Markov Random Field & Belief Propagation

--- Probability Inference for Network

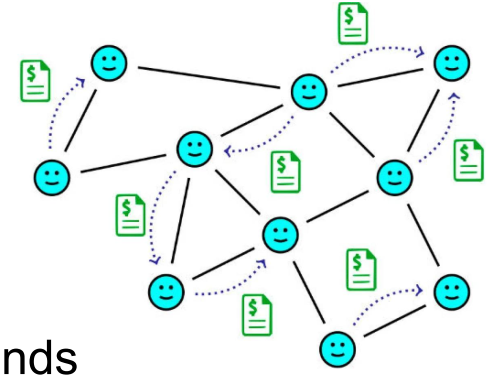
Duruo Li

Northwestern

Why do we need Markov Random Field?

Answer: A quantifiable framework to **undertand** the “**complex** world”

- complex: “network” rather than “chain”
- understand: make inference (oberserved & hidden)



Examples:

- *local* image pixels <=> *global* people's pose
- *local* individuals' trades <=> *global* market's performance
- *local* private communication <=> *global* public opinion trends
- ......

observed states=> infer hidden distribution

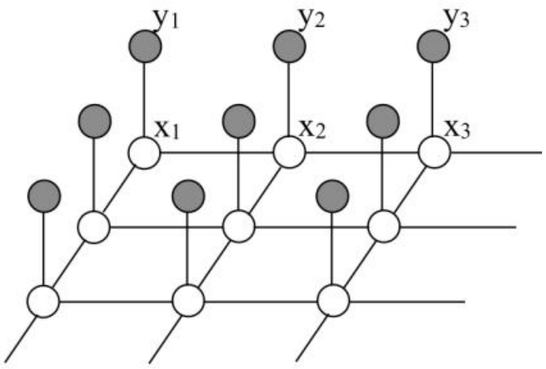
P.S. Not necessarily have different “levels”, e.g., decoding of error-correcting codes (observed: codes received => hidden: codes sent)

Markov Random Field (MRF)

<u>Why?</u>

- Describe complicated relationships between r.v.
- Infer hidden r.v. from observed r.v.

P.S "relationships" and "inference" are dsecribed by joint, marginal, condional distributions (e.g., P(X=x))



<u>What?</u>

Def: A set of **random variables** with **Markov** property described by an **undirected** graph

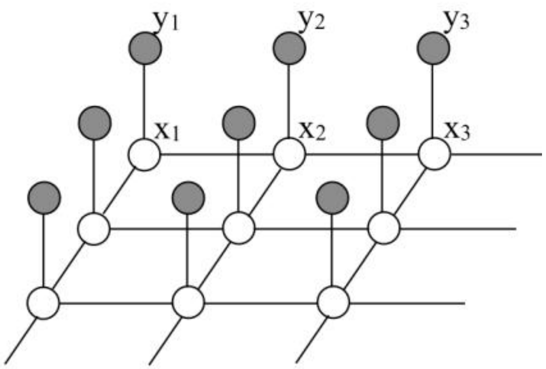
Property: conditional independence

- Pairwise: For any $i, j \in V$ not equal or adjacent, $X_i \perp\!\!\!\perp X_j | X_{V \setminus \{i,j\}}$.
- Local: For any $i \in V$ and $J \subset V$ not containing or adjacent to i, $X_i \perp\!\!\!\perp X_J | X_{V \setminus (\{i\} \cup J)}$.
- Global: For any $I, J \subset V$ not intersecting or adjacent, $X_I \perp\!\!\!\perp X_J | X_{V \setminus (I \cup J)}$.

Markov Random Field (MRF): Quantification

Quantify the process:

1. **present state**: $p(\{x\},\{y\}) = \frac{1}{Z}\prod_{(ij)}\psi_{ij}(x_i,x_j)\prod_i \phi_i(x_i,y_i)$



Z: normalized constant, yi: observed r.v., xi: hidden r.v.
Φi(xi,yi): local "evidence" for xi (given yi), be shortened as Φi(xi) if consider yi as fixed
Ψi,j(xi,xj): inherent "structure" of x, ("transition matrix")

2. **Inference**

Given pre-knowledge and observations, find the most possible/mean value of a hidden r.v. => Given a posterior distribution P({x}|{y}), find the **marginal** distribution P(xi|{y}) of the hidden r.v. , i.e., "beliefs"

$$p(x_N) = \sum_{x_1}\sum_{x_2}\cdots\sum_{x_{N-1}} p(x_1, x_2, x_3, \ldots, x_N)$$

