Efficient Distributed Estimation using Adaptive Value of Information based Self-censoring

Beipeng Mu, Girish Chowdhary, Jonathan P. How

Laboratory for Information and Decision Systems
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology

Jun. 18th, 2013

Outline

Distributed Sensing Under Uncertainties and Communication Constraints

- Variety of Multi-agent Distributed Systems
  - Military: formation controls
  - Robotics: team cooperation
  - Agriculture: soil condition monitoring

Multi-agent Distributed Missions

- **Problem**: estimate global parameters to maintain situation awareness and consistency
- **Challenge**
  - dynamic system, uncertain environments, constrained resources

Literature Review

- **Full-Relay**[?]: all measurements are broadcast or relayed
  - Comparable to centralized estimation
  - Inefficient: communication cost very high

- **Consensus** [? ? ? ? ?]: agents average parameters with neighbors
  - Example: Consensus; Gossip
  - Works for arbitrary connected network
  - Comm. cost lower than FR, but agents still communicate at all times
  - Purposefully censoring agents will lead to bias [?]

  - **Random Relay** [?]: randomly censor sensors
    - Communication cost reduced and no bias
    - Randomness in performance and longer convergence time

Literature Review

- **Full-Relay**[**?**]: all measurements are broadcast or relayed
  - Comparable to centralized estimation
  - Inefficient: communication cost very high

- **Consensus** [**? ? ? ? ?**]: agents average parameters with neighbors
  - Example: Consensus; Gossip
  - Works for arbitrary connected network
  - Comm. cost lower than FR, but agents still communicate at all times
  - Purposefully censoring agents will lead to bias [**?**]
  - **Random Relay** [**?**]: randomly censor sensors
    - Communication cost reduced and no bias
    - Randomness in performance and longer convergence time

Literature Review

- **Graphical Models** [? ? ? ?]: agent correlate local distributions
  - Example: channel filters, Bayesian network, Markov random field

  - **Censoring** [? ? ? ? ?]: compute *Value of Information* (VoI) and censor uninformative measurements/agents
    - Comm. cost significantly reduced without much performance loss

  - Not easily scalable to cyclic graphs
    - Multiple paths between two agents $\Rightarrow$ duplicate messages
    - Approximate algorithms leads to bias
    - Exact methods have high overhead computation and communication

Main Result (Mu et al. [? ?])

- Developed Value of Information based Distributed Sensing (VoIDS)
  - Differentiate highly informative agents from less informative ones
  - Agents self-censor when measurements have low-value information
  - Works for arbitrary connected network topologies
  - Trade-off between comm. cost and performance
- Developed **Adaptive VoIDS** (A-VoIDS)
  - Strikes **balance between comm. cost and estimation error**
- Gave theoretical bounds on performance of VoIDS and A-VoIDS

Main Result (Mu et al. [? ?])

- Developed Value of Information based Distributed Sensing (VoIDS)
  - Differentiate highly informative agents from less informative ones
  - Agents self-censor when measurements have low-value information
  - Works for arbitrary connected network topologies
  - Trade-off between comm. cost and performance
- Developed **Adaptive VoIDS** (A-VoIDS)
  - Strikes **balance between comm. cost and estimation error**
- Gave theoretical bounds on performance of VoIDS and A-VoIDS

Main Result (Mu et al. [? ?])

- Developed Value of Information based Distributed Sensing (VoIDS)
  - Differentiate highly informative agents from less informative ones
  - Agents self-censor when measurements have low-value information
  - Works for arbitrary connected network topologies
  - Trade-off between comm. cost and performance
- Developed **Adaptive VoIDS** (A-VoIDS)
  - Strikes **balance between comm. cost and estimation error**
- Gave theoretical bounds on performance of VoIDS and A-VoIDS

Bayesian Inference and Conjugacy

- **Bayesian update** (θ: parameters; ω: hyperparameters)

$$p(\theta|z,\omega) = \frac{p(z|\theta)p(\theta|\omega)}{\int p(z|\theta)p(\theta|\omega)\,d\theta}$$

