Challenges and Results in Active Sensing

David A. Castañón

Boston University

Special thanks to collaborators D. Hitchings, K. Jenkins, H. Ding, V. Saligrama, K. Trapeznikov and J. Wang

Acknowledgements

- **Thanks to sponsoring agencies**
  - NSF, AFOSR, DHS
- **Thanks to Secretary Kelly and President Trump**
  - Laptop on plane still allowed ➔ this talk exists!
- **And thanks to NCI for hosting my sabbatical**
  - We like X-rays!

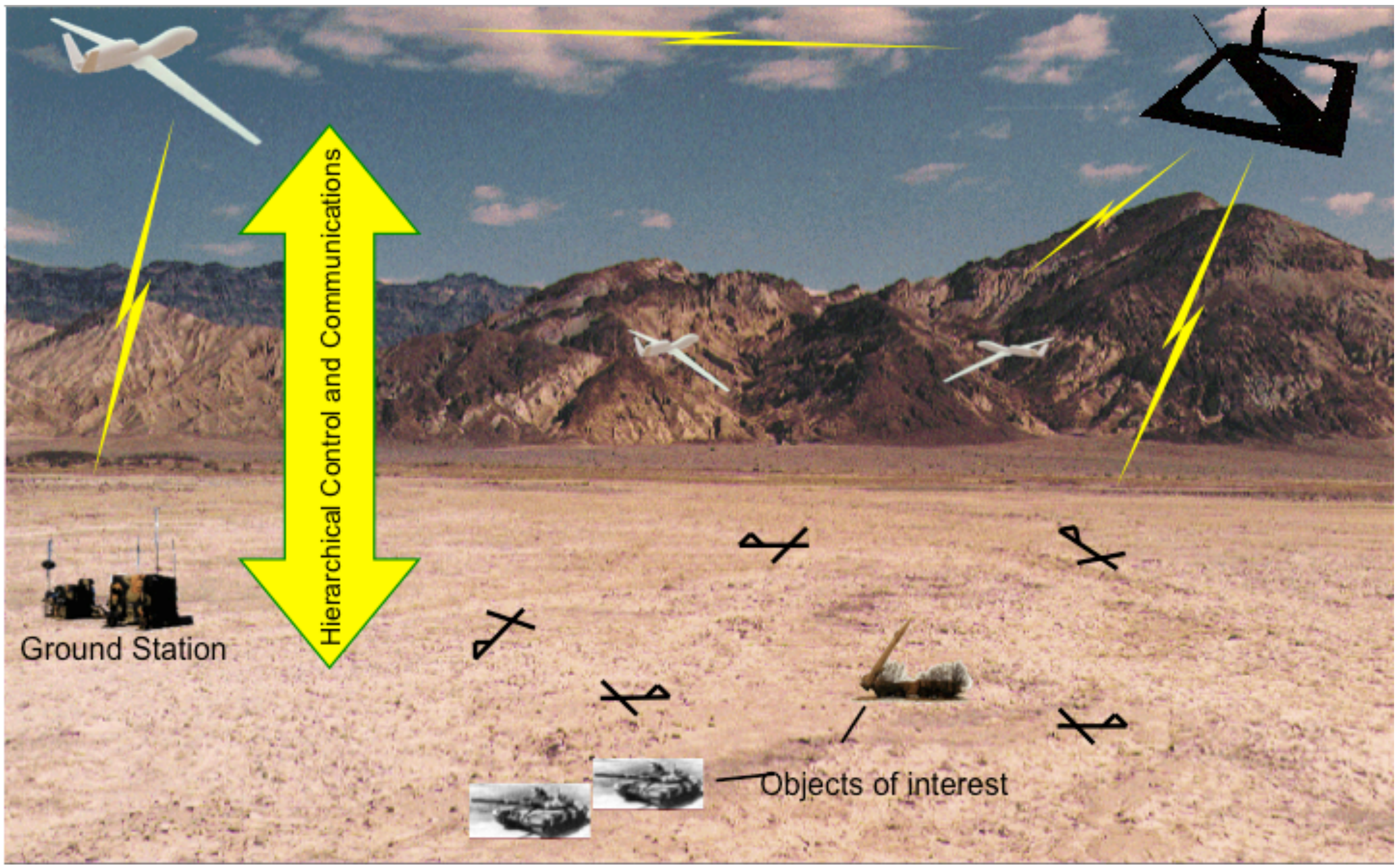
NCI's checkpoint of the future?

Motivation: Search

- **Classic problem in WW II: submarine search**
  - Assignment of search patrols (air, surface) to locate in suspected areas
  - Key problem: not guaranteed to find when searching an area
    - Limited visibility, range, intelligent adversary
  - Book: Search and Screening – B. Koopman 1946

Motivation: Surveillance

- **Unmanned and manned vehicles in coordinated monitoring**
  - Detection, tracking, classification of objects, activities, …



Motivation: Diagnosis

- **Medical diagnosis, fault detection in components, …**
  - Key aspect: Imperfect tests
  - Need to interpret collected measurements, identify what additional information is needed
  - Information costs matter

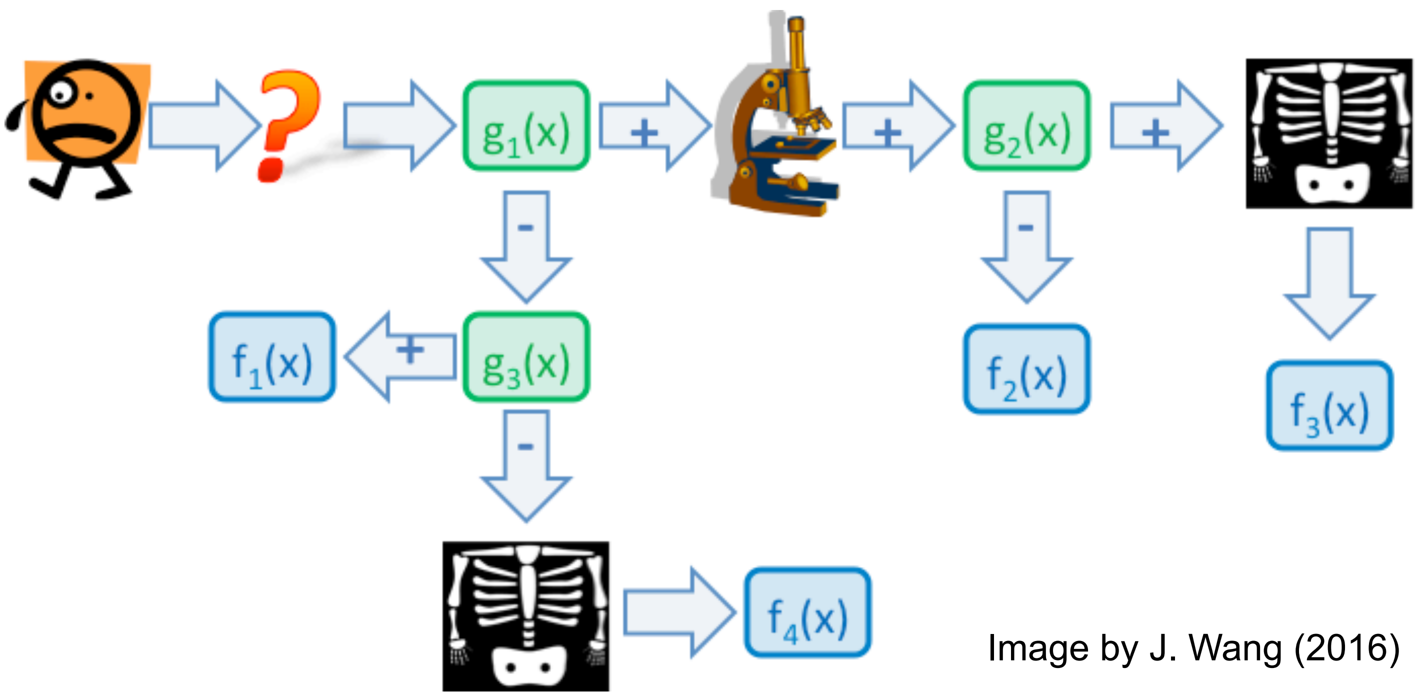
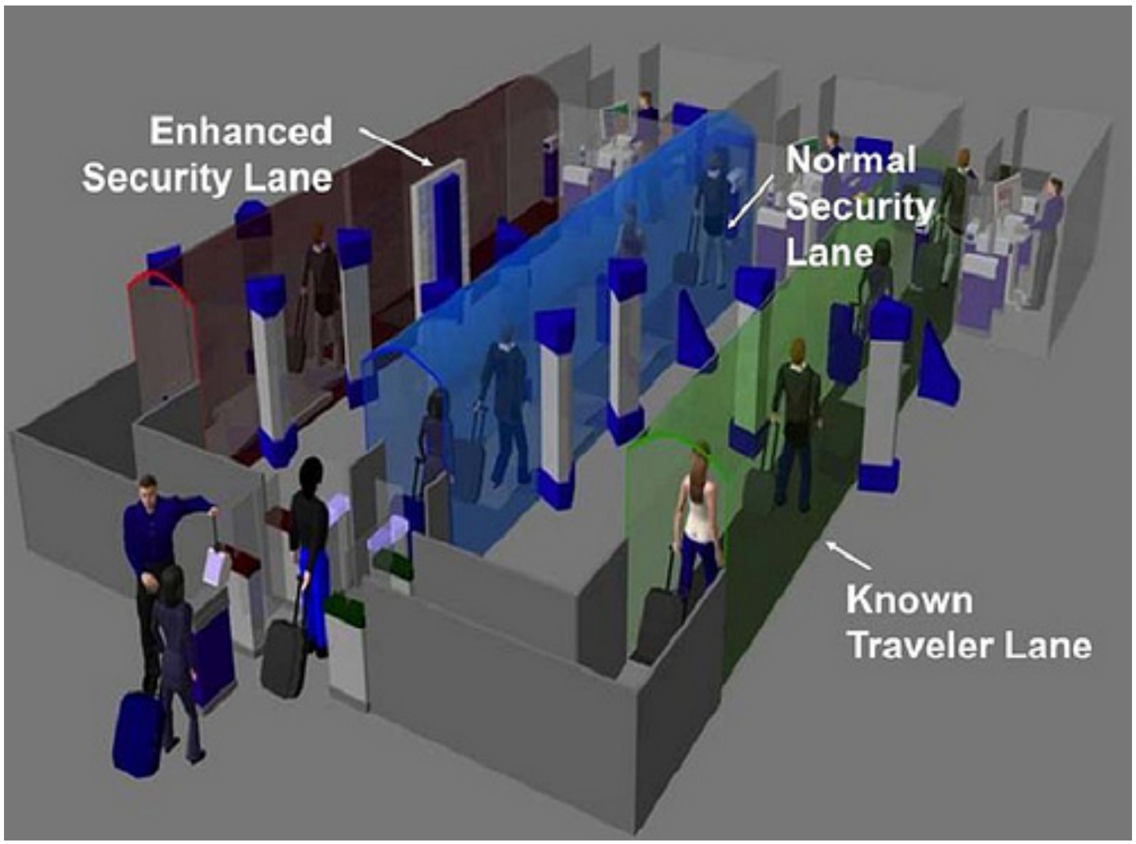