Note that O(|x|^(N-1)) terms need to be computed
|xi|: number of states for xi

Markov Random Field (MRF): Inference

Can we alleviate the computational burden?
Sure! **Belief Propagation**

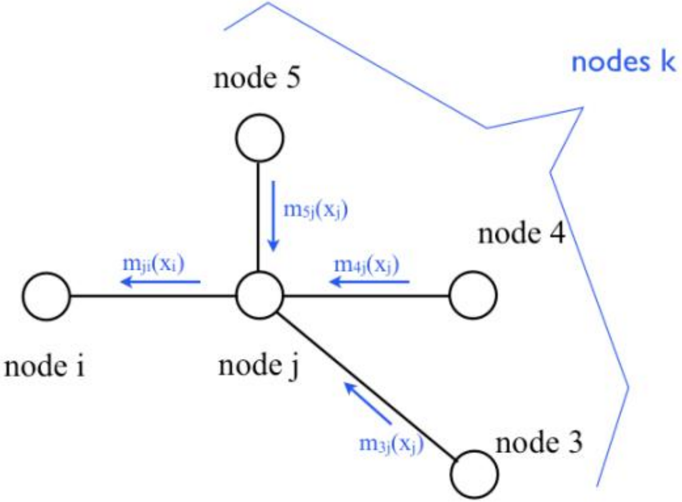
Some concepts:

- **Message** mji(xi): "information" sent from j to i about what state node i should be in

Essence: a re-usable partial sum
*Later we only consider i,j being hidden nodes

- **Belief** bi(xi): proportional to the product of local evidence Φi(xi) and all the messages coming in to node i mji(xi)

Essence: marginal distribution



$$m_{ij}(x_j) \leftarrow \sum_{x_i} \phi_i(x_i)\psi_{ij}(x_i, x_j) \prod_{\boxed{k \in N(i) \backslash j}} m_{ki}(x_i).$$

$$b_i(x_i) = k\phi_i(x_i) \prod_{j \in N(i)} m_{ji}(x_i)$$

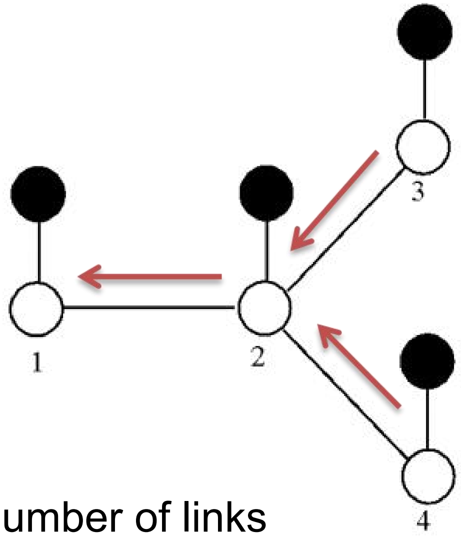
Belief Propagation (Pairwise MRF, No loop)

No loop i.e., singly-connected

An example:

$$b_1(x_1) = k\phi_1(x_1)\sum_{x_2}\psi_{12}(x_1,x_2)\phi_2(x_2)\sum_{x_3}\phi_3(x_3)\psi_{23}(x_2,x_3)\sum_{x_4}\phi_4(x_4)\psi_{24}(x_2,x_4)$$

$$m_{ij}(x_j) \leftarrow \sum_{x_i}\phi_i(x_i)\psi_{ij}(x_i,x_j)\prod_{k\in N(i)\backslash j} m_{ki}(x_i).$$ (update rule)



$$b_1(x_1) = k\phi_1(x_1)\sum_{x_2}\psi_{12}(x_1,x_2)\phi_2(x_2)m_{32}(x_2)m_{42}(x_2)$$

$$b_1(x_1) = k\phi_1(x_1)m_{21}(x_1)$$ |x|^N-1 => k*|x| N: number of nodes, k: number of links

Key Trick: hierarchial, “global” computation => odered “local” computation e.g. from “upstream” node 3,4 to “downstream” node 2 (no replication)

“Messages flow like a river”

Generalized Belief Propagation (Loopy)

<u>Problem</u>: No exact "upstream", it's cyclic, how to start?
<u>Solution</u>: Initialize a set of messages
Belief equation:

$$b_i(x_i) = k\phi_i(x_i) \prod_{j \in N(i)} m_{ji}(x_i)$$

Update rule:

$$m_{ij}(x_j) \leftarrow \sum_{x_i} \phi_i(x_i)\psi_{ij}(x_i, x_j) \prod_{k \in N(i)\backslash j} m_{ki}(x_i).$$

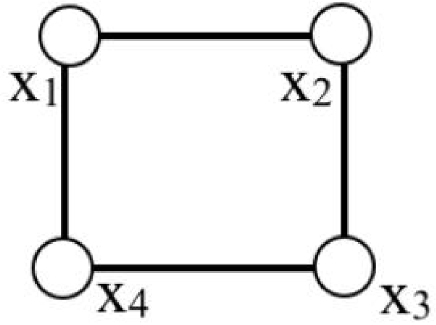


Figure 7.17: A loopy graph

Have nothing to do with the "global" topology, therefore, BP algorithm still works
Converge? Maybe
<u>Empirical evidence:</u> usually works very well!!
Why?

