



In-situ Training and Integration of New Sensors for a Sensor-net in variable operational environments: Machine Learning Advancements

DDDAS Conference, MIT, Cambridge, Ma., 7-9, Aug 2017

Dr. Shashi Phoha; PI
Graduate Faculty of EE, and
Director, Information Sciences and Technology
Center
Penn State University/ARL
sxp26@psu.edu; 814 863-8005
<http://www.personal.psu.edu/sxp26>

Student Contributors:
Michael Hauser (presenter)
Samer Saab Jr.
Nurali Virani



OUTLINE

- **DDDAS Motivations**
 - Dynamically incorporate data into an executing application
 - Dynamically steer the measurement process
- **Theoretical Framework for In-situ Learning**
 - New Sensor: Situational Model Prediction and Learning
- **Situational Machine Learning Advancements**
 - Extracting actionable intelligence from other sensors
 - Probabilistic Forecasting via deep neural networks
 - Sensor Model Adaptation through density estimation
- **Simulation Based Validation**
- **Field Experiments and Validation: Border Control Application**
 - New Seismic sensor learning from other seismic sensors in the network
 - New Seismic sensor learning from advanced sensors such as camera on a UAV



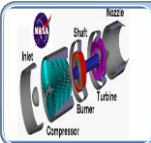
DDDAS: Dynamically steering measurement processes



Intelligent Transportation Systems

Sensors: Camera, LIDAR, IMU, GPS

- Task: Localization, navigation, safety.



Health Monitoring of Electro-mechanical Systems

Sensors: Pressure, temperature, speed

- Task: Fault detection and diagnosis



Area surveillance; Robocop:

- Sensors: Seismic, acoustic, PIR, camera, radar
- Task: Detect, classify, and track movement



Remote Patient Monitoring System

- Heart rate, respiration, muscle activity, blood pressure
- Task: Anomaly detection, Data logging

Social Media



Social Media

- User Communities
- Strength of communications among users
- Task: Community Detection, Marketing Intelligence, etc.

Data-driven Dynamic System Characterization

- **Multiple components interact over a sensing infrastructure**, often mobile, to perceive the evolution of physical dynamic processes,

Sensing Infrastructure

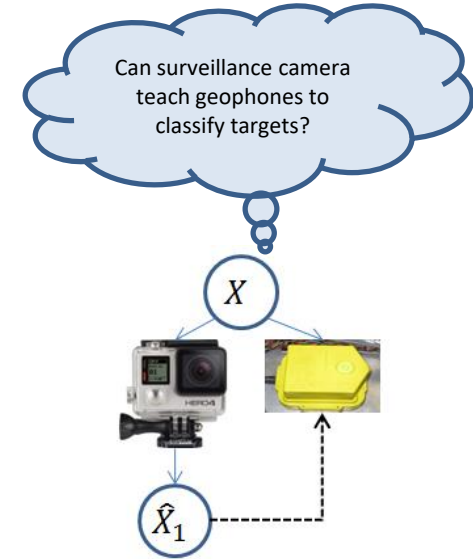
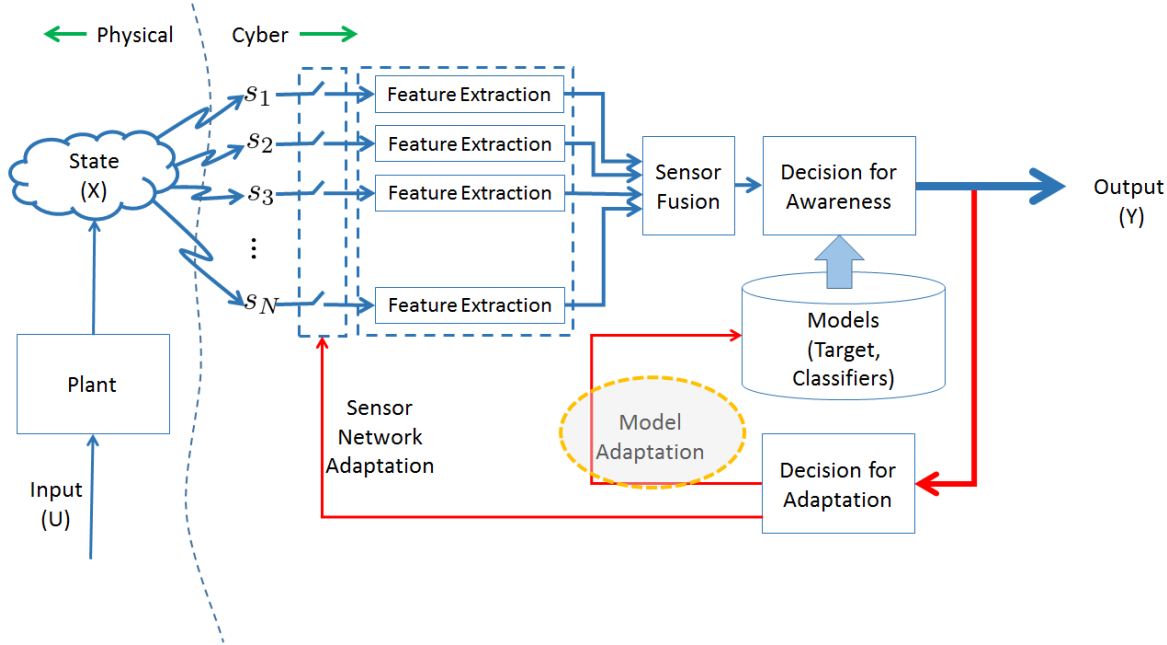
- **Heterogeneous sensors for measurements, uncertainty quantification, inferencing, prediction & control.**
- **Sensors require physical interaction with sensed phenomena and are subject to a number of noise factors**
- **Relative significance of data varies with situational context**
- **High dimensionality, much redundancy**
- **Uncontrolled Physical stimuli in the Operational Environment affect sensor outputs**

Dynamically Steering Measurement Models

- **To get reliable performance from individually less reliable sensors,**
- **Circumvent limitations of sensing, situational effects, etc.**
- **Introducing/replacing sensors.**



Problem of sequential learning from other sensors



Can a sensor sequentially learn its correct measurement model from other imperfect sensors?



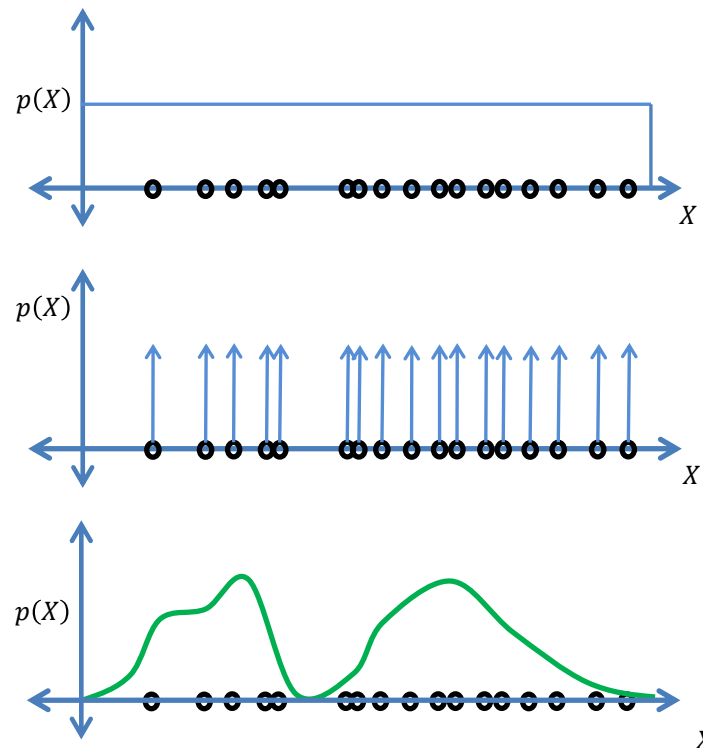
The problem of noisy density estimation

Given a finite set of observations :

