Notes on MAP on MRF

Yuandong Tian

February 22, 2010

1 Definitions

The problem is to maximization the following function:

$$E(\mathbf{x}|\bar{\theta}) = \sum_{s \in V} \bar{\theta}_s(x_s) + \sum_{(s,t)\in E} \bar{\theta}_{st}(x_s, x_t) \tag{1}$$

We denote $\Phi_\infty(\bar{\theta})$ as the optimal value of the function, i.e.:

$$\Phi_\infty(\bar{\theta}) = \max_{\mathbf{x}} E(\mathbf{x}|\bar{\theta}) \tag{2}$$

F^n optimization of multi-label variables with at most size-n cliques; P^n: optimization of multi-label variables with at most size-n cliques.

2 An important LP relaxation

From [7]. Denote

$$\begin{aligned} \bar{\theta}_{s;j} &= \bar{\theta}_s(x_s = j) & (3) \\ \bar{\theta}_{st;jk} &= \bar{\theta}_{st}(x_s = j, x_t = k) & (4) \end{aligned}$$

And introduce new variable $\boldsymbol{\mu}$ whose entries are the marginal distribution of some probability $p(\mathbf{x})$:

$$\begin{aligned} \mu_{s;j} &= \mathbb{E}_p\left[I(x_s = j)\right] & (5) \\ \mu_{st;jk} &= \mathbb{E}_p\left[I(x_s = j, x_t = k)\right] & (6) \end{aligned}$$

In other word, vector $\boldsymbol{\mu}$ is in the following space, known as *marginal polytope*:

$$\mathrm{MARG}(G) = \{\boldsymbol{\mu} | \mu_{s;j} = \mathbb{E}_p\left[I(x_s = j)\right], \mu_{st;jk} = \mathbb{E}_p\left[I(x_s = j, x_t = k)\right] \text{ for some } p(\mathbf{x})\} \tag{7}$$

Then the problem becomes

$$\begin{aligned} \min \quad & E(\mathbf{x}|\theta) = \textstyle\sum_s \sum_j \bar{\theta}_{s;j}\mu_{s;j} + \sum_{(s,t)\in E} \bar{\theta}_{st;jk}\mu_{st;jk} = <\boldsymbol{\phi}, \boldsymbol{\mu}> & (8) \\ s.t \quad & \mu \in \mathrm{MARG}(G) & (9) \end{aligned}$$

It is an LP and all vertices of MARG(G) are integer. Thus we can solve a combinatorial minimization exactly using LP! However, the constraints MARG(G) contains exponential many inequalities.

Instead we can use the following *local polytope* as the outer bound of MARG(G):

$$\mathrm{LOCAL}(G) = \{\boldsymbol{\mu} \geq 0 | \sum_j \mu_{s;j} = 1, \sum_k \mu_{st;jk} = \mu_{s;j}\} \tag{10}$$

Which contains much fewer constraints. However, it contains some fractional vertices.

3 Basic Ideas

3.1 Basic idea #1: min-cut

Binary + submodular = exact solution (st min-cut). Submodular property:

$$f(0,0) + f(1,1) \leq f(0,1) + f(1,0) \tag{11}$$

In one words, it favors similar labels.

α-Expansion/swap. In multilabel case, we can do α-expansion/$\alpha - \beta$ swap. If the pairwise potential is metric, then each step is submodular and gives you exact solutions. Hopefully the overall quality of local minimum is not bad.

The each step of the algorithm is basically a multi-binary (or subset) minimization problem. It is doable in polynomial time if it is submodular. And a subset function is submodular if it is pair-wise submodular: for all index pairs $\{i, j\}$.

Has convergence guarentee (function value is monotonous). No optimality guarentee. May need to invoke many max-flows and each with many argumented paths.

Coded the multi-label into binaries[6]. Encode each multi-label variables in the energy function to be one or more binary variables using battleship coding scheme, so that some of them (submodular ones) can be exactly minimized using st-mincut.

High-order clique[1]. A special kind of high-order clique can be used as a part of energy function with st-mincut method still usable. Here they prove that st-mincut will yield the exact solution for each step (each move), or submodular with respect to two variables.