  - **closed from solution** exists for exponential family distributions
    $\Rightarrow$ efficient Bayesian inference

- **Conjugacy**
  - **Exponential Family likelihood**: $p(\mathbf{x}|\theta) = \exp\left\{\theta^T T(\mathbf{x}) - A(\theta)\right\}$
    - $T(\mathbf{x})$: sufficient statistics; $A(\theta)$: log partition
  - **Conjugate prior**: $p(\theta|\omega,\nu) = \exp\left\{[\omega,\nu]^T[\theta,-A(\theta)] - \Lambda(\omega,\nu)\right\}$
    - ω, ν: hyperparameters; Λ: log partition of conjugate prior
  - **Posterior**: same form as prior; with additive update to hyperparameters

$$\omega \leftarrow \omega + T(\mathbf{z}), \quad \nu \leftarrow \nu + n$$

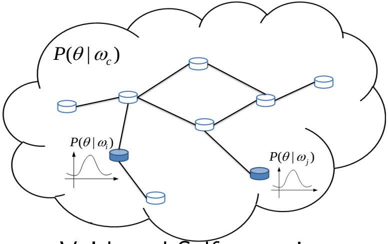
Two Compared Method

- **Hyperparameter Consensus (HPC)** [?]: Linear consensus on hyperparameters:
  - Guaranteed to asymptotically converge to centralized Bayesian estimate
  - Agents continuously communicate with neighbors

- **Random Relay** [?]:
  - **Censoring + Broadcasting**
    - Agents randomly censor themselves
    - Uncensored agents form **Active Set** C, while self-censored agents act as relays (**Relay Set**)
  - Works for dynamic, cyclic network topologies, comm. cost reduced
  - Wide variance in performance, longer convergence time
  - Approach ignores **Value of Information (VoI)**

Key idea: VoI based Self-censoring

- Agents compute VoI of local measurements, compare with **threshold** V^*
- **Informative Set** C
  - An agent declares itself as informative when VoI $> V^*$
  - Informative agents broadcast their messages to neighbors
- **Relay Set**
  - An agent becomes a relay if VoI $\leq V^*$
  - Uninformative agents censor themselves but relay messages for others



VoI based Self-censoring

- Agents estimate global parameters $P(\theta|\omega_c)$
- Dark ones are informative agents with higher VoI
- White ones with lower VoI censor themselves

Value of Information Metric

- **Divergence**: dissimilarity between two distributions
  - Many possible measures [**? ? ?**], usually hard to compute
  - **Kullback-Leibler (KL) divergence** [**? ?**]

$$\mathrm{D}_{KL}(P\|Q) = \int \ln \frac{P(x)}{Q(x)}\, dQ(x)$$

    - Closed-form solution for exponential family distributions
    - Can get exact value with little computation

- **Value of Information**: KL divergence between prior and posterior

$$\mathrm{VoI}(\omega, \nu, \mathbf{z}) = \mathrm{D}_{KL}(\ p(\theta|\omega, \nu)\ \|\ p(\theta|\mathbf{z}, \omega, \nu)\ )$$

  - Closed-form for exponential family distributions [**?**]

VoI Realized Distributed Sensing (VoIDS)

Algorithm 1 VoIDS

initiate hyperparameters $\omega[0], \nu[0]$
for t **do**
  for each agent i **do**
    take measurement, compute local VoI
    if $V_i[t] > V^*$ **then**
      agent i is informative, broadcasts messages
    end if
  end for
  Relay message for informative agents
  for each broadcast message **do**
    update the global posterior
  end for
end for

Performance Guarantees on VoIDS

- Communication Cost

Theorem

Communication frequency of agents → 0 a.s. when $t \to \infty$.

  - Incremental communication cost → 0 a.s when $t \to \infty$.

