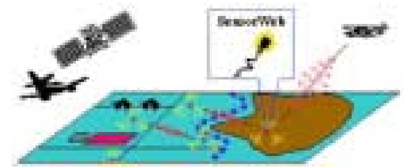
Robust fusion and acquisition of information

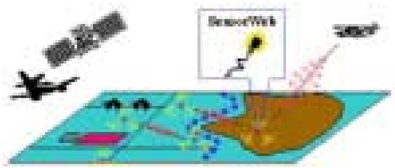
Tommi S. Jaakkola
MIT AI Lab

June 18, 2001



Research topics

- Consistency of information in large sensor networks (with M. Wainwright and A. Willsky); IT 1, RCA 5
- Robust fusion of partial information sources (with A. Corduneanu); IT 1&2, RCA 5 (& 6)
- Optimal acquisition of information through multi-resolution sensors (with H. Siegelman); IT 2, RCA 4&5

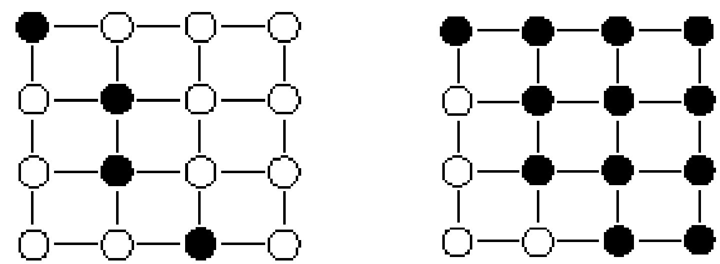


Talk plan

- PART I: Robust fusion of partial information sources
- PART II: acquisition of information from multi-resolution sensors with resource constraints
- Discussion

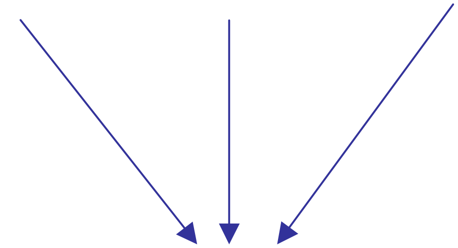
PART I: Robust fusion of partial information sources

- Multiple partial sources of information (e.g., heterogeneous sensors, fragmented databases)
- Sources provide different quality and quantity of useful information
- The problem is to find a robust estimate of the model

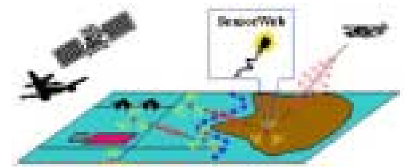


Source 1 Source 2

Source 1 Source 2 Source 3 ...

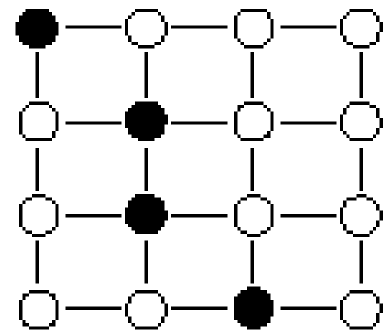


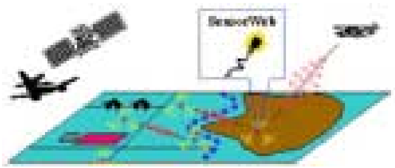
Model estimation



Robust estimation

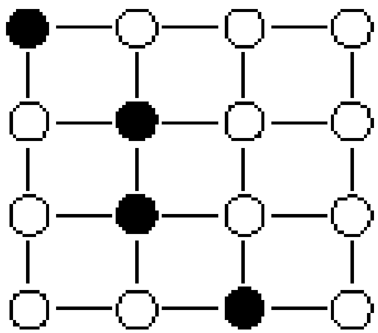
- The problem features:
  - A structured graph model over the domain
  - A likelihood based estimation criterion
  - Incomplete data sources (e.g., missing variables)
- Relevant questions:
  - How should we balance the sources?
  - Are the standard algorithms stable?

