The $\epsilon - cutset$ effect in Bayesian networks

Bozhena Bidyuk
ICS Department
University Of California Irvine
Irvine, CA 92697-3425
bbidyuk@ics.uci.edu

Rina Dechter
ICS Department
University Of California Irvine
Irvine, CA 92697-3425
dechter@ics.uci.edu

Abstract

The paper investigates the behavior of iterative belief propagation algorithm (IBP) in Bayesian networks with loops. In multiply-connected network, IBP is only guaranteed to converge in linear time to the correct posterior marginals when evidence nodes form a loop-cutset. We propose an ϵ-cutset criteria that IBP will converge and compute posterior marginals close to correct when a single value in the domain of each loop-cutset node receives very strong support compared to other values thus producing an effect similar to the observed loop-cutset. We investigate the support for this criteria analytically and empirically and show that it is consistent with previous observations of IBP performance in multiply-connected networks.

1 Introduction

The paper investigates the correctness of iterative belief propagation (IBP) algorithm in Bayesian networks with loops. Pearl [10] proposed iterative belief propagation algorithm for singly connected Bayesian networks and demonstrated that algorithm converges in the number of iterations equal to the diameter of the network to the correct posterior values. In general, IBP does not always converge and does not produce correct posterior values for Bayesian networks with loops. It can be applied to networks with loops to derive approximate inference where exact methods such as bucket-elimination [3] and tree-clustering [4] become impractical due to exponential growth in time and memory required.

Empirically, IBP was demonstrated to perform very well for several classes of loopy networks, including noisy-or networks (used in diagnostics), pyramid networks (used in image recognition) [8], and especially coding networks [11, 9, 5] where it was shown to outperform other approximation schemes such as variational decoder [5] and mini-bucket approximation algorithm [7].

Our theoretical understanding of the behavior of IBP in networks with loops is limited. For general types of networks, the main contribution has been made by Weiss [13, 12] who proved (for Markov model) that IBP always converges in a single-loop networks, defined the error in posterior marginals produced by IBP as a function of eigenvalues of a matrix computed from conditional probability tables (CPTs) of all the variables in a loop, and established the correlation between the accuracy of IBP and its convergence rate. However, the error expression proposed by Weiss is hard to interpret in terms of basic network attributes.

We investigate the accuracy of IBP in a directed Bayesian networks with binary nodes as a function of loop size, CPT values, prior beliefs, and evidence support. Analytically, we derive an expression for the error value in the posterior belief of a sink node in a single-loop Bayesian network without evidence (section 4) and then extend our conclusions empirically to several classes of loopy networks with evidence (section 5). As a result, we propose an ϵ-cutset criteria that guarantees convergence of IBP and a certain level of accuracy in posterior marginals computed by IBP in Bayesian networks with loops when ϵ-support is provided for loop-cutset nodes(section 3).

2 Background

DEFINITION **2.1 (graph concepts)** *We will skip the definitions of basic belief network concepts (which can be found in [10]) and belief networks due to lack of space.*

A node X in a directed graph D is called a root *if no edges are directed into X. A node X in a directed graph D is called a* leaf *if all of its adjacent edges are directed into X. The underlying graph G of a directed graph D is the undirected graph formed by ignoring directions of edges in D. A* cycle *in G is a path whose two end-points coincide. A* loop *in D is a subgraph of D whose underlying graph is a cycle. A vertex v is a* sink *with respect to loop $\mathcal{L}$ if two edges adjacent to v in $\mathcal{L}$ are directed into v. A vertex that is not a sink with respect to a loop $\mathcal{L}$ is called an* allowed *vertex with respect to $\mathcal{L}$. A* loop-cutset *of a directed graph D is a set of vertices that contains at least one allowed vertex in each loop in D. (We borrowed loop-cutset definition from [1]). A graph is* singly connected *(also called a* polytree*), if its underlying undirected graph has no cycles. Otherwise, it is* multiply connected*.*

