Look Ma, No Latent Variables: Accurate Cutset Networks via Compilation

Tahrima Rahman, Shasha Jin, Vibhav Gogate

The University of Texas at Dallas

ICML 2019

What is the paper about?

Probabilistic Graphical Models (**PGMs**) vs Latent Tractable Probabilistic Models (**LTPMs**):

Criteria	**Who wins?**	**Why?**
Test set log-likelihood	PGMs	Expressive model
Marginal (MAR) estimates	LTPMs	Reliable Exact Inference
Maximum-a-posteriori (MAP) estimates	No Clear Winner	Both use unreliable approximate inference approaches

Tractability helps achieve better marginal predictions even though the model fit is inferior. **Can we do the same for MAP inference?**

What is the paper about?

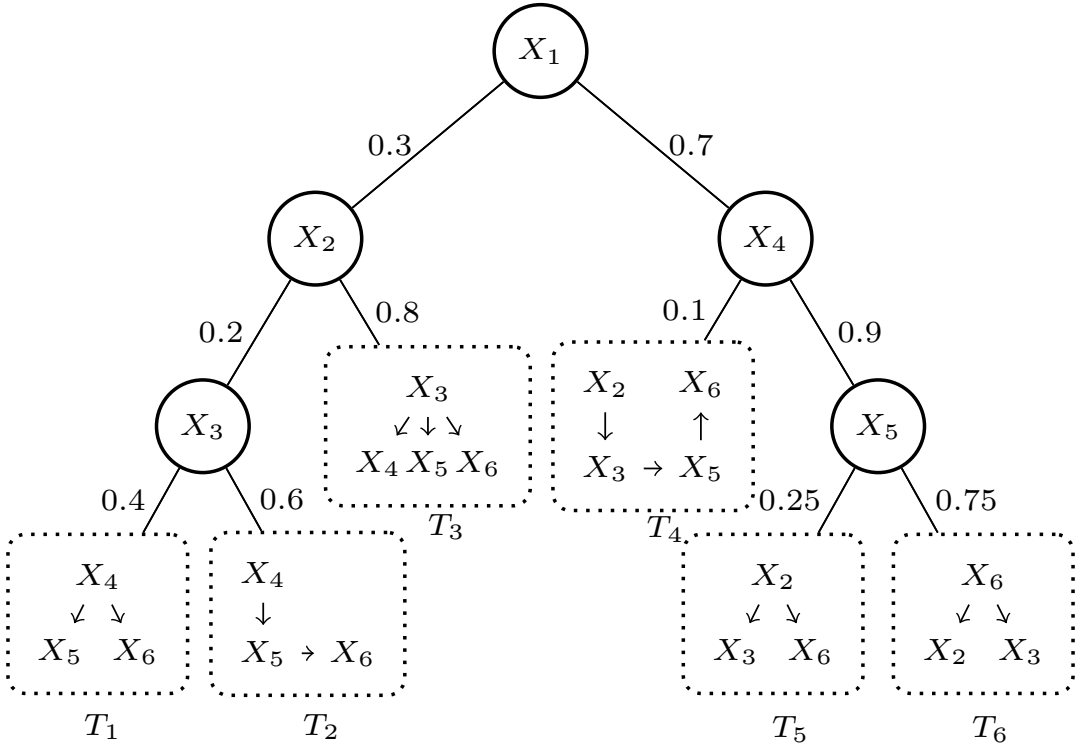
Probabilistic Graphical Models (**PGMs**) vs Latent Tractable Probabilistic Models (**LTPMs**):

Criteria	**Who wins?**	**Why?**
Test set log-likelihood	PGMs	Expressive model
Marginal (MAR) estimates	LTPMs	Reliable Exact Inference
Maximum-a-posteriori (MAP) estimates	No Clear Winner	Both use unreliable approximate inference approaches

Tractability helps achieve better marginal predictions even though the model fit is inferior. **Can we do the same for MAP inference?**

- **YES: If we use MAP-tractable cutset networks and improve their fit**.

Cutset Networks



- OR tree (probabilistic decision trees) with tree Bayesian networks at leaves
- MAP tractable and Interpretable unlike sum-product networks

Improving Accuracy of Cutset Networks

Issue: When learned just from data the test set LL score of cutset networks is much smaller than latent variable models

Our approach to improve accuracy: Compile cutset networks from a more accurate latent variable model.

1. First learn a latent variable model $\mathcal{M}$ that admits tractable posterior marginal inference
2. Compute the sufficient statistics used by the classic structure learning algorithm for cutset networks by performing marginal inference over $\mathcal{M}$

Experiments

Hypothesis: If we improve the fit (test set LL) of cutset networks, because of tractability, they will yield better MAP estimates as compared with latent variable models

Experiments verify our hypothesis

Dataset	Before Compilation	After Compilation
DNA	-78.33	**-74.25**
Movie	-41.38	**-38.14**
BBC	-160.27	**-158.26**

Table: Conditional Log-Likelihood of the MAP Estimates

More details at the poster...