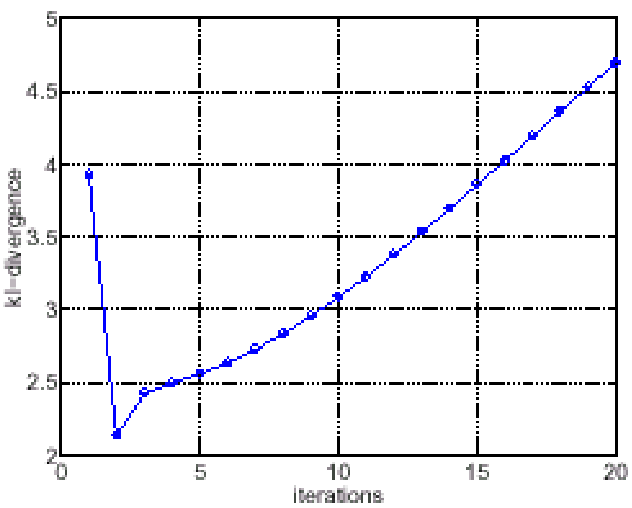


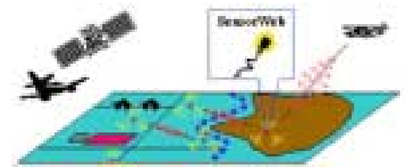


Robust estimation

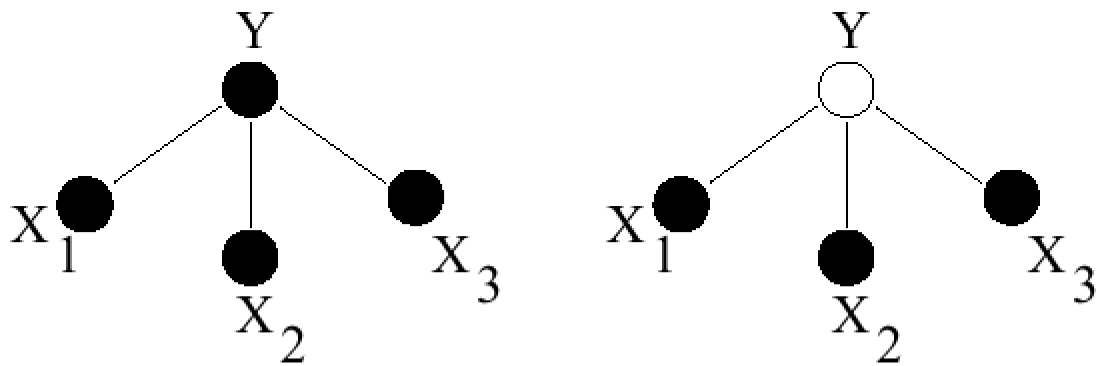
- The problem features:
  - A structured graph model over the domain
  - A likelihood based estimation criterion
  - Incomplete data sources (e.g., missing variables)
- Relevant questions:
  - How should we balance the sources?
  - Are the standard algorithms stable?







Example setting, the EM algorithm



Source 1 (few) Source 2 (many)

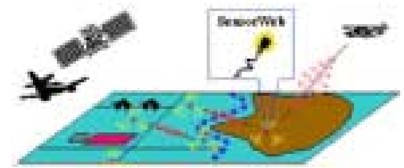
E-step: $P(x,y) \leftarrow (1-\lambda)P_1(x,y) + \lambda Q(y|x)P_2(x)$

M-step: $Q_i(x_i,y) \leftarrow \sum_{x \setminus x_i} P(x,y)$

where λ balances the information sources

- We can collapse these updates into a single operator

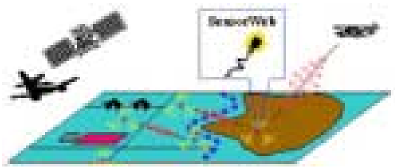
$$Q \leftarrow EM_\lambda(Q)$$



The EM algorithm: why not?