- Error $e[t]$: KLD between global estimate by agents and centralized Bayes estimate

$$e[t] = D_{\mathrm{KL}}\left(\text{global estimate}\|\text{centralized estimate}\right)$$

Theorem

VoIDS estimation error *is bounded above by* $f(N)\mathrm{V}^*$

Performance Guarantees on VoIDS

- Communication Cost

Theorem

Communication frequency of agents → 0 a.s. when $t \to \infty$.

  - Incremental communication cost → 0 a.s when $t \to \infty$.

- Error $e[t]$: KLD between global estimate by agents and centralized Bayes estimate

$$e[t] = D_{\text{KL}}\left(\text{global estimate}||\text{centralized estimate}\right)$$

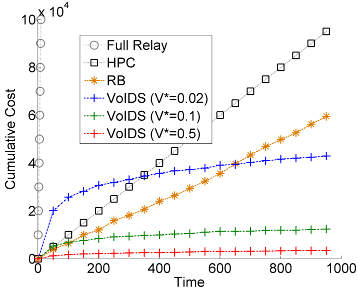
Theorem

VoIDS estimation error *is bounded above by* $f(N)\mathrm{V}^*$

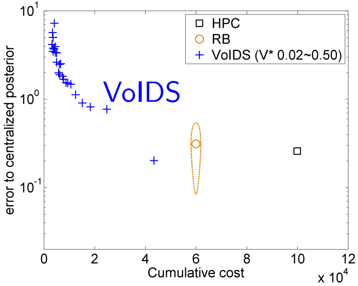
Example: Poisson Distribution

▶ Simulation settings

- Likelihood: Poisson distribution $\mathrm{Poi}(\lambda)$
- Conjugate prior: gamma distribution $\Gamma(\lambda|\alpha,\beta)$
  - Update law: $\alpha \leftarrow \alpha + \mathbf{z}, \quad \beta \leftarrow \beta + n$
- $N = 100$ agents measure corrupted λ: $\lambda_i \sim U(4,6)$



Total Communication Cost



Cost-Error Summary

▶ Communication cost gradually levels off, but error persists

$\Rightarrow$ **trade-off between cost and accuracy**

Adaptive VoI Realized Distributed Sensing (A-VoIDS)

- Control frequency of communication by **adjusting $V^\star$ in response to communication load**

Adaptive VoI

Given c as the desired communication cost,

$$\begin{aligned} &\text{communication cost} < c: && \text{decrease } V^\star \\ &\text{communication cost} \geq c: && \text{increase } V^\star \end{aligned}$$

  - c: targeted communication cost, can be tuned to reflect available communication bandwidth

- Intuition:
  - If many nodes are informative, increase $V^\star$ to reduce communication load
  - If communication load is low, decrease $V^\star$ to increase accuracy

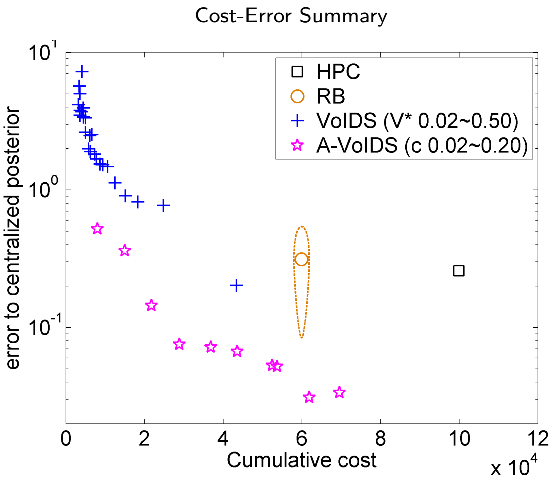
Performance Guarantees of A-VoIDS

Theorem

Estimation error $e[t]$ *satisfies:* $\lim_{t\to\infty} e[t] = 0$ *a.s.*

- Error is asymptotically decreasing $\Rightarrow$ algorithm asymptotically converges to true parameters
- Communication cost in each step is tunable $\Rightarrow$ communication bandwidth can be fully utilized
- Balance between comm. cost and inference error

