Data → Information → Action

Reasoning, Uncertainty and Resource Limitations

John Fisher

Sensing, Learning, & Inference Group
Computer Science & Artificial Intelligence Laboratory
Massachusetts Institute of Technology
http://groups.csail.mit.edu/vision/sli/

January 25, 2019

35,000

35,000

# of decisions per day
source: "the internet"

773,618

200

35,000

# of decisions per day
source: "the internet"

70

143,262

773,618

# of decisions per lifetime
source: Daily Mirror, 2011

of **FOOD** decisions per day
source: Waninski, Sobal, 2007

200

35,000

# of decisions per day
source: Iyengar, Lepper, 2000

70

# of decisions per day
source: "the internet"

# of decisions we **REGRET** per lifetime
source: Daily Mirror, 2011

143,262

Decision Making is Hard

Why?

- Accurately assessing risk is challenging.
- Multiple sources of uncertainty.
- Gathering information requires resource expenditures.

Decision Making is Hard

Why?

- Accurately assessing risk is challenging.
- Multiple sources of uncertainty.
- Gathering information requires resource expenditures.

Decision Fatigue

- Decision quality declines under repetitive decision making.
- Reasoning over many options causes stress.
- We are better at relative rather than absolute comparisons.

Decision Making is Hard

Why?

- Accurately assessing risk is challenging.
- Multiple sources of uncertainty.
- Gathering information requires resource expenditures.

Decision Fatigue

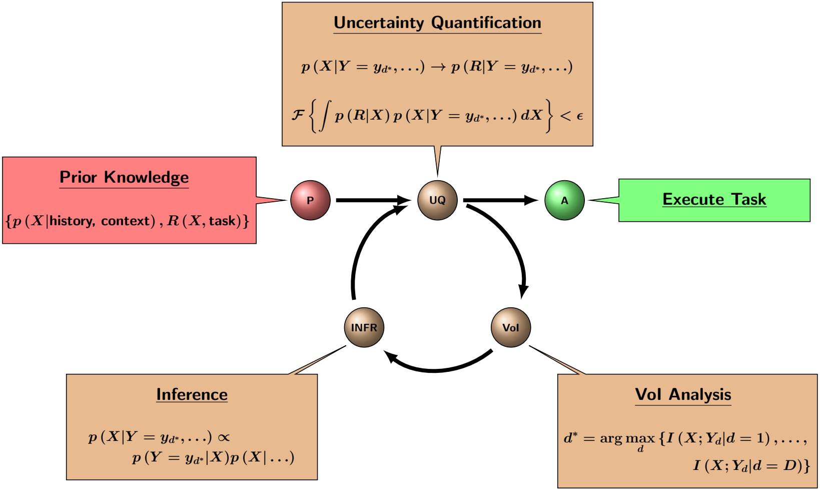
- Decision quality declines under repetitive decision making.
- Reasoning over many options causes stress.
- We are better at relative rather than absolute comparisons.

All of which leads to ...


,1-¿

Information Planning

Bayesian Experimental Design



Begin at the End

Simple Wagers

sometimes not so simple

Assume the wager amount is paid prior to the coin flip.

$b =$ amount wagered $\qquad w =$ win multiple $\qquad p =$ success probability

Reward

$$r = \begin{cases} wb - b & \text{; success} \\ -b & \text{; failure} \end{cases}$$

$$\mathbb{E}\{r\} = (pw - 1)\, b$$

Winning...in expectation

$$\mathbb{E}\{r\} > 0 \rightarrow \begin{cases} p > \frac{1}{w} \\ w > \frac{1-p}{p} + 1 \end{cases}$$

So, the win multiple must be greater than the odds against winning (plus one to account for the prepaid wager amount).

Simple Wagers

sometimes not so simple

You have saved up ~~$1,000,000~~, I mean ¥100,000,000. It is your life savings. With the previous analysis in hand, you're confident risking it all on a favorable wager.

$b = 100,000,000$ ¥ $\quad w = 21 \quad p = 1/20$

Simple Wagers

sometimes not so simple

You have saved up ~~\$1,000,000~~, I mean ¥100,000,000. It is your life savings. With the previous analysis in hand, you're confident risking it all on a favorable wager.

$b = 100,000,000$ ¥ $\qquad w = 21 \qquad p = 1/20$

Careful Analysis

$$w = 21 > \frac{19}{1} + 1 = \left(\frac{1-p}{p} + 1\right)$$

$$\mathbb{E}\{r\} = (pw - 1)\,b = ¥5,000,000$$

Simple Wagers

sometimes not so simple

You have saved up ~~$1,000,000~~, I mean ¥100,000,000. It is your life savings. With the previous analysis in hand, you're confident risking it all on a favorable wager.