- The problem with the EM-algorithm is the existence of multiple fixed points

$$Q_\lambda^i = EM_\lambda(Q_\lambda^i),\ i = 1, \ldots, L$$

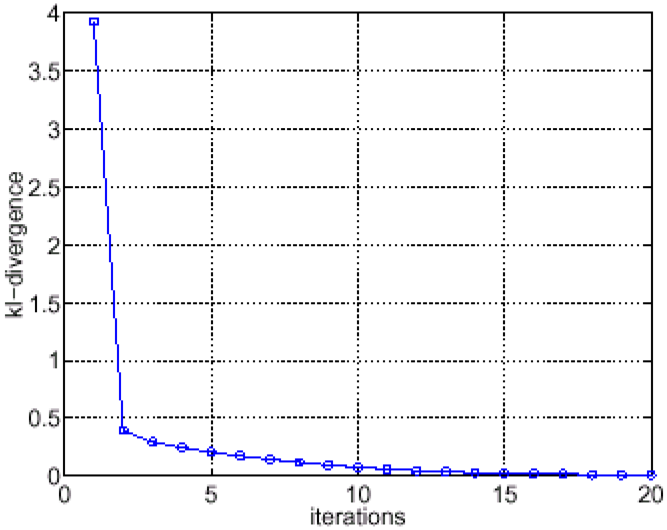


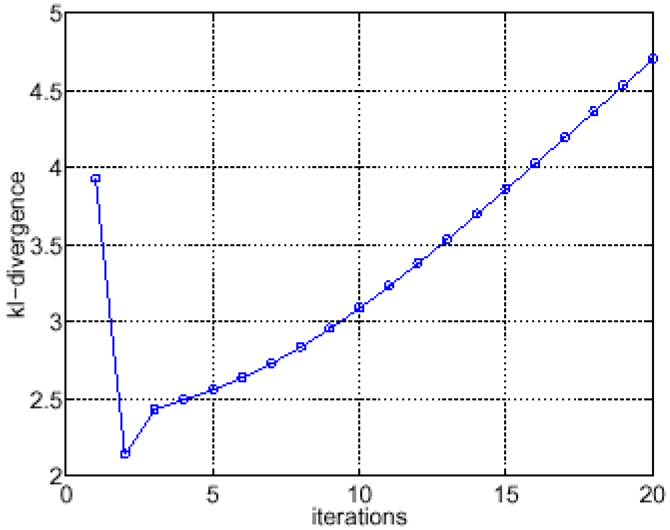
The EM algorithm: why not?

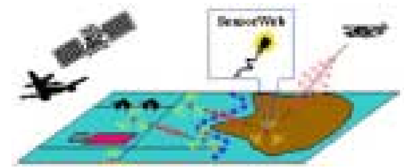
- The problem with the EM-algorithm is the existence of multiple fixed points

