Loop Corrected Belief Propagation

Joris Mooij\textsuperscript{1} B. Wemmenhove\textsuperscript{1} B. Kappen\textsuperscript{1} T. Rizzo\textsuperscript{2}

\textsuperscript{1}Department of Biophysics
Radboud University Nijmegen, The Netherlands
\{j.mooij|b.wemmenhove|b.kappen\}@science.ru.nl

\textsuperscript{2}E. Fermi Center
Rome, Italy
tommaso.rizzo@inwind.it

AISTATS 2007
Problem setting

Let $\mathcal{V} := \{1, \ldots, N\}$. Consider a probability distribution on $N$ discrete random variables $x = (x_1, \ldots, x_N)$ that factorizes as follows:

$$P(x_1, \ldots, x_N) = \frac{1}{Z} \prod_{K \in \mathcal{F}} \psi_K(x_K)$$

where $\mathcal{F} \subseteq \mathcal{P}(\mathcal{V})$. 

Example A Bayesian network or Markov Random Field.

Objective Calculate single node marginals $P(x_i) = \frac{1}{Z} \sum_{x_{\mathcal{V} \setminus i}} \prod_{K \in \mathcal{F}} \psi_K(x_K)$
Let $\mathcal{V} := \{1, \ldots, N\}$. Consider a probability distribution on $N$ discrete random variables $x = (x_1, \ldots, x_N)$ that factorizes as follows:

$$P(x_1, \ldots, x_N) = \frac{1}{Z} \prod_{K \in \mathcal{F}} \psi_K(x_K)$$

where $\mathcal{F} \subseteq \mathcal{P}(\mathcal{V})$.

**Example**

A Bayesian network or Markov Random Field.
Problem setting

Let $\mathcal{V} := \{1, \ldots, N\}$. Consider a probability distribution on $N$ discrete random variables $x = (x_1, \ldots, x_N)$ that factorizes as follows:

$$P(x_1, \ldots, x_N) = \frac{1}{Z} \prod_{K \in \mathcal{F}} \psi_K(x_K)$$

where $\mathcal{F} \subseteq \mathcal{P}(\mathcal{V})$.

Example

A Bayesian network or Markov Random Field.

Objective

*Calculate single node marginals*

$$P(x_i) = \frac{1}{Z} \sum_{x_{\mathcal{V}\setminus i}} \prod_{K \in \mathcal{F}} \psi_K(x_K)$$
Some definitions

**Definition**

To the probability distribution $P \propto \prod_K \psi_K(x_K)$ corresponds a factor graph, a bipartite graph with variable nodes $i, j, \ldots$ (circles) and factor nodes $K, L, \ldots$ (rectangles) with an edge between variable $i$ and factor $K$ iff $i \in K$. 

Example

$\partial_i$ is the Markov blanket of $i$, i.e. all neighboring variables of $i$. 

$\Delta_i := \partial_i \cup \{i\}$. 

---

Joris Mooij et al. (RU Nijmegen)  
Loop Corrected Belief Propagation  
AISTATS’07
Some definitions

Definition

To the probability distribution $P \propto \prod_K \psi_K(x_K)$ corresponds a factor graph, a bipartite graph with variable nodes $i, j, \ldots$ (circles) and factor nodes $K, L, \ldots$ (rectangles) with an edge between variable $i$ and factor $K$ iff $i \in K$.

Example

[Diagram of a factor graph with variable and factor nodes and edges.]
Some definitions

Definition

To the probability distribution \( P \propto \prod_K \psi_K(x_K) \) corresponds a factor graph, a bipartite graph with variable nodes \( i, j, \ldots \) (circles) and factor nodes \( K, L, \ldots \) (rectangles) with an edge between variable \( i \) and factor \( K \) iff \( i \in K \).