$$\left\{ x_i : i \in \{1, 2, \dots, N\} \right\},$$

construct an estimate of the underlying probability density function :

$$p(x) = \frac{d}{dx} F(x) = \frac{d}{dx} p(X \leq x)$$

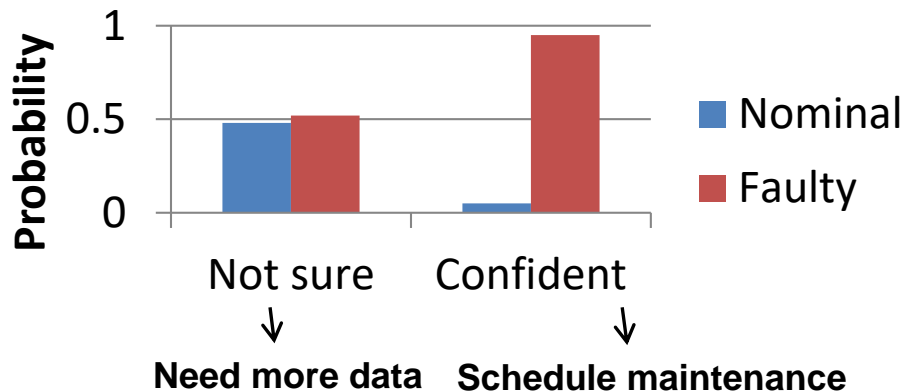




Why do we need uncertainty quantification?

“Generating a classification result is usually not the end goal (in machine monitoring and medicine).”

Machine fault detection for
condition-based maintenance



Knowledge of uncertainty in estimation can help to choose better control actions and eventually incorporate safety in AI and control systems.



Learning with Noisy Labels

- Formulate student-instructor relation as a Bayes network
- Factorize the network based on conditional probabilities:

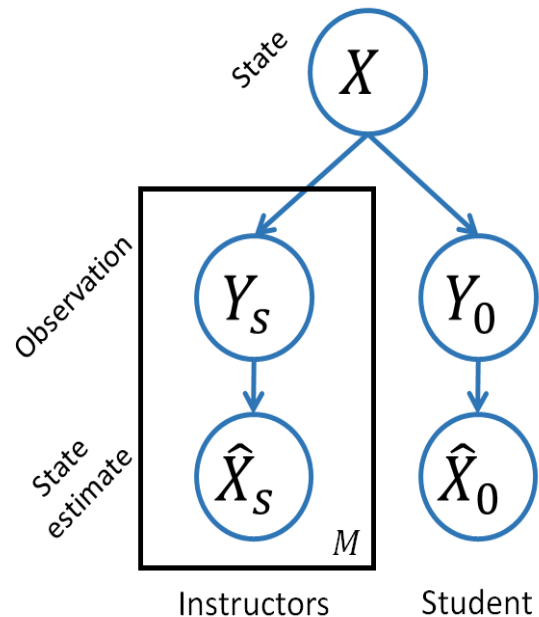
$$\begin{aligned}
 p(Y_0 | \hat{X}_1 = i) &= \sum_{j=1}^L p(Y_0, X = j | \hat{X}_1 = i) \\
 &= \sum_{j=1}^L p(Y_0 | X = j, \hat{X}_1 = i) p(X = j | \hat{X}_1 = i) \\
 &= \sum_{j=1}^L p(Y_0 | X = j) p(X = j | \hat{X}_1 = i) \\
 &= \sum_{j=1}^L \alpha_{ij}^1 p(Y_0 | X = j)
 \end{aligned}$$

where $\alpha_{ij}^1 = p(X = j | \hat{X}_1 = i) = \frac{c_{ij}^1 w_j}{\sum_{k=1}^L c_{ik}^1 w_k}$

c_{ij}^1 - classification performance $p(\hat{X}_1 = i | X = j)$
 w_j - prior probabilities $p(X = j)$

- If α_{ij} invertible, sensors are qualified to be instructors and the density (given noisy labels) may be fed into the following recursive density estimator:

$${}^* \hat{p}_j^t(y; 1) = \frac{n_j^t - 1}{n_j^t} \hat{p}_j^{t-1}(y; 1) + \frac{1}{n_j^t h(n_j^t)^d} K\left(\frac{y - y^t}{h(n_j^t)}\right)$$

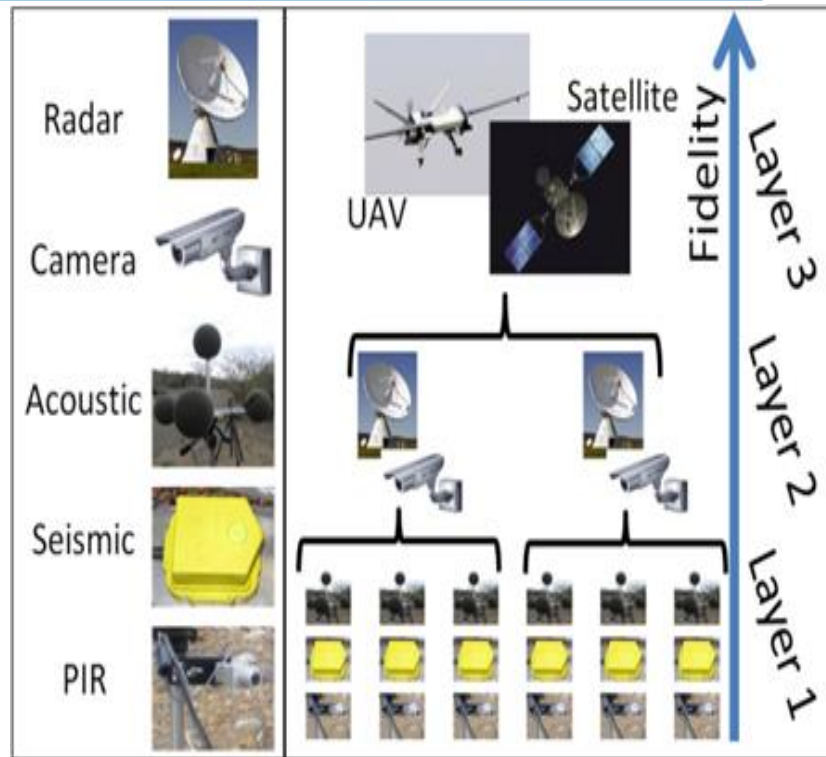




Multi-layer Multimodal Sensor Network

Hierarchical sensor network

- ❑ We use a 3-tier hierarchy of high-to-low fidelity sensors, for detecting and classifying border crossing events.
- ❑ Low-fidelity (low-layer) sensors are **cheap**, yet may generate **false alarms**. High-fidelity (high-layer) sensors have **expensive operating costs**, yet have **high classification accuracy**.