$$Q_\lambda^i = EM_\lambda(Q_\lambda^i), \; i = 1, \ldots, L$$

- Some of the fixed points are good, others may be terrible





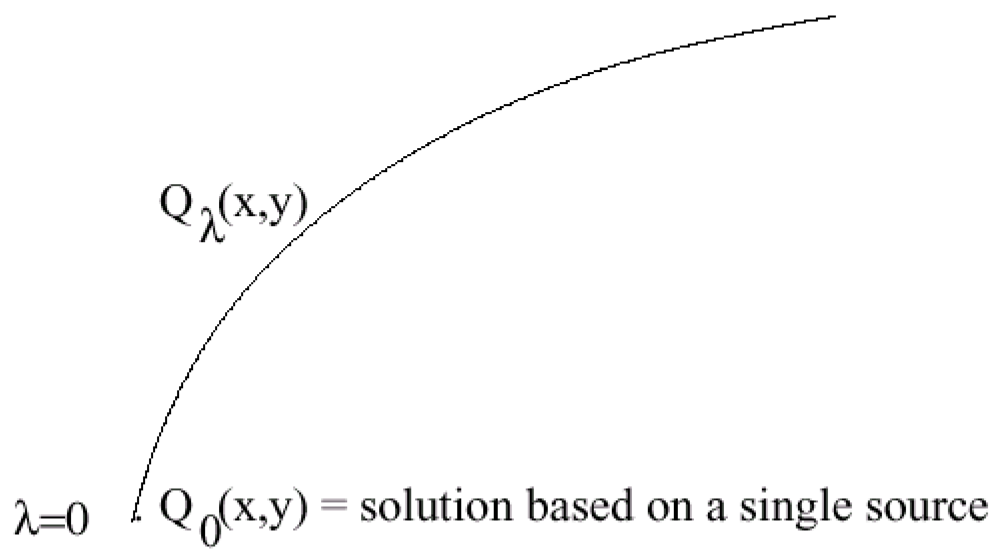


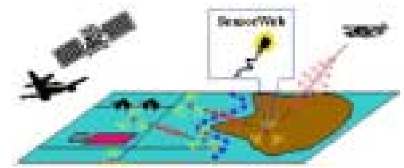
Controlled evolution of fixed points

- Instead of finding a single fixed point, we can trace a continuous path of fixed points starting from a single source (complete)

$$Q_\lambda = EM_\lambda(Q_\lambda), \ \ \lambda \in [0,1)$$

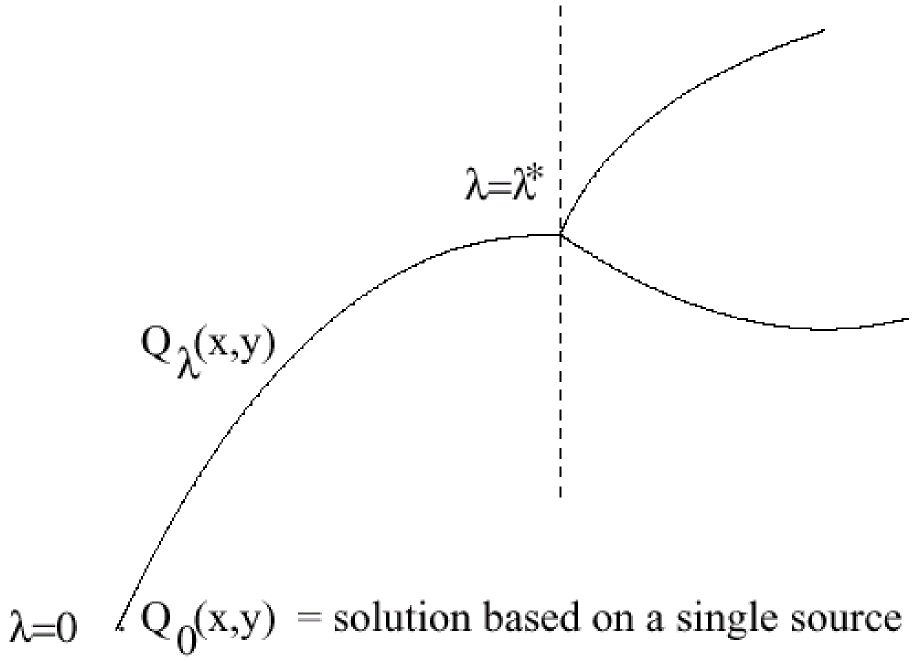
- Each fixed point in this curve is firmly rooted in the maximum likelihood solution based on the initial source or Q_0



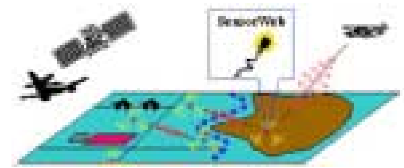


Evolution of fixed points cont'd

- We can explicitly identify any critical points, i.e., points where multiple fixed points begin to emerge



- Although we are typically only interested in solutions before the critical points (predictability), such bifurcation events can be traced further