Example

Definition

\( \partial i \) is the Markov blanket of \( i \), i.e. all neighboring variables of \( i \).
Some definitions

Definition

To the probability distribution \( P \propto \prod_K \psi_K(x_K) \) corresponds a factor graph, a bipartite graph with variable nodes \( i, j, \ldots \) (circles) and factor nodes \( K, L, \ldots \) (rectangles) with an edge between variable \( i \) and factor \( K \) iff \( i \in K \).

Example

\[ \partial i \] is the Markov blanket of \( i \), i.e. all neighboring variables of \( i \).

\[ \Delta i := \partial i \cup \{i\}. \]
Existing solutions for calculating single node marginals

- Exact methods (e.g. junction trees)
- Sampling methods
- “Deterministic” approximate methods, e.g.
  - Belief Propagation (BP)
  - Generalized Belief Propagation (GBP)
  - TreeEP
Existing solutions for calculating single node marginals

- Exact methods (e.g. junction trees)
- Sampling methods
- “Deterministic” approximate methods, e.g.
  - Belief Propagation (BP)
  - Generalized Belief Propagation (GBP)
  - TreeEP

Belief Propagation

Belief Propagation yields exact results on tree structured factor graphs. However, if the factor graph contains one or more loops, results are approximate and typically are worse for denser graphs.
Existing solutions for calculating single node marginals

- Exact methods (e.g. junction trees)
- Sampling methods
- “Deterministic” approximate methods, e.g.
  - Belief Propagation (BP)
  - Generalized Belief Propagation (GBP)
  - TreeEP

Generalized Belief Propagation

GBP can handle short loops more precisely by combining variables into clusters that subsume the loops.
Existing solutions for calculating single node marginals

- Exact methods (e.g. junction trees)
- Sampling methods
- “Deterministic” approximate methods, e.g.
  - Belief Propagation (BP)
  - Generalized Belief Propagation (GBP)
  - TreeEP

**TreeEP**

TreeEP improves over BP by performing exact inference over a spanning tree and can handle loops that consist of part of the tree and one additional factor.
Existing solutions for calculating single node marginals

- Exact methods (e.g. junction trees)
- Sampling methods
- “Deterministic” approximate methods, e.g.
  - Belief Propagation (BP)
  - Generalized Belief Propagation (GBP)
  - TreeEP

Problem

Presence of strong loops typically results in low quality approximations.
Existing solutions for calculating single node marginals

- Exact methods (e.g. junction trees)
- Sampling methods
- “Deterministic” approximate methods, e.g.
  - Belief Propagation (BP)
  - Generalized Belief Propagation (GBP)
  - TreeEP

Problem

Presence of strong loops typically results in low quality approximations.

Our solution

We propose a method that corrects BP for the presence of loops in the factor graph; it typically obtains significant improvements in accuracy.
Existing solutions for calculating single node marginals

- Exact methods (e.g. junction trees)
- Sampling methods
- “Deterministic” approximate methods, e.g.
  - Belief Propagation (BP)
  - Generalized Belief Propagation (GBP)
  - TreeEP

**Problem**

Presence of strong loops typically results in low quality approximations.

**Our solution**

We propose a method that corrects BP for the presence of loops in the factor graph; it typically obtains significant improvements in accuracy.

Definition

The cavity graph of $i$ is the factor graph obtained by removing variable $i$ together with all its neighboring factors.

Example

![Diagram showing the cavity graph of variable $i$ with its neighbors removed]
Cavity graphs

Definition

The cavity graph of \( i \) is the factor graph obtained by removing variable \( i \) together with all its neighboring factors.

Example

![Diagram showing cavity graph with nodes and edges]
Cavity graphs

**Definition**

The cavity graph of $i$ is the factor graph obtained by removing variable $i$ together with all its neighboring factors.

**Example**

![Diagram of a cavity graph]

**Definition**

The cavity distribution of $i$ is the marginal of the cavity graph on $\partial i$:

$$P_{c}^{i}(x_{\partial i}) := \frac{1}{Z_{i}} \sum_{x_{\Delta i}} \prod_{K \in F, i \notin K} \psi_{K}(x_{K}).$$
Cavity graphs

**Definition**
The cavity graph of \( i \) is the factor graph obtained by removing variable \( i \) together with all its neighboring factors.