$b = 100,000,000$ ¥ $\qquad w = 21 \qquad p = 1/20$

Careful Analysis

✓ $$w = 21 > \frac{19}{1} + 1 = \left(\frac{1-p}{p} + 1\right)$$

✓ $$\mathbb{E}\{r\} = (pw - 1)\, b = ¥5,000,000$$

Not bad, a 5% return!

Simple Wagers

sometimes not so simple

You have saved up ~~\$1,000,000~~, I mean ¥100,000,000. It is your life savings. With the previous analysis in hand, you're confident risking it all on a favorable wager.

$$b = 100,000,000 \text{ ¥} \qquad w = 21 \qquad p = 1/20$$

Careful Analysis

✓ $$w = 21 > \frac{19}{1} + 1 = \left(\frac{1-p}{p} + 1\right)$$

✓ $$\mathbb{E}\{r\} = (pw - 1)\, b = \text{¥}5,000,000$$

Not bad, a 5% return!

Sharing the Wealth

Out of generosity, you share your analysis and convince 19 of your closest friends to make the same wager with their life savings.

Simple Wagers

sometimes not so simple

You have saved up ~~$1,000,000~~, I mean ¥100,000,000. It is your life savings. With the previous analysis in hand, you're confident risking it all on a favorable wager.

$b = 100{,}000{,}000$ ¥ $\quad w = 21 \quad p = 1/20$

Careful Analysis

✓ $$w = 21 > \frac{19}{1} + 1 = \left(\frac{1-p}{p} + 1\right)$$

✓ $$\mathbb{E}\{r\} = (pw - 1)\, b = ¥5{,}000{,}000$$

Not bad, a 5% return!

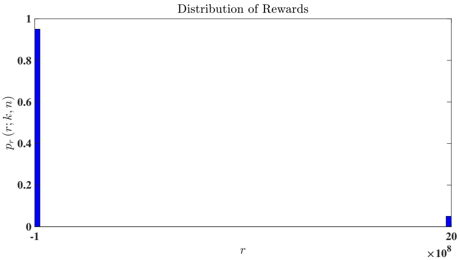
Sharing the Wealth

Out of generosity, you share your analysis and convince 19 of your closest friends to make the same wager with their life savings.

There is a greater than 33% chance that not only will you lose all of your money, but ALL of your friends will lose their money, as well! ☹

Simple Wagers

the distribution of risk/rewards matters too!



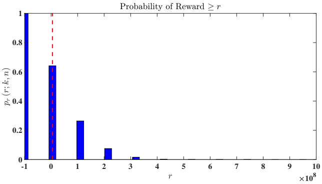
- For a one-time wager, the expected reward is not useful.
- The distribution of risk is a better descriptor.

Simple Wagers

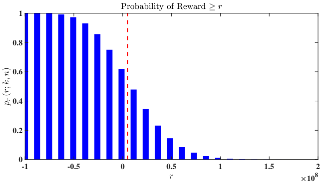
the distribution of risk/rewards matters too!

- Why not amortize over multiple trials?
- Even at 170 trials the probability of a loss is great than 50%.

20 trials



170 trials

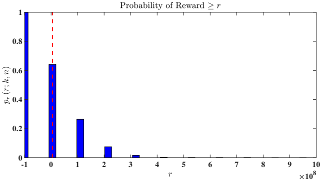


Simple Wagers

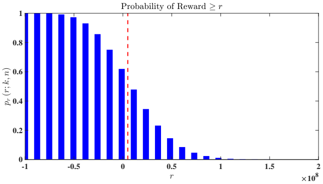
the distribution of risk/rewards matters too!

- Why not amortize over multiple trials?
- Even at 170 trials the probability of a loss is great than 50%.

20 trials



170 trials



- At this point, our best "bet" is to improve our odds.
- Information becomes an important and quantifiable factor.

Actionable Information

Actionable Information

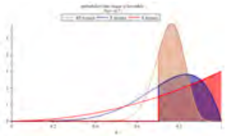
Why are we collecting data?

...information is actionable if it is prescriptive of actions that can be taken to either improve upon the state of uncertainty for a particular task or allow one to accurately evaluate the cost of ancillary decisions related to the task.

-original source in dispute

The perfect is the enemy of the good.

