Computational Intelligence Seminar (8 November, 2005, Waseda University, Tokyo, Japan)

Linear Programming and Belief Propagation with Convex Free Energies

Applications in Computer Vision and Computational Biology

Yair Weiss and Talya Meltzer and Chen Yanover [1]

School of Computer Science and Engineering,

The Hebrew University of Jerusalem

The task of finding the most probable explanation (or MAP) in a graphical model comes up in a wide range of applications including image understanding [9], error correcting codes [2] and protein folding [11]. For an arbitrary graph, this problem is known to be NP hard [8] and various approximation algorithms have been proposed (see. e.g [5] for a recent review).

Linear Programming (LP) Relaxations are a standard method for approximating combinatorial optimization problems in computer science [1]. They have been used for approximating the MAP problem in a general graphical model by Santos [7]. More recently LP relaxations have been used for error-correcting codes [2], and for protein folding [4]. LP relaxations have an advantage over other approximate inference schemes in that they come with an optimality guarantee - when the solution to the linear program is integer, then the LP solution is guaranteed to give the global optimum of the posterior probability.

The research described here grew out of our experience in using LP relaxations for problems in computer vision, computational biology and statistical physics. In all these fields, the number of variables in a realistic problem may be on the order of 10^6 or more. We found that using standard, off-the-shelf LP solvers these problems cannot be solved using standard desktop hardware. A second problem, is that even when we worked on toy problems in which the number of variables was much smaller, we found that a large fraction of the variables in the LP solution were non-integer. This means that the guarantee of optimality is lost, and in practice thresholding the fractional LP solutions gave poor results.

The fact that *general purpose* LP solvers are not suitable for these problems is not that surprising. The linear programs that arise out of LP relaxations for graphical models have a common structure and are a small subset of all possible linear programs. The challenge is to find a LP solver that (1) takes advantage of this special structure and (2) can be used to solve a hierarchy of tighter relaxations in case the standard LP relaxation fails. In this paper, we show that generalized belief propagation with a convex free energy provides such a solver. We show that tree reweighted BP suggested by Wainwright and colleagues [10] is a special case of GBP with a convex free energy but there are many convex free energies that cannot be represented as a tree reweighted free energy. This result has theoretical implications since it shows that the property of solving the LP is distinct from the property of providing a rigorous bound

[1] E-mail: yweiss@cs.huji.ac.il

on the free energy, as well as practical implications since it provides a much expanded family of possible LP algorithms.

1 MAP and LP relaxation

Given an observation vector y, we wish to perform inference on $\Pr(x|y)$ which is assumed to factorize into a product of potential functions:

$$\Pr(x|y) = \frac{1}{Z}\prod_{\alpha}\Psi_{\alpha}(x_{\alpha}) \tag{1}$$

$$= \frac{1}{Z}e^{-\sum_{\alpha}E_{\alpha}(x_{\alpha})} \tag{2}$$

where x_{α} is the domain of the potential Ψ_{α} (the set of all variables that participate in the potential) and we define the "energy" $E_{\alpha}(x_{\alpha})$ as the negative logarithm of the potential.

The MAP, is the vector x^* which maximizes the posterior probability:

$$x^* = \arg\max_{x}\prod_{\alpha}\Psi_{\alpha}(x_{\alpha}) \tag{3}$$

$$= \arg\min_{x}\sum_{\alpha}E_{\alpha}(x_{\alpha}) \tag{4}$$

To define the LP relaxation, we first reformulate the MAP problem as one of integer programming. We introduce indicator variables $q_i(x_i)$ for each individual variable and additional indicator variables $q_{\alpha}(x_{\alpha})$ for all the potential domains. Using these indicator variables we define the integer program:

Minimize:

$$\sum_{\alpha}\sum_{x_{\alpha}}q_{\alpha}(x_{\alpha})E_{\alpha}(x_{\alpha}) \tag{5}$$

Subject to:

$$q_{\alpha}(x_{\alpha}) \in \{0,1\} \tag{6}$$

$$\sum_{x_{\alpha}}q_{\alpha}(x_{\alpha}) = 1 \tag{7}$$

$$\sum_{\alpha\setminus i}q_{\alpha}(x_{\alpha}) = q_i(x_i) \tag{8}$$

where the last equation, enforces the consistency of indicator variables for different potential domains.

This integer program is completely equivalent to the original MAP problem, and is hence computationally intractable. We can obtain the linear programming relaxation by allowing the indicator variables to take on non-integer values. This leads to the following problem:

The LP relaxation of MAP: Minimize:

$$\sum_{\alpha}\sum_{x_{\alpha}}q_{\alpha}(x_{\alpha})E_{\alpha}(x_{\alpha}) \tag{9}$$

Subject to:

$$q_{\alpha}(x_{\alpha}) \in [0,1] \tag{10}$$

$$\sum_{x_\alpha} q_\alpha(x_\alpha) = 1 \tag{11}$$

$$\sum_{\alpha \setminus i} q_\alpha(x_\alpha) = q_i(x_i) \tag{12}$$

2 Solving LP relaxations using Belief Propagation

.

As shown in Yedidia et al. [12] we can construct a belief propagation algorithm for any approximate free energy. We consider approximate free energies with the following entropy term:

$$\tilde{H} = \sum_\alpha c_\alpha H_\alpha + \sum_i c_i H_i \tag{13}$$

As we explain in the next section, this approximate entropy term may or may not be convex as a function of the beliefs.

Convex BP=LP Theorem: Let b_α, b_i be fixed-point beliefs from running belief propagation with a convex entropy approximation at temperature T. For sufficiently small T these fixed-point beliefs are a solution to the linear program.