Image by J. Wang (2016)

Motivation: Security

- **Checkpoint of the future: many possible tests**
  - Exploit real-time information for flexible routing through stations
  - New tests under development



Checkpoint of the future (IATA 2010)

Problem Features

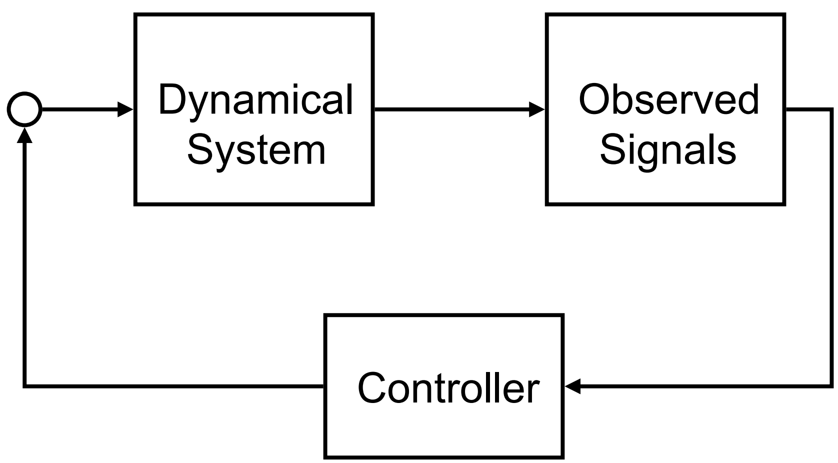
- **Opportunity for selecting measurements sequentially**
- **Information processed from previous measurements to select future measurements**
  - Feedback
- **Meaningful mission objectives to guide selection of measurements**
  - Correct diagnosis
  - Detection
  - Estimation accuracy
  - Classification accuracy
- **Important issue at the heart of the problem: Value of information**
- **So, what do we know about these problems?**

Focus on 3 problems

- **Discrete search**
  - Finding object of interest in extensions of classical search theory models
- **Dynamic search with information theoretic objectives**
  - New class of search models with full adaptive solutions
- **Adaptive test sequencing for detection**
  - Learning adaptive decision rules from training data

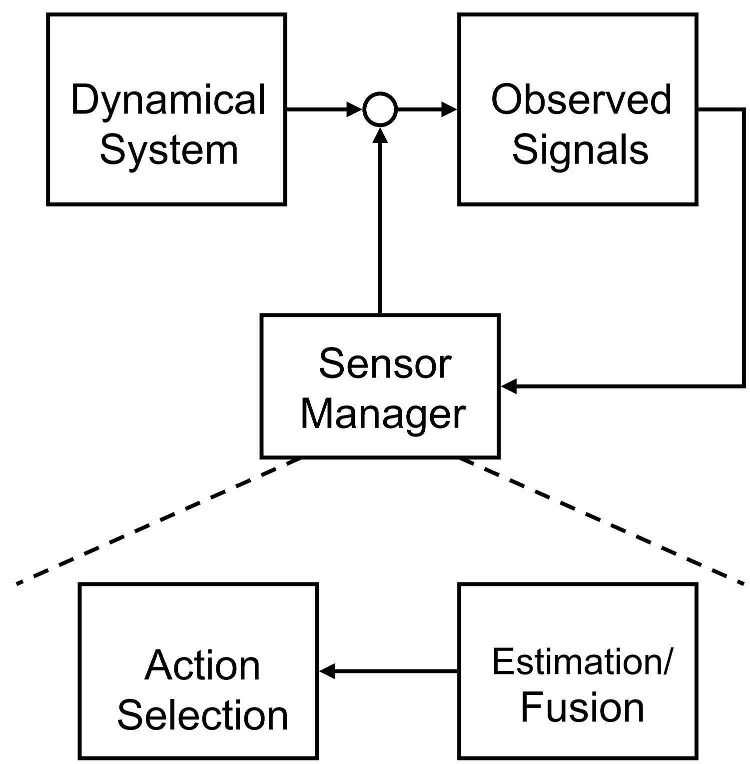
Feedback to control information

- **Feedback Control**
  - Focus on changing dynamics



- Implicit Assumption
  - Rapid, automated processing of observations to generate “state” information

- **Active Sensing**
  - Focus on changing observations



Search Theory

- **Simple model: Stationary object, finite locations, prior probability of object location (Stone, Kadane, …)**
- **Simple action model: look only in one place**
- **Simple sensor model**
  - Search of an area yields detect or not
  - $P_d < 1$, but no prob. false alarm
  - Conditionally independent detections; no switching or travel costs

P_1	P_2	P_3
P_4	P_5	P_6

- **Objectives**
  - Maximize probability of detection/minimize time to detection
  - Whereabouts search: Maximize probability of identifying correct location after fixed effort

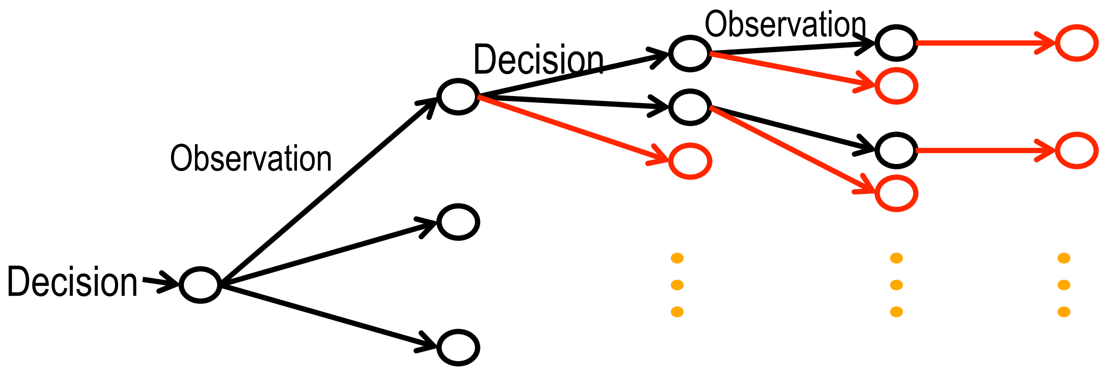
Problem setup

- **Notation**
  - Locations i = 1, …, N; object location x ε {1, …., N}
  - p_i is probability of detection when searching location i if object is there
  - Initial probability distribution $\pi_i(0) = P(x = i)$
  - Decision u(t) ε {1, …., N} → search location i
  - Information after measurement at t: I(t) = {u(1), u(2), …, u(t)} **(!!!)**
- **Bayesian information dynamics**

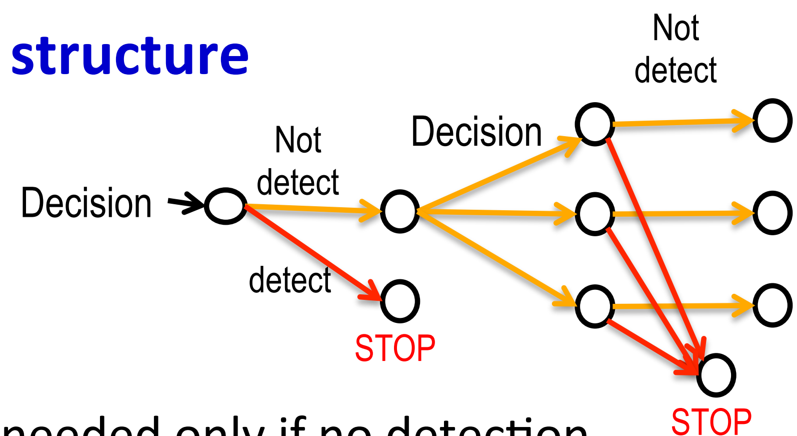
$$P(x|I(t)) \equiv \pi_i(t) = \begin{cases} \frac{\pi_i(t-1)(1-p_i)}{1-\pi_i(t-1)+\pi_i(t-1)(1-p_i)} & u(t) = i \\ \frac{\pi_i(t-1)}{1-\pi_j(t-1)+\pi_j(t-1)(1-p_j)} & u(t) = j \neq i \end{cases}$$

No feedback needed!

- **Typical feedback structure: decision tree**



- **Search feedback structure**



  - Future decisions needed only if no detection
  - Solve as a sequence of actions, not a feedback law
  - Operations research vs stochastic control

Optimization Form

- **Prob. object is at i and is not detected after observations I(t):**

$$\pi_i(0)(1-p_i)^{n_i(t)} \text{ where } \; n_i(t) = \sum_{k=1}^{t} \mathcal{I}(u(k)=i)$$

- **Objective: minimize probability of no detection with searches up to T:**

$$\min_{u(1),\ldots,u(T)} \sum_{i=1}^{N} \pi_i(0)(1-p_i)^{n_i(T)} \text{ such that } n_i(T) = \sum_{t=1}^{T} \mathcal{I}(u(t)=i)$$

- **Solution:**
  - Feedback form: At time t, search the location with highest $\pi_i(t-1)$.
  - Open-loop: Sort $\{\pi_i(0)p_i(1-p_i)^k, i=1,\ldots,N; k=0,\ldots,T\}$ in decreasing order and select the T largest. Search the locations in that order.

