Loop Corrected Belief Propagation

Joris Mooij[1] B. Wemmenhove[1] B. Kappen[1] T. Rizzo[2]

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

[2]E. Fermi Center
Rome, Italy
tommaso.rizzo@inwind.it

AISTATS 2007

Problem setting

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

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

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

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.

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

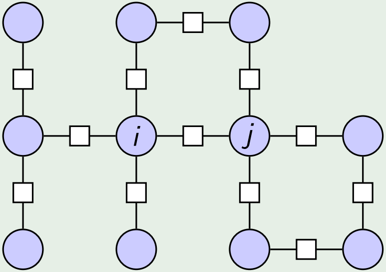
To the probability distribution $P \propto \prod_K \psi_K(x_K)$ corresponds a factor graph, a bipartite graph with variable nodes $i, j, \dots$ (circles) and factor nodes $K, L, \dots$ (rectangles) with an edge between variable i and factor K iff $i \in K$.

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 $i, j, \dots$ (circles) and factor nodes $K, L, \dots$ (rectangles) with an edge between variable i and factor K iff $i \in K$.

Example

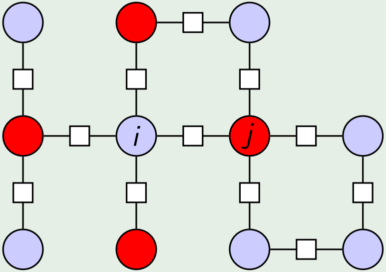


Some definitions

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

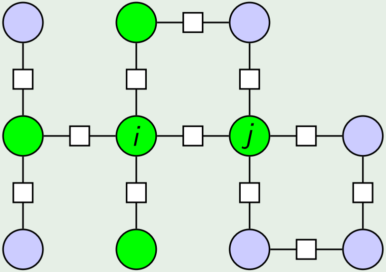
∂i is the Markov blanket of i, i.e. all neighboring variables of i.

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



Definition

∂i is the Markov blanket of i, i.e. all neighboring variables of i.

Definition

$\Delta i := \partial i \cup \{i\}$.

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.

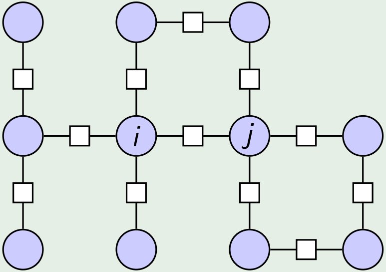
Related work: Montanari & Rizzo (2005), Parisi & Slanina (2005), Chertkov & Chernyak (2006)

Cavity graphs

Definition

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

Example

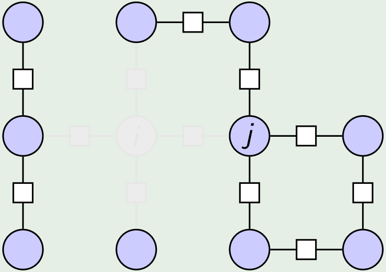


Cavity graphs

Definition

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

Example



Cavity graphs

Definition

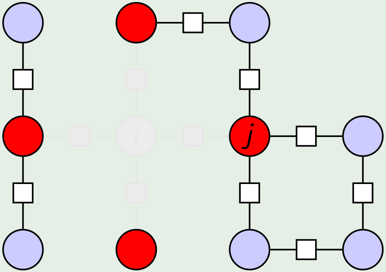
The cavity graph of i is the factor graph obtained by removing variable i together with all its neighboring factors.

Definition

The cavity distribution of i is the marginal of the cavity graph on ∂i:

$$P^{\setminus i}(x_{\partial i}) := \frac{1}{Z_{\setminus i}} \sum_{x_{\setminus \Delta i}} \prod_{\substack{K \in \mathcal{F} \\ i \notin K}} \psi_K(x_K).$$

Example

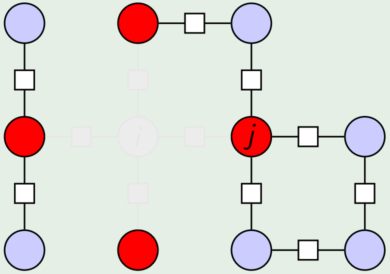


Cavity graphs

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 ∂i:

$$P^{\setminus i}(x_{\partial i}) := \frac{1}{Z_{\setminus i}} \sum_{x_{\setminus \Delta i}} \prod_{\substack{K \in \mathcal{F} \\ i \notin K}} \psi_K(x_K).$$

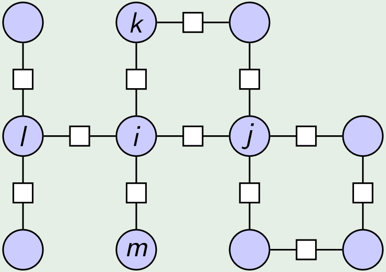
Proposition

$$P(x_{\Delta i}) \propto P^{\setminus i}(x_{\partial i}) \Psi_i(x_{\Delta i})$$

where

$$\Psi_i(x_{\Delta i}) := \prod_{\substack{K \in \mathcal{F} \\ i \in K}} \psi_K(x_K).$$