Critical points

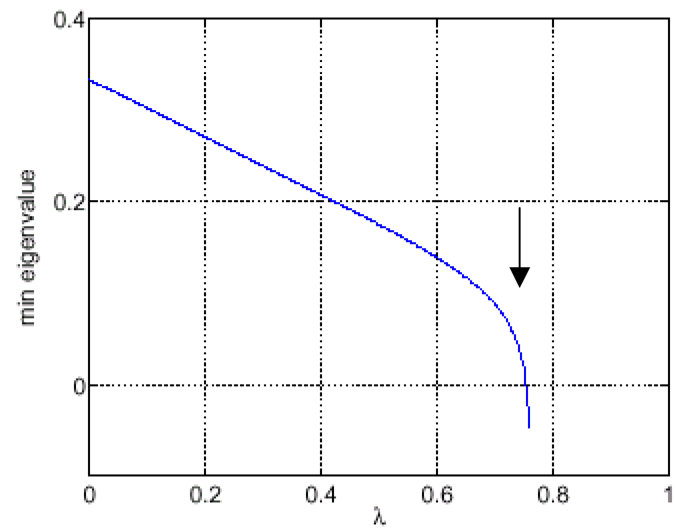
- The differential equation governing the fixed points is

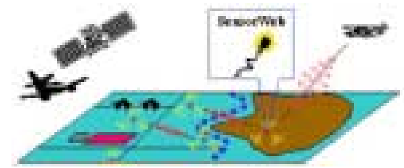
Current solution — EM operator — Initial solution

$$\frac{\partial}{\partial \lambda} Q_\lambda = \left(I - \lambda \nabla_Q EM_1(Q_\lambda)\right)^{-1} \times \left(EM_1(Q_\lambda) - Q_0\right)$$

- The critical points are realized as zero eigenvalues of the (transformed) Jacobian

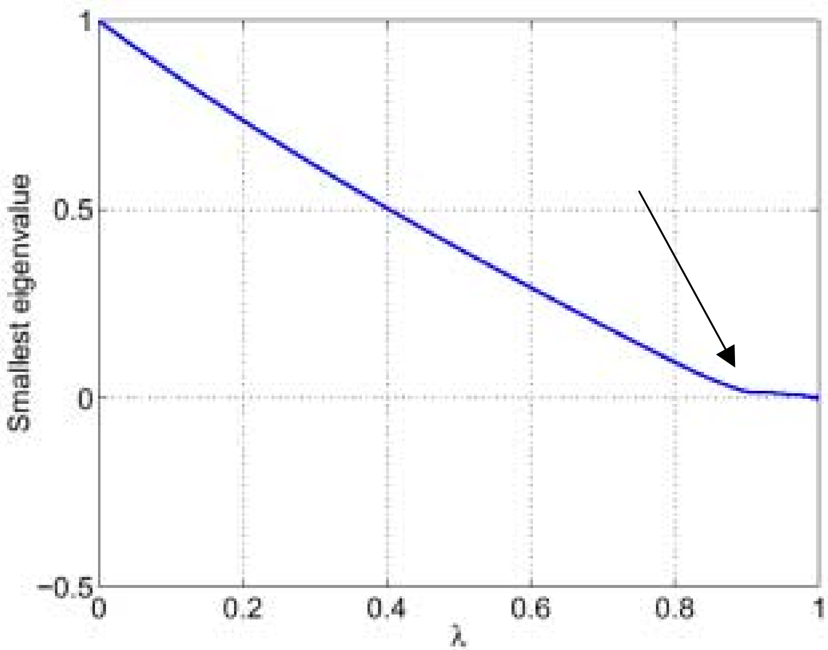
$$\left(I - \lambda \nabla_Q EM_1(Q_\lambda)\right)$$