-Voltaire, 1764 (though, he probably said it in French)



Information and Risk

Kelly [1956]

$B = \sum_i b_i$ is the total sum invested in different outcomes.

b_i = wager i $\quad$ w_i = win multiple i $\quad$ p_i = success probability i

Reward

$$r_i = w_i b_i - B$$

$$\mathbb{E}\{r\} = \sum_i p_i r_i = \sum_i p_i w_i b_i - B$$

Some observations

- If all $p_i w_i = 1$ then the game is fair, $\mathbb{E}\{r\} = 0$ for any choice of b_i.
- If any $p_i w_i > 1$ then allocating everying to the $\max_i p_i w_i$ maximizes $\mathbb{E}\{r\}$.

- Maximizing the expected reward over repeated trials $\rightarrow$ Gamblers ruin.
- Maximizing the rate of reward over repeated trials avoids ruin, but requires information to succeed.

Information and Risk

Repeated trials and rates Kelly [1956]

X_n is the outcome of the nth investment. It is random with distribution p.

$b(X) =$ allocations $\quad r(X) = b(X)w(X)$ relative wealth increase

Growth Rate

$$r_n = \prod_{k=1}^{n} r_k(X_k)$$
$$\doteq 2^{nW(b,p)}$$
$$W(b,p) = \mathbb{E}\left(\log r(x)\right)$$
$$= \sum_i p_i \log b_i w_i$$

Optimal Rates & Information

- $b_i = p_i$ optimizes $W(b,p)$ (growth rate is zero for a fair game).
- But p an estimate, information yields an edge.
- With data $p\left(X\right) \rightarrow p\left(X|Y\right)$.

$$W\left(p\left(X|Y\right), b\left(X|Y\right)\right) = I\left(X;Y\right)$$

Representations: More than Answers

The Blind Men and the Elephant

The Blind Men and the Elephant
John Godfrey Saxe (1816-1887)

It was six men of Indostan
To learning much inclined,
Who went to see the Elephant
(Though all of them were blind),
That each by observation
Might satisfy his mind

...six stanzas in which
each demonstrates that they have
questionable judgement...[a]

And so these men of Indostan
Disputed loud and long,
Each in his own opinion
Exceeding stiff and strong,
Though each was partly in the right,
And all were in the wrong!

[a] minor paraphrase

The Blind Men and the Elephant

The measurement is rarely the answer

The Blind Men and the Elephant
John Godfrey Saxe (1816-1887)

It was six men of Indostan
To learning much inclined,
Who went to see the Elephant
(Though all of them were blind),
That each by observation
Might satisfy his mind

...six stanzas in which
each demonstrates that they have
questionable judgement...[a]

And so these men of Indostan
Disputed loud and long,
Each in his own opinion
Exceeding stiff and strong,
Though each was partly in the right,
And all were in the wrong!

[a] minor paraphrase

They end up with this...

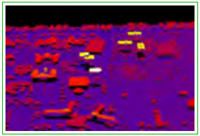
Multi-modal Data Fusion

LiDAR/EO/Semantic Fusion Example [Cabezas et al., 2015]

EO/WAMI

geo-tagged text

LiDAR



How do we mediate the transition from data to reasoning?

Photo Reconstruction

Semantic Reconstruction

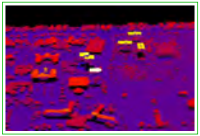
Multi-modal Data Fusion

LiDAR/EO/Semantic Fusion Example [Cabezas et al., 2015]

EO/WAMI

geo-tagged text

LiDAR



Representation

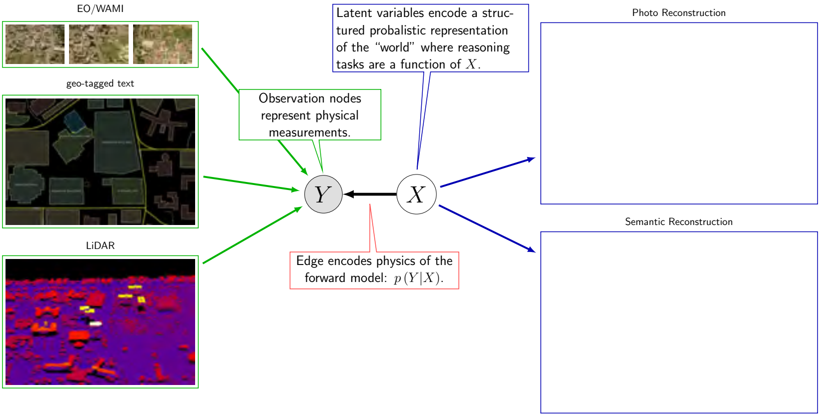
- Quantify uncertainty.
- Explain the sensor physics.
- Accommodate new sensor physics.
- Robust to missing data.
- Support multiple reasoning tasks.
- Facilitate Value of Information reasoning.
- ...