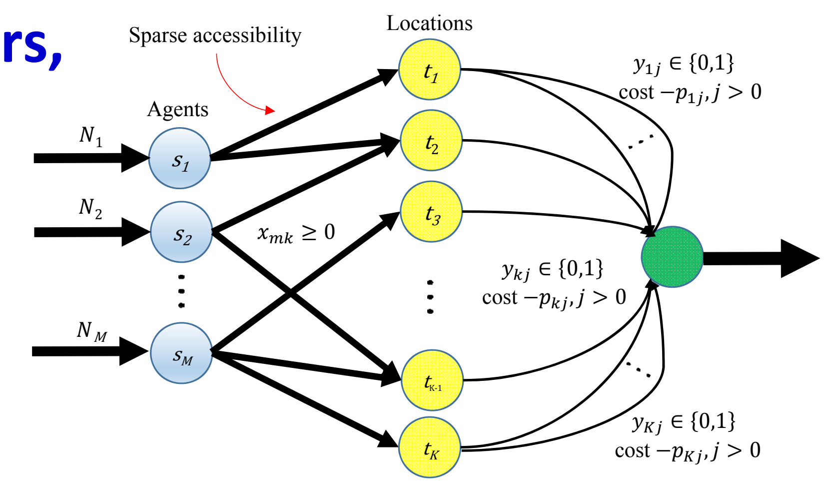
Extensions

- **Whereabouts search**

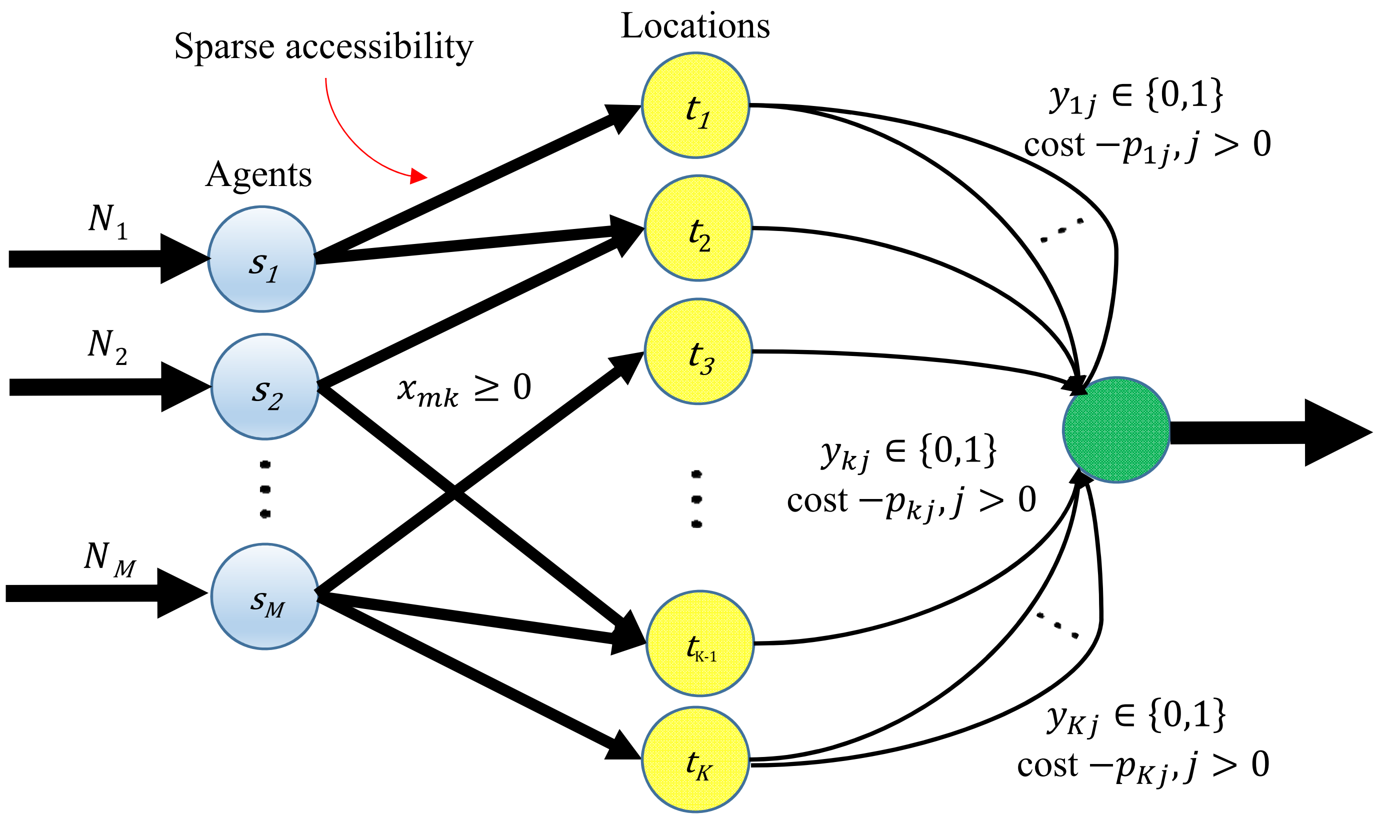
$$\max_{u(1),\ldots,u(T)} \left\{ \sum_{i=1}^{N} \pi_i(0) \left(1 - (1-p_i)^{n_i(T)}\right) + \max_i \pi_i (1-p_i)^{n_i(T)} \right\}$$

- **Optimal strategy:**
  - Feedback form: At time t, search the location with second highest $\pi_i(t-1)$.
  - Open-loop form available: search optimally among all except highest $\pi_i(0)$

- **Other extensions: multiple sensors, sparse coverage constraints**



Extensions



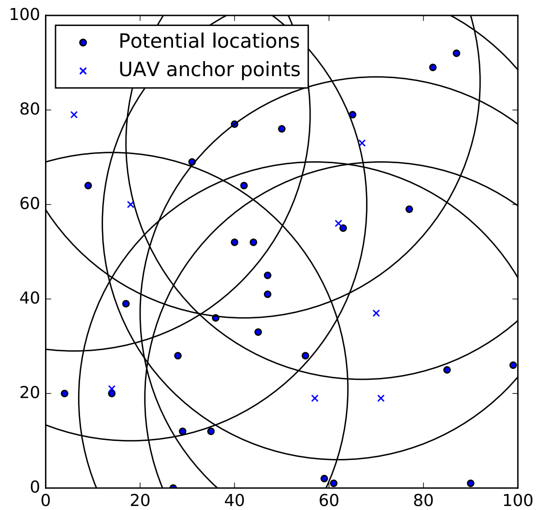
- **Convert to network optimization**
  - Integer program with unimodular constraints
  - Fast algorithm developed Ding-C. '17 -- complexity

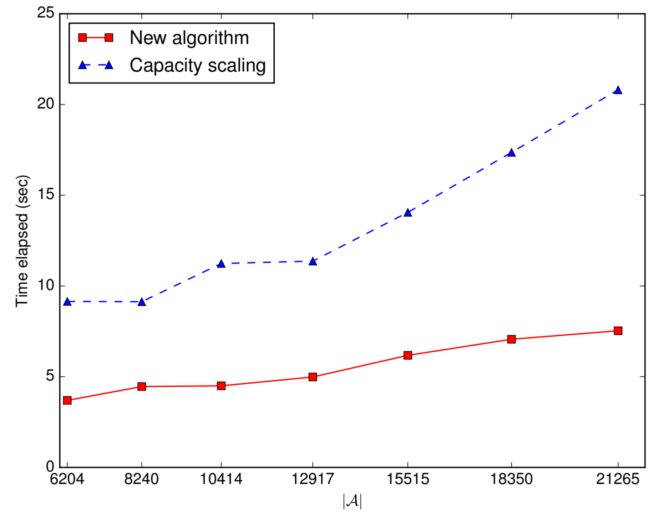
$$O(N|\mathcal{A}|) \text{ where } N = \sum_{m=1}^{M} N_m$$

Results

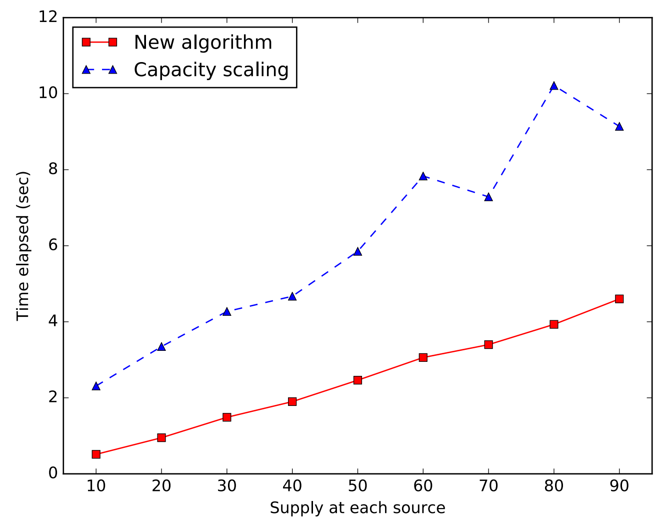
- **Sample search: 9 UAVs with limited field of regard**
  - P_ds from 0.7 to 0.9

Illustration of coverage and points of interest





Run-time vs commercial alternative with Increasing density



Run-time vs commercial alternative with Increasing supply

Can we improve sensor model?

- **Assume we have both false alarms and missed detections**
  - Observation y in {0,1}
  - P(y = 1 | object present) = p_i, P(y = 1|object absent) = q_i
  - Action u(t) yields measurement y(t)
  - Information after measurement at t: I(t) = {u(1),y(1), …, u(t), y(t)}
- **Bayesian dynamics**

$$P(x=i|I(t)) \equiv \pi_i(t) = \begin{cases} \frac{\pi_i(t-1)p_i}{(1-\pi_i(t-1))q_i+\pi_i(t-1)p_i} & u(t)=i, y(t)=1 \\ \frac{\pi_i(t-1)(1-p_i)}{(1-\pi_i(t-1))(1-q_i)+\pi_i(t-1)(1-p_i)} & u(t)=i, y(t)=0 \\ \frac{\pi_i(t-1)q_j}{(1-\pi_j(t-1))q_j+\pi_j(t-1)p_j} & u(t)=j\neq i, y(t)=1 \\ \frac{\pi_i(t-1)(1-q_j)}{(1-\pi_j(t-1))(1-q_j)+\pi_j(t-1)(1-p_j)} & u(t)=j\neq i, y(t)=0 \end{cases}$$

…with great difficulty!

