Energy Efficient Routing for Statistical Inference of Markov Random Fields

A. Anandkumar[1] **L. Tong**[1] **A. Swami**[2]

Cornell University

[1]School of Electrical and Computer Engineering
Cornell University, Ithaca, NY 14853

[2]Army Research Laboratory, Adelphi MD 20783

Conference on Information Sciences and Systems 2007

Supported by the Army Research Laboratory CTA

Classical Distributed Detection and Routing

Distributed Detection

- Quantization rule @ sensors
- Inference rule @ fusion center
- Conditionally IID sensor data
- Communication
  - Perfect reception
  - Rate constraints

Classical Routing

- Generic Performance Metric
  - Throughput, Avg. delay
- Layered architecture separates from application : Suboptimal
- Modular, simple to implement

Issues in Wireless Sensor Networks

Sensor Signal Field

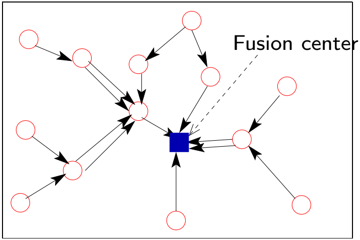
- Large coverage area
- Large number of sensors
- Correlated sensor readings
- Arbitrary sensor placement

Sensor Characteristics

- Limited battery
- Limited processing capability
- Limited transmission range
- Prone to failures

Design of Routing for Detection in Wireless Sensor Networks

Minimum Energy Routing for Inference



Transmission graph

Setup

- Correlated sensor readings
- Optimal detection at fusion center
- Minimum total energy consumption

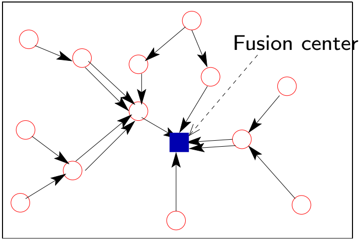
Theorem (Dynkin)

Likelihood function is minimal sufficient statistic for inference

Minimum Energy Routing for Inference

Minimize total energy of routing such that the sequence of transmissions ensures that likelihood function is delivered to fusion center

Minimum Energy Routing for Inference



Transmission graph

Setup

- Correlated sensor readings
- Optimal detection at fusion center
- Minimum total energy consumption

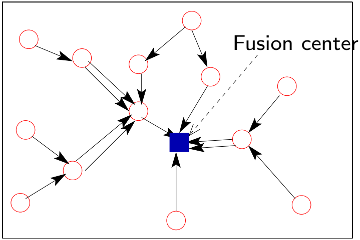
Theorem (Dynkin)

Likelihood function is minimal sufficient statistic for inference

Minimum Energy Routing for Inference

Minimize total energy of routing such that the sequence of transmissions ensures that likelihood function is delivered to fusion center

Minimum Energy Routing for Inference



Transmission graph

Setup

- Correlated sensor readings
- Optimal detection at fusion center
- Minimum total energy consumption

Theorem (Dynkin)

Likelihood function is minimal sufficient statistic for inference

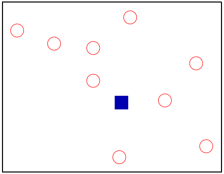
Minimum Energy Routing for Inference

Minimize total energy of routing such that the sequence of transmissions ensures that likelihood function is delivered to fusion center

In-Network Processing

Categories

- Forwarding: Transmission of raw data without processing
- Fusion: Intermediate processing before reaching fusion center



Data Forwarding

- Direct transmission
- Shortest path

Data Fusion

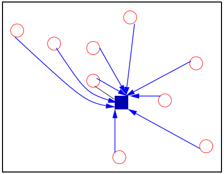
- Energy efficient
- Depends on data model

Influence of Correlation Structure of Data on Fusion Mechanism

In-Network Processing

Categories

- Forwarding: Transmission of raw data without processing
- Fusion: Intermediate processing before reaching fusion center



Data Forwarding

- Direct transmission
- Shortest path

Data Fusion

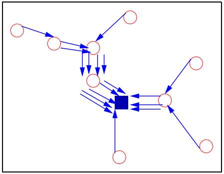
- Energy efficient
- Depends on data model

Influence of Correlation Structure of Data on Fusion Mechanism

In-Network Processing

Categories

- Forwarding: Transmission of raw data without processing
- Fusion: Intermediate processing before reaching fusion center