Validation

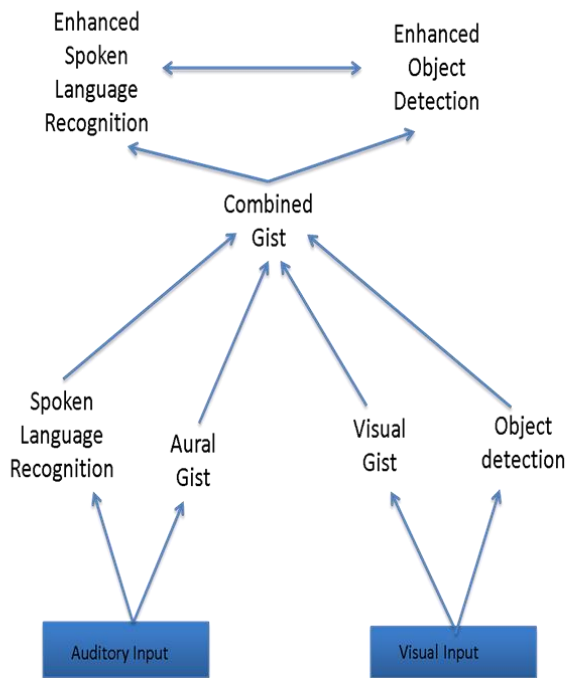
- Simulation Results
- Field experimental set-up for border control sensor-fence
 - Learning from similar sensors
 - Learning from an oracle
 - Learning from multiple instructors



FUTURE RESEARCH: Machine Perception and Learning Methods

Current Limitations: LACK SITUATIONAL AWARENESS

Audio-visual Disambiguation



MAJOR CONTRIBUTIONS

- Development of the problem of **learning from other sensors** as a sequential learning algorithm using recursive kernel-density estimation
- Formulation of a **conditional density estimation** problem to learn measurement models from data
- Statistical Forecasting using deep neural networks for real-time situational learning

FUTURE WORK

SELF-ORGANIZING SENSOR NETWORKS

- **Develop robust situational awareness to sensor measurement models**
- **Add/replace failed sensors**
- **Dynamically add data of high relative significance**--Building on this analytical framework, operate adaptive sensor networks that cluster in space-time neighborhoods of emergent hot-spots for progressively fine grained sampling, fusion, event classification, and prediction.

MULTI-MODAL SITUATION AWARENESS:

- ▣ **Spatial-temporal Environment**
- ▣ **Semantic Context**
- ▣ **Visual Context**
- ▣ **Neural Context**

Audio-visual Disambiguation



Relevant Publications

- **Journal Publications:**
- S. Phoha, "Grand Challenge: Situational Intelligence using cross-sensory fusion", *Frontiers in Robotics and AI, Sensor Fusion and Machine Perception*, August 2014.
- S. Phoha and E. Blasch, Guest Editors, "Special Issue: Dynamic Data-driven Dynamic Systems (DDDAS) Concepts in Signal Processing", *Springer Journal in Signal Processing*, May 2017: 1-2.
- N. Virani, S. Phoha, and A. Ray, "Learning from Multiple Imperfect Instructors in Sensor Networks," in *IEEE Trans. on Neural Networks and Learning Systems*, under review.
- M. Hauser, Y. Fu, S. Xiong, S. Phoha, and A. Ray, "Neural Probabilistic Forecasting of Symbol Sequences with Long Short-Term Memory," in *ASME JDSMC*, under revision.
- N. Virani, D. K. Jha, Z. Yuan, I. Shekhawat, and A. Ray, "Imitation of Demonstrations using Bayesian Filtering with Nonparametric Data-Driven Models," in *ASME JDSMC (Special Issue for Commemorating the life, achievements and impact of Rudolph E. Kalman)*, in press.
- **Conference Papers:**
- M. Hauser, Y. Fu, Y. Li, S. Phoha, and A. Ray, "Probabilistic Forecasting of Symbol Sequences with Deep Neural Networks," in *American Control Conference*, Seattle, Washington, May 2017.
- N. Virani, D. K. Jha, and A. Ray, "Sequential Hypothesis Tests Using Markov Models of Time-Series Data," in *ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Workshop on Machine Learning for Prognostics and Health Management*, 2016. (**Best Student Paper Award**)
- N. Virani, J.-W. Lee, S. Phoha, and A. Ray, "Information-Space Partitioning and Symbolization of Multi-Dimensional Time-Series Data using Density Estimation," in *American Control Conference*, pp. 3328-3333, IEEE, 2016.
- **Book Chapters:**
- N. Virani, P. Chattopadhyay, S. Sarkar, B. Smith, J.-W. Lee, S. Phoha, and A. Ray, "A Context-aware Multi-layered Sensor Network for Border Surveillance," in *Dynamic Data-driven Application Systems*, to be edited by F. Darema, Springer, accepted.
- **Other DDDAS Related Publications:**
- N. Virani, "Learning Data-driven Models for Decision-making in Intelligent Physical Systems," *Ph.D. dissertation*, 2017.



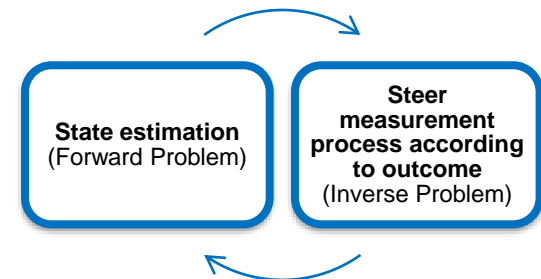
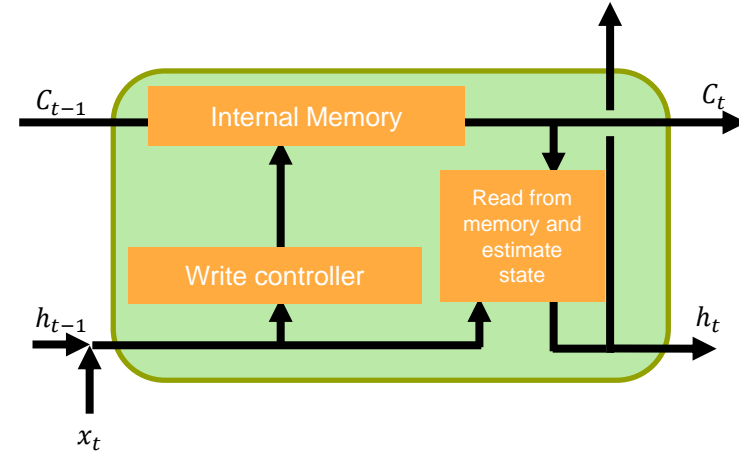
Probabilistic Forecasting within DDDAS

Long Short-Term Memory neural networks as a DDDAS

- LSTM is implemented to perform state estimation.
- Dynamic input data can write-to and read-from an internal memory device designed into the LSTM.
- Measurement process automatically controlled by the internal memory state as well as physical system state.

Benefits of Probabilistic forecasting

- Can forecast system state before measuring system state.
- Can implement control action to steer measurement process sooner.
- Prefer to predict probability densities over states, as opposed to deterministic values.





Algorithm Design

Feed forward and Long Short-Term Memory networks

- Formulate forecasting in probabilistic sense: $h_{\theta}: X \times Y \rightarrow [0,1]$ without assumption on data distribution.
- Nested compositions of affine transformations followed by simple non-linear activations.
 - $x^{(l+1)} = \tanh(W^{(l)}x^{(l)} + b^{(l)})$
 - Nested compositions of non-linear coordinate transformations can form very non-linear coordinates which can organize very non-linear data.
- Goal: Given a past history of data, predict future symbol with softmax probability distribution.

$$P(Y = y_i | X = x^{(L-1)}) = \frac{\exp(W_i^{(L-1)} x^{(L-1)})}{\sum_j \exp(W_j^{(L-1)} x^{(L-1)})}$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$$