Photo Reconstruction

Semantic Reconstruction

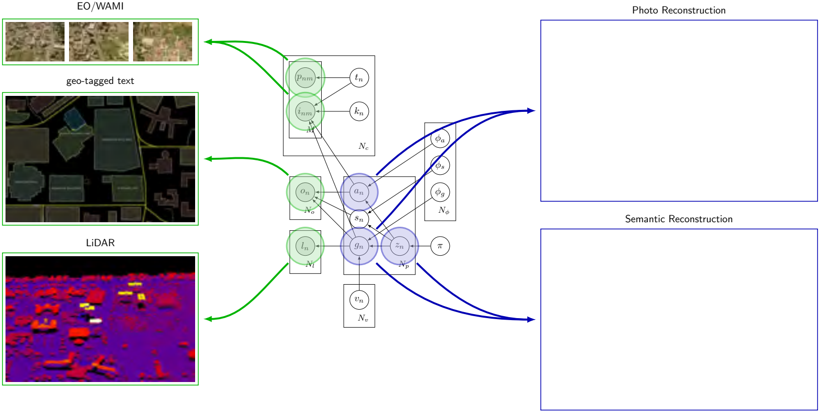
Multi-modal Data Fusion

LiDAR/EO/Semantic Fusion Example [Cabezas et al., 2015]



Multi-modal Data Fusion

LiDAR/EO/Semantic Fusion Example [Cabezas et al., 2015]

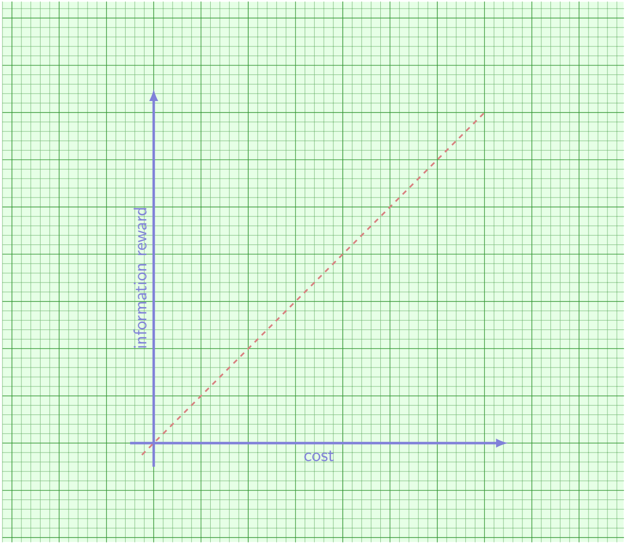


Interesting models generally have complex structure described by a graph. Measurements (shaded nodes) depend on different aspects of the latent representation (unshaded nodes). Reasoning often involves functions of a subset of the latent variables.