Data Forwarding

- Direct transmission
- Shortest path

Data Fusion

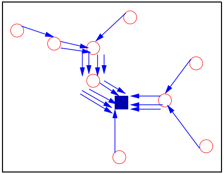
- Energy efficient
- Depends on data model

Influence of Correlation Structure of Data on Fusion Mechanism

In-Network Processing

Categories

- Forwarding: Transmission of raw data without processing
- Fusion: Intermediate processing before reaching fusion center



Data Forwarding

- Direct transmission
- Shortest path

Data Fusion

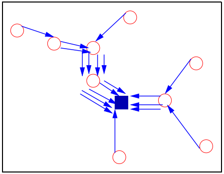
- Energy efficient
- Depends on data model

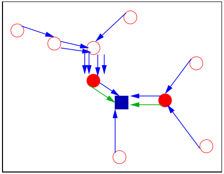
Influence of Correlation Structure of Data on Fusion Mechanism

In-Network Processing

Categories

- Forwarding: Transmission of raw data without processing
- Fusion: Intermediate processing before reaching fusion center





Data Forwarding

- Direct transmission
- Shortest path

Data Fusion

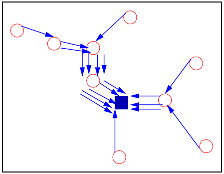
- Energy efficient
- Depends on data model

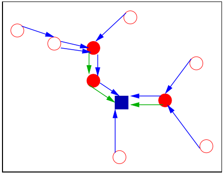
Influence of Correlation Structure of Data on Fusion Mechanism

In-Network Processing

Categories

- Forwarding: Transmission of raw data without processing
- Fusion: Intermediate processing before reaching fusion center





Data Forwarding

- Direct transmission
- Shortest path

Data Fusion

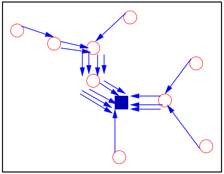
- Energy efficient
- Depends on data model

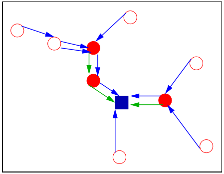
Influence of Correlation Structure of Data on Fusion Mechanism

In-Network Processing

Categories

- Forwarding: Transmission of raw data without processing
- Fusion: Intermediate processing before reaching fusion center





Data Forwarding

- Direct transmission
- Shortest path

Data Fusion

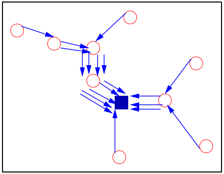
- Energy efficient
- Depends on data model

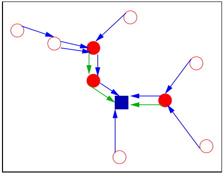
Influence of Correlation Structure of Data on Fusion Mechanism

In-Network Processing

Categories

- Forwarding: Transmission of raw data without processing
- Fusion: Intermediate processing before reaching fusion center





Data Forwarding

- Direct transmission
- Shortest path

Data Fusion

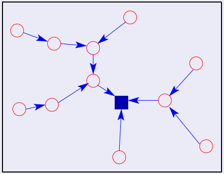
- Energy efficient
- Depends on data model

Influence of Correlation Structure of Data on Fusion Mechanism

Related Work

Data Aggregation

- Special form of fusion
- Incoming packets to 1 packet
- Compute special fn.: sum, max
- Survey (Rajagopalan & Varshney 2006, Giridar & Kumar 2006)
- Cond. Independent: LLR is a sum
- Minimum energy routing: MST, directed towards fusion center



Correlated Data Gathering

- Joint-Entropy based Coding: Cristescu *et al.* , 2006
- LEACH, PEGASIS, LEGA *etc.*,

Fusion in MRF

- Belief Propagation (Pearl 1986): Dist. Comp. of marginals
- Dynamic Prog. to tracking (Williams *et al.* 2006)
- Inference with 1-bit comm. (Kreidl *et al.* 2006)

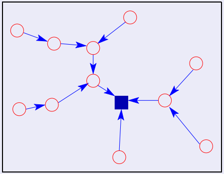
Chernoff Routing (Sung *et al.*)

- Link-metric for detection
- 1-D Gauss-Markov process

Related Work

Data Aggregation

- Special form of fusion
- Incoming packets to 1 packet
- Compute special fn.: sum, max
- Survey (Rajagopalan & Varshney 2006, Giridar & Kumar 2006)
- Cond. Independent: LLR is a sum
- Minimum energy routing: MST, directed towards fusion center