3 Iterative Belief Propagation Algorithm

Iterative Belief Propagation (IBP) computes an approximate belief $BEL(x) = P(X = x|E)$, where E is observed evidence, for every variable X in the network. Belief is propagated by sending messages between the nodes. During each iteration $(t+1)$, each node X sends *causal* support messages $\pi_{Y_j}^{(t+1)}(x)$ to each child Y_j:

$$\pi_{Y_j}^{(t+1)}(x) = \alpha \lambda_X(x) \prod_{k \neq j} \lambda_{Y_k}^t(x) \pi^t(X) \tag{1}$$

and *diagnostic* support messages $\lambda_x^t(u_i)$ to each parent u_i:

$$\lambda_X^{(t+1)}(u_i) = \beta \sum_x \lambda^t(x) \sum_{u_k, k \neq i} P(x|u) \prod_{k \neq i} \pi_X^t(u_k) \tag{2}$$

where

$$\lambda^t(x) = \lambda_X(x) \prod_j \lambda_{Y_j}^t(x) \tag{3}$$

$$\pi^t(x) = \sum_u P(x|u) \prod_i \pi_X^t(u_i) \tag{4}$$

In equations 1 and 2, α and β are normalization constants such that $\sum_{u_i} \lambda_X(u_i) = 1$ and $\sum_X \pi_{Y_j}(X) = 1$. Normalization is recommended to avoid numerical underflow although it does not affect the results produced by IBP ([10, 2]). Message $\lambda_X(X)$ is introduced to incorporate evidence information into the equation (similar to [8]). The posterior belief is computed for each node X by combining $\pi_X(u_i)$ messages received from its parents u_i and $\lambda_{Y_j}(X)$ messages received from its children:

$$BEL(x) = \alpha \lambda^t(X) \pi^t(X) \tag{5}$$

An *activation schedule* (variable ordering) A specifies the order in which the nodes are processed (activated). After all nodes are processed, the next iteration of belief propagation begins.

Belief propagation was proposed by Pearl [10] for singly-connected networks where it is guaranteed to converge to correct posterior marginals for all nodes in linear time. In multiply-connected networks, IBP is only guaranteed to converge to correct posterior marginals if evidence nodes form a loop-cutset. If node X is observed, $\lambda^t(X) = 0$ for all values of X except its observed value. As a result, $\pi_{Y_j}(X)$ messages become independent from any other messages received by X from its parents and children and $\lambda_X(u_i)$ messages become independent from the λ messages X received from its children. Thus, an observed loop-cutset node X effectively breaks the loop by stopping the flow information between its parents and its children.

Now, assume that node X is not observed but receives very strong support for one value in its domain, x', from prior beliefs $|1 - P(X = x')| < \epsilon$ and/or support messages from its children $|1 - \lambda_{Y_k}(X = x')| < \epsilon$ (assume that λ has been normalized). We will call it ϵ-support. Then $lim_{\epsilon \to 0} \lambda^t(X = x') \to 1$ generating effect similar to the observed nodes. Therefore, we propose following conjecture:

Conjecture 1 ($\epsilon - cutset$**)** *Let G be a multiply-connected Bayesian network. Let* X_i, $i = 1, ..., m$, *form a loop-cutset in G. Then, for any* $\delta \in (0, 1)$, *there exists such value* $\epsilon_0 \in (0, 1)$ *that if each loop-cutset node* X_i *receives an* ϵ_0*-support* $|1 - P(X_i = x_i')| < \epsilon_0$ *or* $|1 - \lambda_{Y_k}(X_i = x_i')| < \epsilon_0$ *for normalized* $\lambda_{Y_k}(X_i)$, *then IBP will converge and compute posterior marginals within specified error limit* δ.

4 Single Loop Bayesian Network without Evidence

In this section, we will analyze the performance of belief propagation algorithm in a single-loop Bayesian network with binary nodes as shown in fig 1. Without evidence, the posterior marginals of all nodes except the sink node will be computed correctly ([2]). Thus, we focus here on computing the error produced by IBP in the posterior marginal values of sink node D. Let $G(D)$ be correct posterior belief:

$$G(D) = \sum_{B_n, C_m} P(D|B_n, C_m) \sum_A P(B_n|A) P(C_m|A) P(A) \tag{6}$$

In the absence of evidence, one iteration of belief propagation is sufficient for convergence given activation schedule such that parents of a node are processed first and normalization of the belief $BEL(D)$ is not required ($\alpha = 1$) ([2]). Then, we can easily derive the posterior marginal belief $G^*(D)$ computed by IBP:

$$G^*(D) = \sum_{B_n, C_m} P(D|B_n, C_m) (\sum_A P(B_n|A) P(A)) (\sum_A P(C_m|A) P(A)) \tag{7}$$