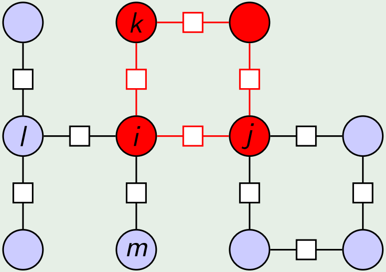
Example



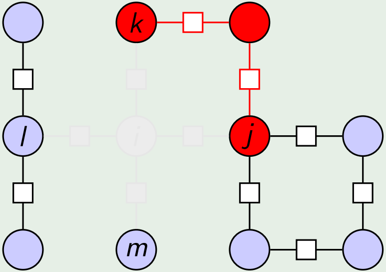
Cavities and loops

What is the relationship between loops and cavity distributions?

Example



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 $P^{\setminus i}$ of i.

$$P^{\setminus i}(x_{\partial i}) = P^{\setminus i}(x_j, x_k) P^{\setminus i}(x_l) P^{\setminus i}(x_m).$$

LCBP in a nutshell

In practice, exact cavity distributions are unavailable. Instead, we use approximate cavity distributions $Q^{\setminus i} \approx P^{\setminus i}$.

LCBP in a nutshell

1. Calculate *initial* approximate cavity distributions $\{Q_0^{\setminus i}\}_{i \in \mathcal{V}}$;
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 $\{Q_\infty^{\setminus i}\}_{i \in \mathcal{V}}$.

Consistency of single node marginals

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

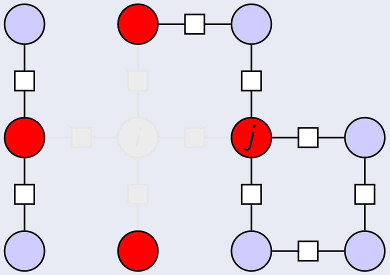
$$\Psi_i^{\setminus K}(x_{\Delta i}) := \frac{\Psi_i(x_{\Delta i})}{\psi_K(x_K)} = \prod_{\substack{L \in \mathcal{F} \\ i \in L, L \neq K}} \psi_L(x_L) \qquad \text{(and similarly for } j\text{).}$$

Consistency of single node marginals

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

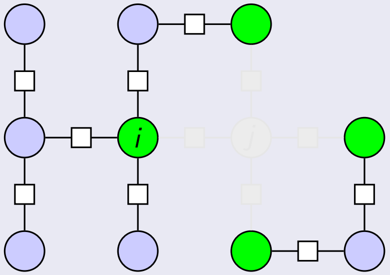
$$\Psi_i^{\setminus K}(x_{\Delta i}) := \frac{\Psi_i(x_{\Delta i})}{\psi_K(x_K)} = \prod_{\substack{L \in \mathcal{F} \\ i \in L, L \neq K}} \psi_L(x_L) \qquad \text{(and similarly for } j\text{)}.$$

Cavity graph of i



$P^{\setminus i}(x_{\partial i})$

Cavity graph of j



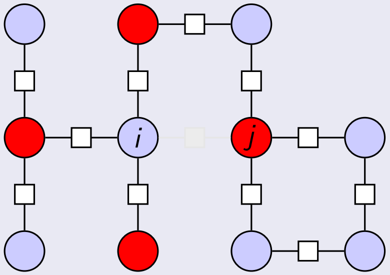
$P^{\setminus j}(x_{\partial j})$

Consistency of single node marginals

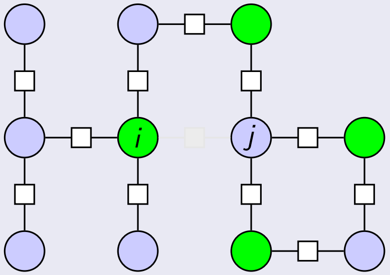
Let i, j be two neighboring variables with common factor K. Define

$$\Psi_i^{\setminus K}(x_{\Delta i}) := \frac{\Psi_i(x_{\Delta i})}{\psi_K(x_K)} = \prod_{\substack{L \in \mathcal{F} \\ i \in L, L \neq K}} \psi_L(x_L) \qquad \text{(and similarly for } j\text{)}.$$

Cavity graph of i



Cavity graph of j

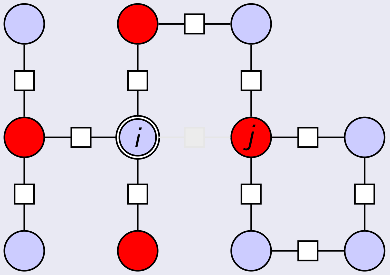


$$P^{\setminus i}(x_{\partial i})\Psi_i^{\setminus K}(\Delta i) \qquad P^{\setminus j}(x_{\partial j})\Psi_j^{\setminus K}(\Delta j)$$

