
- Detecting weak seismic waves / faint arrivals is tough.

Goal of this work

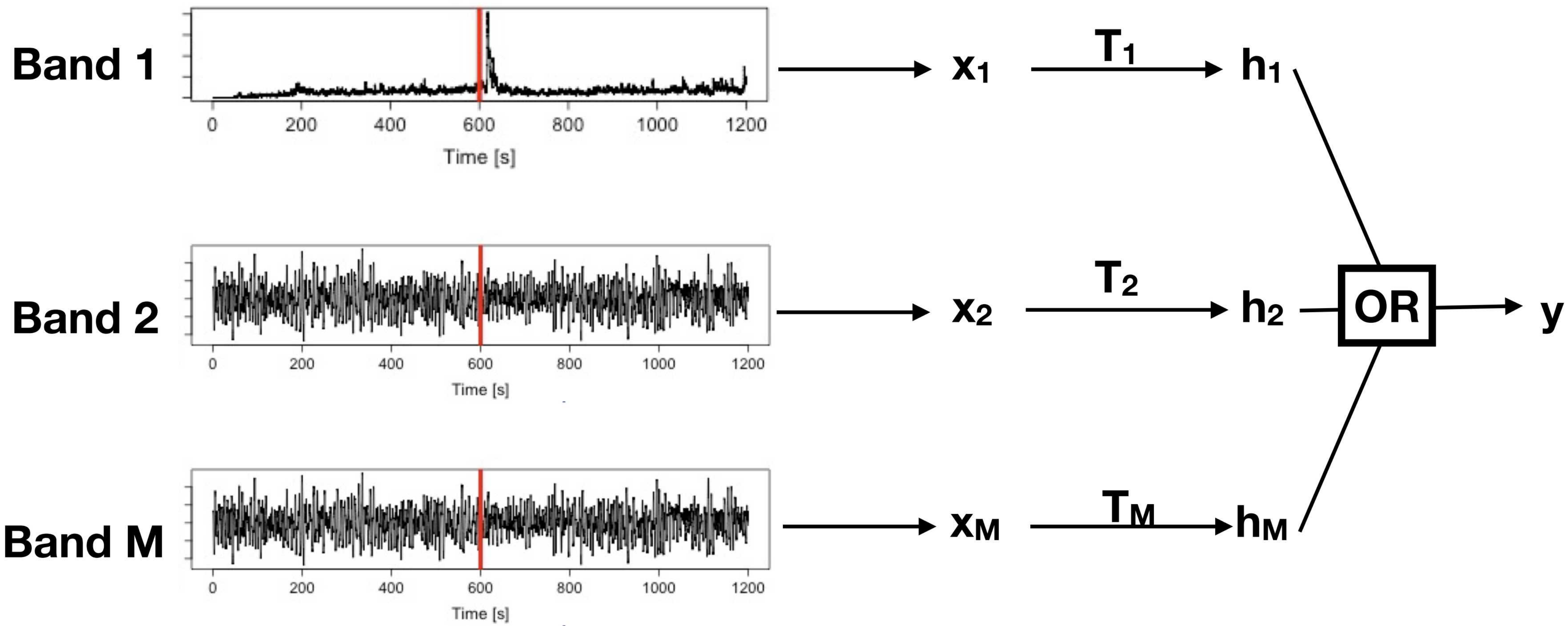
- Seismic waves currently detected using manually chosen rules of thumb.
- Want to replace rules of thumb with classifier learned from data.
- Data has shortcomings that require use of ML method called multiple instance learning.

This is a time series classification problem



- Need to filter raw signal to the frequency band in which event occurs in order to detect it.

Detection is currently done without ML using rules of thumb

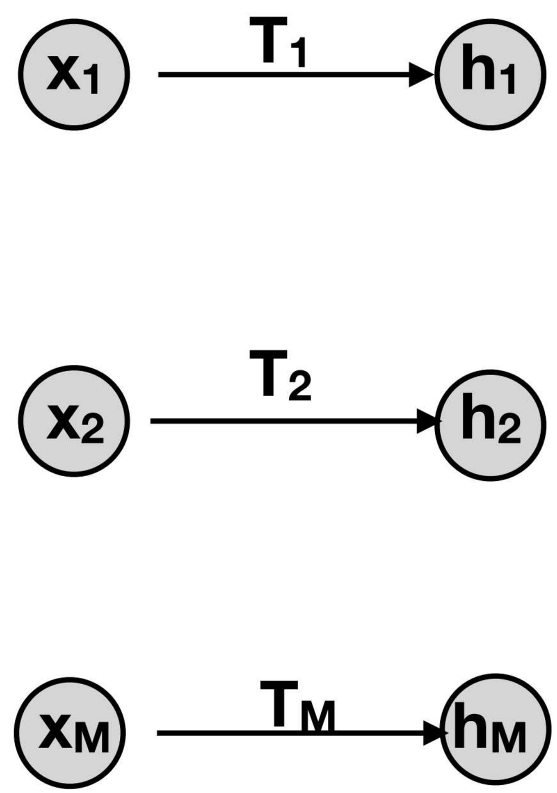


- Rule of thumbs $T_1 \dots T_M$ currently chosen manually!
- $x_i = (x_{i1}, x_{i2}, \dots, x_{iD})$
- $T_i(x_i) = 1$ if $x_{i1} > \theta_{i1}$ or $x_{i2} > \theta_{i2}$ … or $x_{iD} > \theta_{iD}$. M x D thresholds