**Example**

**Definition**
The cavity distribution of \( i \) is the marginal of the cavity graph on \( \partial i \):

\[
P_{\setminus i}(x_{\partial i}) := \frac{1}{Z_{\setminus i}} \sum_{x_{\Delta i}} \prod_{K \in F} \psi_K(x_K).
\]

**Proposition**

\[
P(x_{\Delta i}) \propto P_{\setminus i}(x_{\partial i}) \psi_i(x_{\Delta i})
\]

where

\[
\psi_i(x_{\Delta i}) := \prod_{K \in F \setminus \{i\}} \psi_K(x_K).
\]
Cavities and loops
What is the relationship between loops and cavity distributions?

Example
Cavities and loops
What is the relationship between loops and cavity distributions?

Example

\[ \begin{align*}
&\text{Example} \\
&\begin{array}{c}
\text{\includegraphics[width=\textwidth]{example_diagram.png}}
\end{array}
\end{align*} \]
Cavities and loops
What is the relationship between loops and cavity distributions?

Example

The loop through $x_i$, $x_j$ and $x_k$ results in a dependency between $x_j$ and $x_k$ in the cavity distribution $P^i$ of $i$.

$$P^i(x_{\partial i}) = P^i(x_j, x_k)P^i(x_l)P^i(x_m).$$
In practice, exact cavity distributions are unavailable. Instead, we use approximate cavity distributions $Q^i \approx P^i$.

1. Calculate initial approximate cavity distributions $\{Q_0^i\}_{i \in V}$;
2. Cancel out errors in the approximate cavity distributions by demanding consistency of single node marginals;
3. Calculate final single node marginals from corrected cavity distributions $\{Q_\infty^i\}_{i \in V}$. 
Consistency of single node marginals

Let $i, j$ be two neighboring variables with common factor $K$. Define

$$
\psi_i^K(x_{\Delta i}) := \frac{\psi_i(x_{\Delta i})}{\psi_K(x_K)} = \prod_{L \in \mathcal{F}} \psi_L(x_L) \quad \text{(and similarly for } j).$$

$$
\psi_j^K(x_{\Delta j}) := \frac{\psi_j(x_{\Delta j})}{\psi_K(x_K)} = \prod_{L \in \mathcal{F}} \psi_L(x_L) \quad \text{(and similarly for } j).$$

$$
C_1 \quad \text{Cavity graph of } i
$$

$$
C_2 \quad \text{Cavity graph of } j
$$
Let $i, j$ be two neighboring variables with common factor $K$. Define

\[
\Psi_i^K(x_{\Delta i}) := \frac{\Psi_i(x_{\Delta i})}{\Psi_K(x_K)} = \prod_{L \in F, \ i \in L, L \neq K} \psi_L(x_L) \quad \text{(and similarly for $j$)}.
\]
Consistency of single node marginals

Let \( i, j \) be two neighboring variables with common factor \( K \). Define

\[
\Psi_i^K(x_{\Delta i}) := \frac{\Psi_i(x_{\Delta i})}{\Psi_K(x_K)} = \prod_{L \in \mathcal{F}} \psi_L(x_L) \quad \text{(and similarly for \( j \))}
\]

Cavity graph of \( i \)

Cavity graph of \( j \)

\[
P^{\setminus i}(x_{\partial i}) \Psi_i^K(\Delta i) \quad P^{\setminus j}(x_{\partial j}) \Psi_j^K(\Delta j)
\]
Consistency of single node marginals

Let $i, j$ be two neighboring variables with common factor $K$. Define

$$
\psi_i^K(x_{\Delta i}) := \frac{\psi_i(x_{\Delta i})}{\psi_K(x_K)} = \prod_{L \in F, i \in L, L \neq K} \psi_L(x_L)
$$

(and similarly for $j$).