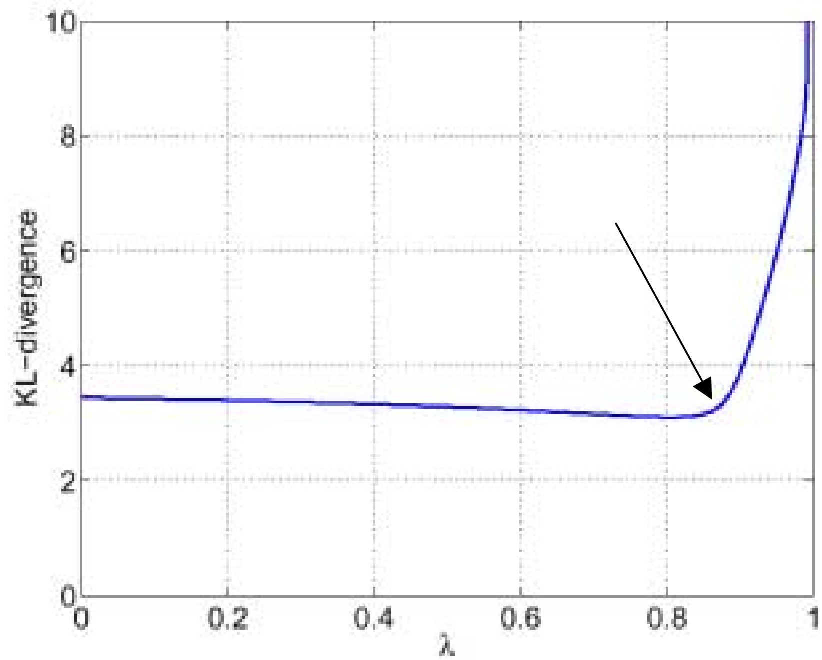


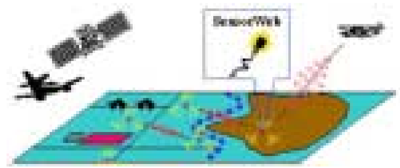


Preliminary fusion results:

- By identifying “critical points” in estimation, we can prevent dramatic loss of accuracy

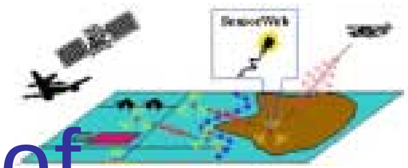




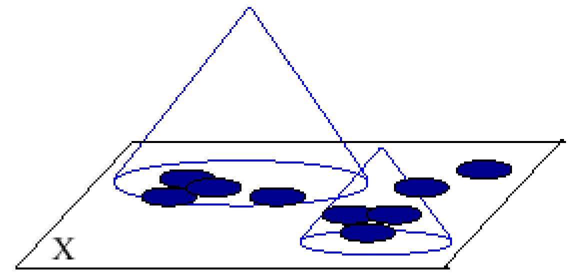


Fusion of partial information sources

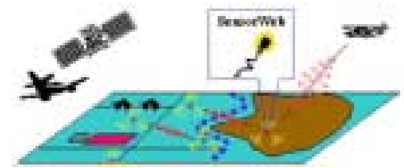
- The standard algorithms can lead to dramatically worse results as we include more incomplete information
- Our alternative approach concerns with a controlled evolution of differential equations governing locally optimal solutions
- By explicitly identifying critical points we can avoid unstable and unpredictable solutions
- Remaining questions:
  - Optimal allocation of sources based on additional information (e.g., uncertainty)
  - Computationally efficient algorithms



PART II: optimal acquisition of information

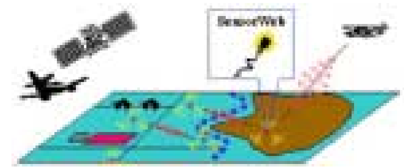


- The setting:
  - Multi-resolution (dynamically defined) sensors
  - Sensor queries governed by resource constraints (bandwidth, cost of deployment, etc.)
- Problems to address:
  - Find (near) optimal strategies for querying the sensors under such resource constraints to
    - quickly locate features or
    - maintain an accurate model of the domain
  - Characterize the inherent trade-offs between cost, expected completion time, sensor types, etc.