GBP (Loopy): Why it works?

Answer: stationary point of BP (if existed) = minima of **Bethe approximation**

Concepts:

- **Joint/two-node beliefs**:

$$b_{ij}(x_i, x_j) = k \boxed{\psi_{ij}(x_i, x_j)} \boxed{\phi_i(x_i)\phi_j(x_j)} \boxed{\prod_{k \in N(i)\backslash j} m_{ki}(x_i) \prod_{l \in N(i)\backslash i} m_{lj}(x_j)}$$

$$b_i(x_i) = \sum_{x_j} b_{ij}(x_i, x_j)$$

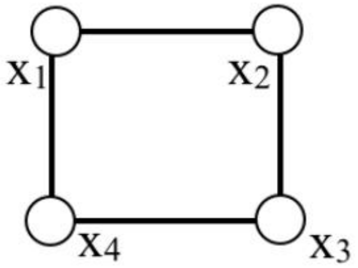
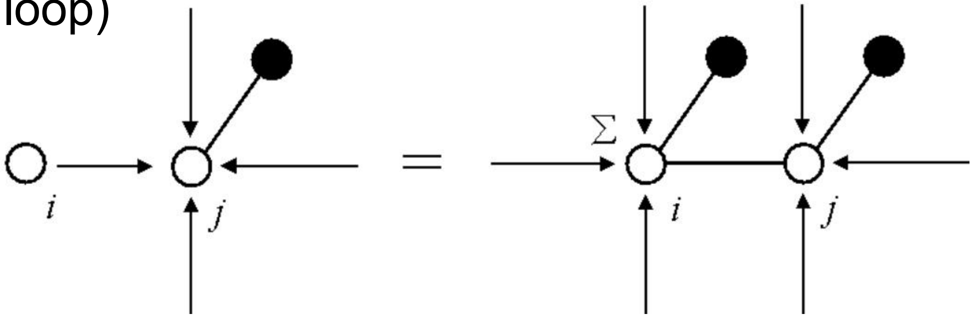


Figure 7.17: A loopy graph

Justified by joint marginal distrbution (no loop)

$$p_{ij}(x_i, x_j) \equiv \sum_{z : z_{ij} = (x_i, x_j)} p(\{z\}).$$



Generalized Belief Propagation (Loopy)

- **Gibbs free energy**:

1) KL distance between “fixed” p({x}) and “changing ”b({x}) (due to iterations)

$$D\left(b(\{x\})||p(\{x\}\right) = \sum_{\{x\}} b(\{x\}) \ln \frac{b(\{x\})}{p(\{x\})}$$

Essence: measure the difference between two distributions

D>=0; D≈0 when converged

2) Boltzmann's law: $p(\{x\}) = \frac{1}{Z} e^{-E(\{x\})/T}$

3) Gibbs free energy: (T=1)

$$D\left(b\{x\}||p(\{x\}\right) = \sum_{\{x\}} b(\{x\}) E(\{x\}) + \sum_{\{x\}} b(\{x\}) \ln b(\{x\}) + \ln Z$$

$$G\left(b(\{x\})\right) = \sum_{\{x\}} b(\{x\}) E(\{x\}) + \sum_{\{x\}} b(\{x\}) \ln b(\{x\}) = U(b\{x\}) - S(b\{x\})$$

=> Find b{x} s.t. G reaches the minima

Generalized Belief Propagation (Loopy)

b{x}, Analytically intractable

$$G(b(\{x\})) = \sum_{\{x\}} b(\{x\})E(\{x\}) + \sum_{\{x\}} b(\{x\}) \ln b(\{x\}) = U(b\{x\}) - S(b\{x\})$$

Approximation for Gibbs free energy

- Mean-field approximation

$$b(\{x\}) = \prod_i b_i(x_i)$$

$$E(\{x\}) = -\sum_{(ij)} \ln \psi_{ij}(x_i, x_j) - \sum_i \ln \phi_i(x_i)$$

(Energy of pairwise MRF)

$$U_{MF}(\{b_i\}) = -\sum_{(ij)} \sum_{x_i, x_j} b_i(x_i) b_j(x_j) \ln \psi_{ij}(x_i, x_j) - \sum_i \sum_{x_i} b_i(x_i) \ln \phi_i(x_i)$$

$$S_{MF}(\{b_i\}) = -\sum_i \sum_{x_i} b_i(x_i) \ln b_i(x_i)$$

- Bethe approximation
- Kikuchi approximation*

Generalized Belief Propagation (Loopy+pairwise)