Proof: We know that the BP beliefs are constrained stationary points of the free energy:

$$\begin{aligned} F_\beta &= \sum_\alpha \sum_{x_\alpha} b_\alpha(x_\alpha) E_\alpha(x_\alpha) \\ &+ T \left(\sum_\alpha \sum_{x_\alpha} c_\alpha b_\alpha(x_\alpha) \ln b_\alpha(x_\alpha) \right) \\ &- T \left(\sum_i c_i \sum_{x_i} b_i(x_i) \ln b_i(x_i) \right) \end{aligned} \tag{14}$$

The minimization of F_β is done subject to the following constraints:

$$b_\alpha(x_\alpha) \in [0,1] \tag{15}$$

$$\sum_{x_\alpha} b_\alpha(x_\alpha) = 1 \tag{16}$$

$$\sum_{\alpha \setminus i} b_\alpha(x_\alpha) = b_i(x_i) \tag{17}$$

The energy term is exactly the LP problem. As we decrease the temperature the Bethe free energy approaches the LP problem. If we assume the entropy function to be convex then the Bethe free energy is convex and hence any fixed point corresponds to the global minimum.

3 Convex Free energies versus Tree Reweighted Free Energies

Heskes [3] has shown the following conditions for convexity of the entropy term.

Definition: An approximate entropy term of the form:

$$H = \sum_\alpha c_\alpha H_\alpha + \sum_i c_i H_i \tag{18}$$

is said to be *provably convex* if there exist *positive* numbers $c_{i\alpha}, d_\alpha, d_i$ such that

$$H = \sum_{i\alpha} c_{i\alpha}(H_\alpha - H_i) + \sum_i d_\alpha H_\alpha + \sum_i d_i H_i \tag{19}$$

where the sum is only over pairs $i\alpha$ that are connected in the region graph.

Tree-reweighted free energies [10] use entropy terms of the form:

$$H_{TRW}(\mu) = \sum_T \mu_T H_T \tag{20}$$

where T is a spanning tree in the graph, μ_T defines a distribution over spanning trees and H_T is the entropy of that tree. It is easy to show that all tree-reweighted entropies are convex however *tree reweighted entropies are a tiny fraction of all convex entropies*. To see this, note that any tree reweighted entropy can be rewritten:

$$H_{TRW}(\mu) = \sum_{<ij>} \rho_{ij} H_{ij} + \sum_i (1 - deg(i)) H_i \tag{21}$$

where ρ_{ij} is the edge appearance probability defined by μ and $deg(i)$ is the average degree of node i: $deg(i) = \sum_j \rho_{ij}$. In comparing this to the general entropy approximation (equation 18) we see that we tree reweighted entropies are missing a degree of freedom (with c_i). In fact, for any TRW entropy we can add an infinite number of possibile positive combination of single node entropies and still maintain convexity. Thus TRW entropies are a measure zero set of all convex entropies.

In some cases, we can even subtract single node entropies from a TRW entropy and still maintain convexity. For example, the Bethe free energy for a single cycle can be shown to be convex but *it cannot be represented as tree-reweighted entropy* [10]. In particular, it does not give rise to a bound on the free energy. This shows that the family of BP algorithms that provide a bound on the free energy is a strict subset of the family of BP algorithms that can be used to solve linear programs.

4 Experiments

We have empirically compared convex BP and powerful general-purpose LP solvers (CPLEX) on relaxations of real-world graphical models from the fields of computer vision and computational biology. We find that BP almost always finds the solution significantly faster than all the solvers in CPLEX and more importantly, BP can be applied to large scale problems for which the solvers in CPLEX cannot be applied. Using BP we can find the MAP configurations in a matter of minutes for a large range of real world problems. In particular, for the stereo vision problem we can find the global optimum for the standard benchmark images ,whereas all previous approaches only guaranteed a local optimum [6].

References

[1] D. Bertismas and J. Ttsitskikilis. *Introduction to linear optimization.* Athena Scientific, 1997.

[2] J. Feldman, D. Karger, and M. J. Wainwright. LP decoding. In *Allerton Conference on Communication, Control, and Computing*, 2003.

[3] T. Heskes. On the uniqueness of loopy belief propagation fixed points. *Neural Computation*, 16:2379–2413, 2004.

[4] Carl Kingsford, Bernard Chazelle, and Mona Singh. Solving and analyzing side-chain positioning problems using linear and integer programming. *Bioinformatics*, 2005. in press.

[5] Radu Marinescu, Kalev Kask, and Rina Dechter. Systematic vs. non-systematic algorithms for solving the MPE task. In *Proceedings of Uncertainty in Artificial Intelligence (UAI 2003)*, 2003.

[6] T. Meltzer, C. Yanover, and Y. Weiss. Globally optimal solutions for energy minimization in stereo vision using reweighted belief propagation. In *Proceedings International Conference on Computer Vision (ICCV)*, 2005.

[7] E.G. Santos. On the generation of alternative explanations with implications for belief revision. In *UAI*, 1991.

[8] Y. Shimony. Finding the maps for belief networks is np-hard. *Aritifical Intelligence*, 68(2):399–410, 1994.

[9] M. Tappen and W.T. Freeman. Graph cuts and belief propagation for stereo, using identical MRF parameters. In *ICCV*, 2003.

[10] M. J. Wainwright, T. Jaakkola, and A. S. Willsky. Map estimation via agreement on (hyper)trees: Message-passing and linear programming approaches. In *Allerton Conference on Communication, Control, and Computing*, 2002.

[11] C. Yanover and Y. Weiss. Approximate inference and protein folding. *Advances in Neural Information Processing Systems*, 2002.

[12] J.S. Yedidia, W.T. Freeman, and Yair Weiss. Constructing free energy approximations and generalized belief propagation algorithms. *IEEE Transactions on Information Theory*, 51(7):2282–2312, 2005.