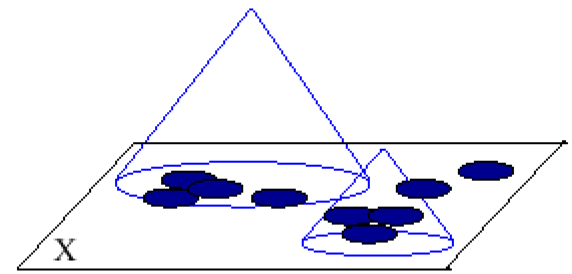
Problem simplification

- We start with a simpler problem
  - E.g., battlefield scenario ←→ database
  - Multi-resolution sensors ←→ cluster hierarchy
  - Sensor measurements ←→ selection/annotation of cluster
  - Resource constraints ←→ clusters/query
- The goal is to locate features in the database with minimum number of queries
- We still have
  - Multiple levels of abstraction
  - Resource constraints
  - The same theoretical questions



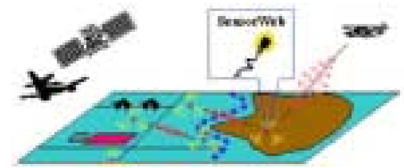
Information, queries, responses

- Information we are after
  - Unknown “weights” associated with the elements in the database (relevance; feature distribution)



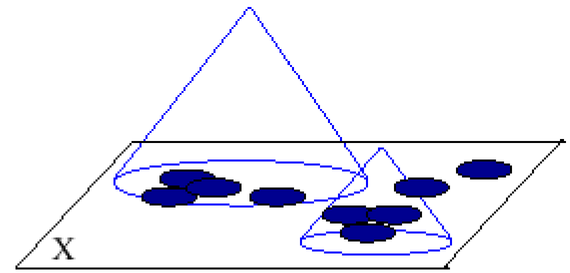
$$\{\theta_x\}, \quad \sum_{x\in X}\theta_x = 1$$

- Permitted queries:
  1. Annotation of a subset of clusters
  2. Cluster choice
- Interpretation of the responses:
  1. Annotations based on thresholded cluster weights
  2. Stochastic weight based selection of clusters

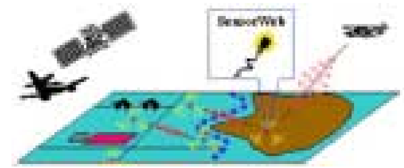


The problem

- Components
  - Available clusters $\{C_1, C_2, \ldots, C_m\}$
  - Response model
  - Query limitations (k clusters)
  - Initial model $P(\theta)$ (Dirichlet)
- We wish to recover the underlying weight distribution with the minimum number of iterations

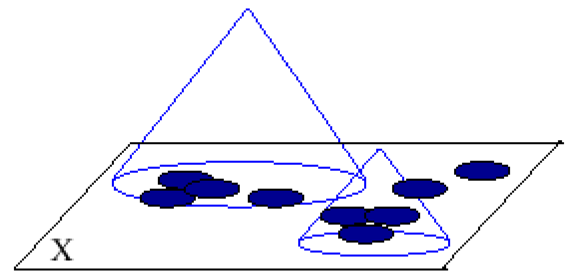


$$\{\theta_x\}, \quad \sum_{x \in X} \theta_x = 1$$

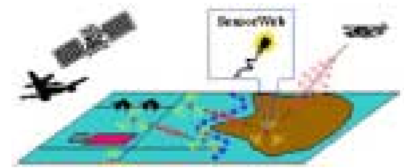


Computational problems

- Algorithmic issues:
  - How to find the optimal query set
  - How to maintain an accurate estimate $P(\theta)$
- Theoretical questions:
  - Bounds on the expected interaction length
  - Trade-offs between the query set size, response model, and the expected interaction time
  - Robustness against structural errors