Correlated Data Gathering

- Joint-Entropy based Coding: Cristescu *et al.* , 2006
- LEACH, PEGASIS, LEGA *etc.*,

Fusion in MRF

- Belief Propagation (Pearl 1986): Dist. Comp. of marginals
- Dynamic Prog. to tracking (Williams *et al.* 2006)
- Inference with 1-bit comm. (Kreidl *et al.* 2006)

Chernoff Routing (Sung *et al.*)

- Link-metric for detection
- 1-D Gauss-Markov process

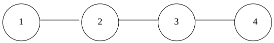
Our Approach and Contributions

- Routing correlated data for detection not dealt before
- Employ the Markov-random field model for correlation
- Single-shot scheme (not flow-based)
- Formulate minimum energy problem
- Provide a simple algorithm with approx. bound of 2

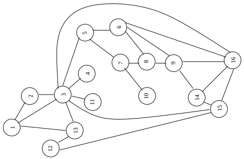
Outline

1. Introduction
2. Markov Random Field
3. Statistical Inference
4. Routing in MRF

Model for Correlated Data : Graphical Model



Linear graph corresponding to autoregressive process of order 1



Graph of German states and states with common borders are neighbors

Temporal signals

- Conditional independence based on ordering
- Fixed number of neighbors
- Causal (random processes)

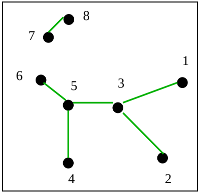
Spatial signals

- Conditional independence based on (undirected) Dependency Graph
- Variable set of neighbors
- Maybe acausal

Remark

Dependency Graph is NOT related to Communication Capabilities, but to the Correlation Structure of Data!

Markov Random Field



MRF with Dependency Graph $\mathcal{G}_d(\mathcal{V}, \mathcal{E})$

$\mathbf{Y}(\mathcal{V}) = \{Y_i : i \in \mathcal{V}\}$ is MRF with $\mathcal{G}_d(\mathcal{V}, \mathcal{E})$ if PDF is positive and it satisfies Markov property

In figure

- Components of DG are independent
- $Y_3 \perp Y_{\mathcal{V}\backslash\{1,2,5\}} | Y_{1,2,5}$
- $Y_1 \perp Y_2$ given rest of network

Equivalent Properties

- Global Markov $\mathbf{Y}_A \perp \mathbf{Y}_B | \mathbf{Y}_C$,
  A, B, C are disjoint, C separates A, B
- Local Markov
  $A = \{i\}, B = \mathcal{V}\backslash\{i, \mathcal{N}_e(i)\}, C = \mathcal{N}_e(i)$
- Pairwise Markov
  $Y_i \perp Y_j | \mathbf{Y}_{\mathcal{V}\backslash\{i,j\}} \iff (i, j) \notin \mathcal{E}$

Likelihood Function of MRF

Hammersley-Clifford Theorem (1971)

For a MRF $\mathbf{Y}$ with dependency graph $\mathcal{G}_d(\mathcal{V}, \mathcal{E}_d)$,

$$\log \mathbb{P}(\mathbf{Y}; \mathcal{G}_d) = Z + \sum_{c \in \mathcal{C}} \Psi_c(\mathbf{Y}_c), \;\; Z \stackrel{\Delta}{=} e^{-\int\limits_{\mathbf{Y}} \prod\limits_{c \in \mathcal{C}} \Psi_c(\mathbf{Y}_c)},$$

where $\mathcal{C}$ is the set of all cliques in $\mathcal{G}_d$ and Ψ_C the clique potential

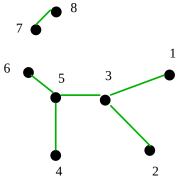
Besag's Auto Model (1974): only 2-clique potentials

$$\log \mathbb{P}(\mathbf{Y}; \mathcal{G}_d) = Z + \sum_{(i,j) \in \mathcal{E}_d} \Psi_{i,j}(Y_i, Y_j) + \sum_{i \in \mathcal{V}} \Psi_i(Y_i)$$

Special Case : Gauss-Markov Random Field

Besag's model : Potential Matrix of GMRF

- Non-zero elements of Potential matrix correspond to graph edges
- Inverse of covariance matrix of a GMRF