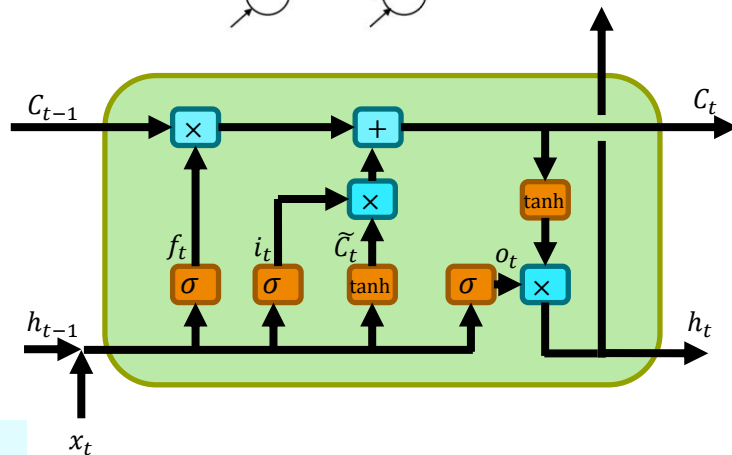
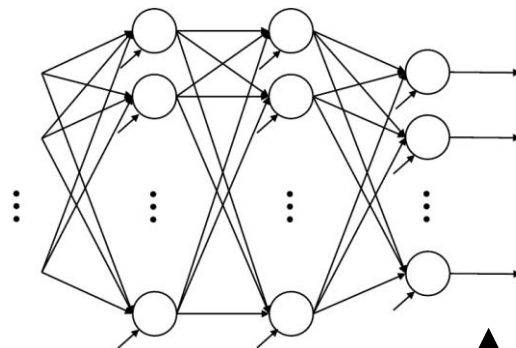
$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c)$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)$$

$$C_t = i_t * \tilde{c}_t + f_t C_{t-1}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o)$$

$$h_t = o_t * \tanh(C_t)$$

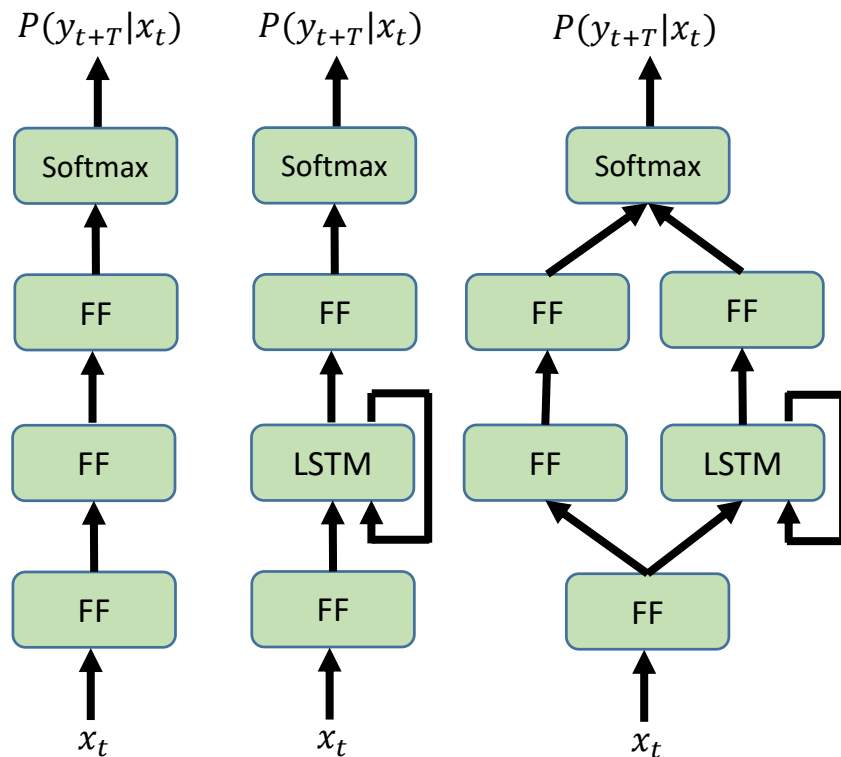




Algorithm Design

Feed forward and Recurrent Neural Networks

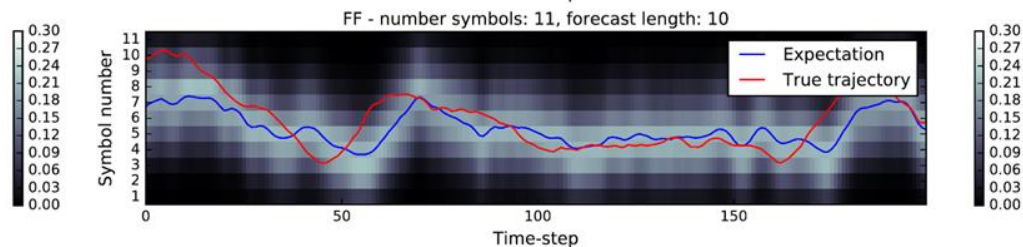
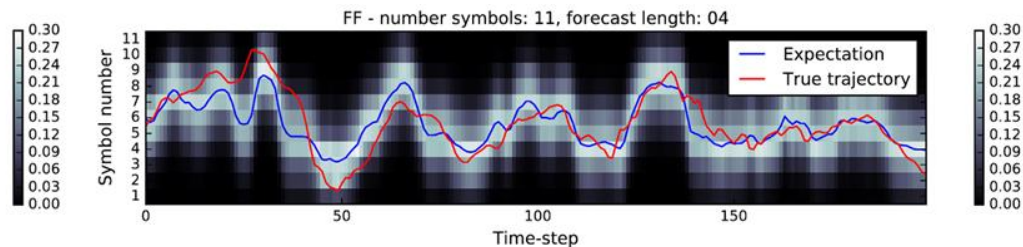
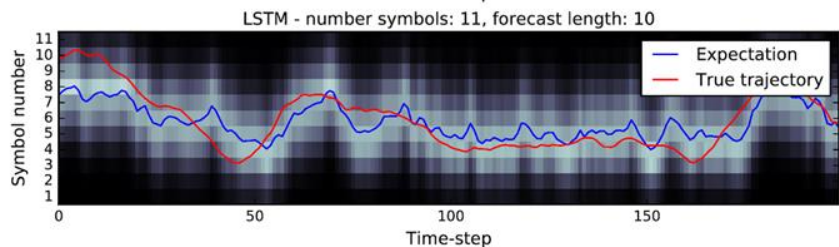
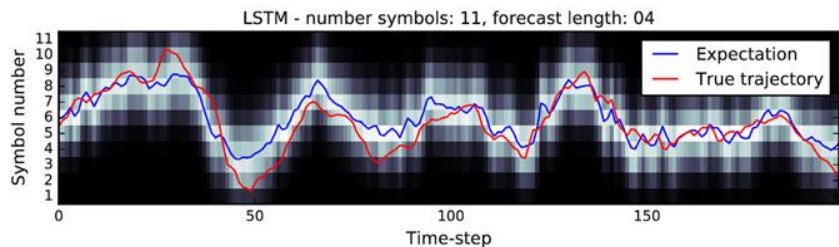
- Three network architectures compared.
 - Feed forward (FF)
 - LSTM
 - FF - LSTM
- Intuition:
 - LSTM works very well for structured time series data such as speech.
 - Tested on chaotic time series data collected from an experimental combustion system (not nearly as structured as speech).
 - Therefore intuition suggests they should all give similar results.