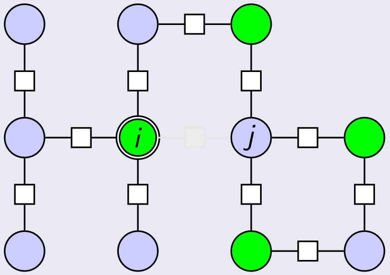
Let i, j be two neighboring variables with common factor K. Define

$$\Psi_i^{\setminus K}(x_{\Delta i}) := \frac{\Psi_i(x_{\Delta i})}{\psi_K(x_K)} = \prod_{\substack{L \in \mathcal{F} \\ i \in L, L \neq K}} \psi_L(x_L) \qquad \text{(and similarly for } j\text{).}$$

Cavity graph of i



Cavity graph of j



$$\sum_{x_{\partial i}} P^{\setminus i}(x_{\partial i}) \Psi_i^{\setminus K}(\Delta i) = \sum_{x_{\Delta j \setminus i}} P^{\setminus j}(x_{\partial j}) \Psi_j^{\setminus K}(\Delta j)$$

We modify the initial approximations $\{Q_0^{\setminus i}\}_{i\in\mathcal{V}}$ by changing single variable interactions but keeping higher order interactions fixed:

$$Q^{\setminus i}(x_{\partial i}) := Q_0^{\setminus i}(x_{\partial i}) \prod_{j\in\partial i} \phi_j^{\setminus i}(x_j),$$

where the factors $\phi_j^{\setminus i}$ are chosen such that:

$$\sum_{x_{\partial i}} Q^{\setminus i}(x_{\partial i})\Psi_i^{\setminus K}(\Delta i) = \sum_{x_{\Delta j \setminus i}} Q^{\setminus j}(x_{\partial j})\Psi_j^{\setminus K}(\Delta j) \qquad \forall_{i\in\mathcal{V}}\forall_{j\in\partial i}$$

This can be solved using simple fixed point iteration of the $\phi_j^{\setminus i}$ factors.

LCBP

1. Calculate *initial* approximate cavity distributions $\{Q_0^{\setminus i}\}_{i \in \mathcal{V}}$
2. 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: $$Q_{t+1}^{\setminus j} \propto Q_t^{\setminus j} \frac{\sum_{x_{\partial i}} Q_t^{\setminus i} \; \Psi_i^{\setminus K}}{\sum_{x_{\Delta j \setminus i}} Q_t^{\setminus j} \; \Psi_j^{\setminus K}}$$
    5: **end for**
    6: $t \leftarrow t + 1$
    7: **until** convergence

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

$$q_i(x_i) \propto \sum_{x_{\partial i}} Q_\infty^{\setminus i}(x_{\partial i}) \Psi_i(x_{\Delta i}).$$

Take uniform distributions. . .

Take uniform distributions...

Theorem

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

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:

Possible ways of calculating initial cavity distributions
High accuracy

A high accuracy initialization scheme:

for all $i \in \mathcal{V}$ **do**
  for all $x_{\partial i}$ **do**
    calculate $F^{\setminus i}_{Bethe}(x_{\partial i})$, the Bethe free energy corresponding to the cavity graph of i clamped in state $x_{\partial i}$, using BP
  end for
  $Q_0^{\setminus i}(x_{\partial i}) \leftarrow e^{-F^{\setminus i}_{Bethe}(x_{\partial i})}$
end for

Possible ways of calculating initial cavity distributions

High accuracy

A high accuracy initialization scheme:

for all $i \in \mathcal{V}$ **do**

  for all $x_{\partial i}$ **do**

    calculate $F^{\setminus i}_{Bethe}(x_{\partial i})$, the Bethe free energy corresponding to the cavity graph of i clamped in state $x_{\partial i}$, using BP

  end for

  $Q_0^{\setminus i}(x_{\partial i}) \leftarrow e^{-F^{\setminus i}_{Bethe}(x_{\partial i})}$

end for

Theorem

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

A high accuracy initialization scheme:

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

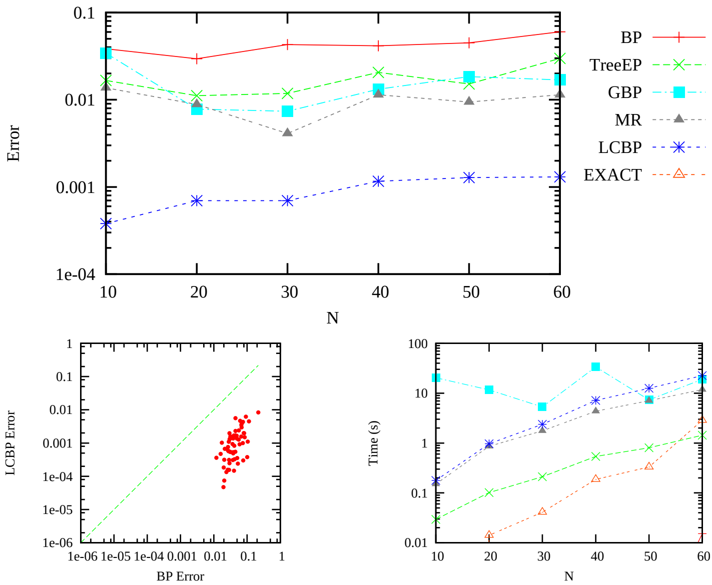
Theorem

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

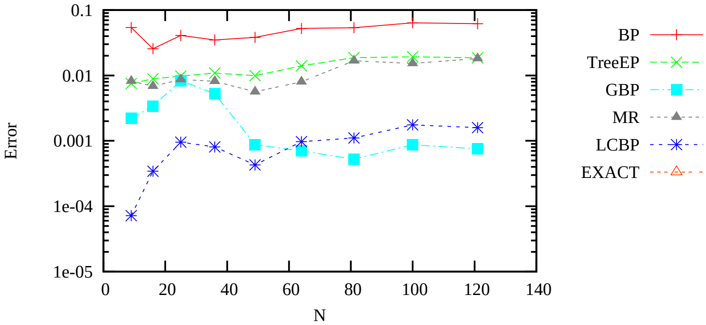
In general, this yields high accuracy approximations.

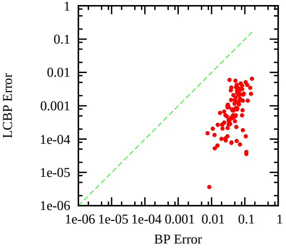
Experiments on random graphs

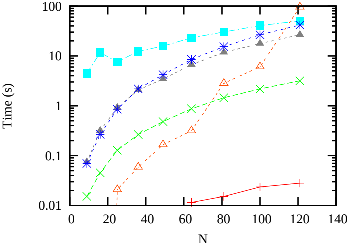
with binary variables and random pairwise interactions (fixed degree $|\partial i| = 5$)

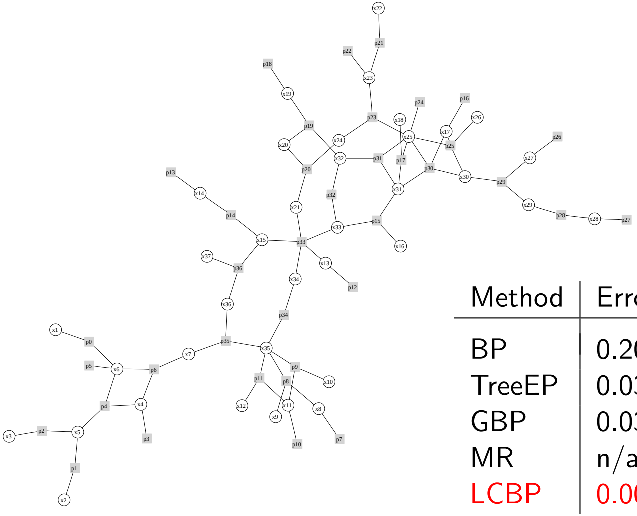


Experiments on periodic grids
with binary variables and random pairwise interactions









Method	Error	Time (s)
BP	0.203	0.00
TreeEP	0.039	0.22
GBP	0.035	161.0
MR	n/a	n/a
LCBP	0.00054	23.4

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

- Andrea Montanari and Tommaso Rizzo, *JSTAT* 2005(10):P10011, 2005.
- Giorgio Parisi and Frantisek Slanina, *arXiv.org preprint* `cond-mat/0512529`.
- Michael Chertkov and Vladimir Y Chernyak, *JSTAT* 2006(06):P06009, 2006.
- J M Mooij and H J Kappen, *arXiv.org preprint* `cs.AI/0612030`.

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

$t \leftarrow 0$
repeat
  for all $i \in \mathcal{V}$ **do**
    for all $K \in N_i$ **do**

$$Q_{t+1}^{\setminus j} \leftarrow Q_t^{\setminus j} \frac{\prod_{j \in K \setminus i} \left(\sum_{\Delta j \setminus (K \setminus i)} Q_t^{\setminus j} \Psi_j^{\setminus K} \right)^{1/|K \setminus i|}}{\sum_{x_{\Delta i \setminus (K \setminus i)}} Q_t^{\setminus i} \Psi_i^{\setminus K}}$$

    end for
  end for
  $t \leftarrow t + 1$
until convergence