- **Objective: Given T observations,**

$$\max_{\gamma_1,\dots,\gamma_T}\left\{\max_i \pi_i(T)\right\} \text{ such that } \gamma_t(\mathcal{I}(t-1)) = u(t)$$

- **Only one known characterization of optimal strategies (C.'95)**
  - Special case: $p_i = 1 - q_i = p$
  - Optimal feedback strategy: at time t, measure the location with either the largest or second largest $\pi_i(t-1)$
  - Myopic strategies optimal for dynamic model

- **Results do not extend to multiple searchers, non-symmetric error probabilities, …**
  - Must resort to stochastic control approaches
  - Stochastic dynamic programming for partially observed process
    - C.'97, Evans and Krishnamurthy '01, Wintenby and Krishnamurthy'06, Kreucher et al '06: Bashan et al'08, C.'05, C.-Hitchings '10, '11…
  - Or alternative formulations – information theory, e.g.

Information Theory and Search

- **Information theory: quantitative measures of information and uncertainty**
  - Given conditional distribution $\pi_i(t-1)$, can measure entropy

$$J(\pi(t-1)) = \sum_{i=1}^{N} -\pi_i(t-1)\log(\pi_i(t-1))$$

  - Heuristic strategy for active sensing: Select location to search that maximizes expected reduction in entropy (Kastella '95, many others)

$$u(t) = \arg\min_{i=1,\ldots,n} \big(J(\pi(t-1)) - E_{y(t)} J(\pi(t)|u(t)=i, y(t))\big)$$

  - Equivalent to maximizing expected KL divergence between $\pi(t-1)$, $\pi(t)$
  - Computable for simple problems (e.g. search, not too many locations), but no guarantee of dynamic optimality (myopic, one-stage lookahead)
  - And weak correlation with mission metrics (entropy vs location)

Change search problem

- **Object located in compact subset of Euclidean space**
  - x is now continuous-valued in domain ***A***
  - Prior information $p_0(x)$ given as a density (absolutely continuous w.r.t. Lebesgue measure)
- **Sensor model motivated by group testing, compressive sensing**
  - Sensor observes subset of domain ***A***
  - If x in ***A***, then observe measurement y distributed as $f_1(y)$; else observe y distributed as $f_0(y)$
  - Multiple measurements are conditionally independent given x
- **Objective: use controls over multiple time windows to minimize the differential entropy of for evolution of information dynamics**

Background

- **Noisy decoding (Horstein ‘63, Burnashev ’73, …)**
  - Probabilistic bisection search
  - Decode continuous signal using quantized binary measurements
  - No dynamic optimality, no performance guarantees (error bounds)
- **20 question search…inspired by probabilistic binary search**
  - Jedinak, Frazier, Sznitman et al ‘11, ’13, …Single sensor
  - Tsigliradis, Sadler, Hero ‘14, ’15, C.-Ding ‘15, ‘16: Multiple sensors
  - Dynamic optimality (!!!)
  - Performance guarantees (some, in a simple case)
  - Extends to costly sensing (with limited models…)

Formulation

- **M sensors searching for a single object located at unknown X present in compact region *A* in R^n**
  - Discrete stages: at each stage sensor m chooses A^m a Borel subset of *A* to observe, receives discrete-valued observation Y^m

$$P(Y^m = y|A^m, X) = \begin{cases} f_1^m(y) & X \in A^m \\ f_0^m(y) & X \notin A^m \end{cases}$$

  - Y^m_k discrete, assumed conditionally independent over sensors, time given X

- **Information history for decisions at stage k:**
  - $D_k = \{(A^1_1, y^1_1), \ldots, (A^M_1, y^M_1), \ldots, (A^1_{k-1}, y^1_{k-1}), \ldots, (A^M_{k-1}, y^M_{k-1})\}$
  - Information state: probability density $p_n(X) = p(X|D_k)$ (prior information $p_1(X)$ assumed absolutely continuous with respect to Lebesgue measure)
  - Evolution using Bayesian dynamics of inference yields a measure-valued state process

Formulation - 2

- **Admissible strategies:**
  - Each sensor m: map conditional probability densities on *A* to actions A^m
  - Unusual action space…no clear topological structure
  - Π denotes space of admissible joint strategies for M sensors, over N stages
- **Information Dynamics: Bayes' rule**

$$p(x|D_{n+1}) \equiv p_{n+1}(x) = \frac{p_n(x) \sum_{i_1,\ldots,i_M=0}^{1} \prod_{k=1}^{M} f_{i_k}^k(y^k) \mathcal{I}[X \in \cap_{k=1}^{M} (A^k)^{i_k}]}{\int p_n(\sigma) \sum_{i_1,\ldots,i_M=0}^{1} \prod_{k=1}^{M} f_{i_k}^k(y^k) \mathcal{I}[\sigma \in \cap_{k=1}^{M} (A^k)^{i_k}] d\sigma}$$

- **Objective: Minimize differential entropy after observations at stage T**

$$\inf_{\pi \in \Pi} H(p_{T+1}(x)) \equiv \inf_{\pi \in \Pi} E[-\int_{x \in A} p_{T+1}(x) \log p_{T+1}(x) dx]$$

Solution

- **Stochastic control:**
  - Notation: $\boldsymbol{A}^k = \{A^1_k, \ldots, A^M_k\}$; $Y^k = \{y^1_k, \ldots, y^M_k\}$
- **Bellman's equation for optimal cost**

$$V(p_n, n) = \inf_{\boldsymbol{A}_n} \Big(E_{\mathbf{Y}_n} [V(p_{n+1}, n+1) | \boldsymbol{A}_n, p_n] \Big)$$

- **Verification: A strategy that achieves optimal cost is an optimal strategy**
  - Measure valued state process, non-metric action space requires special considerations

Backward Induction

- **One-stage problem**
  - Let $i_{1:M}$ be a Boolean vector indicating the possible conditions of how X relates to the set of queries A by the M sensors
    - e.g. $i_1 = 0$ if X is not in A^1, $i_2 = 1$ if X is in A^2, …
  - Notation: $(A)^0 = A^c$, $(A)^1 = A$. Then, for $i_{1:M}$, A, define

$$P(\boldsymbol{Y} = \boldsymbol{y}|\boldsymbol{A}, X) = \sum_{i_1,\ldots,i_M=0}^{1} \prod_{k=1}^{M} f_{i_k}^k(y^k)\mathcal{I}[X \in \cap_{k=1}^M (A^k)^{i_k}]$$

$$u_{i_{1:M}}(\boldsymbol{A}, p_N) \equiv u_{i_{1:M}} = \int_{\cap_{k=1}^M (A^{k*})^{i_k}} p_N(\sigma)d\sigma \ \geq 0$$

- **Bayes' rule simplifies:**

$$p_{n+1}(x) = \frac{p_n(x)\sum_{i_1,\ldots,i_M=0}^{1} \prod_{k=1}^{M} f_{i_k}^k(y^k)\mathcal{I}[X \in \cap_{k=1}^M (A^k)^{i_k}]}{\sum_{i_1,\ldots,i_M=0}^{1} u_{i_{1:M}} \prod_{k=1}^{M} f_{i_k}^k(y^k)}$$

Solution - 2

- **One-stage problem solution**
  - Expected differential entropy of decisions ***A*** (after standard info-theory manipulations)

$$E_{\boldsymbol{Y}_N}[H(p_{N+1})|\boldsymbol{A}, p_N)] = H(p_N) - \Big[\mathcal{H}\Big(\sum_{i_1=0,\dots,i_M=0}^{1} u_{i_{1:M}} \prod_{k=1}^{M} f_{i_k}^k(y^k)\Big) - \sum_{i_1=0,\dots,i_M=0}^{1} u_{i_{1:M}} \mathcal{H}\Big(\prod_{k=1}^{M} f_{i_k}^k(y^k)\Big)\Big]$$

  - Shannon entropy for discrete variables: $\mathcal{H}(f(\boldsymbol{y})) = -\sum_{\boldsymbol{y}} p(\boldsymbol{y}) \ln p(\boldsymbol{y})$
  - Dependence on $p_N(x)$, ***A*** only through scalars $\{u_{\underline{I}}, \underline{\boldsymbol{I}} = (i_1, \dots, i_M)\} = \boldsymbol{u}$
  - Note: term in brackets $G(\boldsymbol{u})$ is mutual information of the variable X conditioned on D_N and y
  - We want to select ***A*** to maximize mutual information between them

Solution - 3

- **One-stage problem solution (cont)**
  - Lemma: $G(\boldsymbol{u})$ is strictly concave in $\boldsymbol{u}$.
  - $\boldsymbol{u}$ is a probability vector (sums to 1, non-negative)
  - Maximization is computation of a channel capacity
  - $\boldsymbol{u}^*$ = arg max $G(\boldsymbol{u})$, **and** does not depend on $p_n(x)$, $\boldsymbol{A}$

  - Theorem: For any $\boldsymbol{u}^*$, there exists a set of queries by the sensors $\boldsymbol{A}^*$ such that

$$u^*_{i_{1:M}} = \int_{\cap_{k=1}^{M}(A^{k*})^{i_k}} p_N(\sigma)\,d\sigma$$

    - Proof exploits existence of conditional density…

  - Corollary: Optimal cost $V(p_N, N) = H(p_N) - G(\boldsymbol{u}^*)$
    - Note that the cost-to-go is again the differential entropy and a constant

Solution - 4

- **N-stage problem solution**
  - Lemma: For any density $p_n(x)$, we can find $\boldsymbol{A}$ such that

$$u_{i_{1:M}}(\boldsymbol{A}, p_n) = u^*_{i_{1:M}} \text{ for all } \; i_{1:M}$$

  - Theorem: Optimal cost $V(p_n, n) = H(p_n) - (N+1-n)G(\boldsymbol{u}^*)$
  - Corollary: the following strategies are optimal:

$$\boldsymbol{A}_n \text{ such that } \; u_{i_{1:M}}(\boldsymbol{A_n}, p_n) = u^*_{i_{1:M}} \text{ for all } \; i_{1:M}$$

  - General result for correlated errors among sensors
  - Computation of $\boldsymbol{u}^*$ is still large: 2^M variables concave maximization problems
    - Simplify? Exploit conditional independence…

Solution - 5

- **Single Sensor Problem (Jedinak, Frazier, Sznitman '13)**
  - Define for sensor *m*:

$$G^m(u) = \mathcal{H}(uf_1^m(y) + (1-u)f_0^m(y)) - u\mathcal{H}(f_1^m(y)) - (1-u)\mathcal{H}(f_0^m(y))$$
$$u^{m*} = \arg\max_u G^m(u)$$