Dependency Graph

$$\begin{bmatrix} \times & & \times & & & & & \\ & \times & \times & & & & & \\ \times & \times & \times & & \times & & & \\ & & & \times & \times & & & \\ & & \times & \times & \times & \times & & \\ & & & & \times & \times & & \\ & & & & & & \times & \times \\ & & & & & & \times & \times \end{bmatrix}$$

$\times$: Non-zero element of Potential Matrix

Form of Log-Likelihood Function in GMRF

$$\log \mathbb{P}(\mathbf{Y}_n; \mathcal{G}_d) = Z + \sum_{(i,j)\in\mathcal{E}_d} M_{i,j} Y_i Y_j + \sum_{i\in\mathcal{V}} M_{i,i} Y_i^2$$

Outline

Inference in MRF

Detection of dependency graph and independent parameter θ of MRF

$\mathcal{G}_{d,0}, \theta = \theta_0$ vs. $\mathcal{G}_{d,1}, \theta = \theta_1$. Set $\mathcal{G}_d = \mathcal{G}_{d,0} \cup \mathcal{G}_{d,1}$

Form of Log-Likelihood Ratio

$$\mathrm{LLR}(\mathbf{Y}_n; \mathcal{G}_d) = \log \frac{f(\mathbf{Y}_n|\mathcal{V};\mathcal{H}_0)}{f(\mathbf{Y}_n|\mathcal{V};\mathcal{H}_1)} = Z' + \sum_{(i,j)\in\mathcal{E}_d} \Phi_{i,j}(Y_i, Y_j) + \sum_{i\in\mathcal{V}} \Phi_i(Y_i)$$

Dependency Graph Model for $\mathcal{G}_d$

- Irregular lattice (arbitrary placements of nodes) ?
- Dependency graph is a proximity graph (edges between nearby points)
- Simplest proximity graph: nearest-neighbor graph

Definition

In NNG, (i, j) is an edge if i is nearest neighbor of j or vice versa

Inference in MRF

Detection of dependency graph and independent parameter θ of MRF

$$\mathcal{G}_{d,0}, \theta = \theta_0 \text{ vs. } \mathcal{G}_{d,1}, \theta = \theta_1. \quad \text{Set } \mathcal{G}_d = \mathcal{G}_{d,0} \cup \mathcal{G}_{d,1}$$

Form of Log-Likelihood Ratio

$$\mathrm{LLR}(\mathbf{Y}_n; \mathcal{G}_d) = \log \frac{f(\mathbf{Y}_n|\mathcal{V};\mathcal{H}_0)}{f(\mathbf{Y}_n|\mathcal{V};\mathcal{H}_1)} = Z' + \sum_{(i,j)\in\mathcal{E}_d} \Phi_{i,j}(Y_i, Y_j) + \sum_{i\in\mathcal{V}} \Phi_i(Y_i)$$

Dependency Graph Model for $\mathcal{G}_d$

- Irregular lattice (arbitrary placements of nodes) ?
- Dependency graph is a proximity graph (edges between nearby points)
- Simplest proximity graph: nearest-neighbor graph

Definition

In NNG, (i, j) is an edge if i is nearest neighbor of j or vice versa

Inference in MRF

Detection of dependency graph and independent parameter θ of MRF

$$\mathcal{G}_{d,0}, \theta = \theta_0 \text{ vs. } \mathcal{G}_{d,1}, \theta = \theta_1. \quad \text{Set } \mathcal{G}_d = \mathcal{G}_{d,0} \cup \mathcal{G}_{d,1}$$

Form of Log-Likelihood Ratio

$$\mathrm{LLR}(\mathbf{Y}_n; \mathcal{G}_d) = \log \frac{f(\mathbf{Y}_n|\mathcal{V};\mathcal{H}_0)}{f(\mathbf{Y}_n|\mathcal{V};\mathcal{H}_1)} = Z' + \sum_{(i,j)\in\mathcal{E}_d} \Phi_{i,j}(Y_i, Y_j) + \sum_{i\in\mathcal{V}} \Phi_i(Y_i)$$

Dependency Graph Model for $\mathcal{G}_d$

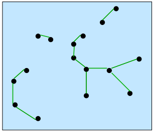
- Irregular lattice (arbitrary placements of nodes) ?
- Dependency graph is a proximity graph (edges between nearby points)
- Simplest proximity graph: nearest-neighbor graph