Probabilistic Forecasting

Typical results of probabilistic forecasting

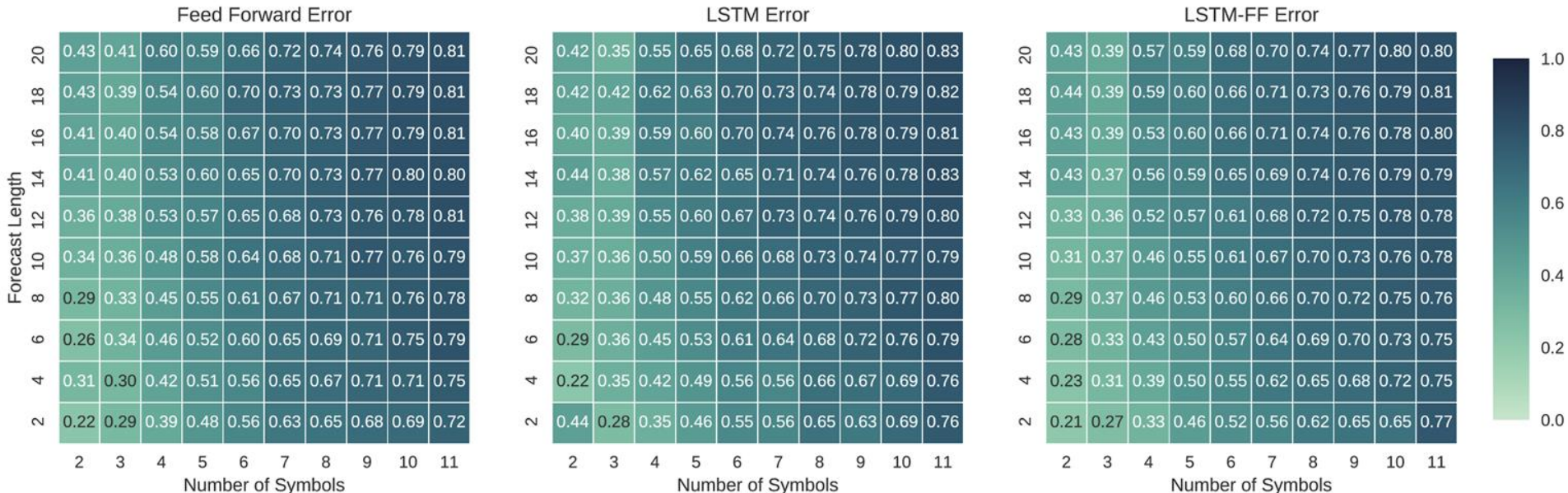




Probabilistic Forecasting

Thorough evaluation of relative performances - Error

LSTM / FF	LSTM / LSTM-FF	LSTM-FF / FF
2.57%	0.587%	1.99%

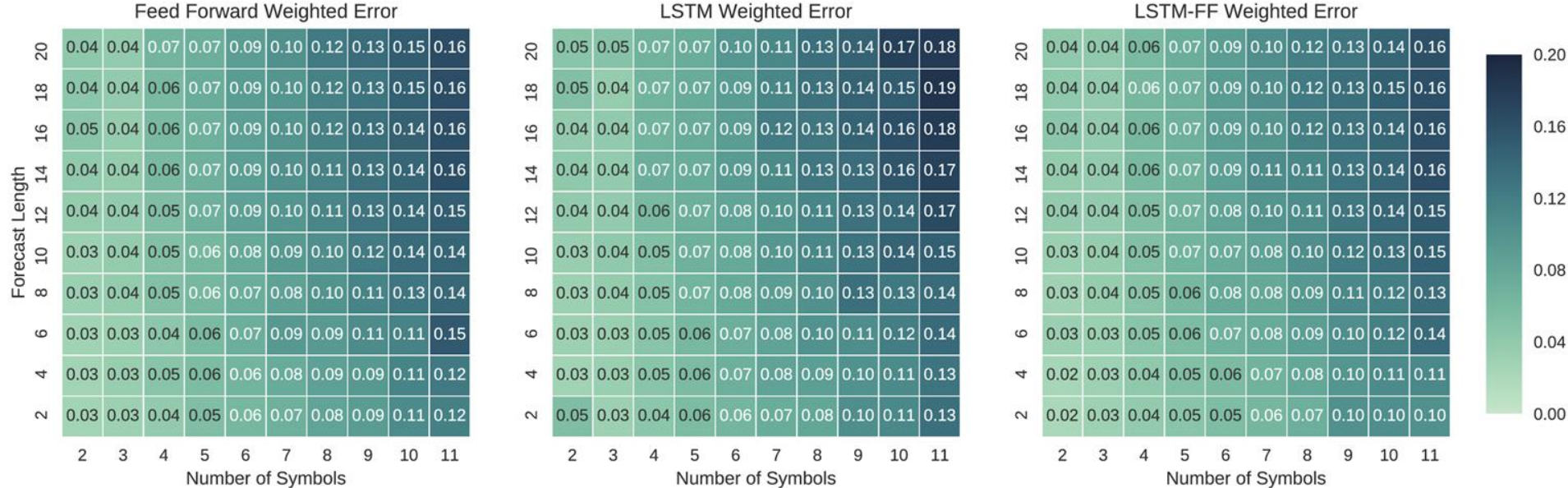




Probabilistic Forecasting

Thorough evaluation of relative performances – Weighted Error

LSTM / FF	LSTM / LSTM-FF	LSTM-FF / FF
3.94%	0.997%	2.98%





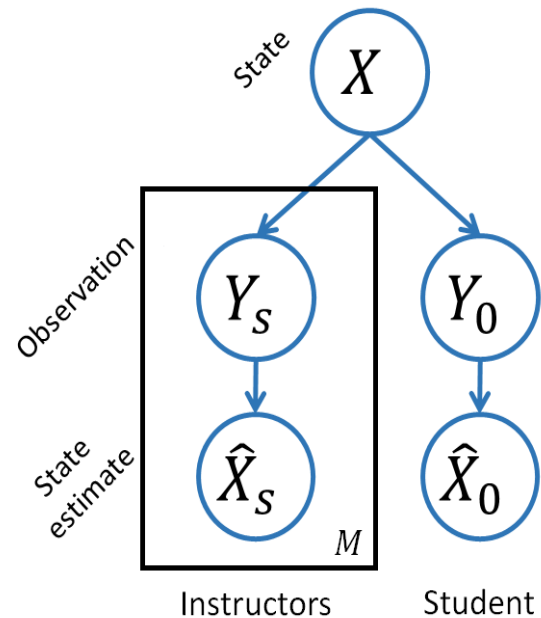
In-situ Training and Integration of New Sensors

Can “noisy” sensors train other sensors in variable environments?

Having models which forecast measurements is desired when state dynamics occur at a faster time scale than model control actions.

Yes, by developing a relationship, defined by the Bayes network, of probabilistic models between instructor generated (i.e. noisy) labels and the student's measurement model. The trained neural probabilistic forecasting models can act as measurement models to train new student model.

- ☐ A recursive nonparametric kernel density estimator is used to obtain measurement models of new sensors
- ☐ Specified sensors are selected to assume roles of instructors
- ☐ Typically sensors neighboring the new sensor
- ☐ Considered noisy due to imperfect classification performance
- ☐ Newly added sensors may assume the roles of students





Notation

$\mathcal{X} = \{1, 2, \dots, L\}$ - Set of hypothesis of random state X

$\mathcal{S} = \{1, 2, \dots, M\}$ - Set of existing sensors

Y_s - Observation from sensor $s \in \mathcal{S}$

\hat{X}_s - Estimated state (label) from sensor $s \in \mathcal{S}$

$p(Y_s|X)$ - Measurement model

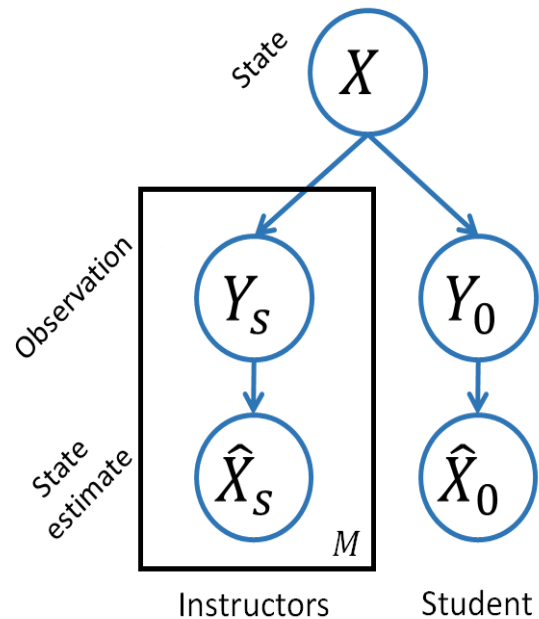
\mathcal{Y}_0 - Set of (new) student sensors

y_0^t - observation of student sensor at time instant t

At every time instant, the new sensor has access to:

$$z^t = (y_0^t, \{\hat{x}_1^t, \hat{x}_2^t, \dots, \hat{x}_M^t\})$$

From $\{z^t\}$, we can obtain the estimate, $\hat{p}(Y_0|X = x)$, of the measurement model of the new sensor





