Correctness of Local Probability Propagation in Graphical Models with Loops

by

Yair Weiss

Ben Eyal

November 16, 2016

Outline

Outline

1. Introduction
   - Definitions
   - The Problem

2. Belief Propagation on PGMs with a Single Loop
   - Correctness

3. Conclusion
   - What Have We Achieved?
   - The End

Breaking Down the Title - Part I

Correctness of Local Probability Propagation in Graphical Models with Loops

Breaking Down the Title - Part I

Correctness of Local Probability Propagation in Graphical Models with Loops

Probabilistic Graphical Models

Probabilistic Graphical Models

Definition

A Probabilistic Graphical Model (PGM) is a graph, either directed or undirected, in which the nodes correspond to random variables, and the edges correspond to direct probabilistic interactions between them.

Probabilistic Graphical Models

Probabilistic Graphical Models

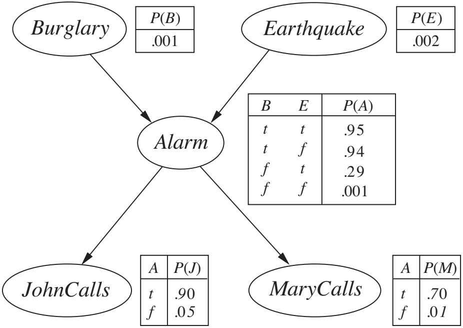
Definition

A **Bayes network** is a directed acyclic PGM whose edges can be seen as "cause" and "effect", e.g. an edge $X \to Y$ means that X directly influences Y, or "X is the cause of Y".

Probabilistic Graphical Models

Definition

A **Bayes network** is a directed acyclic PGM whose edges can be seen as "cause" and "effect", e.g. an edge $X \to Y$ means that X directly influences Y, or "X is the cause of Y".



Probabilistic Graphical Models

Probabilistic Graphical Models

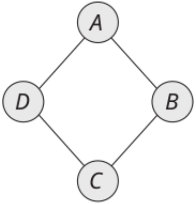
Definition

A Markov Random Field (MRF), or a **Markov network**, is an undirected PGM which is used when the relations between random variables are symmetric, rather than hierarchical, e.g. pixels in an image.

Probabilistic Graphical Models

Definition

A Markov Random Field (MRF), or a **Markov network**, is an undirected PGM which is used when the relations between random variables are symmetric, rather than hierarchical, e.g. pixels in an image.



$\phi_1(A,B)$

a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

$\phi_2(B,C)$

b^0	c^0	100
b^0	c^1	1
b^1	c^0	1
b^1	c^1	100

$\phi_3(C,D)$

c^0	d^0	1
c^0	d^1	100
c^1	d^0	100
c^1	d^1	1

$\phi_4(D,A)$

d^0	a^0	100
d^0	a^1	1
d^1	a^0	1
d^1	a^1	100

Breaking Down the Title - Part II

Correctness of Local Probability Propagation in Graphical Models with Loops

Breaking Down the Title - Part II

Correctness of Local Probability Propagation in Graphical Models with Loops

Reminder

Conditional Probability

$$\Pr(A \mid B) = \frac{\Pr(A, B)}{\Pr(B)}$$

Reminder

Conditional Probability

$$\Pr(A \mid B) = \frac{\Pr(A, B)}{\Pr(B)}$$

Bayes' Theorem

Bayes' theorem gives us the relation between the posterior probability, $\Pr(A \mid B)$, and the prior probability, $\Pr(A)$.

$$\Pr(A \mid B) = \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B)}$$

Reminder

Reminder

Marginal Probability

Given a joint probability distribution $\Pr(X, Y)$,

$$\begin{aligned}\Pr(X = x) &= \sum_y \Pr(X = x, Y = y) \\ &= \sum_y \Pr(X = x \mid Y = y) \Pr(Y = y)\end{aligned}$$

is the marginal distribution of X.

Belief Propagation

The Task of Inference

Inference is divided into three subtasks:

Belief Propagation

The Task of Inference

Inference is divided into three subtasks:

- *Marginalization*: Calculating the marginal probability of a variable given the observed variables – e.g. $\Pr\left(C = c \mid \mathcal{O}\right)$.