There are a lot of features

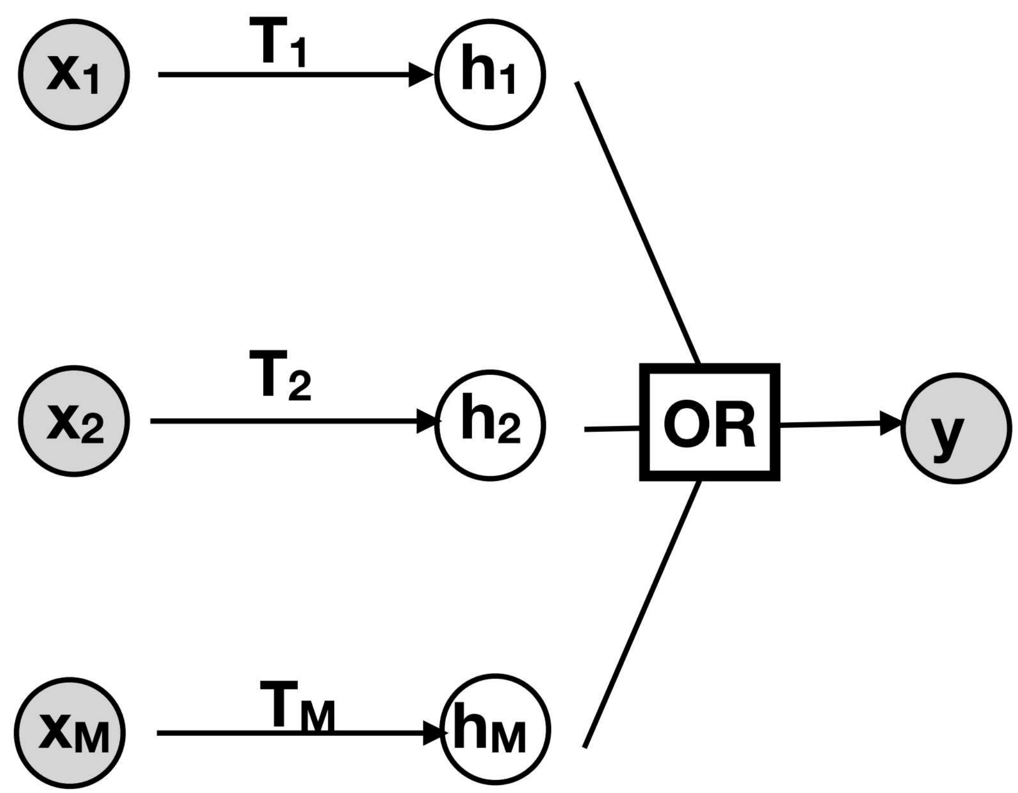
- The filtered signal has 3 channels
- For each channel, apply different transforms (LTA/STA, recursive LTA_STA), and features are the maximum value of those features in the time series.
- > 6 features
- More if we consider the 95-th percentile of a transform, 50th percentile, etc.
- This leads to a high false positive rate.

How to learn rule of thumbs in ideal scenario?



- Fit $T_i()$ using many samples of (x_i,h_i). Easy.

But we have a data problem

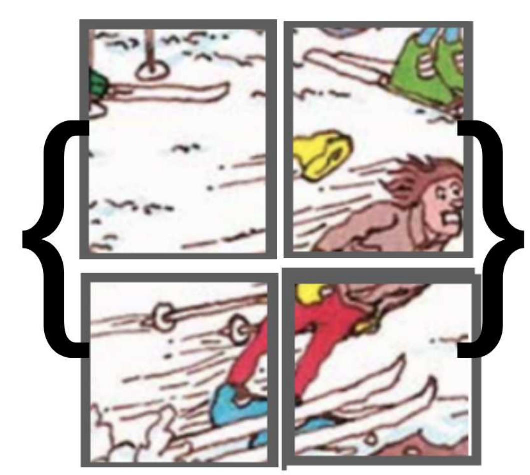


- No one bothered recording the frequency band in which past events occurred.
- Can't impute h_i. Events occur in different frequency bands.
- Need Multiple Instance Learning

What is the Multiple Instance Learning Scenario?

- Training Data is Weakly Labelled
- Consists of bags of instances
- Unobserved: instance labels
- Observed: whether each bag has some positive instance

Contains a Waldo instance



Has no Waldo instance

Sandia National Laboratories

What are the goals in MIL?

- Learn instance classifier

Is  a Waldo instance?

- Learn bag classifier

Does {  } have a Waldo instance?

- Do both: **interpretability**

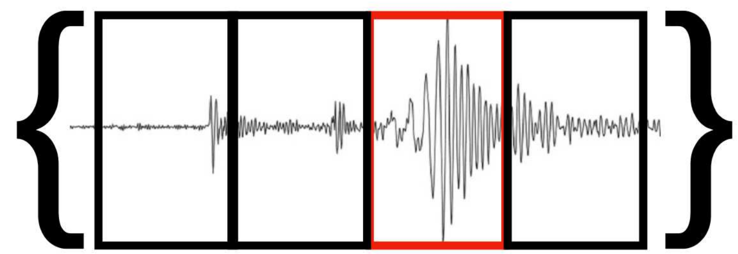
Does {  } have a Waldo instance?