Learning with Noisy Labels via Bayes Network Factorization

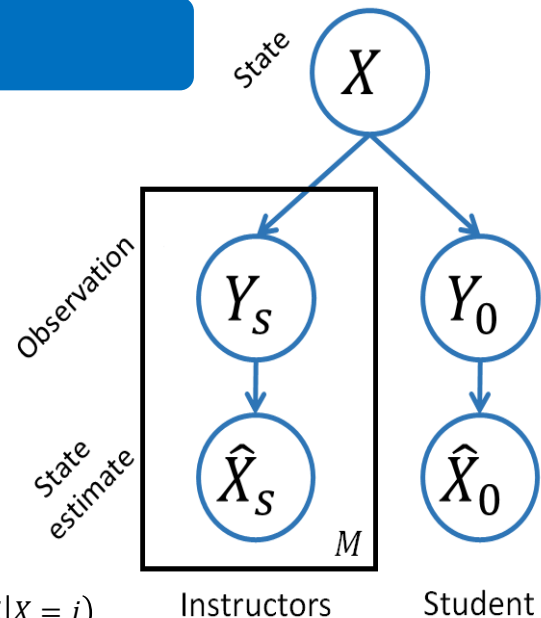
Student-Instructor Bayes Network

Factorize based on conditional probabilities:

$$\begin{aligned}
 p(Y_0 | \hat{X}_1 = i) &= \sum_{j=1}^L p(Y_0, X = j | \hat{X}_1 = i) \\
 &= \sum_{j=1}^L p(Y_0 | X = j, \hat{X}_1 = i) p(X = j | \hat{X}_1 = i) \\
 &= \sum_{j=1}^L p(Y_0 | X = j) p(X = j | \hat{X}_1 = i) \\
 &= \sum_{j=1}^L \alpha_{ij}^1 p(Y_0 | X = j)
 \end{aligned}$$

$$\alpha_{ij}^1 = p(X = j | \hat{X}_1 = i) = \frac{c_{ij}^1 w_j}{\sum_{k=1}^L c_{ik}^1 w_k}$$

c_{ij}^1 - classification performance $p(\hat{X}_1 = i | X = j)$
 w_j - prior probabilities $p(X = j)$



If A_1 is invertible, existing sensor qualifies to be instructor, and we get following equation: $P = A_1^{-1} \tilde{P} = B_1 \tilde{P}$



Generalization to Multiple Instructors

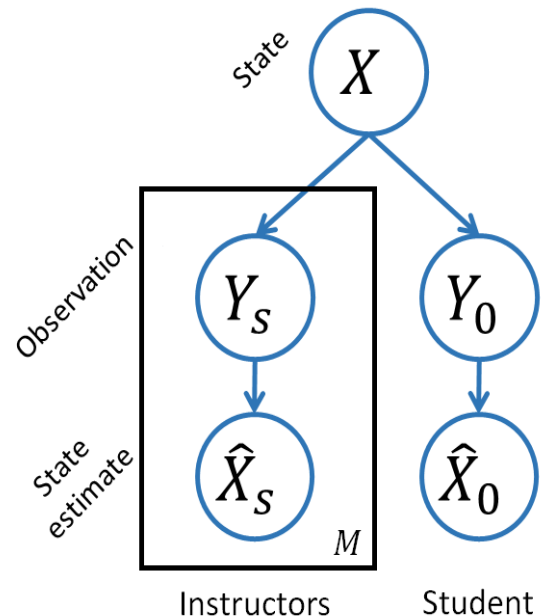
Generalize through matrix concatenation

$$\begin{bmatrix} \tilde{p}_1^t(y; 1) \\ \vdots \\ \tilde{p}_L^t(y; 1) \\ \tilde{p}_1^t(y; 2) \\ \vdots \\ \tilde{p}_L^t(y; 2) \\ \vdots \\ \tilde{p}_1^t(y; M) \\ \vdots \\ \tilde{p}_L^t(y; M) \end{bmatrix} = \begin{bmatrix} \alpha_{11}^1 & \cdots & \alpha_{1L}^1 \\ \vdots & \ddots & \vdots \\ \alpha_{L1}^1 & \cdots & \alpha_{LL}^1 \\ \hline \alpha_{11}^2 & \cdots & \alpha_{1L}^2 \\ \vdots & \ddots & \vdots \\ \alpha_{L1}^2 & \cdots & \alpha_{LL}^2 \\ \hline \vdots \\ \hline \alpha_{11}^M & \cdots & \alpha_{1L}^M \\ \vdots & \ddots & \vdots \\ \alpha_{L1}^M & \cdots & \alpha_{LL}^M \end{bmatrix} \begin{bmatrix} p(y|X=1) \\ \vdots \\ p(y|X=L) \end{bmatrix}$$

Pseudo-inverse to solve for student density given true labels:

$$\tilde{P} = AP,$$

$$P = (A^T A)^{-1} A^T \tilde{P} = B\tilde{P}$$





Sequential Update Rule

Recursive Density Estimator (RDE)

$$* \tilde{p}_j^t(y; 1) = \frac{n_j^t - 1}{n_j^t} \tilde{p}_j^{t-1}(y; 1) + \frac{1}{n_j^t h(n_j^t)^d} K\left(\frac{y - y^t}{h(n_j^t)}\right)$$

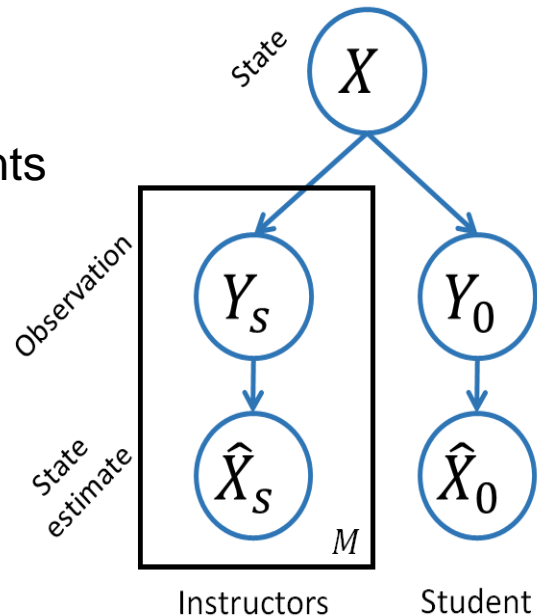
n_j^t - number of observations assigned label j in t time instants

$\tilde{p}_j^t(y; 1)$ - denotes $p(y | \hat{X}_1 = j)$

$K(\cdot)$ - kernel function

$h(n)$ - kernel width

d - dimensionality of observation

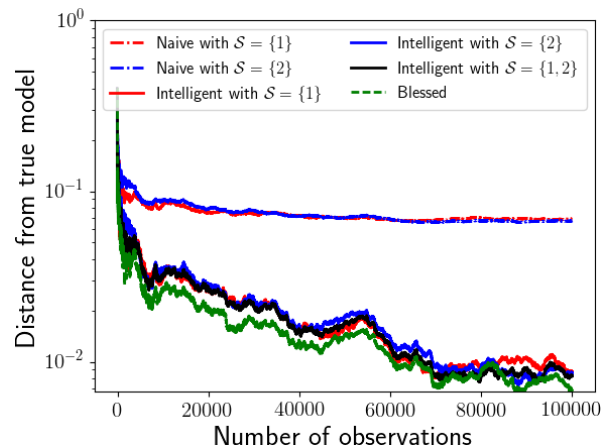
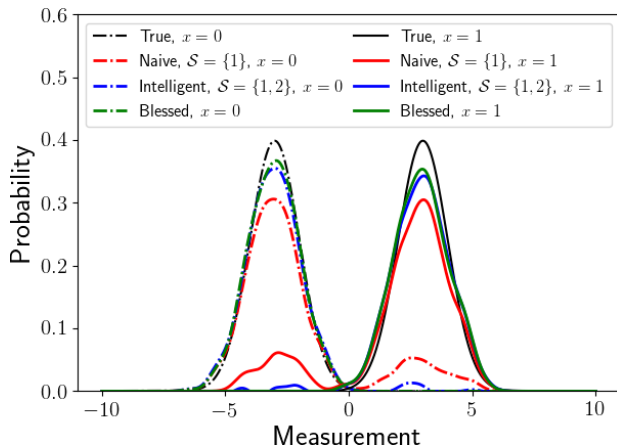




