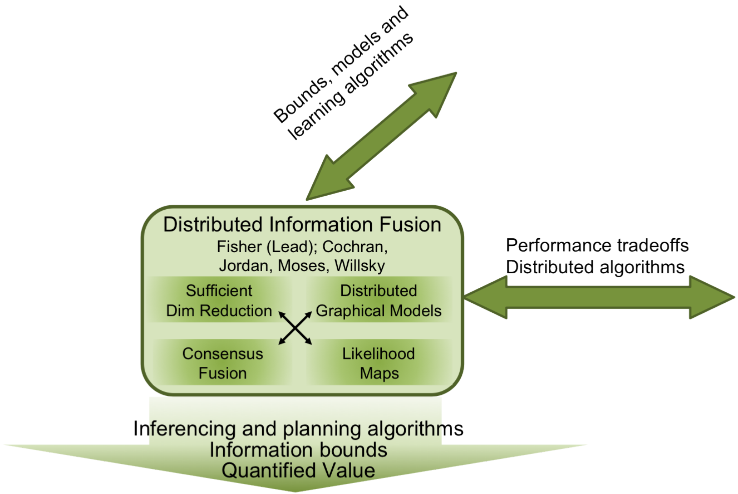
Efficient Information Planning in Graphical Models
computational complexity considerations

John Fisher & Giorgos Papachristoudis, MIT

VITALITE Annual Review 2013

September 9, 2013

Distributed Information Fusion



Established Results

1. A broad class of information measures - f-divergences – are fundamentally linked to bounds on risk. Bartlett et al. [2003], Nguyen et al. [2009]
   - f-divergence $\rightarrow$ ϕ-risk $\rightarrow$ bound on excess risk
2. submodularity – as applied to information measures – is a key enabler. Krause and Guestrin [2005], Williams et al. [2007], Papachristoudis and Fisher III [2012]
   - off-line and on-line performance bounds
   - guarantees on tractable planning methods
   - incorporations of inhomogenous resource constraints
3. Submodular properties are intimately related to the structure of graphical models. Williams et al. [2007]
   - local properties (and computations) yield global properties

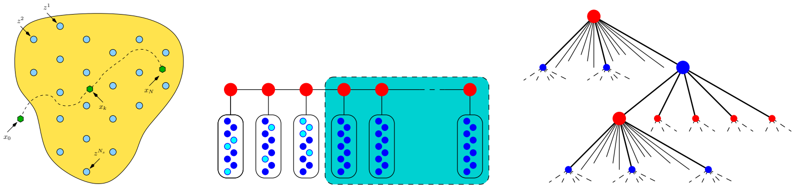
Key Ideas

Information planning posed as combinatorial selection problem over sequential consideration of groups of measurements

1. bounds apply to all sequences (visit paths)
2. information rewards vary across walks
3. evaluation of multiple walks leads to increased information rewards with diminishing probability
4. evaluation of multiple walks leads to tighter upper bound also with reduced probability

Some Context

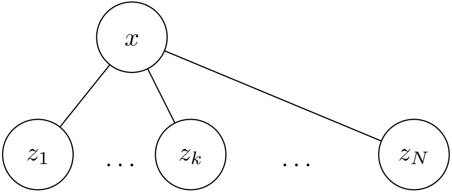
Distributed Sensing



Computational Hurdles

- Evaluating information measures for complex sensors induces a computational bottleneck.
- Evaluating information measures for simple sensors and complex graphs (or even simple graphs) induces a computational bottleneck.
- Due to the branching structure (*i.e.*,, dependence on prior sensor actions), optimal plans are intractable due to exponential complexity.

