Tree-based Message Passing Algorithms for Loopy Graphs, Bethe-Kikuchi

Nasim Zolaktaf

University of British Columbia

September 2, 2015

Exact Variational Principal

- Convex conjugate of A:

$$A^*(\mu) = sup_{w \in \mathcal{W}}\{\mu^T w - A(w)\} \quad (1)$$

- Convex conjugate of A^*:

$$A(w) = sup_{\mu \in \mathcal{U}}\{w^T \mu - A^*(\mu)\} \quad (2)$$

Exact Variational Principal

- Convex conjugate of A:

$$A^*(\mu) = sup_{w \in \mathcal{W}}\{\mu^T w - A(w)\} \tag{1}$$

- Convex conjugate of A^*:

$$A(w) = sup_{\mu \in \mathcal{U}}\{w^T \mu - A^*(\mu)\} \tag{2}$$

- $P(X) = \frac{exp(W^T F(X))}{Z(w)} = exp(w^T F(X) - A(W))$, where $A(w) = log(Z(w))$:

Exact Variational Principal

- Convex conjugate of A:

$$A^*(\mu) = sup_{w \in \mathcal{W}} \{\mu^T w - A(w)\} \tag{1}$$

- Convex conjugate of A^*:

$$A(w) = sup_{\mu \in \mathcal{U}} \{w^T \mu - A^*(\mu)\} \tag{2}$$

- $P(X) = \frac{exp(W^T F(X))}{Z(w)} = exp(w^T F(X) - A(W))$, where $A(w) = log(Z(w))$:

$$\mathcal{A}^*(\mu) = -H(p_\mu) \tag{3}$$

Exact Variational Principal

- Convex conjugate of A:

$$A^*(\mu) = sup_{w \in \mathcal{W}}\{\mu^T w - A(w)\} \quad (1)$$

- Convex conjugate of A^*:

$$A(w) = sup_{\mu \in \mathcal{U}}\{w^T \mu - A^*(\mu)\} \quad (2)$$

- $P(X) = \frac{exp(W^T F(X))}{Z(w)} = exp(w^T F(X) - A(W))$, where $A(w) = log(Z(w))$:

$$\mathcal{A}^*(\mu) = -H(p_\mu) \quad (3)$$

- Inference as a convex optimization problem:

$$log(\mathcal{Z}(w)) = sup_{\mu \in \mathcal{U}}\{w^T \mu - A^*(\mu)\} = sup_{\mu \in \mathcal{U}}\{w^T \mu + H(p_\mu)\} \quad (4)$$

Marginal Polytope $\mathbb{M}(G)$

$$\mathbb{M}(G) := \{\mu \in \mathbb{R}^d | \exists \text{p with marginals } \mu_s(x_s), \mu_{s,t}(x_s, x_t)\} \tag{5}$$

- Node-based marginal:

$$\mu_s(x_s) := \sum_{j \in \mathcal{X}_s} \mu_{s;j} \mathbb{I}_{s;j}(x_s)$$

- Edge-based marginal:

$$\mu_{s,t}(x_s, x_t) := \sum_{(j,k) \in \mathcal{X}_s \times \mathcal{X}_t} \mu_{st;jk} \mathbb{I}_{st;jk}(x_s, x_t)$$

where:

$$\mu_{s;j} = P[x_s = j] \text{ and } \mu_{st;jk} = P[x_s = j, x_t = k]$$

Locally Consistent Marginal Distribution $\mathbb{L}(G)$

$$\mathbb{L}(G) := \{\tau \geq 0 | \text{Condition 7 holds for all nodes and conditions 8 and 9 hold for all edges.}\} \tag{6}$$

- A set of non-negative node-based functions $\{\tau_s, s \in V\}$, where .