If yes, where is Waldo?

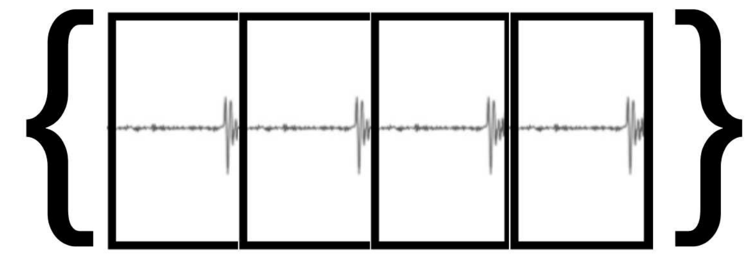
Sandia National Laboratories

SPAN Application: Seismic Event Detection

- Training data:

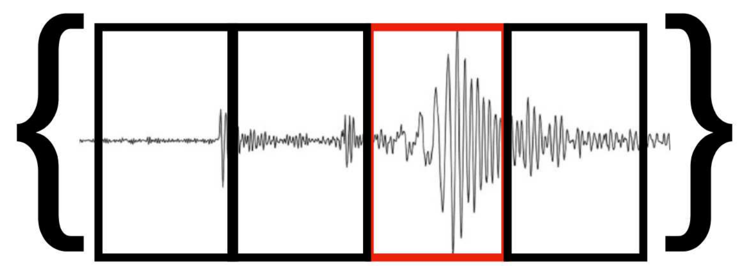


Contains frequency with event



Contains no frequency with event

- Goal:

Does  contain frequency w/ event?

If yes, which frequency is the event?

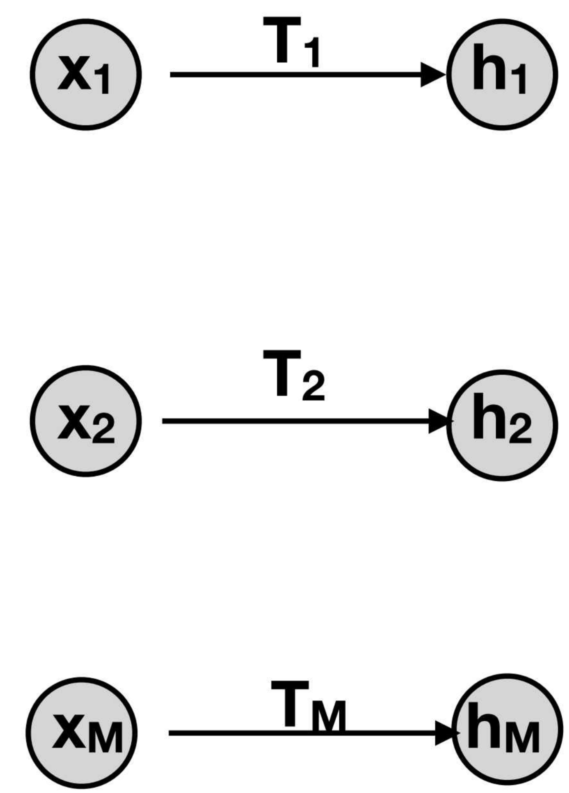
- Knowing frequency band would inform the event type

Sandia National Laboratories

How to model the ideal scenario?

- Use probabilistic model

- Variables for each bag:



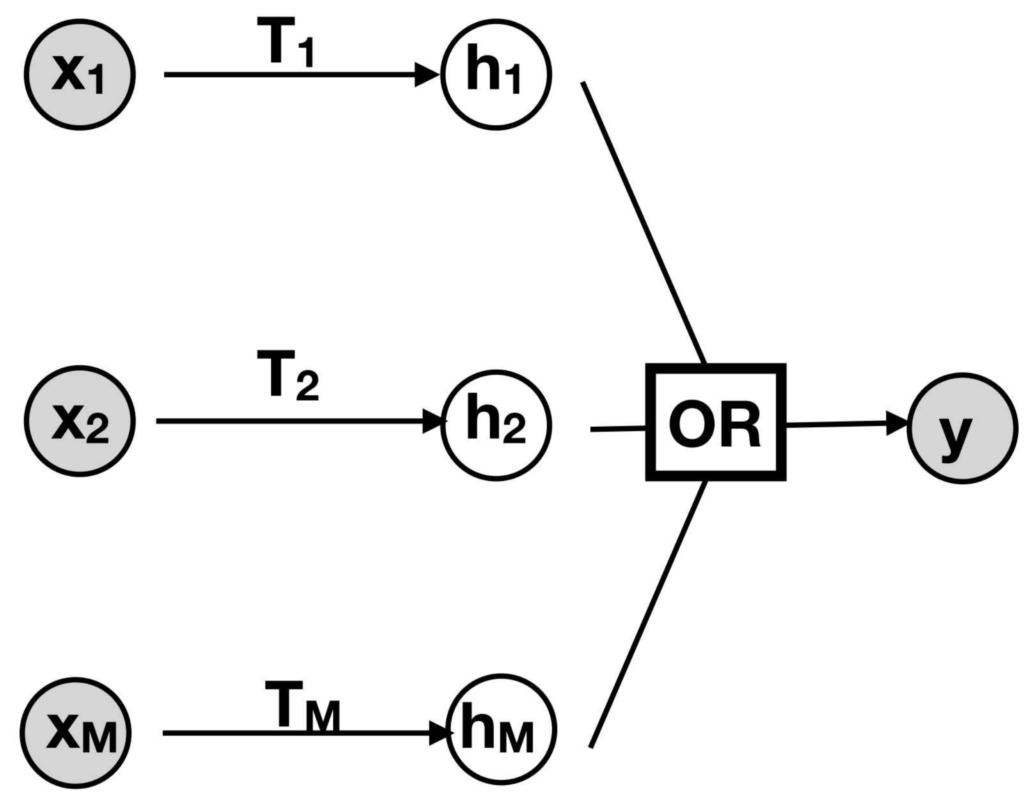
  - Instance feature vectors $x_1..x_M$

  - instance labels $h_1..h_M$ (0/1)

- Learn probabilistic rule of thumb: $T_i(x_i) = P(h_i=1|x_i)$

How to do multiple instance learning?

- Use probabilistic **latent** variable model
- Variables for each bag:
  - Bag label y (0/1)
  - Instance feature vectors $x_1..x_M$
  - Latent instance labels $h_1..h_M$ (0/1)



- Multiple instance assumption: $y = 1$ if some $h_i = 1$
- Learn probabilistic rule of thumb: $T_i(x_i) = P(h_i{=}1|x_i)$

Options for rules of thumb

- Logistic regression "MI-Logreg"
    - $P(h_i=1|x_i;B_i) = \text{sigmoid}(B_i^T x_i)$
- Threshold function (similar to heuristic) "MI-Thresh"
    - $P(h_i=1|x_i; \theta_i) = 1$ if $x_{i1} > \theta_{i1}$ or $x_{i2} > \theta_{i2}$ … or $x_{iD} > \theta_{iD}$.

Data Preparation

- Available data: raw signal from seismometers and timestamps of known seismic events
- To generate bags: ~ raw signal divided into 20 minute windows. Windows that contain a known seismic event are positive bags.
- To generate instances: each 20 minute raw signal (bag) decomposed into contributions from M non-overlapping frequency bands to obtain M instances per bag.

Sandia National Laboratories

Experiment #1: Overview

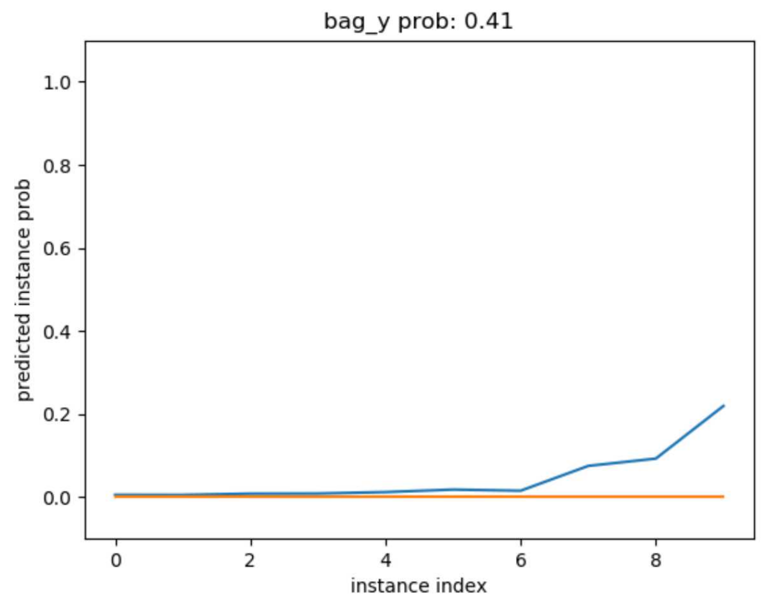
- Goal: Compare AUC’s of MI-Logreg and vanilla classifiers
- Vanilla classifiers directly classify a bag by representing it as the concatenation of the feature vectors of the instances the bag contains
- Features: for each channel, (100, 95)-th percentiles of lta/sta, 100-th percentile of sta
- Frequency bands: 10 bands spanning 0-5 hz.