Belief Propagation

The Task of Inference

Inference is divided into three subtasks:

- *Marginalization*: Calculating the marginal probability of a variable given the observed variables – e.g. $\Pr(C = c \mid \mathcal{O})$.
- *Maximum a posteriori (MAP) assignment*: Finding assignments to the unobserved variables that are most probable given the observed variables – e.g. finding (b, c, d, e) such that $\Pr(B = b, C = c, D = d, E = e \mid \mathcal{O})$ is maximized.

Belief Propagation

The Task of Inference

Inference is divided into three subtasks:

- *Marginalization*: Calculating the marginal probability of a variable given the observed variables – e.g. $\Pr(C = c \mid \mathcal{O})$.
- *Maximum a posteriori (MAP) assignment*: Finding assignments to the unobserved variables that are most probable given the observed variables – e.g. finding (b, c, d, e) such that $\Pr(B = b, C = c, D = d, E = e \mid \mathcal{O})$ is maximized.
- *Maximum Marginal (MM) assignment*: Finding assignments to the unobserved variables that maximize the marginal probability of the assignment – e.g. finding (b, c, d, e) such that $\Pr(B = b \mid \mathcal{O})$, $\Pr(C = c \mid \mathcal{O})$, $\Pr(D = d \mid \mathcal{O})$, $\Pr(E = e \mid \mathcal{O})$ are all maximized.

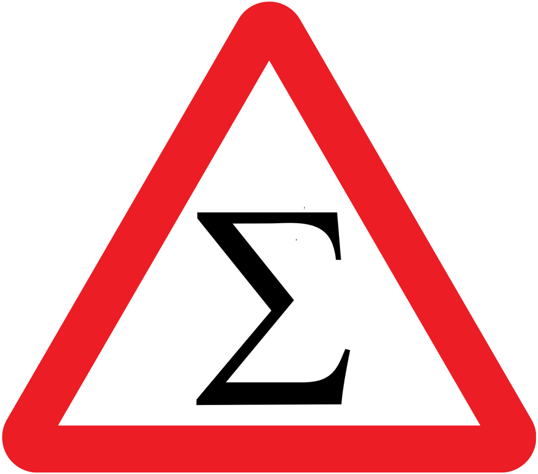
Belief Propagation

Definition

Belief Propagation (Pearl, 1982) is an iterative message-passing algorithm used to perform inference on PGMs.
The algorithm gives exact marginals on polytrees (a directed graph with only a single path between any two nodes).

Example

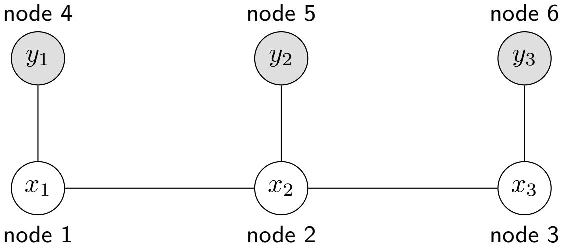
Example



Example

Example

We would like to compute the maximum marginal probability $\Pr(X = x_1 \mid Y)$ given the following Markov network:



Example

Example

If we knew nothing about the structure of the joint probability:

$$\Pr(X = x_1 \mid Y) = \sum_{x_2} \sum_{x_3} \Pr(x_1, x_2, x_3 \mid Y)$$

Example

If we knew nothing about the structure of the joint probability:

$$\Pr(X = x_1 \mid Y) = \sum_{x_2} \sum_{x_3} \Pr(x_1, x_2, x_3 \mid Y)$$

This might not look like much, but the time complexity increases exponentially with the number of unobserved variables.

Example

If we knew nothing about the structure of the joint probability:

$$\Pr(X = x_1 \mid Y) = \sum_{x_2} \sum_{x_3} \Pr(x_1, x_2, x_3 \mid Y)$$

This might not look like much, but the time complexity increases exponentially with the number of unobserved variables.
Did I mention that both exact and approximate inference are hard?

Example

If we knew nothing about the structure of the joint probability:

$$\Pr(X = x_1 \mid Y) = \sum_{x_2} \sum_{x_3} \Pr(x_1, x_2, x_3 \mid Y)$$

This might not look like much, but the time complexity increases exponentially with the number of unobserved variables.
Did I mention that both exact and approximate inference are hard?
As in $\mathcal{NP}$-hard?

Example

Lucky for us, we do know something about the structure of the joint probability!