Validation Through Simulation

Simulation configuration

- Two instructors and one student.
 - Instructors have distributions $\sim N(-1,1)$ (event 0) and $\sim N(+1,1)$ (event 1)
 - Student has true distributions $\sim N(-3,1)$ (event 0) and $\sim N(+3,1)$ (event 1)
- Kullback-Leibler divergence of “intelligent” student closely matches that of “blessed” student.
 - Blessed student receives noise-free labels.
 - Intelligent student learns from noisy –labels via the aforementioned Bayes network.





Example: DDDAS Border Control

Border Control Objective

- ☐ Detect and classify targets crossing the border using fusion of data from a multilayered multimodal sensor network
- ☐ Sensors collocated in a mutual space-time neighborhood collectively predict the target passing through



(a)



(b)

Targets of interest could be e.g. (a) a person walking or (b) a vehicle



Border Control Test-bed

Experimental configuration

Fixed Sensors:

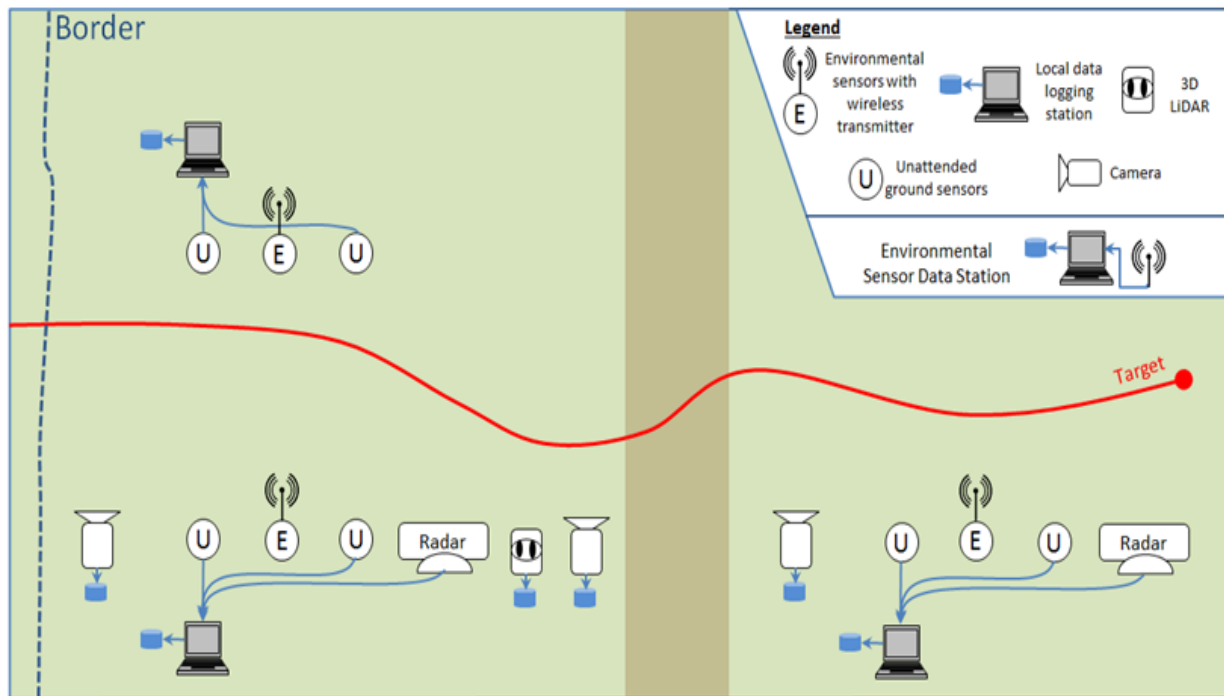
- 6 x UGS (Seismic, Acoustic, PIR)
- 3 x camera
- 2 x radar
- 1 x 3D LiDAR
- 3 x environmental sensors

Targets:

- Human walking/running
- Human with robot

Test-bed Location:

Penn Transport. Instit. Test Track, PennState Univ.

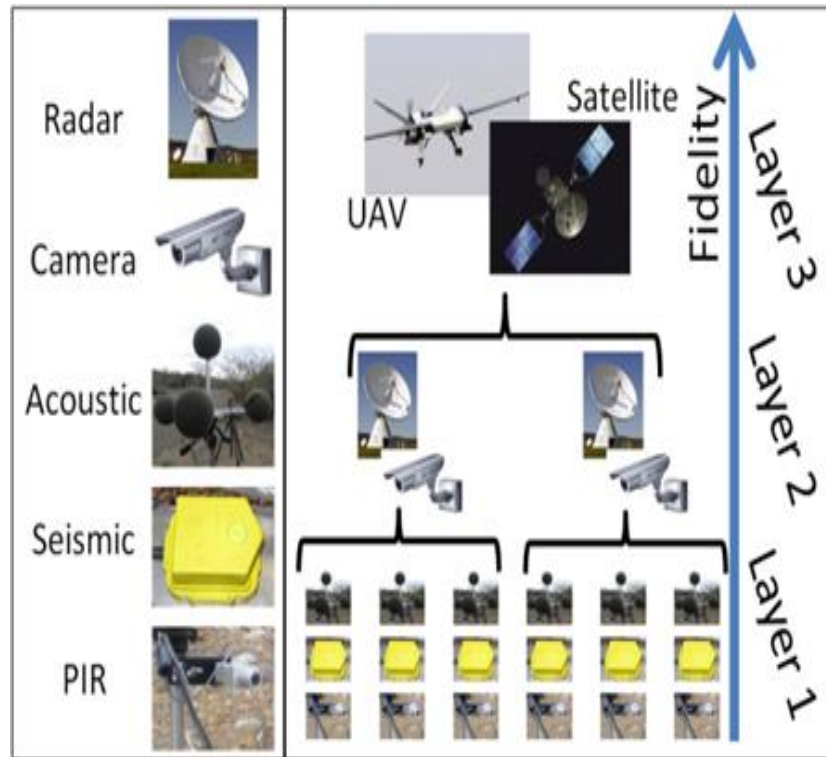




Multi-layer Multimodal Sensor Network

Hierarchical sensor network

- ❑ We use a 3-tier hierarchy of high-to-low fidelity sensors, for detecting and classifying border crossing events.
- ❑ Low-fidelity (low-layer) sensors are **cheap**, yet may generate **false alarms**. High-fidelity (high-layer) sensors have **expensive operating costs**, yet have **high classification accuracy**.