  - Scalar strictly concave maximization for each of *M* sensors

- **Multisensor problem**
  - Theorem: An optimal solution of the multisensor problem at stage *n* from state $p_n(x)$ is given by

$$u_{i_{1:M}}(\boldsymbol{A}^*, p_n) = \prod_{m=1}^{M} (u^{m*})^{i_m}(1-u^{m*})^{1-i_m}; \; G(\boldsymbol{u}^*) = \sum_{m=1}^{M} G(u^{m*})$$

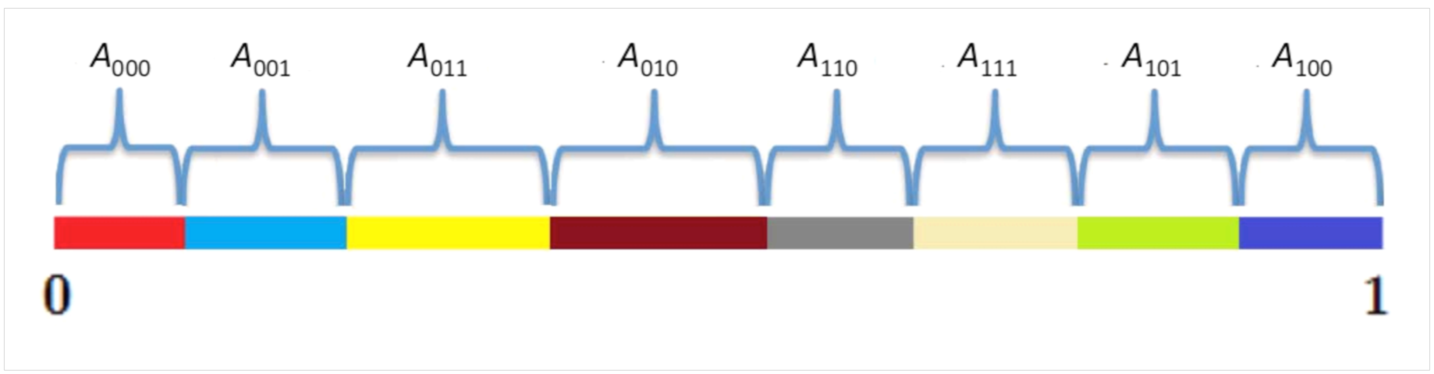
  - Complexity M scalar concave maximization problems

Finding the query regions

- **Approach: Common approach at coding regions**
  - Compute $u_{i_{1:M}}(\boldsymbol{A}^*, p_n) = \prod_{m=1}^{M} (u^{m*})^{i_m} (1 - u^{m*})^{1-i_m}$
  - Order indices $i_{1:M}$ in a linear, total order
  - Allocate regions in same order with probabilities satisfying

$$\int_{A_{i_{i:M}}} p_n(x)dx = u^*_{i_{1:M}}$$

  - Construct A^m *as* $A^m = \cup_{\{i_{1:M} | i_m = 1\}} A_{i_{1:M}}$



Performance bounds

- **Does minimizing entropy guarantee good localization?**
  - Not necessarily. 2-D differential entropy goes to neg. infinity if error goes to 0 in 1 dimension
- **Lower bound:**
  - If $H(p_0)$ is finite, then for any optimal strategy, we have

$$E[||X - \hat{X}_n||_2^2] \geq \frac{d\sqrt[d]{C_0}}{2\pi e} 2^{-\frac{2n\varphi^*}{d}}$$

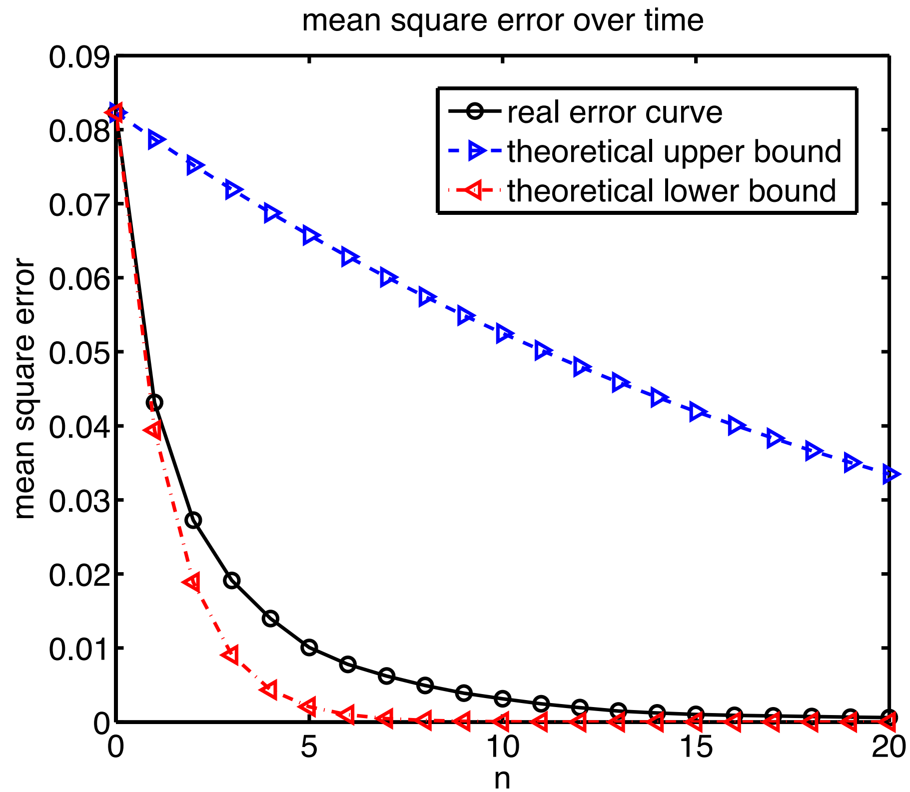
  - Proof from property that Gaussians have maximal entropy for given error covariance
- **Upper bound: Hard! Will depend on specific coding strategy**
  - Can show some optimal strategies have finite lower bounds that do not decay to 0
  - One result (Waeber-Frazier '15): For binary symmetric channels, one sensor, there exists a constant c(p) > 1 such that

$$E[|X - \hat{X}_n|] \in o(c(p)^{-n})$$

  - No results for asymmetric channels or non-binary or multi-sensor

Experiments

- **Illustration of bounds**
  - Single sensor, binary symmetric probability of error 0.1



Multisensor example

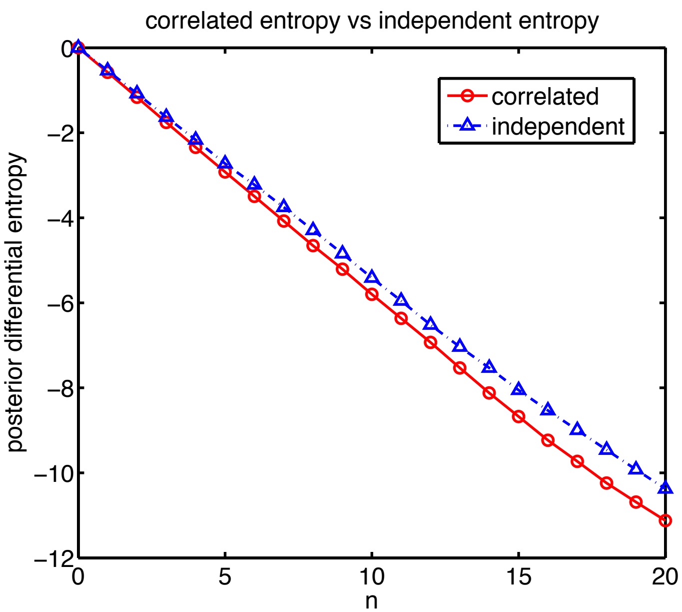
- **Two sensors, binary asymmetric channel**
  - Probability tables:

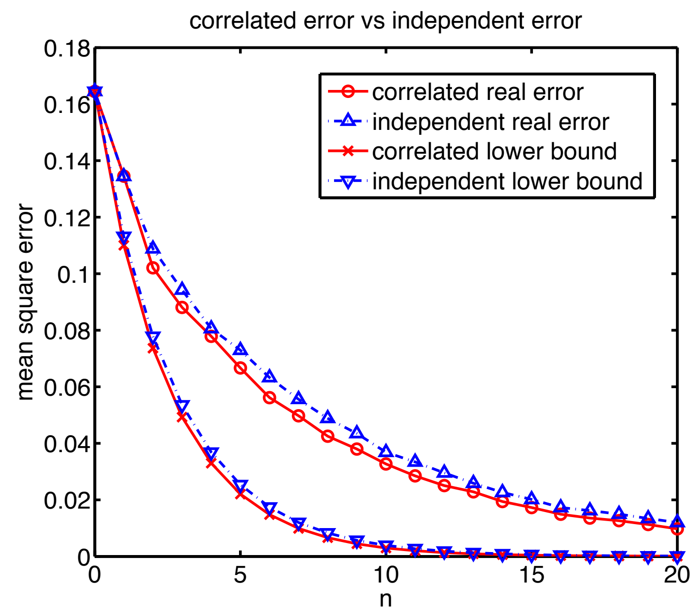
$y =$	0,0	0,1	1,0	1,1
0, 0	0.62	0.17	0.17	0.04
0, 1	0.21	0.57	0.06	0.16
1,0	0.11	0.03	0.68	0.18
1,1	0.11	0.02	0.16	0.71