Example

Lucky for us, we do know something about the structure of the joint probability!

$$\Pr(x_1|Y)=\frac{1}{\Pr(Y)}\sum_{x_2}\sum_{x_3}\phi_{12}(x_1,x_2)\phi_{23}(x_2,x_3)\psi_1(y_1,x_1)\psi_2(y_2,x_2)\psi_3(y_3,x_3)$$

Example

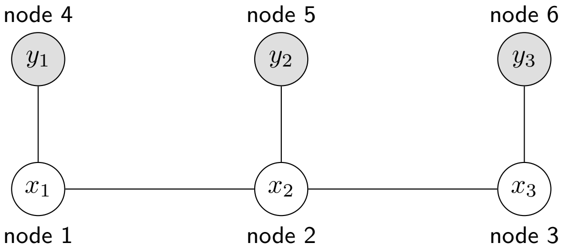
Lucky for us, we do know something about the structure of the joint probability!

$$\Pr(x_1|Y)=\frac{1}{\Pr(Y)}\sum_{x_2}\sum_{x_3}\phi_{12}(x_1,x_2)\phi_{23}(x_2,x_3)\psi_1(y_1,x_1)\psi_2(y_2,x_2)\psi_3(y_3,x_3)$$

Do not despair! This is going to be very clear – very quick.

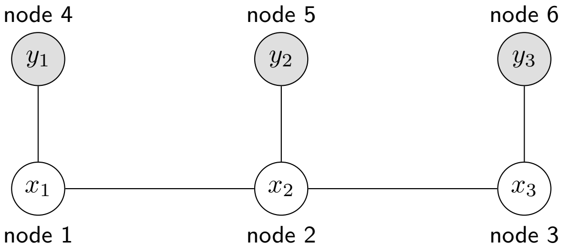
Example

$$\Pr(x_1|Y) = \frac{1}{\Pr(Y)} \sum_{x_2} \sum_{x_3} \phi_{12}(x_1,x_2)\phi_{23}(x_2,x_3)\psi_1(y_1,x_1)\psi_2(y_2,x_2)\psi_3(y_3,x_3)$$



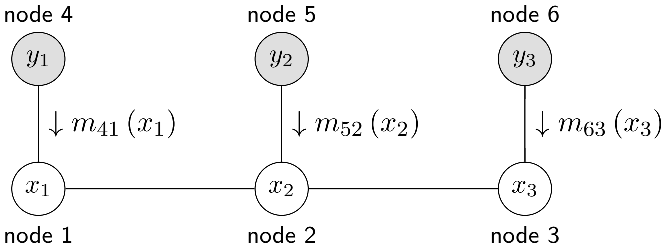
Example

$$\Pr(x_1|Y) = \frac{1}{\Pr(Y)} \psi_1(y_1,x_1) \sum_{x_2} \phi_{12}(x_1,x_2)\psi_2(y_2,x_2) \sum_{x_3} \phi_{23}(x_2,x_3)\psi_3(y_3,x_3)$$



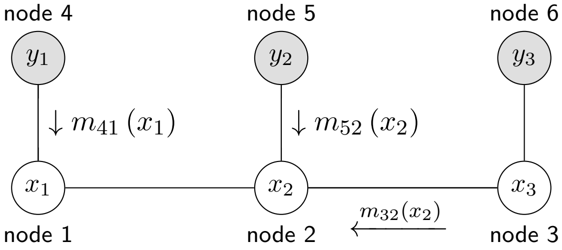
Example

$$\Pr(x_1|Y) = \frac{1}{\Pr(Y)} m_{41}(x_1) \sum_{x_2} \phi_{12}(x_1, x_2) m_{52}(x_2) \sum_{x_3} \phi_{23}(x_2, x_3) m_{63}(x_3)$$



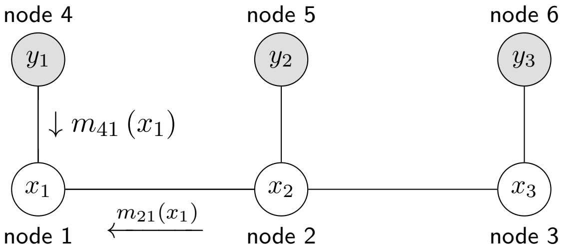
Example

$$\Pr(x_1|Y)=\frac{1}{\Pr(Y)}m_{41}(x_1)\sum_{x_2}\phi_{12}(x_1,x_2)m_{52}(x_2)m_{32}(x_2)$$