Simulation Result of Adaptive VoI



- A-VoIDS's performance curve dominates those of other algorithms considered

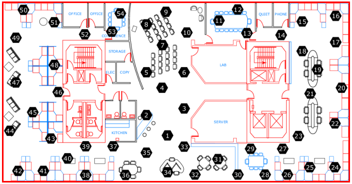
Performance Comparison

Algorithm	**Comm. cost per step**	**KLD Error**
Full Relay	**fixed, high**	**0**
HPC	**fixed, high**	**converge to 0**
Random Broadcast	tunable	converge to 0, randomness
VoIDS	**converge to 0**	**bounded**
A-VoIDS	tunable	**converge to 0**

Intel Lab Dataset [? ? ?]

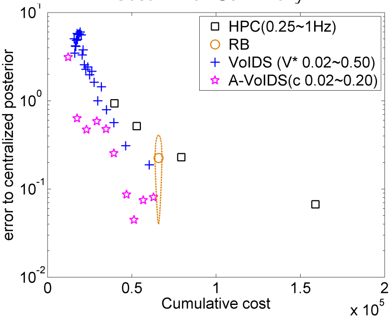
- **Goal**: estimate room temperature distribution
  - Likelihood function: $z \sim \mathcal{N}(\theta, 1)$; conjugate prior: $\theta \sim \mathcal{N}(\mu, \sigma^2)$
  - Update law: $\mu \leftarrow (\mu + \sum_{i=1}^{n} z_i)/(\frac{1}{\sigma^2} + n)$, $\frac{1}{\sigma^2} \leftarrow \frac{1}{\sigma^2} + n$
- Sensors collect data every 30s, update global posterior every 1s.

Sensor Layout



54 sensors in Intel Berkeley Lab.
collect time stamped information
such as temperature, humidity, light

Cost-Error Summary



- A-VoIDS's performance curve closer to bottom-left of plot
  $\Rightarrow$ better balance between accuracy and comm. cost

Conclusions

- Presented Value of Information (VoI) based Distributed Sensing (VoIDS) algorithm
  - Overcome known shortcomings (excessive communication cost and slow convergence speed) of traditional consensus based algorithms
  - Does not require knowledge of network topology
  - Not limited to acyclic networks
  - However, dynamic trade-off exists between estimation accuracy and communication cost
- Presented Adaptive-VoIDS (A-VoIDS) algorithm
  - Adaptively change VoI threshold to better exploit available communication bandwidth
  - Better balance comm. cost and error
- Initial results suggest VoIDS and A-VoIDS work well with real data
- Theoretical bounds on VoIDS and A-VoIDS provided

Future Work

- Consider more complicated situations
  - e.g. multi-variance, dynamic estimation problems
- Other metrics on Value of Information
  - Approximation algorithms and metrics on VoI in no conjugate, no closed form cases
- Non-exponential family distributions
  - e.g. multi-model distributions, non-analytical pdfs
- Information exploitation
  - Mobile agents
  - VoI based planning

References I

VoI Realized Distributed Sensing (VoIDS)

- **Informative Set C** and **Relay Set**
  - Agents locally compute VoI of buffered measurements
  - Broadcast if $\mathrm{VoI} \geq V^\star$; otherwise self-censor local measurements

- **Algorithm**:
  1. Initialize: $\omega_i[0] = \omega, \quad \nu_i[0] = \nu$
  2. Take measurements:

$$S_i[t] = S_i[t-1] + T(z_i[t]), \quad n_i[t] = n_i[t-1] + 1$$

  3. Agents locally check VoI $\boxed{i \in C[t] \quad \text{if} \quad V_i > V^\star}$
  4. Broadcast and relay informative updates $S_{C[t]}[t], \quad n_{C[t]}[t]$
  5. Every node computes posterior

$$\forall i, \, \omega_i[t] = \omega_i[t-1] + \sum_{j \in \nu[t]} S_j[t], \quad \nu_i[t] = \nu_i[t-1] + \sum_{j \in \nu[t]} n_j[t]$$