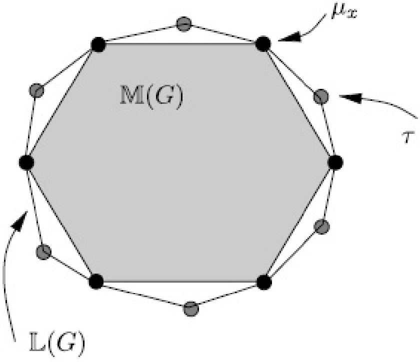
$$\sum_{x_s} \tau_s(x_s) = 1 \tag{7}$$

- A set of non-negative edge-based function $\{\tau_{s,t}, (s,t) \in E\}$, where:

$$\sum_{x'_t} \tau_{st}(x_s, x'_t) = \tau_s(x_s) \tag{8}$$

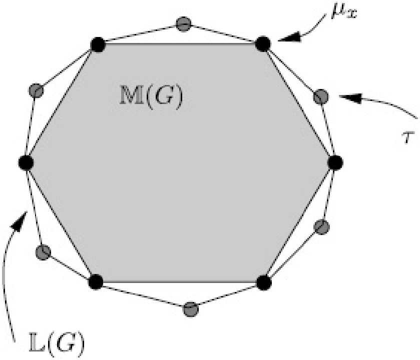
$$\sum_{x'_s} \tau_{st}(x'_s, x_t) = \tau_t(x_t) \tag{9}$$

$\mathbb{M}(G)$ Versus $\mathbb{L}(G)$



- For any graph, then $\mathbb{M}(G) \subseteq \mathbb{L}(G)$.

$\mathbb{M}(G)$ Versus $\mathbb{L}(G)$



- For any graph, then $\mathbb{M}(G) \subseteq \mathbb{L}(G)$.
- By the junction tree theorem, if G is tree-structured, then $\mathbb{M}(G) = \mathbb{L}(G)$.

Entropy for Trees

- By the junction tree theorem, for trees:

$$p_\mu(x) := \prod_{s \in V} \mu_s(x_s) \prod_{(s,t) \in E} \frac{\mu_{st}(x_s, x_t)}{\mu_s(x_s)\mu_t(x_t)}$$

Entropy for Trees

- By the junction tree theorem, for trees:

$$p_\mu(x) := \prod_{s \in V} \mu_s(x_s) \prod_{(s,t) \in E} \frac{\mu_{st}(x_s, x_t)}{\mu_s(x_s)\mu_t(x_t)}$$

- Exact dual function for trees:

$$H(p_\mu) = -A^*(\mu) = \mathop{\mathbb{E}}_{\mu}[-log p_\mu(X)] = \sum_{s \in v} H_s(\mu_s) - \sum_{(s,t) \in E} I_{st}(\mu_{st}) \tag{10}$$

- For each node $s \in V$, singleton entropy:

$$H_s(\mu_s) := -\sum_{x_s \in \mathcal{X}_s} \mu_s(x_s) log \mu_s(x_s)$$

- For each edge $(s,t) \in E$, mutual information:

$$I_{st}(\mu_{st}) := \sum_{(x_s, x_t) \in (\mathcal{X}_s, \mathcal{X}_t)} \mu_{st}(x_s, x_t) log \frac{\mu_{st}(x_s, x_t)}{\mu_s(x_s)\mu_t(x_t)}$$

Bethe Entropy Approximation

- The Bethe approximation to the entropy of an MRF with cycles simply assumes that Equation 10 is valid for a graph with cycles, which yields the Bethe entropy approximation:

$$-A^*(\tau) \approx H_{Bethe}(\tau) := \sum_{s \in V} H_s(\tau_s) - \sum_{(s,t) \in E} I_{st}(\tau_{st}) \qquad (11)$$

Bethe Variational Problem (BVP)

- BVP requires two ingrediants:
  - The set $\mathbb{L}(G)$ is a convex outer bound on the marginal polytope $\mathbb{M}(G)$.
  - The Bethe entropy in Equation 11 is an approximation of the exact dual function $A^*(\tau)$.