a) Correlated

y=	0	1
f_0^1	0.79	0.21
f_1^1	0.14	0.86
f_0^2	0.79	0.21
f_1^2	0.27	0.73

b) Independent





Extensions

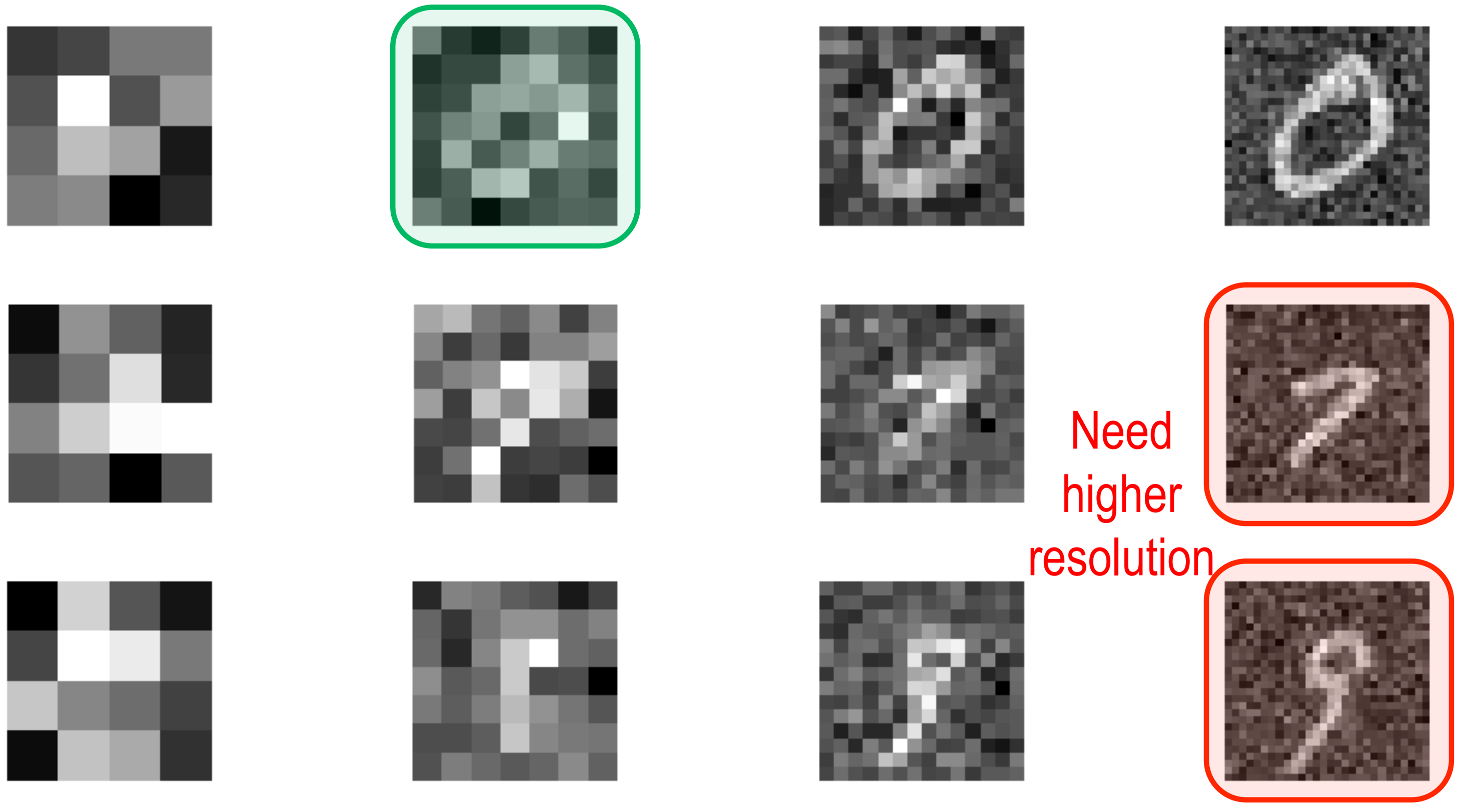
- **Can allow for choice of sensor mode at a cost**
  - Changes measurement distribution of the channel
  - Objective is to minimize expected reduction in differential entropy minus the cost of sensing
  - Snitzman et al '13, C.-Ding '16
- **But, cannot allow for cost to depend on the choice of A**
  - Loses property that cost-to-go in dynamic programming is related to differential entropy
  - Optimal strategy unknown, not likely myopic
  - Counterexamples available
- **No extension to discrete spaces for X**
  - Cannot find query sets to match operating point $\boldsymbol{u}^*$

Data-driven active sensing

- **Previous models require parametric characterizations of uncertainty**
  - Are there theories that `learn' the feedback strategies from data rather than deriving them from models?
- **Study new class of problems: Machine learning with sensing budget**
  - Collection of features is costly
  - Not all features are needed for decisions
  - Deciding to measure a feature depends on what information has been collected to date
  - Recent results by Wang, Saligrama, Trapeznikov, C. ('13-'17)

Motivating Example

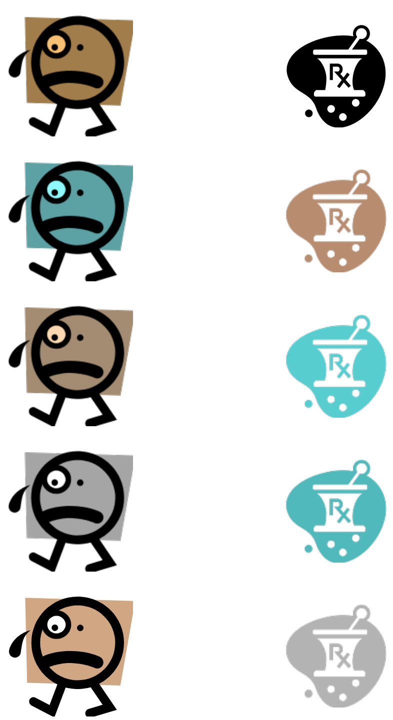
- **Digit recognition: Do we need full resolution?**



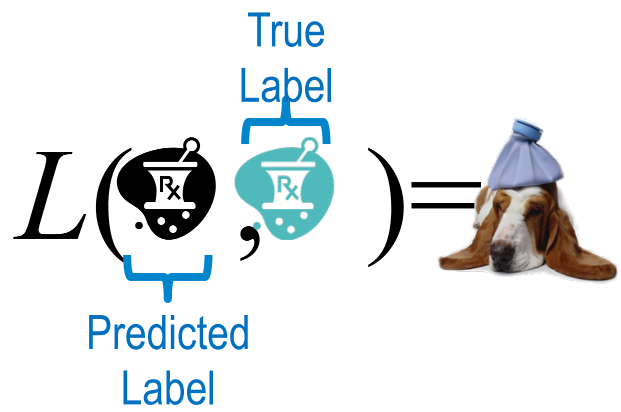
Supervised Learning

Training Data

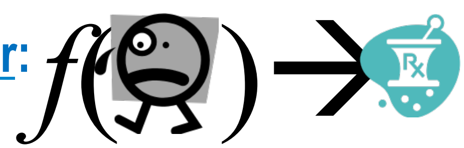
Features: X | **Labels:** Y



Loss:

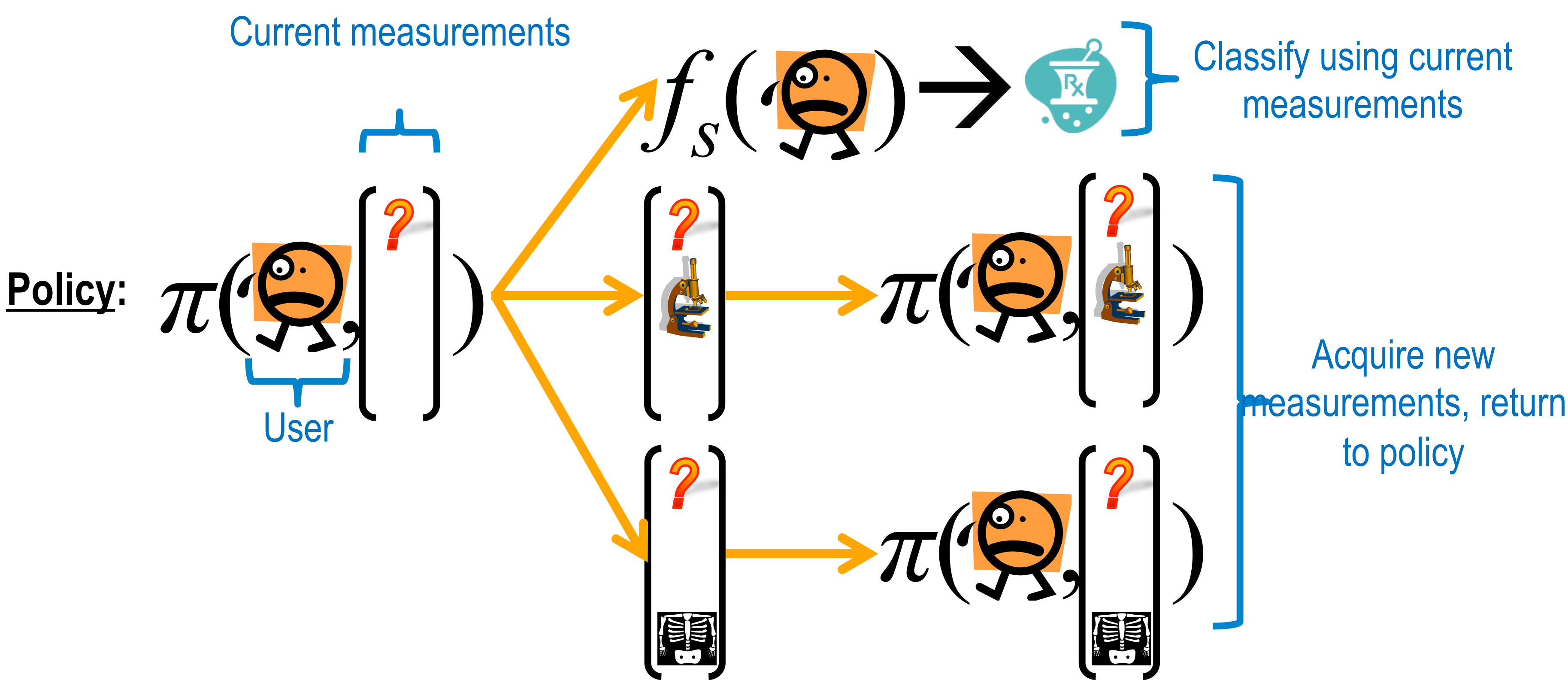


Learn a Classifier:



Adaptive Sensor Selection

- **Goal: Learn a policy π to minimize empirical classification error plus acquisition cost**



Assumption: Have training data

- **Training data with maximal set of features collected**
  - needed to evaluate what can be gained in performance