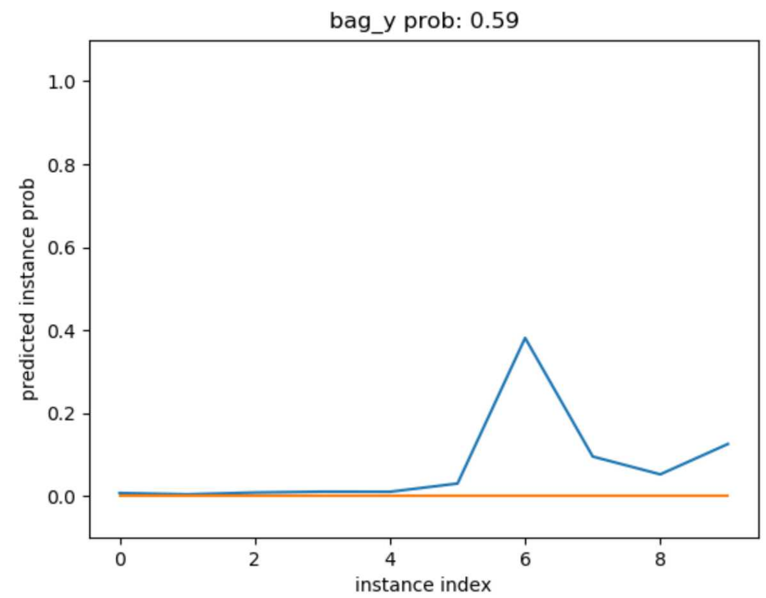
Experiment #1: Results

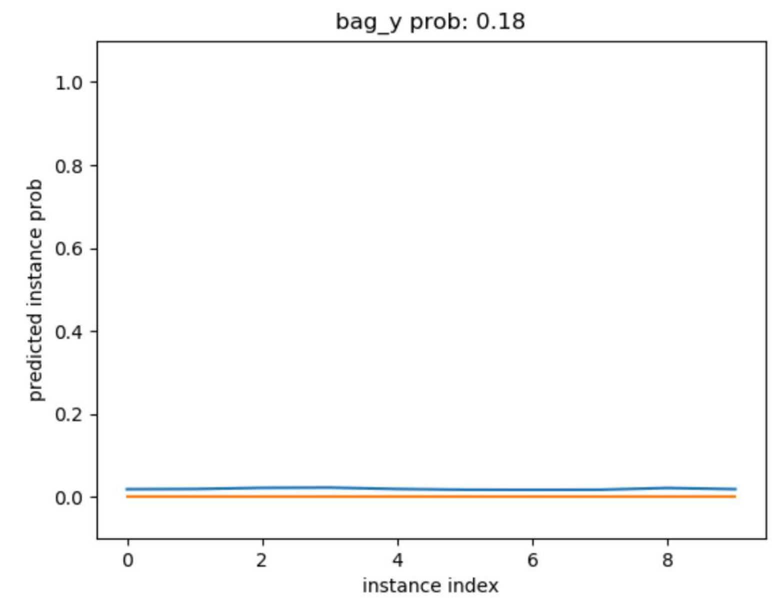
	AUC
MI-Logreg	.77(.01)
Logreg	.71(.01)
RF	.77(.01)

- Logreg and RF (random forest) are vanilla bag classifiers
- No instance labels, so all metrics are bag-level.
- MI-Logreg higher AUC than Logreg (better model, less parameters)
- RF higher AUC than Logreg.

Experiment #1: Multi-instance learning is interpretable by design







Instances ordered by frequency on x-axis.

Blue line indicates probability a frequency contains event $P(h_i=1|x_i)$

Orange line is 0/1 prediction of h_i

Sandia National Laboratories

Experiment #2: Overview

- Goal: How does learning thresholds in threshold instance classifier improve performance compared to not learning them?
- Features: for “merged” channel, 100-th percentile of lta/sta, 100-th percentile of recursive lta/sta
- Frequency bands: 2 bands spanning 0-4 hz.
- Custom loss to minimize: $FNR^2 + FPR^2 + (FNR - FPR)^2$
  - Want roughly equal FPR and FPR

Experiment #2: Results

	FNR	FPR
MI-thresh	27.5%	20%
Naive	43%	4%

- “Naive” method: set θ_{ij} equal to the 95-th percentile of x_{ij} in negative bags (raw time series without event).
- “Naive” method does not balance FPR and FNR.

Experiment #2: Results

	FNR	FPR
LTA/STA, 0-2 hz	35%	9%
LTA/STA, 2-4 hz	53%	7%
Recursive LTA/STA, 0-2 hz	33.5%	13%
Recursive LTA/STA, 2-4 hz	53%	8%
MI-Thresh	27%	20%

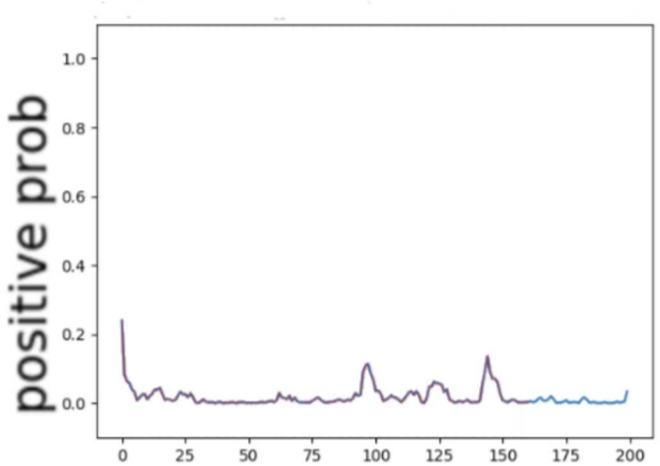
- Trained “MI-Thresh” classifier can be interpreted as “fusing” M x D separate single-feature classifiers.