Inference versus Information



Bayesian Inference

$$p(x|z_1,\ldots,z_k) = p(x)\,\frac{p(z_1|x)}{p(z_1)}\,\frac{p(z_2|x)}{p(z_2|z_1)} \times \ldots \times \frac{p(z_k|x)}{p(z_k|z_1,\ldots,z_{k-1})}$$

Information Gain

$$I(x;z_1,\ldots,z_k) = I(x;z_1) + I(x;z_2) - I(z_1;z_2) + \ldots + \overbrace{I(x;z_k) - \underbrace{I(z_k;z_1,\ldots,z_{k-1})}_{\text{common information}}}^{\text{complementary information}}$$

Submodularity

Given a set V, a real-valued function f on 2^V is submodular if

$$f(A)+f(B) \geq f(A \cup B)+f(A \cap B) \qquad \forall A,B \subseteq V.$$

- Define the set increment function as

$$\rho_S(j) \triangleq f(S \cup j) - f(S).$$

- Equivalently, a real-valued function is submodular if

$$\rho_A(j) \geq \rho_B(j) \quad \forall A \subseteq B \subseteq V \text{ and } j \notin B$$

  that is, the incremental value of j is greater relative to A than to any B which contains A.
- Submodularity captures the notion of "diminishing returns"

Submodularity

- Monotonicity: A real-valued f is monotone if

$$f(A) \leq f(B) \; ; \forall A \subseteq B \text{ , or}$$
$$\rho_S(j) \geq 0 \; ; \forall j \in V, S \subseteq V$$

- Greedy Selection

Batch setting

$$g_j = \arg\max_{u \in V \setminus G^{j-1}} \rho_{G^{j-1}}(u)$$

Sequential setting

$$g_j = \arg\max_{u \in V^{w_j} \setminus G^{j-1}} \rho_{G^{j-1}}(u)$$

- The batch setting chooses from among **all** measurements conditioned on previous selections.
- The sequential setting is restricted to only those available at the current node in the visit walk.

Preliminaries

Notation

- $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ denotes n latent inference variables.
- $\boldsymbol{Z} = \{Z_1, \ldots, Z_n\}$ denotes n measurement vectors.
- Each Z_t is comprised of N_t measurements corresponding to variable X_t.
- $V^t = \{1, \ldots, N_t\}$ indicate measurement indices, *i.e.*, observation sets.
- $Z_i \perp\!\!\!\perp Z_j \mid \boldsymbol{X}$: Measurements are independent given $\boldsymbol{X}$.

Reward function:

- $f : 2^V \rightarrow \mathbb{R}$: a set function that captures the value of sensing actions.

Cost function:

- $c : 2^V \rightarrow \mathbb{R}_+$: a nonnegative set function that
  - quantifies the cost of a subset, and
  - where costs are assumed to be additive over the elements of the subset.

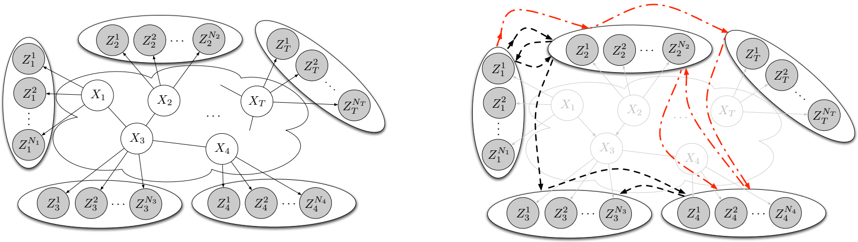
$$c(S) = \sum_{j \in S} c_j$$

.

Sequential Setting

- Goal: Choose k_1 from V^1, …, k_n from V^n:

$$\mathscr{O} = \underset{|S^1|\leq k_1, \ldots |S^n| \leq k_n}{\arg\max} \; f(S) \quad \text{where} \quad S = \bigcup_{t=1}^{n} S^t \text{ and } S^i \cap S^j = \emptyset, \forall i \neq j.$$

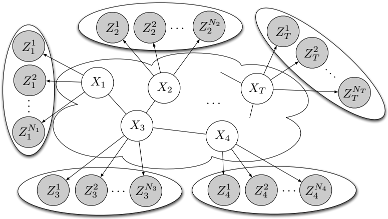


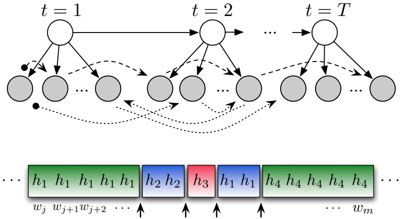
- N_t measurements for each hidden variable X_t.
- Visit walk: define the M-length visit walk as the order $\{w_1, \ldots, w_M\}$ in which we visit observation sets V^t during a selection process.