Example

$$\Pr(x_1|Y) = \frac{1}{\Pr(Y)} m_{41}(x_1) m_{21}(x_1)$$



Belief Propagation

Belief Propagation

Inside Belief Propagation

In Belief Propagation, every iteration is called belief update (BU). In belief update, the message that node X sends to node Y is updated as follows:

Belief Propagation

Inside Belief Propagation

In Belief Propagation, every iteration is called belief update (BU). In belief update, the message that node X sends to node Y is updated as follows:

1. Combine all messages coming into X except for that coming from Y into a vector $\vec{v}$. The combination is done by multiplying all the message vectors element by element.

Belief Propagation

Inside Belief Propagation

In Belief Propagation, every iteration is called belief update (BU). In belief update, the message that node X sends to node Y is updated as follows:

1. Combine all messages coming into X except for that coming from Y into a vector $\vec{v}$. The combination is done by multiplying all the message vectors element by element.
2. Multiply $\vec{v}$ by the matrix M_{XY} corresponding to the link from X to Y.

Belief Propagation

Inside Belief Propagation

In Belief Propagation, every iteration is called belief update (BU). In belief update, the message that node X sends to node Y is updated as follows:

1. Combine all messages coming into X except for that coming from Y into a vector $\vec{v}$. The combination is done by multiplying all the message vectors element by element.
2. Multiply $\vec{v}$ by the matrix M_{XY} corresponding to the link from X to Y.
3. Normalize the product $M_{XY}\vec{v}$ so it sums to 1. The normalized vector is sent to Y.

The belief vector for a node X is obtained by combining all incoming messages to X and normalizing.

Belief Propagation

Maximum A-Posteriori (MAP) Assignment Using Belief Propagation

To find a MAP assignment, we use the algorithm above, with a small change:
Instead of using matrix multiplication (a sum-of-products), we take the argmax, giving us a max-of-products.
This step is called belief revision (BR).

Outline

1. Introduction
   - Definitions
   - The Problem
2. Belief Propagation on PGMs with a Single Loop
   - Correctness
3. Conclusion
   - What Have We Achieved?
   - The End

Breaking Down the Title - Part III

Correctness of Local Probability Propagation in Graphical Models with Loops

Breaking Down the Title - Part III

Correctness of Local Probability Propagation in Graphical Models with Loops

"Loopy" Belief Propagation

Belief Propagation on General Graphs

As we have already seen, Belief Propagation is guaranteed to converge on the exact marginals or MAP assignments given a singly-connected graph.

What happens when we have loops?

"Loopy" Belief Propagation

Belief Propagation on General Graphs

As we have already seen, Belief Propagation is guaranteed to converge on the exact marginals or MAP assignments given a singly-connected graph.

What happens when we have loops?

- Divergence

"Loopy" Belief Propagation

Belief Propagation on General Graphs

As we have already seen, Belief Propagation is guaranteed to converge on the exact marginals or MAP assignments given a singly-connected graph.

What happens when we have loops?

- Divergence
- Fluctuations between different values

“Loopy” Belief Propagation

Belief Propagation on General Graphs

As we have already seen, Belief Propagation is guaranteed to converge on the exact marginals or MAP assignments given a singly-connected graph.
What happens when we have loops?

- Divergence
- Fluctuations between different values
- Convergence on the wrong values

"Loopy" Belief Propagation

Belief Propagation on General Graphs

As we have already seen, Belief Propagation is guaranteed to converge on the exact marginals or MAP assignments given a singly-connected graph.

What happens when we have loops?

- Divergence
- Fluctuations between different values
- Convergence on the wrong values

Theoretically...

"Loopy" Belief Propagation

Belief Propagation on General Graphs

As we have already seen, Belief Propagation is guaranteed to converge on the exact marginals or MAP assignments given a singly-connected graph.

What happens when we have loops?

- Divergence
- Fluctuations between different values
- Convergence on the wrong values

Theoretically...

In practice, "loopy" Belief Propagation is doing surprisingly well!

Outline

Breaking Down the Title - Finale

Correctness of Local Probability Propagation in Graphical Models with Loops

Breaking Down the Title - Finale

Correctness of Local Probability Propagation in Graphical Models with Loops

Theory vs. Practice

Although Loopy Belief Propagation is doing great in some cases, the theoretical problems shown earlier do arise.