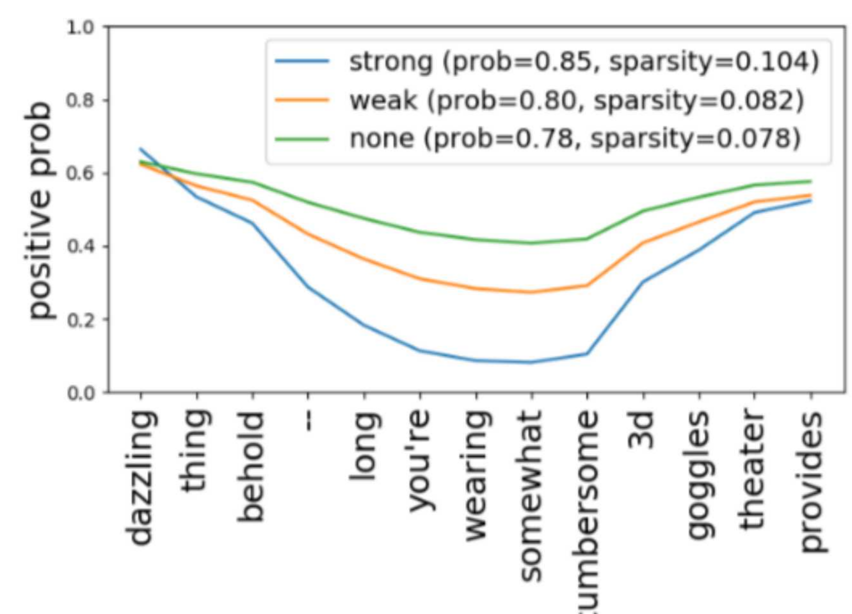
- How would those M x D single-feature classifiers perform?

Challenge: Handling other kinds of data types

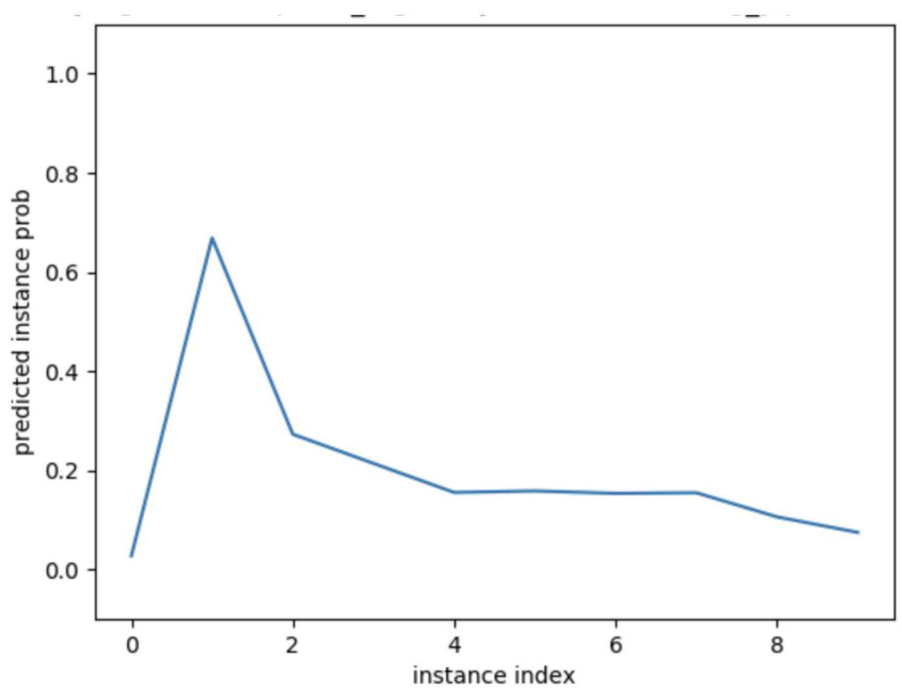
- Nonproliferation data includes structured data.

- Solution: model $P(h_1..h_M \mid x_1..x_M)$ jointly with a conditional random field instead of $P(h_i=1|x_i)$ independently using logistic regression

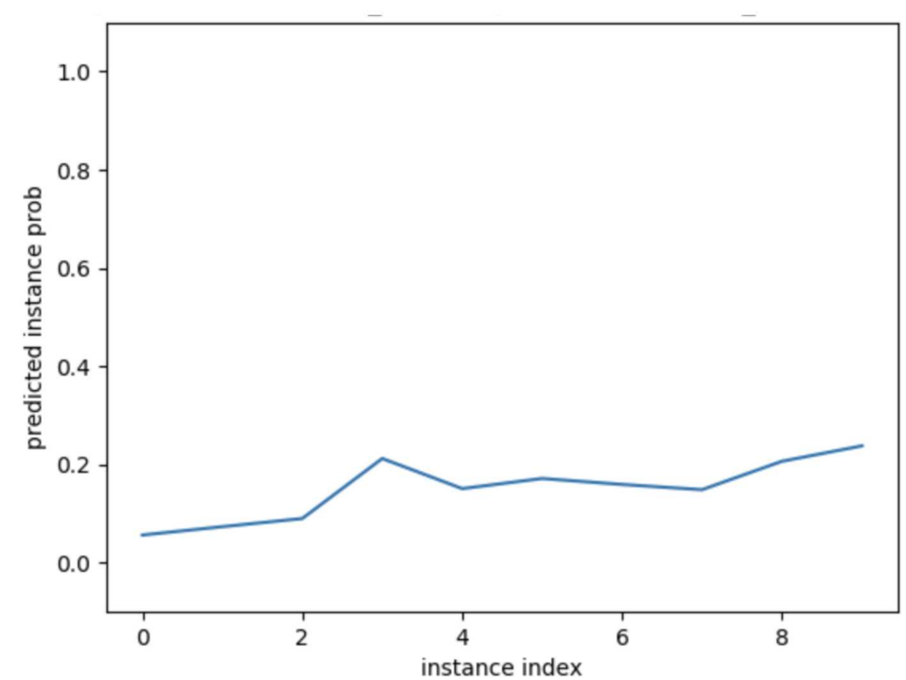




Challenge: instance predictions not always interpretable



Interpretable



Not Interpretable

- Solution: encourage instance prediction to be high for a small number of instances via a regularizer