Value of Information & Diminishing Returns

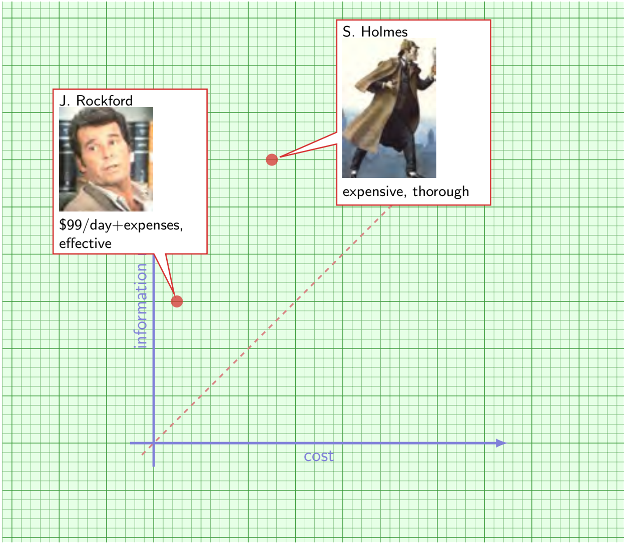
Information Gathering

Quality and Cost of Information Sources



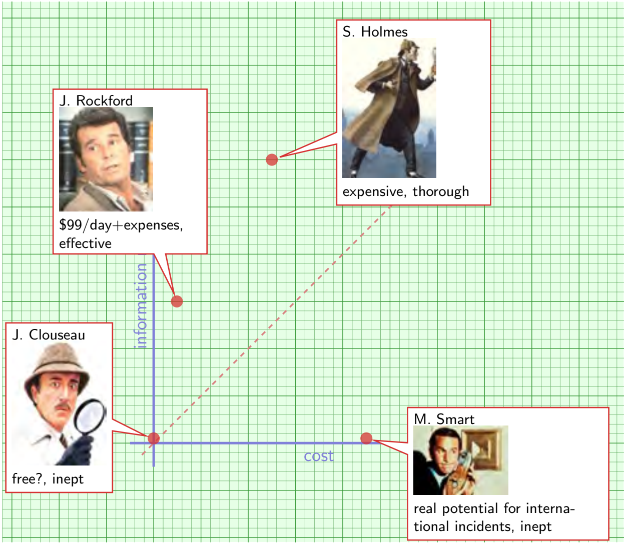
Information Gathering

Quality and Cost of Information Sources



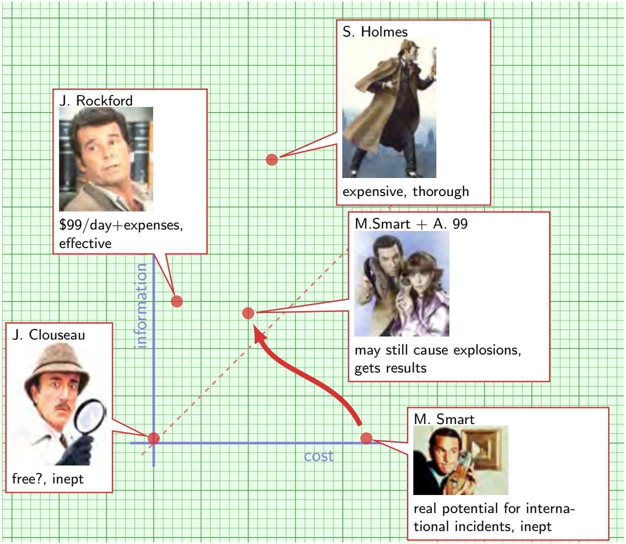
Information Gathering

Quality and Cost of Information Sources



Information Gathering

Quality and Cost of Information Sources



Information Measures and Experimental Design

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-divergences: expectations of convex functions of the likelihood ratio.
    - 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., 2007a, Papachristoudis and Fisher III, 2012]
    - off-line and on-line performance bounds
    - guarantees on tractable planning methods
    - incorporation of inhomogenous resource constraints
3. Submodular properties are intimately related to the structure of graphical models. [Williams et al., 2007a]
    - local properties (and computations) yield global guarantees

Submodularity

Diminishing Returns

- For a set V, a function $f : 2^V \to \mathbb{R}$ is submodular if

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

- The set increment function is defined as

$$\rho_S(j) \triangleq f(S \cup j) - f(S) \quad j \in V, S \subseteq V$$

A real-valued function is submodular if

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

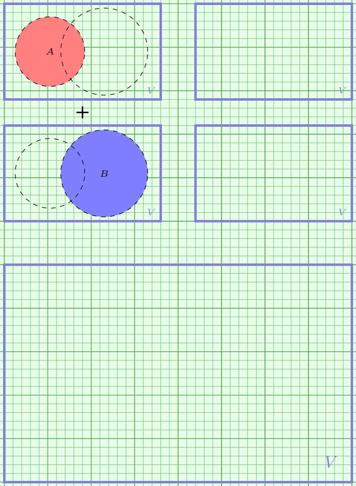
i.e., j has greater incremental value relative to A than to any B containing A.

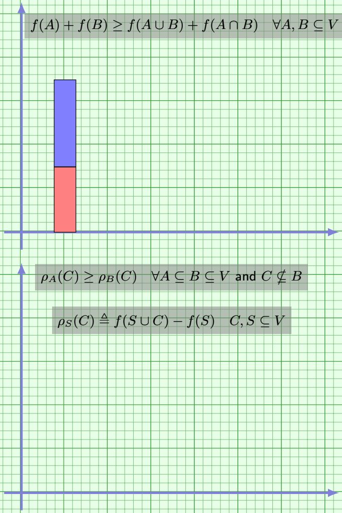
- Monotonicity: A real-valued f is monotone if

