Belief propagation today: What we know, what we don't and what we can do about it

Ravi Palanki

Department of Electrical Engineering
California Institute of Technology
Pasadena, CA 91125, USA
email: ravi@systems.caltech.edu

October 10, 2002

Abstract

Around two months ago, Dr. McEliece gave me a list of open questions, which he probably gave Jonathan and Jeremy too. Today, we have all jointly figured out the answers to many of these questions, and in other cases learnt the answers from someone else (a.k.a Dr. Jonathan Yedidia). In this context, I've decided to briefly answer the answered questions in the open questions list, and state the remaining questions again.

1 What we know

Prove or disprove that stable fixed points of PBP correspond to local minima of BK variational free energy.

A proof of this (available on Tom Heskes' webpage) for ordinary BP/ Bethe free energy exists. Stable fixed points are local minima, but the converse is not necessarily true. We can try working on extensions to PBP.

Can PBP always be made to converge using the smoothing trick?

No. While smoothing helps a lot (Jonathan's experiments), there exist examples (Yedidia's 3D spin glasses) where no PBP algorithm converges. Contrary to what Yedidia said in an eariler draft of his paper, there are cases where PBP fails to converge, but the fixed points are good approximations to the actual beliefs. Therefore alternative BP algorithms can prove useful. I've developed a couple of algorithms called Gradient Descent PBP (converges for any poset, but has higher complexity than PBP) and Cyclic Shift PBP (converges for depth-2 posets, has the same complexity as PBP).

Can other combinatorial optimization methods (e.g. simulated annealing be uses to minimize F_P?

Yes. There are examples of such algorithms, two famous ones being Yuille's double loop algorithm and an algorithm by Welling and Teh. Both these are discrete-time algorithms guaranteed to converge to a local minimum. I do not know of anyone using simulated annealing.

Give a proof that PBP is correct when the Hasse diagram is a tree.

Done. I gave a proof which said that PBP converges in a finite number of parallel updates (equal to the diamater of the tree). Dr. McEliece used Boltzmann distributions to show that PBP converges in a finite number of serial updates (equal to the number of edges in the tree).

What do we know about one cycle posets?

Since PBP on a one cycle poset is a matrix multiplication, we can show that it always converges using a powerful result from matrix theory known as the Perron Frobenius theorem. We also know that one cycle posets have property C.

Is there a general way to find if a poset has property C?

Dr. McEliece has devised a way to do this using "upsets". A similar technique was discovered independently by Pakzad and Anantharam.

What do we know about Boltzmann distributions?

Dr. McEliece showed that the true distribution on a treelike poset is a Boltzmann distribution. Moreover, on a treelike poset, the Boltmann distribution defined by the beliefs at any stage of PBP is invariant. On an unrelated (or maybe not so unrelated) note, I have shown that Boltzmann distributions are the extrema of the function $H_{approx}(X) - H(X)$.

Is it necessary that the i- and R- subposets be trees?

No. There are posets where the i- and R- subposets have short cycles (length 4) which outperform posets with no such short cycles.

Why geometric mean? Why not arithmetic, harmonic etc?

Why not indeed. Yedidia was actually using the arithmetic mean, while we thought he was using the geometric mean.

Can the zero gradient point corresponding to a fixed point of PBP be a local maximum?

Yes. The free energy has a global maximum, and that has to be a fixed point. (This is the same logic used to by Yedidia to show that PBP has a fixed point in the first place)

Are Hasse diagrams better than DAGs?

I think Dr. McEliece and Yedidia jointly concluded that the answer is no.

What happens when $\phi(p) = 0$?

Such a node doesn't contribute to the free energy, and it can be deleted (after connecting its parents to its children) without affecting anything. Yedidia does this in one of the appendices of his paper.

Can we use the smoothing trick on only some of the messages?

Preferably not, according to Yedidia. Empirical evidence shows doing such things hampers convergence.

2 What we don't know

What is the complexity of PBP for a given poset P?

It is very to easy to compute the per-iteration complexity of PBP for any poset. What is hard to compute is the number of iterations required and an optimal smoothing coefficent that minimizes the number of iterations required.

Can be used to design codes?

Hopefully yes. But no one seems to have an idea about what should be done. The best work so far on this subject was using PBP to decode an existing code. (Yedidia, Allerton 2002)

What is the relationship between the BK free energy and the exact Helmholtz free energy?

No clue. See next question.

What can be said about the accuracy of the approximation $H(X) \approx \sum_{\rho \in P} \phi(\rho) H(X_\rho)$*?*

Nothing, except for a way to bound the maximum error of approximation using mutual informations, and some plots by Jonathan.

For which posets does PBP not involve division?

We can't describe the general class of such posets, though we do know examples. But I think the question is not relevant, because we usually have to divide to normalize messages and beliefs.

When can we say that F_P *is strictly convex?*

We don't have an answer yet, but this should be an easy problem.

If PBP converges for a smoothing factor t'*, does it necessarily converge for all* $t \leq t'$*?*

Jonathan has enough evidence to suggest that is the case. But surpisingly we have no proof. I believe that such a result will be found in some math textbook and that we haven't looked hard enough.