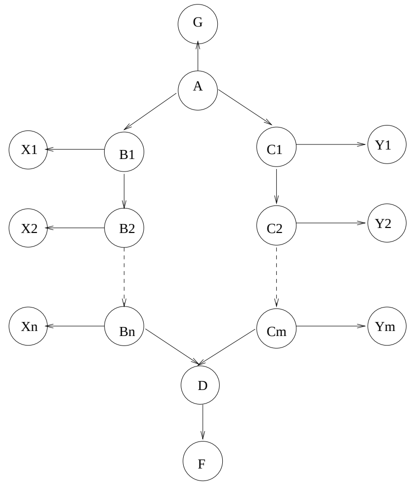

Figure 1: A single loop Bayesian network.

Using notation:

$$P(A) = (\epsilon, 1-\epsilon)$$
$$\beta_0^i = P(B_i = 0|B_{i-1} = 0), \beta_1^i = P(B_i = 0|B_{i-1} = 1)$$
$$\gamma_0^j = P(C_j = 0|C_{j-1} = 0), \gamma_1^j = P(C_j = 0|C_{j-1} = 1)$$

we can express error $\delta = G^*(D) - G(D)$ in the posterior marginals of node D computed by IBP as follows (see [2] for a complete proof):

$$\delta = \epsilon(1-\epsilon)(\sum_{B_n, C_m} (-1)^{B_n}(-1)^{C_m} P(D|B_n, C_m)) \prod_{i=1}^{n} (\beta_0^i - \beta_1^i) \prod_{j=1}^{m} (\gamma_0^j - \gamma_1^j)) \quad (8)$$

From equation 8, it is clear that the accuracy of IBP improves as:

1. Prior beliefs for a root node approach boundary values: $\lim_{\epsilon \to 0,1} \delta = 0$
2. Number of nodes in a loop increases: $\lim_{n\to\infty, m\to\infty} \delta = 0$
3. Conditional probabilities for the same value of X in different rows of CPT get closer: $\lim_{|\beta_0^k - \beta_1^k| \to 0, |\gamma_0^k - \gamma_1^k| \to 0} \delta = 0$.

It is also easy to see that when one of the allowed nodes is observed, $\delta = 0$. When node A is initialized, either $\epsilon = 0$ or $(1-\epsilon) = 0$. When one of the nodes B_i or C_j is observed, it is equivalent to having $\beta_0(B_i) = \beta_1(B_i) = 0|1$ or $\gamma_0(C_j) = \gamma_1(C_j) = 0|1$ yielding $\delta = 0$.

5 Empirical Results

In this section, we empirically investigate accuracy and convergence of IBP in networks with loops. In all experiments, unless specified otherwise, conditional probabilities were represented by noisy-or with 'leak' term fixed at 0.005 and individual noise factors θ_i were chosen uniformly from the range [0,1]. The number of iterations for approximate inference was fixed at 20. All experimental results are consistent with conclusions in section 4 and proposition 1.

5.1 Single Loop with Evidence

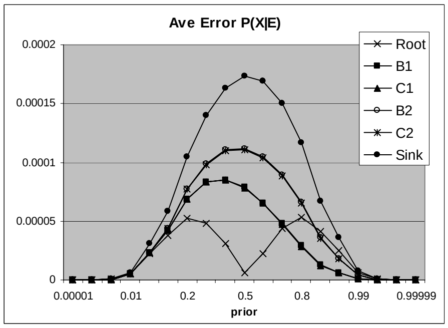

Figure 2: Average error in $P(X|E)$ is plotted against $P(A = 0)$.

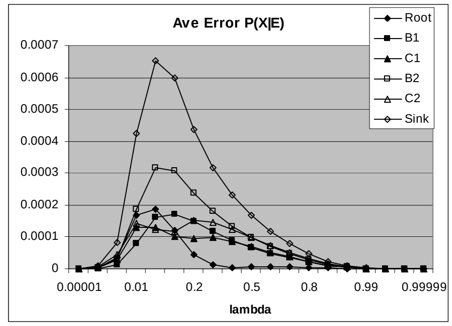

Figure 3: Average error in $P(X|E)$ is plotted against $\lambda_{Y_2}(C_2)$.