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

Submodularity

Graphical Explanation





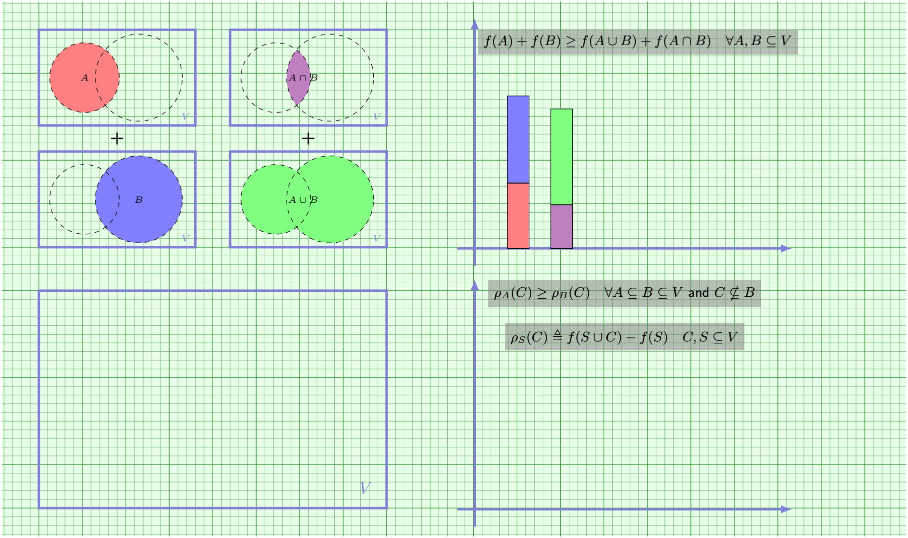
$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B) \quad \forall A, B \subseteq V$$

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

$$\rho_S(C) \triangleq f(S \cup C) - f(S) \quad C, S \subseteq V$$

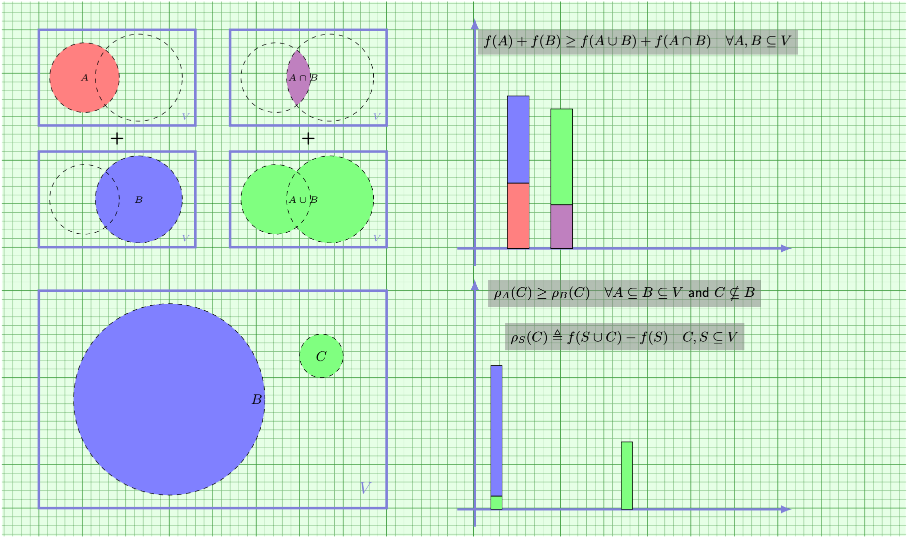
Submodularity

Graphical Explanation



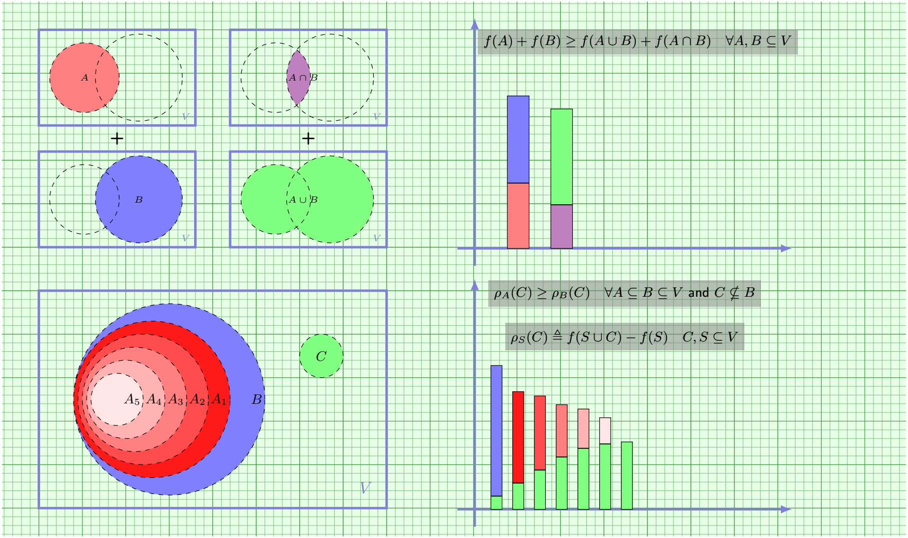
Submodularity

Graphical Explanation



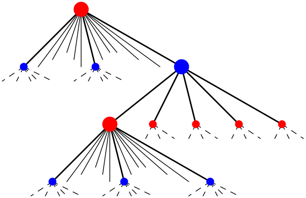
Submodularity

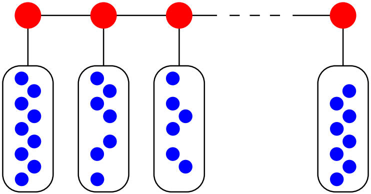
Graphical Explanation



Efficient Information Planning

Tractable greedy selection achieves near-optimal performance.





