Spatial Triplet Markov Trees for auxiliary variational inference in Spatial Bayes Networks
Hugo Gangloff, Jean-Baptiste Courbot, Emmanuel Monfrini, Christophe Collet

To cite this version:
Hugo Gangloff, Jean-Baptiste Courbot, Emmanuel Monfrini, Christophe Collet. Spatial Triplet Markov Trees for auxiliary variational inference in Spatial Bayes Networks. SMTDA 2020: 6th international conference on Stochastic Modeling Techniques and Data Analysis, Jun 2020, Barcelone (online), Spain. pp.237-249. hal-03122810

HAL Id: hal-03122810
https://hal.science/hal-03122810
Submitted on 30 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Spatial Triplet Markov Trees for Auxiliary Variational Inference in Spatial Bayes Networks

Hugo Gangloff¹,²
Jean-Baptiste Courbot³, Emmanuel Monfrini⁴, Christophe Collet¹,

¹Université de Strasbourg - CNRS ICube UMR 7357, Illkirch, France
²GEPROVAS, Strasbourg, France
³IRIMAS UR 7499 Université de Haute-Alsace, Mulhouse, France
⁴SAMOVAR, Télécom SudParis, IP Paris, Évry, France

Stochastic Modeling Techniques and Data Analysis, Virtual Conference, June, 2-5 2020
1 Introduction

2 The probabilistic models
   ■ Markov Tree network
   ■ Spatial Bayes Network
   ■ STMT network

3 Variational Inference
   ■ The method
   ■ Several VIs on a small example
   ■ Experimental results

4 Conclusion & Perspectives
Introduction
Bayesian Networks: main definitions

- Also known as acyclic Directed Graphical Models (DGM)
- Let $\mathcal{G} = (\mathcal{S}, \mathcal{E})$ be a directed graph:
  - Directed edges link parents vertices $\mathcal{P}(s)$ and son vertex $s \in \mathcal{S}$
  - Vertices without any parent are root vertices (denoted $r$)
- Random variables associated to each vertice
- Define local conditional probabilities $p(x_r)$ or $p(x_s | x_{\mathcal{P}(s)})$
  - They define joint distribution on $\mathcal{G}$:

$$p(x) = \prod_{r \in \text{roots}} p(x_r) \prod_{s \in \mathcal{S} \setminus \text{roots}} p(x_s | x_{\mathcal{P}(s)})$$  \hspace{1cm} (1)
Inference basics

Inference tasks:

- compute marginals \( p(x_s) \)
- compute conditionals \( p(x_s | x_{p(s)}) \)
- compute a mode of the distribution
  (e.g. Maximum A Posteriori (MAP): \( \hat{x} = \arg\max_x p(x) \), Maximum Posterior Modes (MPM): \( \hat{x}_s = \arg\max_{x_s} p(x_s) \))

Direct computations in Bayes Networks:

- Forward Backward (Stratonovich 1965) / Belief Propagation (Pearl 1982)

Approximate computations in general DGM:

- Loopy Belief Propagation (Weiss 2000)
- MCMC: Gibbs sampling (Geman et al. 1984)
- Variational Inference (Jordan et al. 1999)
Bayesian Networks: popularity

Hidden Markov Models (Baum et al. 1966) (Dymarski 2011)
Bayesian Networks with latent $X$ and observed variables $Y$. Widely known for:

- Stock index forecasting (Gorynin et al. 2017)
- Speech processing (Toda 2011)
- Gene prediction (Stanke et al. 2006)

Hidden Markov Chain with Independent Noise:

$$p(x, y) = p(x_r) \prod_{s \in \bar{S}} p(x_s | x_{s-1}) \prod_{s \in S} p(y_s | x_s) \quad (2)$$

Hidden Markov Chain (direct computations)
The probabilistic models
Markov Tree (MT) Bayesian Networks

Vertices (except roots) have one parent.

MT joint distribution (Monfrini et al. 2003):

\[
p(x) = p(x_r) \prod_{s \in \bar{S}} p(x_s|x_{s-}). \quad (3)
\]

- Dyadic tree: 1 root and 2 sons/vertex (except last layer).
- Quadtree: 1 root and 4 sons/vertex (except last layer).

Exact inference dyadic/quad trees:

- Upward-Downward (Durand et al. 2004) (principled approach to get the marginals in dyadic MT).

\[
p(x_s) = \sum_{x_{s-}} p(x_{s-})p(x_s|x_{s-}). \quad (4)
\]
Spatial Bayesian Networks (SBN)

A SBN is a Bayesian network with joint distribution (this paper):

\[
p(x) = p(x_r) \prod_{s \in \mathcal{S}} p(x_s | x_{s^-}, x_{v(s)}),
\]  

(5)

\(v\) associates a father’s neighbour: not a Markov Tree!

The spatial context is modeled but the numerous loops makes inference complex and approximate.

\[\text{SBN (based on a dyadic tree) with 4 layers (approximate computations)}\]
Spatial Triplet Markov Tree Bayesian Networks

In STMT, $T = (X, V)$ and $V$ auxiliary process (this paper) and (Courbot et al. 2018).

$T$ is a Markov Tree with joint distribution:

$$p(t) = p(t_r) \prod_{s \in \bar{S}} p(t_s | t_{s-}),$$

(6)

with $t_s$ a triplet $(x_s, v_s^\leftarrow, v_s^\rightarrow)$

$STMT$ (Markov property) 
$(direct computations)$
Spatial Triplet Markov Tree Bayesian Networks

In STMT, $T = (X, V)$ and $V$ auxiliary process (this paper) and (Courbot et al. 2018).