Bethe Variational Problem (BVP)

- BVP requires two ingrediants:
  - The set $\mathbb{L}(G)$ is a convex outer bound on the marginal polytope $\mathbb{M}(G)$.
  - The Bethe entropy in Equation 11 is an approximation of the exact dual function $A^*(\tau)$.
- Exact variational principle (Equation 2):

$$A(w) = sup_{\mu \in \mathcal{U}} \{< \theta, \tau > - A^*(\mu)\}$$

- Bethe variational problem (BVP):

$$max_{\tau \in \mathbb{L}(G)} \{<\theta,\tau> + \sum_{s \in v} H_s(\mu_s) - \sum_{(s,t) \in E} I_{st}(\mu_{st})\} \tag{12}$$

Solving BVP

- Solve BVP with the sum-product algorithm.
- Lagrangian corresponding BVP:

$$\mathcal{L}(\tau, \lambda; \theta) :=< \theta, \tau > + H_{Bethe}(\tau) + \sum_{s \in V} \lambda_{ss} C_{ss}(\tau) \\ + \sum_{(s,t) \in E} [\sum_{x_s} \lambda_{ts}(x_s) C_{ts}(x_s; \tau) + \sum_{x_t} \lambda_{st}(x_t) C_{st}(x_t; \tau)] \tag{13}$$

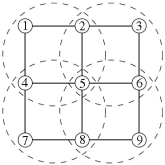
where $C_{ss}(\tau) := 1 - \sum_{x_s} \tau_s(x_s)$ and
$C_{ts}(x_s; \tau) := \tau_s(x_s) - \sum_{xt} \tau_{st}(x_s, x_t)$.

Sum-product algorithm for BVP

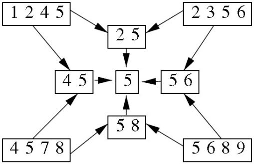
- BVP message update with the sum-product algorithm:

$$M_{t,s}(x_s) \propto \sum_{x_t} \{exp(\theta_{st}(x_s, x_t) + \theta_t(x_t)) \prod_{u \in N(t)/s} M_{ut}(x'_t)\} \tag{14}$$

Kikuchi



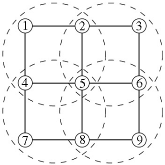
(a)



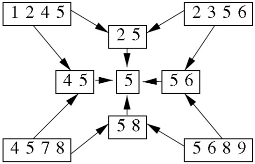
(b)

- Obtain hypergraph with the Kikuchi clustering method.

Kikuchi



(a)



(b)

- Obtain hypergraph with the Kikuchi clustering method.
- Hypertree-based approximation to entropy:

$$H_{app}(\tau) = \sum_{g \in E} c(g) H_g(\tau_g) \qquad (15)$$

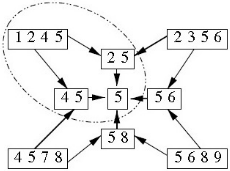
- $H_{app} = [H_{1245} + H_{2356} + H_{4578} + H_{5689}] - [H_{25} + H_{45} + H_{56} + H_{58}] + H_5$

Kikuchi

- Use a generalization of $\mathbb{L}(G)$ (Equation 6) for hypertrees, which is based on marginalization of each hyperedge and any pair of hyperedges.
- Hypertree-based generalization of BVP in Equation 12:

$$max_{\tau \in \mathbb{L}_t(G)} \{ < \theta, \tau > + H_{app}(\tau) \} \tag{16}$$

Parent-to-child Belief Propogation for Kikuchi



$$\tau_{1245} \propto \psi'_{12}\, \psi'_{14}\, \psi'_{25}\, \psi'_{45}\;\; \psi'_{1}\, \psi'_{2}\, \psi'_{4}\, \psi'_{5} \times M_{(2356)\to(25)} M_{(4578)\to(45)} M_{(56)\to 5} M_{(58)\to 5}$$

Illustration of relevant regions for parent-to-child message-passing in a Kikuchi approximation. Message-passing for hyperedge (1245). Set of descendants $\mathcal{D}^{+}\{(1245)\}$ is shown within a dotted ellipse. Relevant parents for τ_{1245} consists of the set $\{(2356),(4578),(56),(58)\}$.

$$\tau_h(x_h) \propto \Big[\prod_{g\in D^+(h)} \psi_g(x_g;\theta)\Big]\Big[\prod_{g\in D^+(h)} \prod_{f\in Par(g)\setminus D^+(h)} M_{f\to g}(x_g)\Big] \qquad (17)$$