- Bethe approximation (pairwise MRF)

Average energy U: (one-node and 2-node distributions are enough)

$$U = -\sum_{(ij)} b_{ij}(x_i, x_j) \ln \psi_{ij}(x_i, x_j) - \sum_i b_i(x_i) \ln \phi_i(x_i)$$

pairwise=>“exact”, free of topology

$$E_i(x_i) = -\ln \phi_i(x_i) \qquad E_{ij}(x_i, x_j) = -\ln \psi_{ij}(x_i, x_j) - \ln \phi_i(x_i) - \ln \phi_j(x_j)$$

$$U = \sum_{(ij)} \sum_{x_i, x_j} b_{ij}(x_i, x_j) E_{ij}(x_i, x_j) + \sum_i (q_i - 1) \sum_{x_i} b_i(x_i) E_i(x_i)$$

qi: number of nodes neighboring i

Entropy S_Bethe:

$$b(\{x\}) = \frac{\prod_{(ij)} b_{ij}(x_i, x_j)}{\prod_i b_i(x_i)^{q_i - 1}}$$

only a approximation (no loop)

$$S_{Bethe} = -\sum_{(ij)} \sum_{x_i, x_j} b_{ij}(x_i, x_j) \ln b_{ij}(x_i, x_j) + \sum_i (q_i - 1) \sum_{x_i} b_i(x_i) \ln b_i(x_i)$$

Generalized Belief Propagation (Loopy)

Bethe Approximation:

$$\begin{aligned} G_{Bethe}(b_i(x_i), b_{ij}(x_i, x_j)) &= \sum_{(ij)} \sum_{x_i, x_j} b_{ij}(x_i, x_j)\left(E_{ij}(x_i, x_j) + \ln b_{ij}(x_i, x_j)\right) \\ &- \sum_i (q_i - 1) \sum_{x_i} b_i(x_i)\left(E_i(x_i) + \ln b_i(x_i)\right) \end{aligned}$$

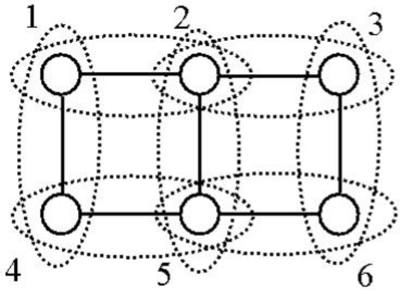
Conclusion: stationary point of BP (if existed) = minima of Bethe approximation

Proof:

See notes

MRF&Belief Propagation: Difficulty & Potential

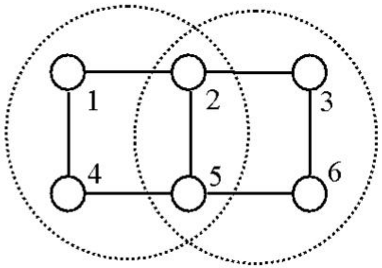
No guarantee for convergence, but according to Murphy et al. (1999):
1. Stop the algorithm after a fixed number of iteration.
2. Stop when no significant difference in belief update.
make good approximation achievable, **AND**
"When the solution converges, it is usually a good approximation."



Techniques to further improve:
message scheduling, residual belief propagation, heuristical initialization and multiple restarts (solve local maximal), etc.

What if not pairwise?
Kikuchi approximation: $G_{Kikuchi} = G_{1245} + G_{2356} - G_{25}$

Do "set algebra"; generalization



Reference:

[1] Yedidia, Jonathan & Freeman, William & Weiss, Yair. (2003). Understanding belief propagation and its generalizations.
[2] J. S. Yedidia, W. T. Freeman and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," in IEEE Transactions on Information Theory, vol. 51, no. 7, pp. 2282-2312
[3] Yedidia, J. S., Freeman, W. T., & Weiss, Y. (2001). Bethe free energy, Kikuchi approximations, and belief propagation algorithms. Advances in neural information processing systems, 13, 689.
[4] Variational Inference: Loopy Belief Propagation https://www.cs.cmu.edu/~epxing/Class/1070814/scribe_notes/scribe_note_lecture13.pdf
[5] Lecture 7: graphical models and belief propagationhttp://helper.ipam.ucla.edu/publications/gss2013/gss2013_11344.pdf

Thank you!