Assume a subset of sensors/ features:

$$s_j \subseteq \{1, \ldots, K\}$$

Training data

$$\overbrace{\begin{bmatrix} x_1^1 \\ x_1^2 \\ x_1^3 \\ \ldots \\ x_1^{K-2} \\ x_1^{K-1} \\ x_1^K \end{bmatrix}, \begin{bmatrix} x_2^1 \\ x_2^2 \\ x_2^3 \\ \ldots \\ x_2^{K-2} \\ x_2^{K-1} \\ x_2^K \end{bmatrix}, \ldots, \begin{bmatrix} x_N^1 \\ x_N^2 \\ x_N^3 \\ \ldots \\ x_N^{K-2} \\ x_N^{K-1} \\ x_N^K \end{bmatrix}}^{\text{Training data}}$$

Fixed a priori

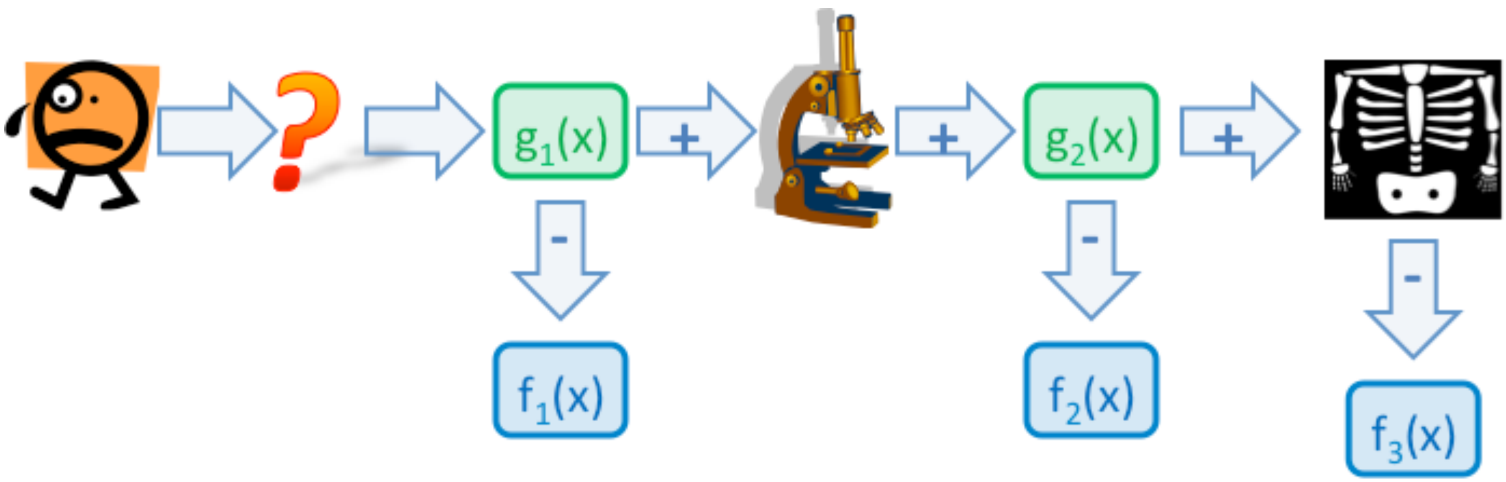
Define the function f_j as the classifier operating on the sensor subset s_j

The cost of using sensor subset s for an example x_i with label y_i can then be defined:

$$L(f_j(x), y) = \underbrace{1_{f_j(x) \neq y}}_{\text{Classification error}} + \overbrace{\sum_{k \in s_j} c_k}^{\text{Cost of sensors in } s}$$

Learning decision strategies in fixed-structures

- **Assume sequence of potential observations is known**
  - Decision is whether to collect more observations or make decision with what is known
  - Generalization of Wald's optimal stopping problem



$$R(g_1, g_2, x, y) = L(f_1(x)) \cdot 1_{g_1(x)<0} + L(f_2(x)) \cdot 1_{g_1(x)\geq 0} 1_{g_2(x)<0} + L(f_3(x)) \cdot 1_{g_1(x)\geq 0} 1_{g_2(x)\geq 0}$$

Empirical Risk Minimization: $\displaystyle \min_{g_1, g_2} \sum_{i=1}^{N} R(g_1, g_2, x_i, y_i)$

Upper bound objectives

- **Bound indicators by convex upper bound surrogates** $\phi(z) \geq 1_{z \leq 0}$

$$\begin{aligned} R(g_1, g_2, x, y) &= L(f_1(x)) \cdot I_{g_i(x) \leq 0} + L(f_2(x)) I_{g_1(x) > 0} I_{g_2(x) \leq 0} + L(f_3(x)) I_{g_1(x) > 0} I_{g_2(x) > 0} \\ &\leq L(f_1(x)) \cdot \phi(g_i(x)) + L(f_2(x)) \phi(-g_1(x)) \phi(g_2(x)) + L(f_3(x)) \phi(-g_1(x)) \phi(-g_2(x)) \end{aligned}$$

- **Problem: non-convex!**
- **Idea: reformulate risk before introducing surrogates**
  - Theorem:

$$R(g_1, g_2, x, y) = C + \max \begin{pmatrix} (\pi_2(x) + \pi_3(x)) \cdot 1_{g_1(x) < 0}, \pi_1(x) \cdot 1_{g_1(x) \geq 0} + \pi_3(x) \cdot 1_{g_2(x) < 0}, \\ \pi_1(x) \cdot 1_{g_1(x) \geq 0} + \pi_2(x) \cdot 1_{g_2(x) \geq 0} \end{pmatrix}$$

  - where $\pi_j(x) = \max_k L(f_k(x)) - L(f_j(x))$

$$C = \max_k L(f_k(x)) - \sum_k \pi_k(x)$$

Now have convex minimization

- **Introducing surrogates leads to a linear program!**

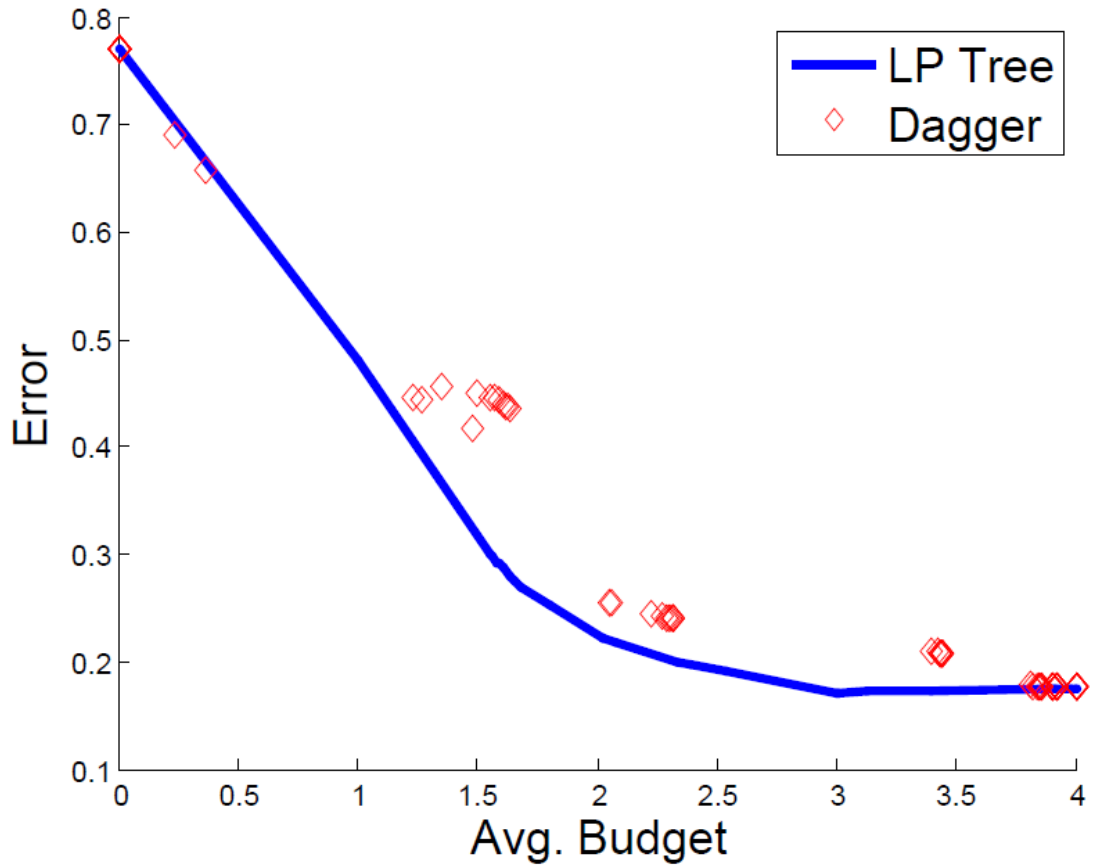
$$\min_{g_1, g_2} \max \begin{pmatrix} (\pi_2(x) + \pi_3(x)) \cdot \phi(g_1(x)), \pi_1(x) \cdot \phi(-g_1(x)) + \pi_3(x) \cdot \phi(g_2(x)), \\ \pi_1(x) \cdot \phi(-g_1(x)) + \pi_2(x) \cdot \phi(-g_2(x)) \end{pmatrix}$$

  - Surrogate $\phi(z) = \max(1-z, 0)$
  - Approach generalizes to arbitrary lengths, as long as order is fixed
  - Approach can also handle tree structures

- **Key idea: Achieve performance close to that of using all the features, while reducing cost of measurement significantly**
  - E.g. risk-based screening at checkpoints to maintain throughput

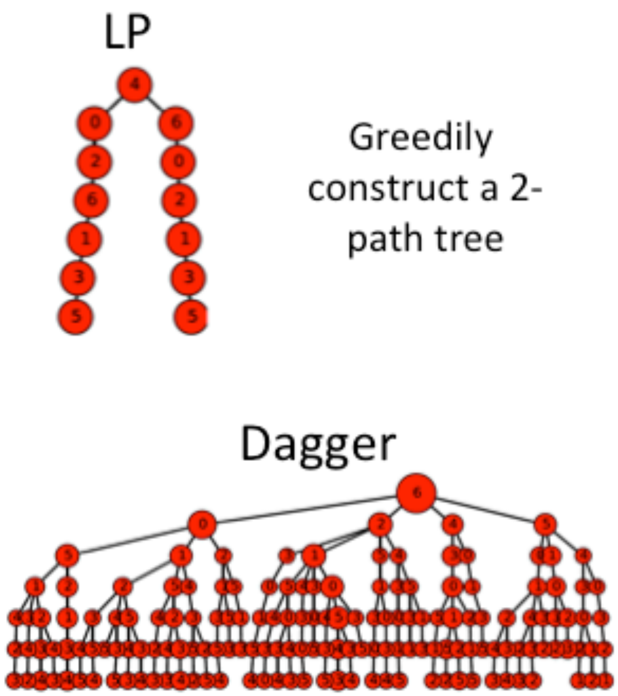
Budget tree experiment

- **Landsat data using 4 spectral bands, each band costs 1**
  - Compare with competing approach (Dagger)