- Overall cost depends on size of informative set, or V^*

Hyperparameter Consensus Algorithm (HPC)

- Fusion can be done by running linear consensus on hyperparameters
  1. Initialization

$$\omega_i[0] = \omega^- + \beta_i(\omega_i - \omega^-), \quad \nu_i[0] = \nu^- + \beta_i(\nu_i - \nu^-)$$

  2. Measurement update

$$\omega_i[k] \leftarrow \omega_i[k] + \beta_i T(\mathbf{z}_i), \quad \nu_i[k] \leftarrow \nu_i[k] + \beta_i n_i$$

  3. Consensus protocol

$$\omega_i[k+1] = \omega_i[k] + \epsilon \sum_{j \in \mathcal{N}_i} (\omega_j[k] - \omega_i[k]), \quad \nu_i[k+1] = \nu_i[k] + \epsilon \sum_{j \in \mathcal{N}_i} (\nu_j[k] - \nu_i[k])$$

- Variables used
  - $\mathcal{N}_i$: neighborhood of agent i
  - β_i: pre-computed weight of agent i, guarantee convergence on sum. Related to network topology
  - $\epsilon \in (0, 1/\max_i |\mathcal{N}_i|)$ is a weighting constant
- **Theorem** [?]: Algorithm guaranteed to asymptotically converge to Bayesian fused estimate over a strongly connected known network

Censoring

- **Censoring**
  - Censoring output of nodes studied for centralized estimation [**? ? ? ? ? ?**]
  - Cetin et al. [**?**] used a VoI metric to censor measurements on a graphical model in the context of a data association problem
  - Censoring is hard to do on consensus based algorithms ⇒ dynamic network topology causes bias [**?**]
- Censoring + Broadcasting
  - Upon getting a new measurement, agent stores them in a local buffer
  - Agent generates a random number, if bigger than threshold, becomes active (**Active Set** C) and broadcast its updates
  - Inactive agents act as relays (**Relay Set**)
- Highly scalable, can function in dynamic, unknown network topologies

Random Broadcast Algorithm

- **Random Broadcast** (inspired by [**?**])
  1. Initialization: $\omega_i[0] = \omega$, $\nu_i[0] = \nu$
  2. Take measurements: $S[t] = S[t-1] + T(\mathbf{z_i})$, $\nu_i[t] = \nu_i[t-1] + n$
  3. If locally generated random number bigger than a threshold: $i \in C[t]$
  4. Broadcast and relay updates: $S_{C[t]}[t]$, $\nu_{C[t]}[t]$
  5. Every node computes the posterior: $\omega_i[t] = \omega_i[t-1] + \sum_{j \in \nu[t]} S_j[t]$, $\nu_i[t] = \nu_i[t-1] + \sum_{j \in C[t]} n_j[t]$
- Communication cost reduced compared to HPC, but reduced frequency of communication leads to slower convergence
- Wide dispersion in performance (censoring based on a random process)
- Approach ignores **Value of Information (VoI)**

Value of Information Metric

- Alternatives: agents determine VoI for broadcast
  - **Divergence** measure of difference between two distributions
  - Many possible divergence measures [? ? ?]:

Metric	Formula
Kullback-Leibler	$D_{\text{KL}}(p\|\|q) = \int \log(\frac{p}{q})dp(x)$
Renyi	$D_{\alpha}(p\|\|q) = \frac{1}{\alpha-1}\log\int p^{\alpha}q^{1-\alpha}dx$
Chernoff	$D_{c}(p\|\|q) = \log\int p^{\alpha}q^{1-\alpha}dx$
f-divergence	$D_{f}(p\|\|q) = \int f(\frac{p}{q})dq(x)$
Variational	$V(p\|\|q) = \int \|p-q\|dx$
Generalized Matusita	$D_{M}(p\|\|q) = \left[\int \|p^{1/r}-q^{1/r}\|^{r}dx\right]^{1/r},\ r>0$

- Typically no closed-form solutions (special case: Renyi divergence for exponential family)