Single loop networks (see fig 1) containing 4, 6, 8, and 10 nodes were constructed. The child of a sink node was always observed. The average error value was averaged over 1000 instances. As the loop size increased, the maximum average error value decreased by the order of magnitude: 0.002 for a 4-node loop, 0.0002 for a 6-node loop, 0.00002 for a 8-node loop, and 0.000002 for a 10-node loop.

For all nodes and all loop sizes, the average error value rapidly decreased as $P(A)$ was approaching 0 and 1 as predicted analytically in equation 8 (results for a 6-node loop are presented in fig 2). For all nodes, except root node, the average error value peaked at $P(A) = 0.5$, as expected, since $\epsilon = 0.5$ is the maximum of the function $f(\epsilon) = \epsilon(1 - \epsilon)$. The average error for a root node consistently had a minimum at $P(A) = 0.5$ which we cannot explain analytically at this point and plan to investigate in the future. For all nodes and all loop sizes, the average error was approaching zero as $\lambda_{Y_m}(C_m) \to 0, 1$ (results for a 6-node loop are presented in fig 3).

5.2 2-layer noisy-or networks

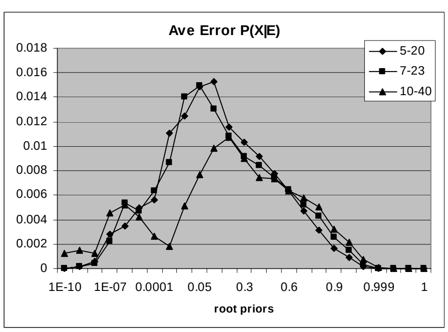

Figure 4: Average error in $P(X|E)$ is plotted against root node priors.

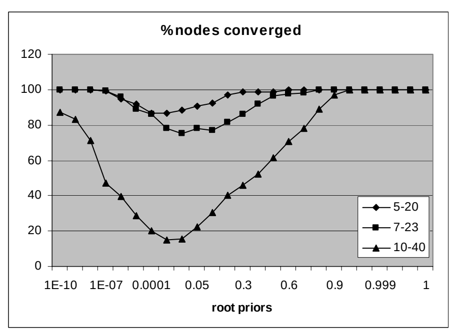

Figure 5: Percent of converged nodes is plotted against root node priors.

We measured the accuracy and convergence of IBP in 2-layer noisy-or networks. Number of root nodes m and total number of nodes n was fixed in each test set (indexed $m - n$).

Each leaf node Y_j was added to the list of children of root node U_i with probability 0.5. All leaf nodes were observed. Results (in figures 4 and 5) were averaged over 100 instances. All data indicate that as priors of loop-cutset nodes approach 0 and 1, the average error value approaches 0 and the number of converged nodes reaches 100%.

5.3 Random noisy-or networks

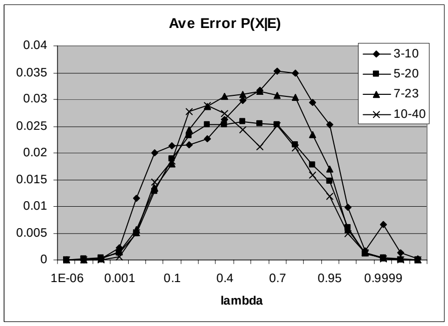

Figure 6: Average error in $P(X|E)$ is plotted against $\lambda_{Y_k}(X_k)$.

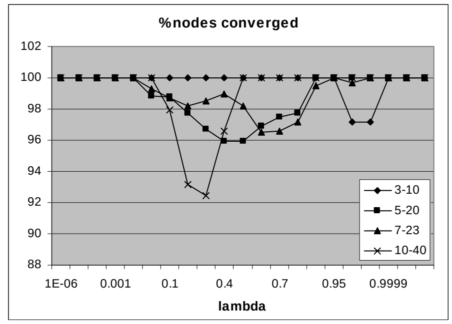

Figure 7: Percent of converged nodes is plotted against $\lambda_{Y_k}(X_k)$.