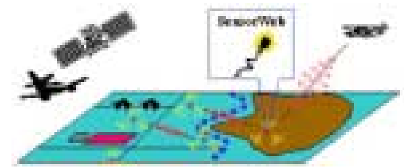


$$\{\theta_x\}, \quad \sum_{x \in X} \theta_x = 1$$



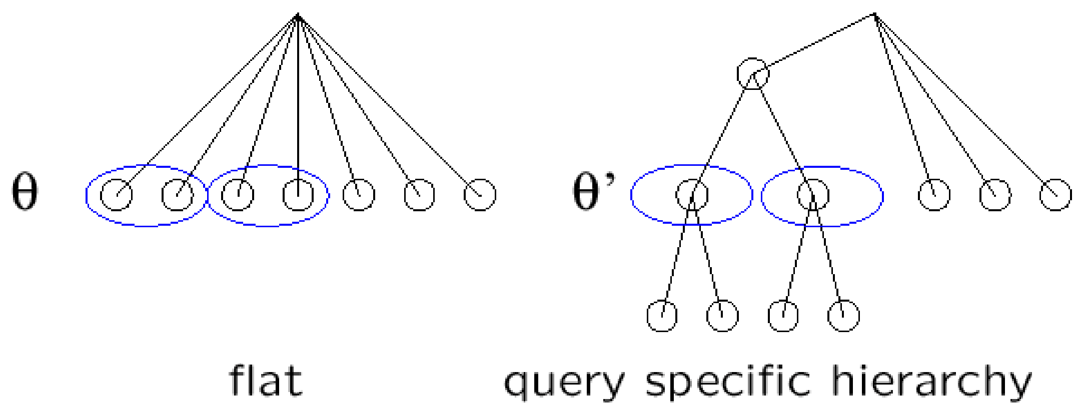
Query set optimization

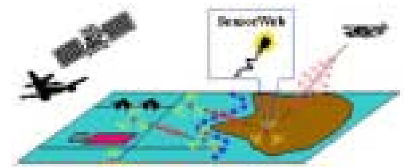
- Criterion: maximize the information we stand to gain from the response, i.e., $I(\theta; y \mid S)$



Query set optimization

- Criterion: maximize the information we stand to gain from the response, i.e., $I(\theta'; y \mid S)$
  - To facilitate the optimization, we transform the current Dirichlet estimate $P(\theta)$ into a hierarchical form



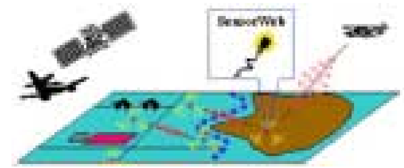


Query set optimization cont'd

- The hierarchical representation leads to an efficient *O(mk)* approximate search algorithm for the query set based on dual "cluster weights"

$$I(\theta'; y \mid S) = \frac{\text{weight' of } S}{\text{weight of } S} + F(\text{weight of } S)$$

(for stochastic responses)



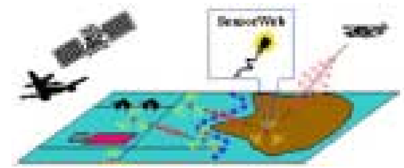
Maintaining the estimate

- The posterior $P(\theta | y)$ no longer remains in the family of interest (e.g., Dirichlet)
- To ensure feasible iterative optimization, we project the posterior back into the family of interest

$$P^{t+1}(\theta) = \arg\min_{Q_\theta} D(P_{\theta|y} \,|\, Q_\theta)$$

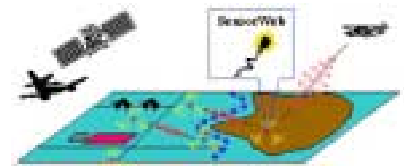
(operation linear in the database size)

- The problem here is closely related to consistent propagation algorithms



Some preliminary results:

- The expected entropy of the projected posterior is still guaranteed to decrease monotonically
- Bounds on the slope of the expected reduction



On-going and future work

- More realistic sensor and sensor response models
- Characterization of the approximation error in the query set optimization
- Theoretical analysis:
  - Bounds on the expected interaction length and the associated trade-offs
  - Robustness against model assumptions