Sequential Setting

- Goal: Choose k_1 from V^1, ..., k_n from V^n:

$$\mathscr{O} = \underset{|S^1|\le k_1,\ldots|S^n|\le k_n}{\arg\max} \; f(S) \quad \text{where} \quad S = \bigcup_{t=1}^{n} S^t \text{ and } S^i \cap S^j = \emptyset, \forall i \neq j.$$





- Analysis specialized to Markov Chains and LQG models
- Extends to trees and polytrees

Gaussian Markov Chains

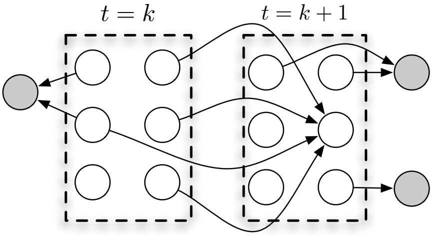
- We consider Gaussian Markov chains for convenience in derivations. The underlying dynamical system is:

$$\mathsf{X}_k = A_{k-1}\mathsf{X}_{k-1} + \mathsf{V}_{k-1}$$
$$\mathsf{Y}_k = C_k\mathsf{X}_k + \mathsf{W}_k,$$

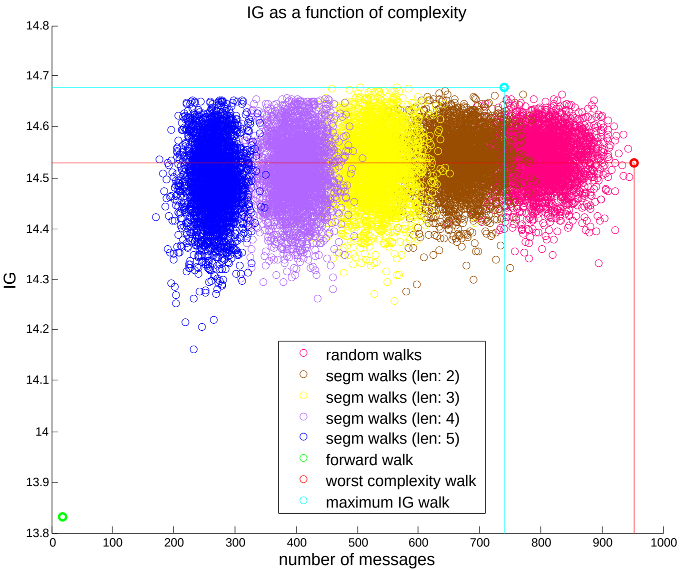
  where $\mathsf{X}_0 \sim \mathcal{N}(\tilde{x}_0, \tilde{\Sigma}_0), \mathsf{V}_{k-1} \sim \mathcal{N}(0, Q_{k-1}), \mathsf{W}_k \sim \mathcal{N}(0, R_k)$.
  (A Markov chain is shown in the upper right figure.)
- Results can be generalized to trees and polytrees.

Sparsity

- Usually, measurements are obtained from a small subset of the underlying process.
- A hidden variable depends only on a restricted set of hidden variables of the previous time point.



Emprical Results



References I

P. L. Bartlett, M. I. Jordan, and J. D. Mcauliffe. Convexity, classification, and risk bounds. Journal of the American Statistical Association, 2003.

A. Krause and C. Guestrin. Near-optimal nonmyopic value of information in graphical models. In Uncertainty in Artificial Intelligence, July 2005.

X. Nguyen, M. J. Wainwright, and M. I. Jordan. On surrogate loss functions and f-divergences. Annals of Statistics, 2009.

G. Papachristoudis and J. W. Fisher III. Theoretical guarantees on penalized information gathering. In Proc. IEEE Workshop on Statistical Signal Processing, August 2012. URL `publications/papers/Papachristoudis12SSPWorkshop.pdf`. **.

J. L. Williams, J. W. Fisher III, and A. S. Willsky. Performance guarantees for information theoretic active inference. In M. Meila and X. Shen, editors, Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics, pages 616–623, March 2007. URL `publications/papers/WilFis07AISTATS.pdf`.

**Outgrowth of Supervised Student Research