Sandia National Laboratories

Challenge: accounting for uncertainty

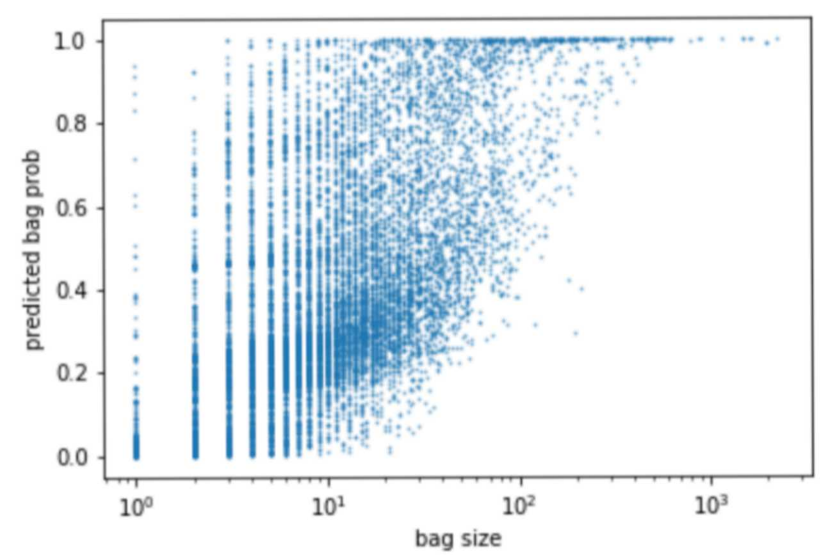
- Weak labels leave room for ambiguity, so that many models (and instance predictions) are plausible.

- Solution: need Bayesian multiple instance models. Currently developing Gaussian Process model which in addition to providing uncertainty estimates, models dependencies between instance labels.

Sandia National Laboratories

Challenge: bag size variability

- You do not get to control the size of bags. There is variability in sentence length, width of time windows in which you know an event occurred.



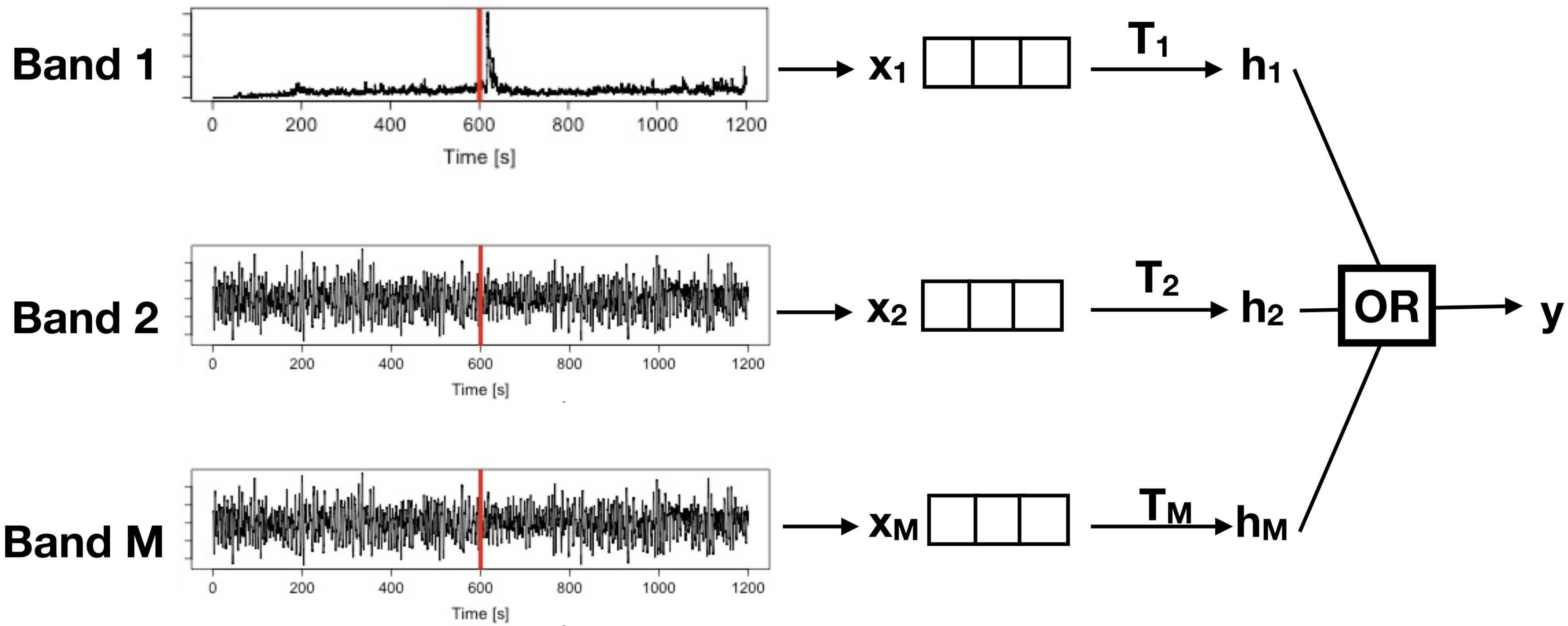
Sandia National Laboratories

Conclusion

- Multiple instance learning can be used to predict the frequency in which seismic events occur despite the training data not containing such information.
- Event detection performance on par with less interpretable random forest
- Multiple instance learning can be useful in many other SPAN applications, with many open challenges.