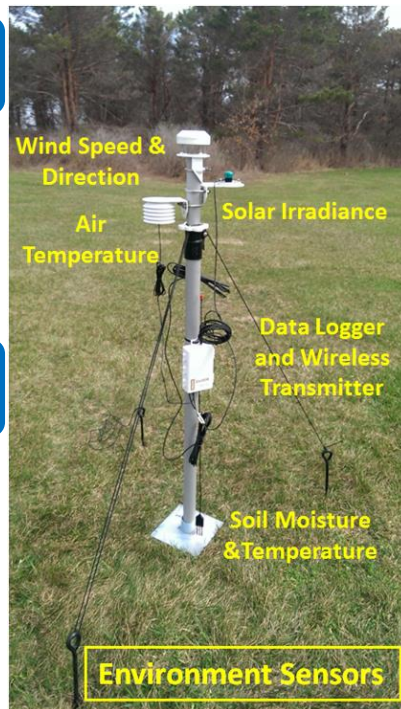
Border control Test-bed

Environmental:

- Wind speed and direction
- Air temperature
- Solar irradiance
- Soil moisture and temperature

Tracking:

- Video
- Acoustic
- Seismic
- PIR
- LiDAR





Experimental Description

Target Detection

- Purpose of this experiment is to classify the difference between a **person** (walking or running) and a **person with a robot**

Video (instructor 1):

- Off the shelf HOG-SVM classifier used.
- Classification accuracy: **93.4%**

Seismic (instructor 2):

- Off the shelf PFSA-SVM classifier used.
- Classification accuracy: **79.3%**

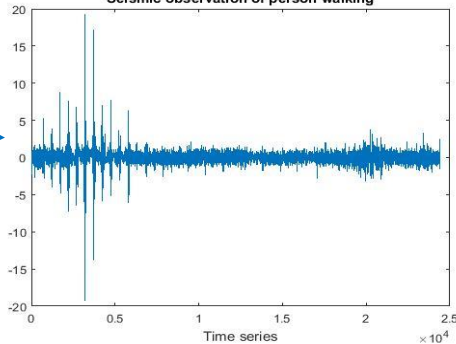


Video

Instructor observations

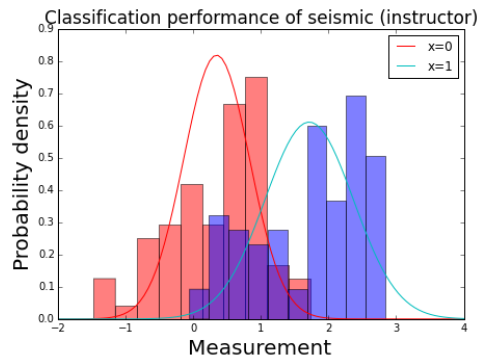
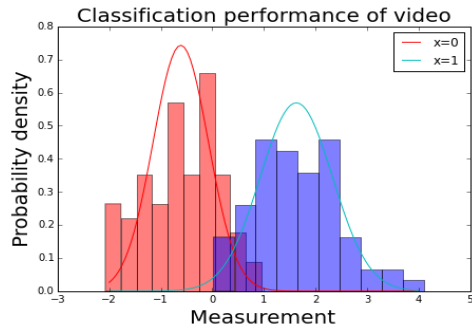


Seismic observation of person walking



Seismic

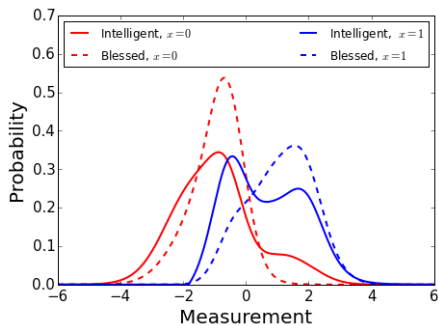
Instructor models



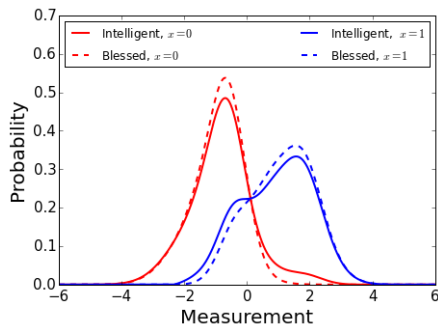


Experimental Results

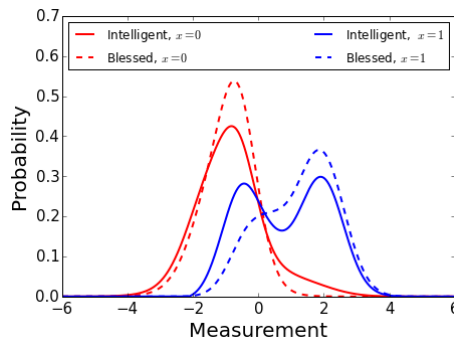
Seismic only



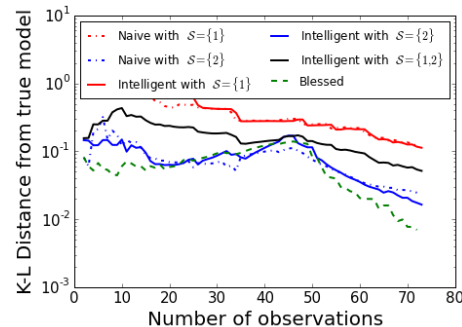
Video only



Seismic and video



KL-Divergence



Single Instructor



- ☐ Video only instructor results in best student measurement model.
- ☐ Video instructor yields KL-distance of 2×10^{-2}
- ☐ Seismic instructor yields KL-distance of 1×10^{-1}

Multiple Instructors



- ☐ Multimodal video-seismic instructor pair similar to single instructors in KL-distance.
- ☐ Multiple instructors yields KL-distance of 5×10^{-2}





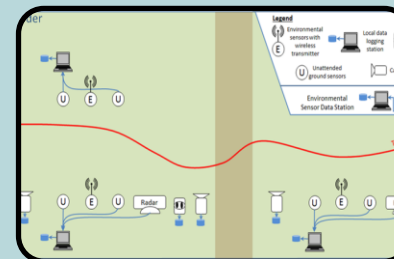
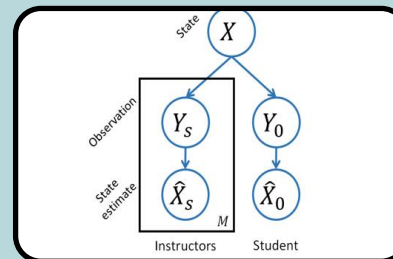
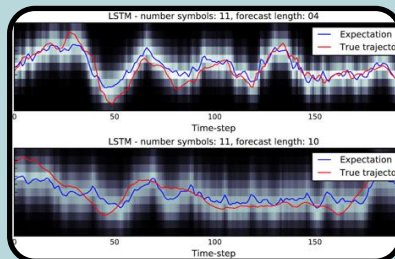
Experimental Results

Student Sensor Performance			
Instructor(s)	Seismic	Video	Both
Classification Accuracy	68.3%	86.68%	74.1%
Error Reduction	3.33%	31.13%	12.10%
KL-Divergence	1×10^{-1}	2×10^{-2}	5×10^{-2}

Note: Error reduction is calculated with respect to baseline approach, which yielded classification accuracy of 66.1% (Baseline approach uses the decision boundary learned from the instructor seismic sensor)



Summary



Motivations

- DDDAS provides flexible framework for integrating tools from machine learning and signal processing into control systems to improve measurement process.
- Flexible framework allows for improvements in accuracy of measurement processes.

Neural Probabilistic Forecasting

- New formulation to probabilistic forecasting with no assumptions on data distributions.
- Feedforward and Long Short-Term Memory neural network architectures used to test formulation.
- Extensive testing performed on experimental chaotic data.

Recursive Density Estimation – Theoretical Formulation

- Instructor-Student Bayes network model.
- Factorized and inverted.
- Recursive density estimator gives student density model.

Recursive density estimation – Results

- Simulation – Gaussian mixture model with K-L divergence shows efficacy of approach.
- Experiment – Border control testbed constructed with multimodal sensing network.
- Video and seismic instructor successfully generated the student seismic measurement model.



Questions?