Sequential Monte Carlo for Graphical Models

Christian A. Naesseth*, Fredrik Lindsten*,† and Thomas B. Schön†
*Linköping University, †The University of Cambridge, ‡Uppsala University

Summary

- Probabilistic Graphical Models (PGM) are routinely used in computer vision, speech recognition, machine learning, etc.
- Cyclic and/or continuous PGMs in general make inference non-trivial and approximate methods are needed.
- Proposed Sequential Monte Carlo (SMC)-based methods:
  - “Standard” SMC: Exploits a sequential decomposition to construct the joint probability of the PGM using “standard” SMC methods.
  - Particle MCMC with Partial Blocking: Blocking and SMC-based kernels for MCMC lets us design efficient and parallelizable high-dimensional MCMC methods.

Graphical Models

We consider models on the form,

\[ p(X_\mathcal{V}) = \frac{1}{Z} \prod_{C \in \mathcal{E}} \psi(X_C), \]

where the graph \( G = (\mathcal{V}, \mathcal{E}) \) has vertex set \( \mathcal{V} = \{x_1, \ldots, x_{|\mathcal{V}|}\} \), edge set \( \mathcal{E} \), cliques \( C \) and \( Z = \int \prod_{C \in \mathcal{E}} \psi(X_C) dX_\mathcal{V} \) is the partition function (normalisation constant).

Sequential Decomposition

Sequential decomposition of a PGM is all about using structure encoded by factors in the graph to construct a valid sequence of target distributions for an SMC sampler.

Sequential Monte Carlo

A standard approach to approximate a sequence of distributions. Denote the newly added random variables \( \xi_k \) and all the added random variables at iteration \( k \), \( X_{\mathcal{V}_k} \subseteq \{x_1, \ldots, x_{|\mathcal{V}|}\} \).

Algorithm 1 SMC for PGM

\[
\text{Perform each step for } i = 1, \ldots, N. \\
\text{Sample } X_{\mathcal{V}_k} \sim p(X_i | \mathcal{V}_{k-1}). \text{ Set } w_k^i = W_k(X_{\mathcal{V}_k}). \\
\text{for } k = 2 \text{ to } K \\
\text{Sample } \xi_k \text{ according to resampling weights: } \pi(\xi_k = j) = \frac{w_k^{i_j}}{\sum_{i=1}^{N} w_k^{i_j}}, \text{ } j \in \{1, \ldots, N\} \\
\text{Sample } \xi_k \sim r_k(\xi_k | X_{\mathcal{V}_{k-1}}) \text{ and set } X_{\mathcal{V}_k} = X_{\mathcal{V}_{k-1}} \cup \xi_k. \text{ Set } w_k^{i_j} = W_k(X_{\mathcal{V}_k}). \\
\text{end for}
\]

Particle MCMC with Partial Blocking

Particle Markov chain Monte Carlo (PMCMC) is a systematic way of constructing Markov kernels for high-dimensional spaces using sequential Monte Carlo techniques.

Standard partially blocked Gibbs updates:

\[
X_1 \sim p(X_1 | X_2, y, \theta), \\
X_2 \sim p(X_2 | X_1, y, \theta), \\
\theta \sim p(\theta | X_1, X_2, y).
\]

When we cannot simulate exactly from the full conditionals above we can do so approximately using Particle Gibbs.

This means we run conditional SMC on the block and sample one of the trajectories \( \{X_{\mathcal{V}_k}\} \) as our new MCMC sample.

Example: Partial blocking of a Markov random field (MRF)

Example: Partial blocking of hidden Markov models

Example: Gaussian MRF (10×10, strong latent interactions)

Figure 1: Example factor graph describing dependencies between random variables in a PDF.

The construction applies to a variety of interesting problems! Factor graphs can in fact be used to represent both Bayesian networks (e.g. state-space models, hierarchical models) and Markov networks (e.g. Restricted Boltzmann Machines, Ising models, ...)

Example: 16×16 (β = 1.1) and 64×64 Classical XY model (β = {0.5, 1.1})

Figure 2: Left: XY model Middle Left: MSE in the estimates of \( \log Z \) for AIS and four different orderings in SMC for PGM. Middle Right: Right: The logarithm of the estimated partition function for the 64×64 XY model with inverse temperature 0.5 (middle right) and 1.1 (right).

Example: Evaluation of LDA topic models (likelihood estimation)

Figure 3: Left: LDA topic model. Middle: Right: Estimates of the log-likelihood of heldout documents for various real datasets using SMC for PGM and Left-Right Sequential (LRS) sampler.


http://www.control.isy.liu.se/