Williams et al. [2007b] reduces complexity of information gathering formulated as a Markov Decision Process.

$$O([N_s 2^{N_s}]^N M^N) \rightarrow O(N N_s^3)$$

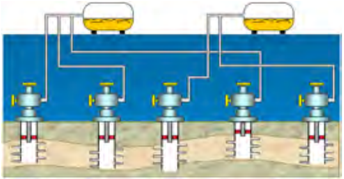
Williams et al. [2007a] the optimal information gathering rate is no greater than twice the greedy information gathering rate.

$$\frac{I\left(X;Z_N^G\right)}{I\left(X;Z_N^*\right)} \geq \frac{1}{2} \quad \forall N$$

N_s: number of sensing actions, N: planning horizon, M: measurement simulation cost.

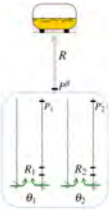
Case Study: Oil & Gas Production

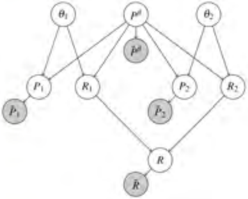
Goal:
- Develop a model and inference algorithm for an off-shore well system.
- Plan a sequence of well tests to estimate unknown properties of the reservoir and each well.

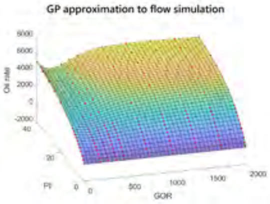


1. Model

- Each well has unknown productivity index (PI), gas-to-oil ratio (GOR), water-cut (WC). ≈ 400 nuisance variables. ≈ 1000 pressure and flow rate measurements.
- Graphical models are used to illustrate structure useful for inference. Gaussian processes are learned to efficiently model multi-phase flow.

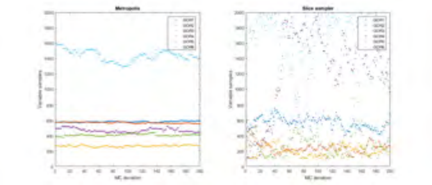






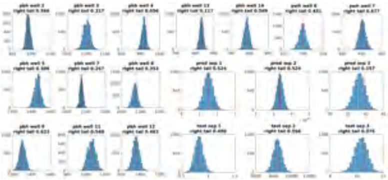
2. Posterior distribution

- Use MCMC to sample from the posterior distribution $p(\theta|y)$.
- Slice sampler has faster mixing; the required extra computation can be parallelized.



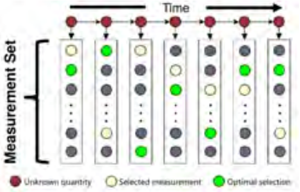
3. Model checking

- Empirically check that measurements from the real system are within the predictive quantiles.



4. Planning

- Taking a measurement is costly and time-consuming, so we would like to select informative measurements.



Planning

- We maximize mutual information between θ and Y:
$I(Y,\theta) = \int p(y) \log p(y)\, dy - \iint p(\theta)\, p(y|\theta) \log p(y|\theta)\, d\theta\, dy$.

- Draw samples $(\theta^{(i)}, y^{(i)})$ from $p(\theta, Y)$, then:

First term

$$\approx \frac{1}{N_y} \sum_{i=1}^{N_y} \log p(y^{(i)})$$

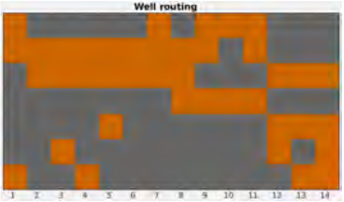
$$\approx \frac{1}{N_y} \sum_{i=1}^{N_y} \log \frac{1}{N_\theta} \sum_{k=1}^{N_\theta} p(y^{(i)} | \, \theta^{(k)})$$

Second term

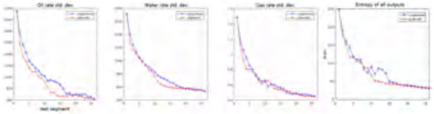
$$\approx \frac{1}{N_\theta} \sum_{i=1}^{N_\theta} H(Y | \, \theta^{(i)})$$

Application: Planning well tests

Well-separator routings can be configured in each test segment. A feasible set of routings are provided by domain experts.