Definition

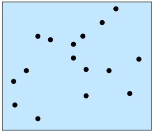
In NNG, (i, j) is an edge if i is nearest neighbor of j or vice versa

Hypothesis testing against Independence[1]

$\mathcal{H}_1$: MRF with $\mathcal{G}_d(\mathcal{V}, \mathcal{E}_d)$

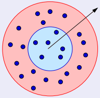


$\mathcal{H}_0$: Independent observations



Additional Assumptions

- Correlation: fn. of edge-length
- Placement :Uniform/ Poisson
- Increase area of coverage with constant sensor density



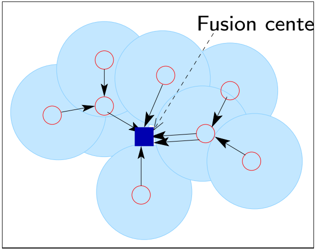
Centralized Error Exponent

- $D = \lim_{n\to\infty} -\frac{1}{n} \log P_M$
- Neyman-Pearson Detection
- $\lim_{n\to\infty} \frac{1}{n}\text{LLR}(\mathbf{Y}_n|\mathcal{V}), \quad \mathcal{H}_0$
- LLN for graph functionals (Penrose & Yukich, 2002)

[1] A.Anandkumar, L.Tong, A. Swami, "Detection of GMRF on nearest-neighbor graph," *Proc. ICASSP*, April 2007

Outline

Network and Energy Model



Minimum Energy Routing

- Data from all nodes needed
- Minimum energy routing reduces to finding $\mathcal{G}^*$

$$\mathcal{G}^*(\mathcal{V}) = \arg \min_{\mathcal{G}_t \subset \text{UDG}(\mathcal{V})} \text{C}(\mathcal{G}_t(\mathcal{V}))$$

- $\mathcal{G}^t$ with a given sequence of transmissions delivers LLR to FC

Network Model

- Unit disk transmission
- Perfect reception of data
- Unit disk graph is connected
- Power control possible
- No quantization error

Energy Model

- Constant Proc. energy
- Transmission energy

$$C_{i,j} = C_t|\text{dist}(i,j)|^\nu, 2 \leq \nu \leq 4$$

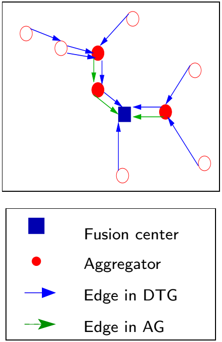
- Transmission graph $\mathcal{G}$

$$\text{C}(\mathcal{G}) = C_t \sum_{e \in \mathcal{E}} r_e^\nu$$

Two Transmission Subgraphs

Form of Log-Likelihood Ratio in MRF

$$\text{LLR}(\mathbf{Y}_n; \mathcal{G}_d) = Z + \sum_{(i,j)\in\mathcal{E}_d} \Phi_{i,j}(Y_i, Y_j) + \sum_{i\in\mathcal{V}} \Phi_i(Y_i)$$



Generalization of Aggregation Problem

- Data transmission graph (DTG): transmission of raw observations
- Likelihood-aggregation graph (AG): transmission of aggregates of LLR
- Aggregator: Nodes processing information from other nodes *i.e.*, those in AG
- Total energy= C(DTG) + C(AG)

Some Intuitions

Formula for log-likelihood ratio in MRF

$$\mathrm{LLR}(\mathbf{Y}_n; \mathcal{G}_d) = Z + \sum_{(i,j)\in\mathcal{E}_d} \Phi_{i,j}(Y_i, Y_j) + \sum_{i\in\mathcal{V}} \Phi_i(Y_i)$$

Data-transmission graph

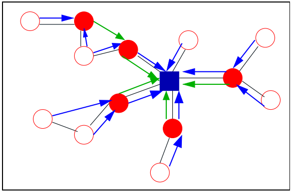
- Edge potential computed locally
- Exist btw. neighbors of DG

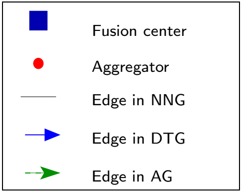
Likelihood-aggregation graph

- If edge pot. computed, sum fn
- Aggregation towards fusion center

Features of optimal graph

Joint design of DTG and AG , No. and type of tx. of every node dependent





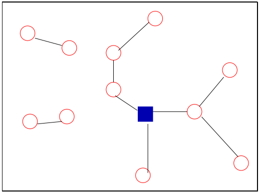
DFMRF: MST-based Transmission Graph

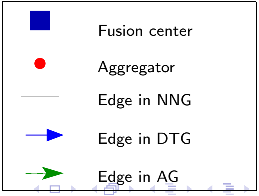
Simple Ideas

- Edge potentials can be calculated if DTG mimics dependency graph
- Since dependency graph is NNG, direct transmission on NNG
- Constructing MST on aggregators