$T$ is a Markov Tree with joint distribution:

$$p(t) = p(t_r) \prod_{s \in \mathcal{S}} p(t_s | t_{s-}), \quad (6)$$

with $t_s$ a triplet $(x_s, v_s^{\leftarrow}, v_s^{\rightarrow})$ such that:

$$p(t_s | t_{s-}) = p(x_s | x_{s-}, v_{s-}) p(v_s^{\leftarrow} | x_{s-}, v_{s-}) p(v_s^{\rightarrow} | x_{s-}, v_{s-}). \quad (7)$$

The transitions are specially designed to model the spatial context.
Sampling from the models

Test the ability to capture spatial context by *clamped* sampling from Markov Random Field (MRF) realizations

Average different rate from MRF
Sampling from the models

Test the ability to capture spatial context by *clamped* sampling from Markov Random Field (MRF) realizations

Average different rate from MRF

→ 0.142
Sampling from the models

Test the ability to capture spatial context by *clamped* sampling from Markov Random Field (MRF) realizations.

Average different rate from MRF:
- MRF: 0.142
- MT: 0.072
- SBN: 0.068

SBN and STMT seem to capture the spatial context.
Sampling from the models
Test the ability to capture spatial context by clamped sampling from Markov Random Field (MRF) realizations

MRF

Average different rate from MRF

MT

→ 0.142

SBN

→ 0.072

STMT

→ 0.068

SBN and STMT seem to capture the spatial context
Variational Inference
The method

VI recasts inference into a maximization problem (Jordan et al. 1999):

$$-\text{KL}(q(x)||p(x)) = \mathbb{E}_q[\log p(x)] - \mathbb{E}_q[\log q(x)],$$

(8)

where:

- $p$ is the target distribution (SBN here)
- $q$ is the variational distribution (structured or not)
The method

VI recasts inference into a maximization problem (Jordan et al. 1999):

\[- \mathbb{KL}(q(\mathbf{x})||p(\mathbf{x})) = \mathbb{E}_q[\log p(\mathbf{x})] - \mathbb{E}_q[\log q(\mathbf{x})], \quad (8)\]

where:

- \(p\) is the target distribution (SBN here)
- \(q\) is the variational distribution (structured or not)

Extended to integrate auxiliary variables (Agakov et al. 2004)

\[- \mathbb{KL}(q(\mathbf{x}, \mathbf{v})||\tilde{p}(\mathbf{x}, \mathbf{v})) = \mathbb{E}_q[\log \tilde{p}(\mathbf{x}, \mathbf{v})] - \mathbb{E}_q[\log q(\mathbf{x}, \mathbf{v})], \quad (9)\]

\(\tilde{p}\) is extended with auxiliary variables (with condition: \(\tilde{p}(\mathbf{x}) = p(\mathbf{x})\))
A small SBN example I

- Target distribution:

Let $\mathbf{x} = (a, a^\leftarrow, a^\rightarrow, b, c, d, e, f, g)$,
$p(\mathbf{x})$ has a SBN factorization

(approximate computations)
A small SBN example

Target distribution:

Let \( x = (a, a^{\leftarrow}, a^{\rightarrow}, b, c, d, e, f, g) \),
\( p(x) \) has a SBN factorization

Variational approximation:

Mean-Field (MF) approximation:

\[
q^{MF}(x) = \prod_i q(x_i) \quad (10)
\]

→ Independent random variables

(apply computations)

MF approximation
(apply computations)
A small SBN example I

- Target distribution:

Let $\mathbf{x} = (a, a^\leftarrow, a^\rightarrow, b, c, d, e, f, g)$, $p(\mathbf{x})$ has a SBN factorization

- Variational approximation:

Markov Tree (MT) approximation:

$$q^{MT}(\mathbf{x}) = q(x_r) \prod_{s \in \bar{S}} q(x_s|x_{s^-}). \quad (10)$$

$\rightarrow q^{MT}$ is MT structured
A small SBN example II

- Target distribution:

\[
\tilde{p}(\mathbf{x}, \mathbf{v}) \text{ is a SBN with auxiliary variables (and } \tilde{p}(\mathbf{x}) = p(\mathbf{x}))
\]

- Variational approximation:

\[
q^{\text{STMT}}(\mathbf{t}) = q(\mathbf{t}_r) \prod_{s \in \tilde{S}} q(\mathbf{t}_s | \mathbf{t}_{s-})
\]

(11)

\[
\rightarrow q^{\text{STMT}} \text{ is STMT structured}
\]

(\emph{approximate computations})
Small network example: model proximity

→ Dispersions of errors for true marginals estimation (1000 trials)
Small network example: model proximity

→ Dispersions of errors for true marginals estimation (1000 trials)

- Error dispersion: MF VI > MT VI > STMT VI
- STMT seems to best capture the enhanced correlations of SBN
Conclusion & Perspectives
Conclusion & Perspectives

Conclusions:

- 2 new tree-structured models with rich correlations
- Importance of the triplet structure for exact approaches in more complex networks.
- Auxiliary variables in probabilistic models can be an efficient asset if the model is well-thought!

Perspectives:

- Analytical comparison STMT/SBN
- Extend the work to quadtrees for image processing
References