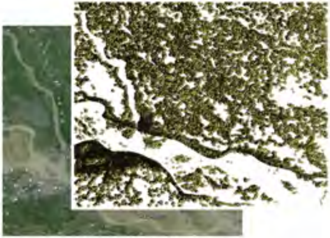


Results - planned vs. expert:



Case Study: Learning to Count

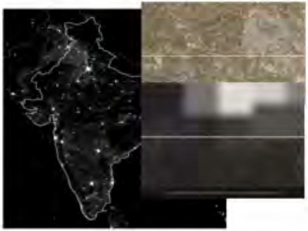
Application: Estimating Electrification Status



Population density

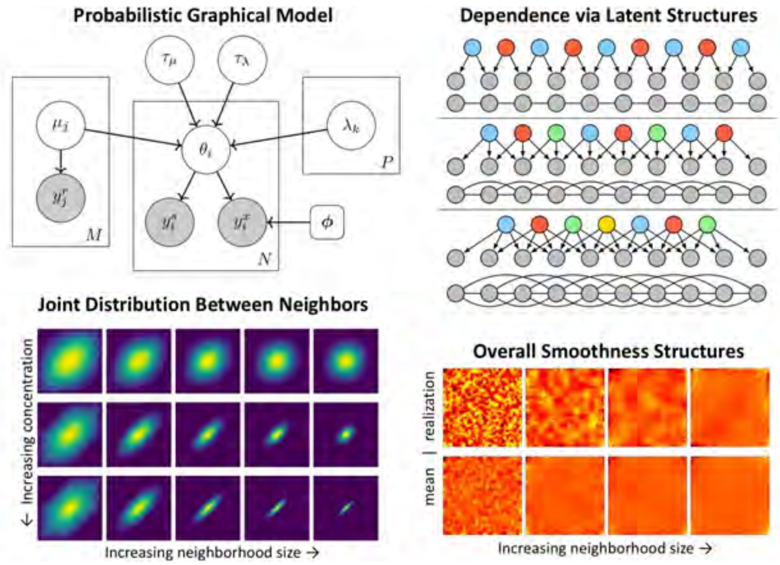
Census and survey data

Existing grid and transformer locations



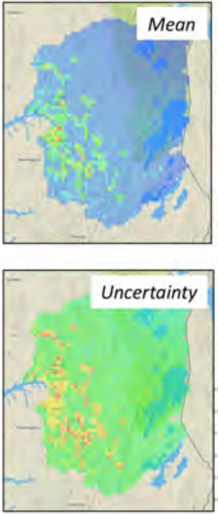
Nighttime lights data

Hierarchical Beta Models with Latent Structure

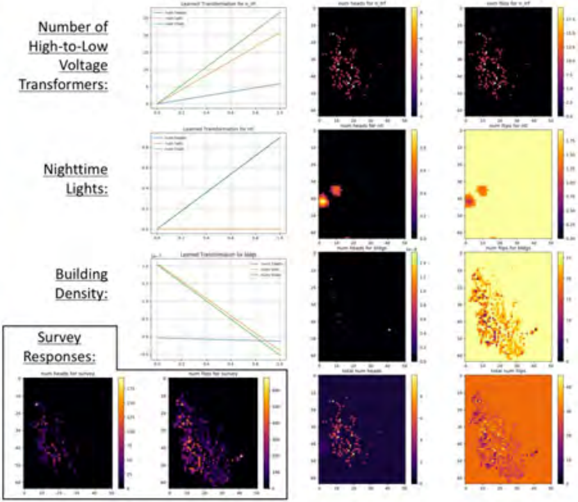


Case Study: Kayonza District in Rwanda

Posterior Analysis



Multi-Modal Data Fusion



Thank you
Questions?
Comments?

References I

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

R. Cabezas, J. Straub, and J. W. Fisher III. Semantically-Aware Aerial Reconstruction from Multi-Modal Data. In International Conference on Computer Vision (ICCV), 2015. URL http://people.csail.mit.edu/rcabezas/pubs/cabezas15_semantics.pdf.

J. Kelly. A new interpretation of information rate. Bell Systems Technical Journal, 1956.

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.

References II

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 2007a. URL publications/papers/WilFis07AISTATS.pdf.

J. L. Williams, J. W. Fisher III, and A. S. Willsky. Approximate dynamic programming for communication-constrained sensor network management. IEEE Transactions on Signal Processing, 55(8):3995–4003, August 2007b. URL publications/papers/williams07a.pdf.

References III