Theory vs. Practice

Although Loopy Belief Propagation is doing great in some cases, the theoretical problems shown earlier do arise.
We wish to know when Loopy Belief Propagation is guaranteed to converge.

Theory vs. Practice

Although Loopy Belief Propagation is doing great in some cases, the theoretical problems shown earlier do arise.
We wish to know when Loopy Belief Propagation is guaranteed to converge.
In his paper, Weiss proves that given a PGM with a single loop:

Theory vs. Practice

Although Loopy Belief Propagation is doing great in some cases, the theoretical problems shown earlier do arise.
We wish to know when Loopy Belief Propagation is guaranteed to converge.
In his paper, Weiss proves that given a PGM with a single loop:

- belief revision in Loopy Belief Propagation will always converge on a correct MAP assignment.

Theory vs. Practice

Although Loopy Belief Propagation is doing great in some cases, the theoretical problems shown earlier do arise.
We wish to know when Loopy Belief Propagation is guaranteed to converge.
In his paper, Weiss proves that given a PGM with a single loop:

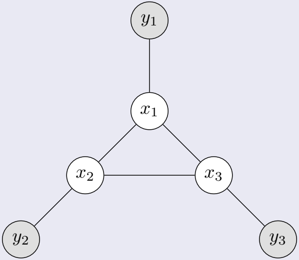
- belief revision in Loopy Belief Propagation will always converge on a correct MAP assignment.
- belief update will converge on a marginal that is not the correct marginal, but is closely related to it.

Motivation

The problem with loops is what Weiss refers to as "double counting":

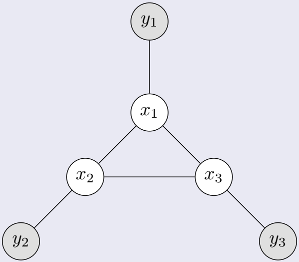
Motivation

The problem with loops is what Weiss refers to as “double counting”:



Motivation

The problem with loops is what Weiss refers to as “double counting”:



In a graph with a single loop, Weiss shows this “double counting” is guaranteed to occur equally among all nodes in the loop, therefore still reaching convergence.

Outline

What Have We Achieved?

- We can define a Probabilistic Graphical Model (PGM), and name two specific types.

- We can define a Probabilistic Graphical Model (PGM), and name two specific types.
- We know what are prior and posterior probabilities, and how Bayes' theorem connects them.

- We can define a Probabilistic Graphical Model (PGM), and name two specific types.
- We know what are prior and posterior probabilities, and how Bayes' theorem connects them.
- We can trace the steps of the Belief Propagation algorithm.

- We can define a Probabilistic Graphical Model (PGM), and name two specific types.
- We know what are prior and posterior probabilities, and how Bayes' theorem connects them.
- We can trace the steps of the Belief Propagation algorithm.
- We are not afraid of phrases such as "sum-product", "max-product", or "message-passing".

- We can define a Probabilistic Graphical Model (PGM), and name two specific types.
- We know what are prior and posterior probabilities, and how Bayes' theorem connects them.
- We can trace the steps of the Belief Propagation algorithm.
- We are not afraid of phrases such as "sum-product", "max-product", or "message-passing".
- We are aware of three problems that may arise in PGMs with loops...

- We can define a Probabilistic Graphical Model (PGM), and name two specific types.
- We know what are prior and posterior probabilities, and how Bayes' theorem connects them.
- We can trace the steps of the Belief Propagation algorithm.
- We are not afraid of phrases such as “sum-product”, “max-product”, or “message-passing”.
- We are aware of three problems that may arise in PGMs with loops...
- ...but we also know that loops are not always a problem in practice.

- We can define a Probabilistic Graphical Model (PGM), and name two specific types.
- We know what are prior and posterior probabilities, and how Bayes' theorem connects them.
- We can trace the steps of the Belief Propagation algorithm.
- We are not afraid of phrases such as "sum-product", "max-product", or "message-passing".
- We are aware of three problems that may arise in PGMs with loops...
- ...but we also know that loops are not always a problem in practice.
- We know what to expect when running Loopy Belief Propagation on a PGM with a single loop.

Outline

The End

Thank you for your time!

Thank you for your time!

Questions?