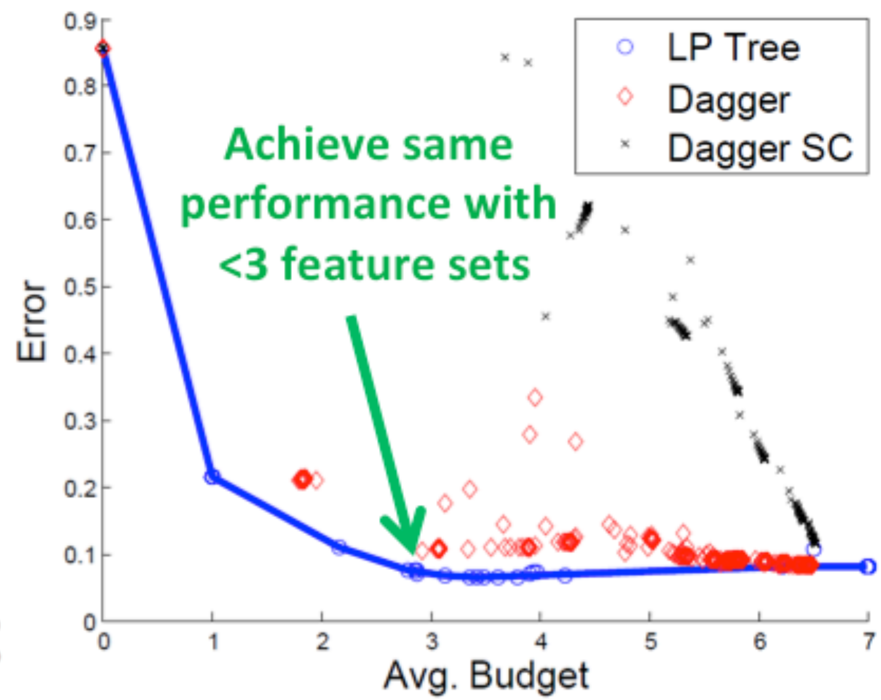


Budget tree experiment 2

- **Image segmentation data set: 7 features**

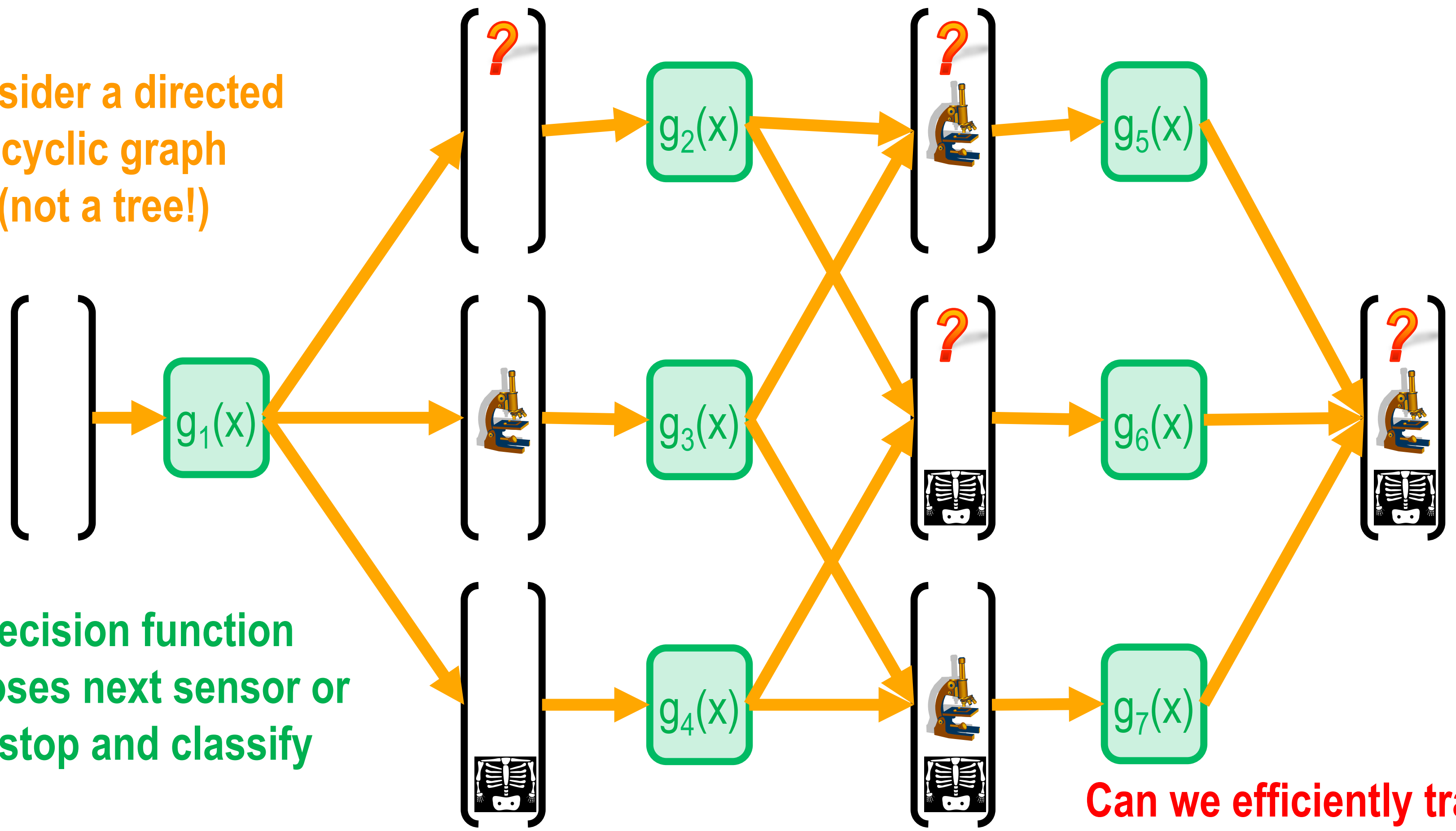


Greedily construct a 2-path tree



What about sensor selection?

Consider a directed acyclic graph (not a tree!)



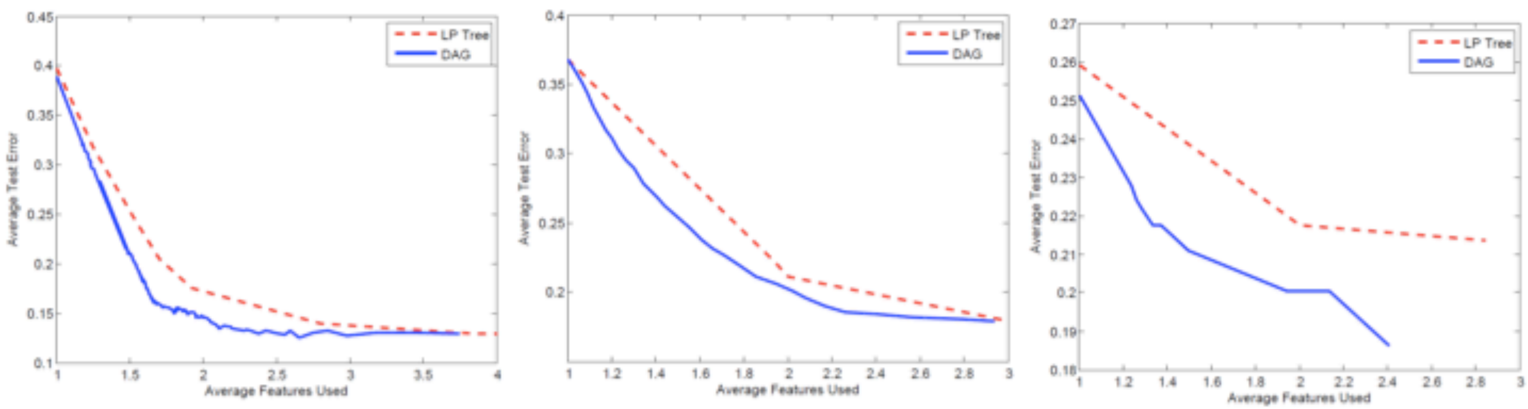
Decision function chooses next sensor or to stop and classify

Can we efficiently train decision functions?

Training decision graphs

- **Yes! Use backward induction (dynamic programming)**
  - Train end classifiers first, use those to get surrogate costs-to-go
  - Recur towards the front in training
  - Theorem: Policy converges to optimal policy as training set grows

Using 3rd order polynomial classifiers and decision functions



landsat   letter   pima

When to do this: small number of sensors, complex functions

When to use LP: simple functions, limited computational resources

Other active sensing problems

- **Active sensing for Gaussian models**
  - Deterministic covariance analysis, mostly using myopic heuristics
  - Long history (Chernoff, Fedorov, Kushner, Athans, …)
- **Active sensing for tracking, classification using approximate stochastic control**
  - Combinatorial dynamic decision problem with uncertainty
  - Approximations and bounds, but not exact results
- **Trajectory optimization for active sensing**
  - Hard! Combines dynamic motion constraints of problems like Traveling salesperson problems with stochastic sequential decision making
- **Guaranteed performance suboptimal algorithms**
  - Exploit structure such as adaptive submodularity, others…
- **Data-driven approaches for test sequencing**
  - Generalization bounds, training data with missing features, deep architectures…

Conclusions

- **Active sensing problems are increasingly important with the deployment of flexible, highly capable sensors**
  - Shared-aperture multifunction RF systems
  - Intelligent UAVs
  - Adaptive diagnosis systems
- **When real-time operation is important, need autonomous decisions rules instead of human-in-the-loop control**
- **Existing theories and results are limited in scope**
  - Can provide some structure and guidance, but hard to guarantee performance
- **Practical solutions will depend on customization of simple models to specific problem instances**
  - Much engineering required…