Transmission Graph DFMRF = DTG $\cup$ AG

- Data transmission graph (DTG): NNG, leaves transmit inwards
- Aggregators: All internal nodes *i.e.*, not the leaves
- Aggregation graph (AG): MST of aggregators, towards FC





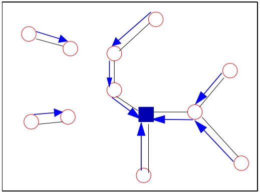
DFMRF: MST-based Transmission Graph

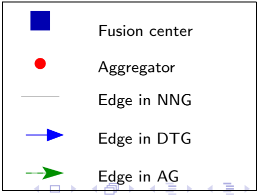
Simple Ideas

- Edge potentials can be calculated if DTG mimics dependency graph
- Since dependency graph is NNG, direct transmission on NNG
- Constructing MST on aggregators

Transmission Graph DFMRF = DTG $\cup$ AG

- Data transmission graph (DTG): NNG, leaves transmit inwards
- Aggregators: All internal nodes *i.e.*, not the leaves
- Aggregation graph (AG): MST of aggregators, towards FC





DFMRF: MST-based Transmission Graph

Simple Ideas

- Edge potentials can be calculated if DTG mimics dependency graph
- Since dependency graph is NNG, direct transmission on NNG
- Constructing MST on aggregators

Transmission Graph DFMRF = DTG $\cup$ AG

- Data transmission graph (DTG): NNG, leaves transmit inwards
- Aggregators: All internal nodes *i.e.*, not the leaves
- Aggregation graph (AG): MST of aggregators, towards FC

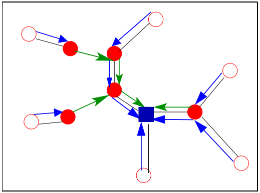
DFMRF: MST-based Transmission Graph

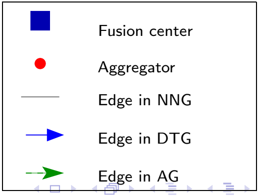
Simple Ideas

- Edge potentials can be calculated if DTG mimics dependency graph
- Since dependency graph is NNG, direct transmission on NNG
- Constructing MST on aggregators

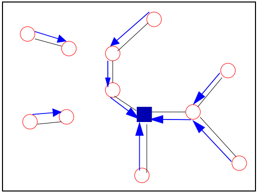
Transmission Graph DFMRF = DTG $\cup$ AG

- Data transmission graph (DTG): NNG, leaves transmit inwards
- Aggregators: All internal nodes *i.e.*, not the leaves
- Aggregation graph (AG): MST of aggregators, towards FC





Algorithm: LLR computation using DFMRF



Data transmission phase

Tx raw data over DTG, compute local contribution

$$m(i) = \Phi_i(Y_i) + \sum_{<j,i> \in \mathrm{DTG}(\mathcal{V})} \Phi_{i,j}(Y_i, Y_j) + \sum_{<j,i> \in \mathrm{DTG}(\mathcal{V}), j \notin \mathcal{V}_{\mathrm{AG}}} \Phi_j(Y_j)$$

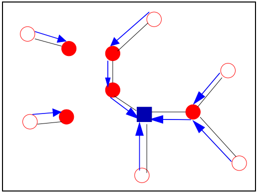
Aggregation phase

- Init: Leaves of AG transmit local contribution
- Recursion: If i has received from all predecessors in AG, transmits $l(i)$

$$l(i) = \sum_{<j,i> \mathrm{AG}(\mathcal{V})} l(j) + m(i)$$

- Stop : Fusion center computes its aggregate

Algorithm: LLR computation using DFMRF



Data transmission phase

Tx raw data over DTG, compute local contribution

$$m(i) = \Phi_i(Y_i) + \sum_{<j,i>\in \mathsf{DTG}(\mathcal{V})} \Phi_{i,j}(Y_i, Y_j) + \sum_{<j,i>\in \mathsf{DTG}(\mathcal{V}), j \notin \mathcal{V}_{\mathsf{AG}}} \Phi_j(Y_j)$$