3.2 Basic idea #2: message passing

Tree + any potential = exact solution (message passing).

Basically it is just dynamic programming.

Note message passing can be regarded as reparameterization, i.e., sending messages is equivalent to changing the parameters $\bar{\theta}$ to $\bar{\theta}'$ so that $E(\mathbf{x}|\bar{\theta}) = E(\mathbf{x}|\bar{\theta}')$ for all $\mathbf{x}$. When message passing converges, the parameters would not change and such parameters are called in *its normal form.*

Tree-reweighted Message Passing[7] (TRW-E) For graph with cycles, by reweighting messages we can obtain TRW. This is done by decomposing a graph into many (spanning) trees with different probabilities $\rho(T)$. By assigning different parameters $\theta(T)$ on trees so that $\sum_T \rho(T)\theta(T)$, we get an upper-bound of the original problem:

$$\Phi_\infty(\bar{\theta}) \leq \sum_T \rho(T)\Phi_\infty(\theta(T)) \tag{12}$$

In order to touch the upper-bound, we need to properly adjust $\theta(T)$ so that label agrees on different trees, or *tree agreement.*

No convergence guarentee (It is a fix-point iteration). However, if it converges (tree agreement is obtained), then the fixed point is global optimal.

Sequential Tree-reweighted Message Passing [2] (TRW-S) It has convergence guarentee. The fixed point has the property of weak tree agreement, or WTA.

3.3 Basic idea #3: Using dual

Primal is hard, let us solve dual.

Primal-Dual Scheme[4, 5]. Simutaneously optimize the primal and dual solution until they meet the *relaxed* complementary slackness property. In each iteration, they run the max-flow and keep some properties correct, and aim to make the other correct when the algorithm converges.

Furthermore, the dual gap is an upper-bound of the number of argumented path in max-flow. So it is much faster than α-expansion.

Has convergence guarentee and suboptimal guarentee (since the dual solution is the upper-bound of the primal).

Dual-decomposition[3] The primal problem $\min_{\mathbf{x}} \sum_i f_i(\mathbf{x})$ is hard to optimize. However if each subproblem $\min_{\mathbf{x}} f_i(\mathbf{x})$ is easy, then we can write the following equalivant problem:

$$\min_{\{\mathbf{x}_i\},\mathbf{x}} \quad \sum_i f_i(\mathbf{x}_i) \tag{13}$$

$$s.t \quad \mathbf{x}_i = \mathbf{x} \tag{14}$$

and use duality to solve it. In its dual form, $\mathbf{x}_i$ are decompled from $\mathbf{x}$ (!!). Then we iteratively solve the multiplier $\boldsymbol{\lambda}_i$ and the $\mathbf{x}_i$ until convergence.

It is possible to decompose a graph into a summation of trees and apply this trick.

Has convergence guarentee and suboptimal guarentee. Very much like TRW.

References

[1] P. Kohli, M.P. Kumar, and P.H.S. Torr. P 3 & Beyond: Solving Energies with Higher Order Cliques. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[2] V. Kolmogorov. Convergent Tree-reweighted Message Passing for Energy Minimization. *PAMI*, 2006.

[3] Nikos Komodakis, Nikos Paragios, and Georgios Tziritas. MRF optimization via Dual Decomposition: Message-Passing Revisited. *ICCV*, 2007.

[4] Nikos Komodakis and Georgios Tziritas. A new framework for approximate labeling via graph-cuts. *ICCV*, 2005.

[5] Nikos Komodakis, Georgios Tziritas, and Nikos Paragios. Fast, Approximately Optimal Solution for Single and Dynamic MRFs. *CVPR*, 2007.

[6] S. Ramalingam, P. Kohli, K. Alahari, and P.H.S. Torr. Exact Inference in Multi-label CRFs with Higher Order Cliques. 2008.

[7] Martin J. Wainwright, Tommi S. Jaakkola, and Alan S. Willsky. MAP estimation via agreement on (hyper)trees: Message-passing and linear programming approaches. *Technical Report No. UCB/CSD-3-1269*, 2003.