Cavity graph of $i$

Cavity graph of $j$

$$
\sum_{x_{\partial i}} P^{i}(x_{\partial i}) \psi_i^K(\Delta i) = \sum_{x_{\Delta j \setminus i}} P^{j}(x_{\partial j}) \psi_j^K(\Delta j)
$$
We modify the initial approximations $\{Q^i_0\}_{i \in \mathcal{V}}$ by changing single variable interactions but keeping higher order interactions fixed:

$$Q^i(x_{\partial_i}) := Q^i_0(x_{\partial_i}) \prod_{j \in \partial i} \phi^i_j(x_j),$$

where the factors $\phi^i_j$ are chosen such that:

$$\sum_{x_{\partial i}} Q^i(x_{\partial i}) \psi^K_i(\Delta i) = \sum_{x_{\Delta j \backslash i}} Q^j(x_{\partial j}) \psi^K_j(\Delta j) \quad \forall i \in \mathcal{V} \forall j \in \partial i$$

This can be solved using simple fixed point iteration of the $\phi^i_j$ factors.
Calculate initial approximate cavity distributions \( \{Q_0^i\}_{i \in \mathcal{V}} \).

Update the approximate cavity distributions:

1. \( t \leftarrow 0 \)
2. repeat
3. for all \( i, j \in \mathcal{V} \) such that \( i, j \in K \) for some \( K \in \mathcal{F} \) do
4. \( Q_{t+1}^\cup \propto Q_t^\cup \frac{\sum_{x_{\partial i}} Q_t^i \psi_i^K}{\sum_{x_{\Delta j \setminus i}} Q_t^\cup \psi_j^K} \)
5. end for
6. \( t \leftarrow t + 1 \)
7. until convergence

Calculate approximate single node marginals \( q_i(x_i) \approx P(x_i) \) using:

\[
q_i(x_i) \propto \sum_{x_{\partial i}} Q_\infty^i(x_{\partial i}) \psi_i(x_{\Delta i}).
\]
Possible ways of calculating initial cavity distributions
BP as a special case of LCBP

Theorem
If the initial cavity distributions factorize completely, fixed points of standard BP are fixed points of the LCBP update algorithm. This justifies the name "Loop Corrected Belief Propagation."

Take uniform distributions...
Possible ways of calculating initial cavity distributions
BP as a special case of LCBP

Take uniform distributions...

**Theorem**

*If the initial cavity distributions factorize completely, fixed points of standard BP are fixed points of the LCBP update algorithm.*
Possible ways of calculating initial cavity distributions
BP as a special case of LCBP

Take uniform distributions…

**Theorem**

*If the initial cavity distributions factorize completely, fixed points of standard BP are fixed points of the LCBP update algorithm.*

This justifies the name “Loop Corrected Belief Propagation”.
A high accuracy initialization scheme:

1. For all $i \in V$
2. For all $x_{\partial_i}$
3. Calculate $F_i^{\text{Bethe}}(x_{\partial_i})$, the Bethe free energy corresponding to the cavity graph of $i$ clamped in state $x_{\partial_i}$, using BP
4. $Q_i^0(x_{\partial_i}) \leftarrow e^{-F_i^{\text{Bethe}}(x_{\partial_i})}$
5. End for
6. End for

Theorem

Using this initialization, LCBP results will be exact if the factor graph contains one loop. In general, this yields high accuracy approximations.
A high accuracy initialization scheme:

1: for all $i \in \mathcal{V}$ do
2:   for all $x_{\partial i}$ do
3:     calculate $F_{\text{Bethe}}^i(x_{\partial i})$, the Bethe free energy corresponding to the cavity graph of $i$ clamped in state $x_{\partial i}$, using BP
4:   end for
5:  $Q_0^i(x_{\partial i}) \leftarrow e^{-F_{\text{Bethe}}^i(x_{\partial i})}$
6: end for
Possible ways of calculating initial cavity distributions

High accuracy

A high accuracy initialization scheme:

1: for all $i \in \mathcal{V}$ do
2: for all $x_{\partial i}$ do
3: calculate $F_{\text{Bethe}}^i(x_{\partial i})$, the Bethe free energy corresponding to the cavity graph of $i$ clamped in state $x_{\partial i}$, using BP
4: end for
5: $Q_0^i(x_{\partial i}) \leftarrow e^{-F_{\text{Bethe}}^i(x_{\partial i})}$
6: end for

Theorem

Using this initialization, LCBP results will be exact if the factor graph contains one loop.
Possible ways of calculating initial cavity distributions
High accuracy

A high accuracy initialization scheme:

1: for all $i \in V$ do
2:    for all $x_{\partial i}$ do
3:      calculate $F_{\text{Bethe}}^i(x_{\partial i})$, the Bethe free energy corresponding to the cavity graph of $i$ clamped in state $x_{\partial i}$, using BP
4:    end for
5:  $Q^i_0(x_{\partial i}) \leftarrow e^{-F_{\text{Bethe}}^i(x_{\partial i})}$
6: end for

Theorem

*Using this initialization, LCBP results will be exact if the factor graph contains one loop.*

In general, this yields high accuracy approximations.
Experiments on random graphs with binary variables and random pairwise interactions (fixed degree $|\partial i| = 5$)

- BP
- TreeEP
- GBP
- MR
- LCBP
- EXACT

Error vs. $N$:

- BP Error
- LCBP Error
- Time (s) vs. $N$
Experiments on periodic grids with binary variables and random pairwise interactions

Joris Mooij et al. (RU Nijmegen) Loop Corrected Belief Propagation AISTATS'07 14 / 18
Experiments on the ALARM network

<table>
<thead>
<tr>
<th>Method</th>
<th>Error</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP</td>
<td>0.203</td>
<td>0.00</td>
</tr>
<tr>
<td>TreeEP</td>
<td>0.039</td>
<td>0.22</td>
</tr>
<tr>
<td>GBP</td>
<td>0.035</td>
<td>161.0</td>
</tr>
<tr>
<td>MR</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>LCBP</td>
<td>0.00054</td>
<td>23.4</td>
</tr>
</tbody>
</table>
Discussion and conclusion

Summary

- We proposed a method to correct BP for the influence of loops in the factor graph, building on the work by Montanari and Rizzo.
- We showed that LCBP can significantly outperform other approximate inference methods in terms of accuracy.
- However, computation time is exponential in the cavity size and application is thus to factor graphs with small cavities.

Future work

- I am currently working on alternative update equations and initialization methods that sacrifice some accuracy in exchange for speed improvements.
- An open question is whether there exists a “free energy” that corresponds to LCBP. That would allow to also compute a loop-corrected version of the Bethe free energy.
Thank you!

- For more details and experiments, see also [Mooij & Kappen, cs.AI:0612030].
- C++ code for all algorithms is available as free/open source software (licensed under the GNU Public License) at my homepage http://www.mbfys.ru.nl/~jorism/libDAI/
- I will graduate in summer and am looking for a post-doc position.

References


Acknowledgments

The research reported here is part of the Interactive Collaborative Information Systems (ICIS) project (supported by the Dutch Ministry of Economic Affairs, grant BSIK03024) and was also sponsored in part by the Dutch Technology Foundation (STW).
Improved LCBP updates
if short loops of length 4 are present

1: \( t \leftarrow 0 \)
2: repeat
3: \quad \text{for all } i \in \mathcal{V} \text{ do}
4: \quad \text{for all } K \in N_i \text{ do}
5: \quad \quad Q^j_{t+1} \leftarrow Q^j_t \frac{\prod_{j \in K \setminus i} \left( \sum_{\Delta j \setminus (K \setminus i)} Q^j_i \psi^K_j \right)^{1/|K \setminus i|}}{\sum_{x \Delta i \setminus (K \setminus i)} Q^i_x \psi^K_i}
6: \quad \text{end for}
7: \text{end for}
8: \quad t \leftarrow t + 1
9: until convergence