Random $m-n$ networks of size n were constructed by designating first m nodes as roots, last m nodes as leaves, and then adding each node X_i, $i > m$, to the list of children of node X_j, $j > i$, with probability 0.2. Results were averaged over 100 instances. Loop-cutset nodes were selected using mga algorithm proposed in [1]. For each loop-cutset node, an extra child node Y_k was added with a symmetrical CPT. To control $\lambda_{Y_k}(X_k)$ messages, $P(Y_k|pa(Y_k))$ was varied. All leaf nodes were observed. As results demonstrate (figures 6 and 7), the average error approaches 0 and convergence of IBP reaches 100% as $\lambda_{Y_j}(X) \to 0, 1$.

6 Related work and conclusions

Empirical study of the performance of belief propagation algorithm [7, 5] in different types of coding networks including Humming codes, low-density parity check, and turbocodes, demonstrated that accuracy of IBP is considerably better when noise level σ is low. Those results correlate very well with proposition 1 since lower noise level means that code nodes receive stronger support for one value from their observed children $\lim_{\sigma\to 0} \lambda(U_i) \to 0, 1$.

An investigation into the distribution of cycle lengths in coding networks has demonstrated that a node has a low probability (less than 0.01) of being in a cycle of length less than or equal to 10 [6]. Furthermore, the CPTs derived for edges with low Gaussian noise σ define very strong correlation between parent/child node values. Thus, observed child node will send quite strong support for the observed value to the parent. Both of the above observations combined with our empirical findings, provide an insight into excellent performance of IBP in coding networks. Coding networks have good parameters for two different factors influencing convergence and accuracy of IBP: large loop size and strong ϵ-support.

The work presented here has two novelties. First, it provides a direct analytical connection between loop size, root priors, and evidence support and the error in the posterior marginals computed by IBP. We derived an exact expression for the error value only for a special case of a node in a single-loop network without evidence. However, the empirical evidence leads us to the same conclusions as our analytical findings which indicates that the mechanics behind the performance of IBP in single-loop network and multiple-loop networks with or without evidence is the same.

The second novelty is extending well-known loop-cutset criteria that guarantees convergence and correctness of IBP in loopy networks when evidence nodes constitute a loop-cutset to instances where loop-cutset nodes are not observed, but receive an ϵ-support. The proposed ϵ-cutset criteria guarantees the convergence and certain degree of accuracy when ϵ is sufficiently small. The next step in our research is to devise means of estimating the threshold ϵ value that will guarantee the convergence of IBP and desired degree of accuracy in posterior marginals computed by IBP.

References

[1] A. Becker, R. Bar-Yehuda, and D. Geiger. Random algorithms for the loop cutset problem. In *Uncertainty in AI (UAI'99)*, 1999.

[2] B. Bidyuk and R. Dechter. Epsilon-cutset effect in bayesian networks of arbitrary topology. Technical Report UCI-ICS 01-01, UCI, 2001.

[3] R. Dechter. Bucket elimination: A unifying framework for reasoning. *Artificial Intelligence*, 113:41–85, 1999.

[4] R. Dechter and J. Pearl. Tree clustering for constraint networks. *Artificial Intelligence*, pages 353–366, 1989.

[5] B. J. Frey and D. J. C. MacKay. A revolution: Belief propagation in graphs with cycles. In *Neural Information Processing Systems*, 1997.

[6] X. Ge, D. Eppstein, and P. Smyth. The distribution of cycle lengths in graphical models for iterative decoding. Technical Report UCI-ICS 99-10, UCI, 1999.

[7] K. Kask I. Rish and R. Dechter. Empirical evaluation of approximation algorithms for probabilistic decoding. In *Uncertainty in AI (UAI'98)*, 1998.

[8] Y. Weiss K. P. Murphy and M. I. Jordan. Loopy belief propagation for approximate inference: An empirical study. In *Uncertainty in AI (UAI'99)*, 1999.

[9] F. R. Kschischang and B. J. Frey. Iterative decoding of compound codes by probability propagation in graphical models. *IEEE Journal on Selected Areas in Communications*, 16:219–230, 1998.

[10] J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, 1988.

[11] D.J.C. MacKay R.J. McEliece and J.-F. Cheng. Turbo decoding as an instance of pearl's belief propagation algorithm. *IEEE J. Selected Areas in Communication*, 1997.

[12] Y. Weiss. Belief propagation and revision in networks with loops. Technical Report AIM-1616, MIT, 1997.

[13] Y. Weiss. Correctness of local probability propagation in graphical models with loops. *Neural Computation 12*, pages 1–41, 2000.