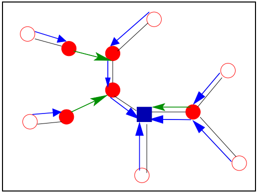
Aggregation phase

- Init: Leaves of AG transmit local contribution
- Recursion: If i has received from all predecessors in AG, transmits $l(i)$

$$l(i) = \sum_{<j,i> \mathsf{AG}(\mathcal{V})} l(j) + m(i)$$

- Stop : Fusion center computes its aggregate

Algorithm: LLR computation using DFMRF



Data transmission phase

Tx raw data over DTG, compute local contribution

$$m(i) = \Phi_i(Y_i) + \sum_{<j,i>\in \mathsf{DTG}(\mathcal{V})} \Phi_{i,j}(Y_i, Y_j) + \sum_{<j,i>\in \mathsf{DTG}(\mathcal{V}), j \notin \mathcal{V}_{\mathsf{AG}}} \Phi_j(Y_j)$$

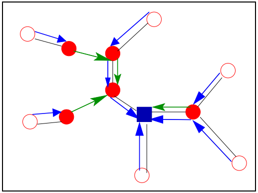
Aggregation phase

- Init: Leaves of AG transmit local contribution
- Recursion: If i has received from all predecessors in AG, transmits $l(i)$

$$l(i) = \sum_{<j,i>\mathsf{AG}(\mathcal{V})} l(j) + m(i)$$

- Stop : Fusion center computes its aggregate

Algorithm: LLR computation using DFMRF



Data transmission phase

Tx raw data over DTG, compute local contribution

$$m(i) = \Phi_i(Y_i) + \sum_{<j,i> \in \mathsf{DTG}(\mathcal{V})} \Phi_{i,j}(Y_i, Y_j) + \sum_{<j,i> \in \mathsf{DTG}(\mathcal{V}), j \notin \mathcal{V}_{\mathsf{AG}}} \Phi_j(Y_j)$$

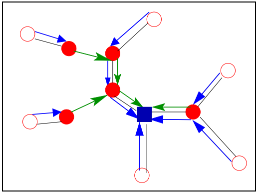
Aggregation phase

- Init: Leaves of AG transmit local contribution
- Recursion: If i has received from all predecessors in AG, transmits $l(i)$

$$l(i) = \sum_{<j,i> \mathsf{AG}(\mathcal{V})} l(j) + m(i)$$

- Stop : Fusion center computes its aggregate

Algorithm: LLR computation using DFMRF



Data transmission phase

Tx raw data over DTG, compute local contribution

$$m(i) = \Phi_i(Y_i) + \sum_{<j,i>\in \mathsf{DTG}(\mathcal{V})} \Phi_{i,j}(Y_i, Y_j) + \sum_{<j,i>\in \mathsf{DTG}(\mathcal{V}), j\notin \mathcal{V}_{\mathsf{AG}}} \Phi_j(Y_j)$$

Aggregation phase

- Init: Leaves of AG transmit local contribution
- Recursion: If i has received from all predecessors in AG, transmits $l(i)$

$$l(i) = \sum_{<j,i>\mathsf{AG}(\mathcal{V})} l(j) + m(i)$$

- Stop : Fusion center computes its aggregate

Performance Analysis

Theorem : bounds for optimal transmission graph $\mathcal{G}_*$

- $\mathrm{C(MST)} \leq \mathrm{C}(\mathcal{G}_*) \leq \mathrm{C(DFMRF)}$
- Lower bound : Pure data aggregation not feasible
- Upper bound : DFMRF scheme

Theorem: Approximation ratio of 2

$$\frac{\mathrm{C}(\mathrm{DFMRF}(\mathcal{V}))}{\mathrm{C}(\mathcal{G}_*(\mathcal{V}))} \leq 2,$$

Other features

- At most 6 transmissions from every node
- Bounded energy and bandwidth at every node
- Single-hop transmission of raw data
- Distributed algorithms to construct the graphs

Conclusion

Summary

- Minimum energy routing for inference of MRF
- Concept of dependency graph based routing
  - Exploit correlation structure to fuse data efficiently
- Proposed DFMRF for NNG dependency
  - 2-approximation, simple construction

Outlook

- Relax assumptions
- Develop better algorithms
- Network lifetime
- Other constraints and costs

Thank You !