Sandia National Laboratories

Detection is currently done without ML using rules of thumb



- Rule of thumbs $T_1 \ldots T_M$ currently chosen manually!
- $T_i(\boxed{x_{i1}}\boxed{x_{i2}}\boxed{x_{iD}}) = 1$ if $x_{i1} > \theta_{i1}$ or $x_{i2} > \theta_{i2} \ldots$ or $x_{iD} > \theta_{iD}$
- M x D different thresholds to choose manually.

Ideal SPAN Scenario is not possible

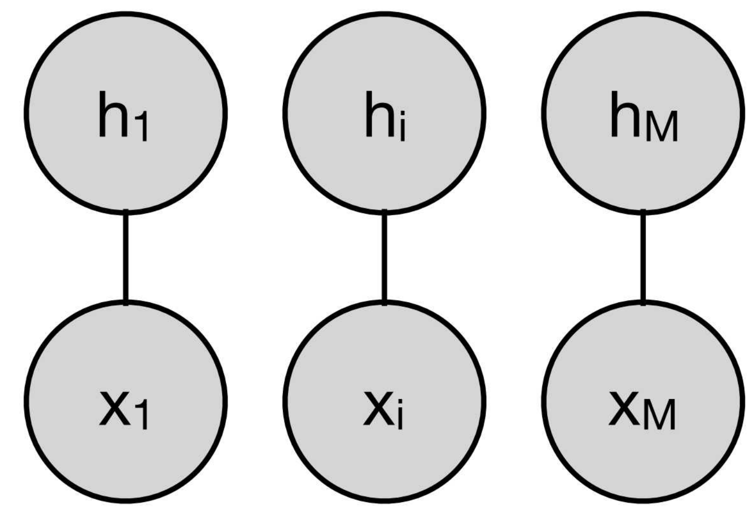
- Observed: instance labels

- Not possible - only observe event timestamps. Need multiple instance learning.

Sandia National Laboratories

How to do model the ideal scenario?

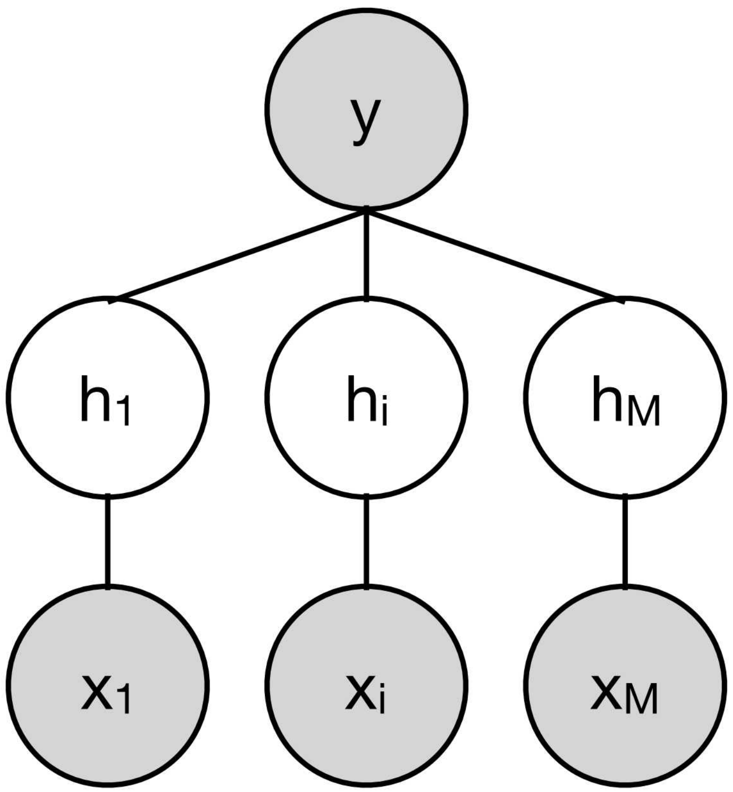
- Use probabilistic model
- Variables for each bag:
  - Instance feature vectors $x_1..x_M$
  - instance labels $h_1..h_M$ (0/1)



- Have probabilistic rule of thumb:
  - $T_i(x_i) = P(h_i=1|x_i)$

How to do multiple instance learning?

- Use probabilistic **latent** variable model
- Variables for each bag:
  - Bag label y (0/1)
  - Instance feature vectors $x_1..x_M$
  - Latent instance labels $h_1..h_M$ (0/1)



- Multiple instance assumption: $y = 1$ if some $h_i = 1$
- Have probabilistic rule of thumb:
  - $T_i(x_i) = P(h_i=1|x_i)$