On Resource-Efficient Bayesian Network Classifiers and Deep Neural Networks

Wolfgang Roth, Günther Schindler,
Holger Fröning, Franz Pernkopf

Signal Processing and Speech Communication Laboratory

January 15, 2020

Outline

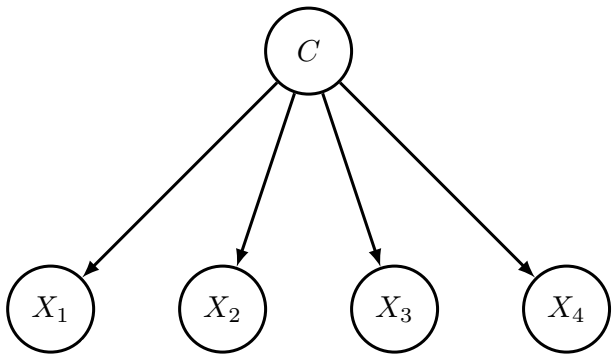
Model-Size-Aware TAN Structure Learning

Quantization-Aware Training for Bayesian Networks

Comparing Bayesian Networks and Neural Networks

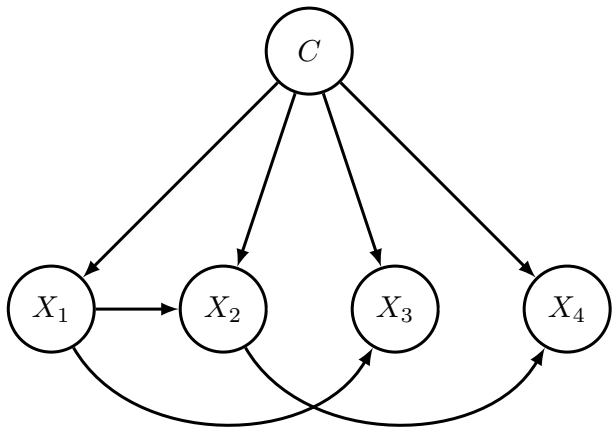
Naive Bayes vs. Tree-Augmented Naive Bayes (TAN)

Naive Bayes



- ✓ Few parameters
- ✓ Few operations
- ↯ Accuracy

TAN



- ✓ Accuracy
- ✓ Few operations
- ⚠ Structure learning: Model size depends on structure

Model-Size-Aware TAN Structure Learning

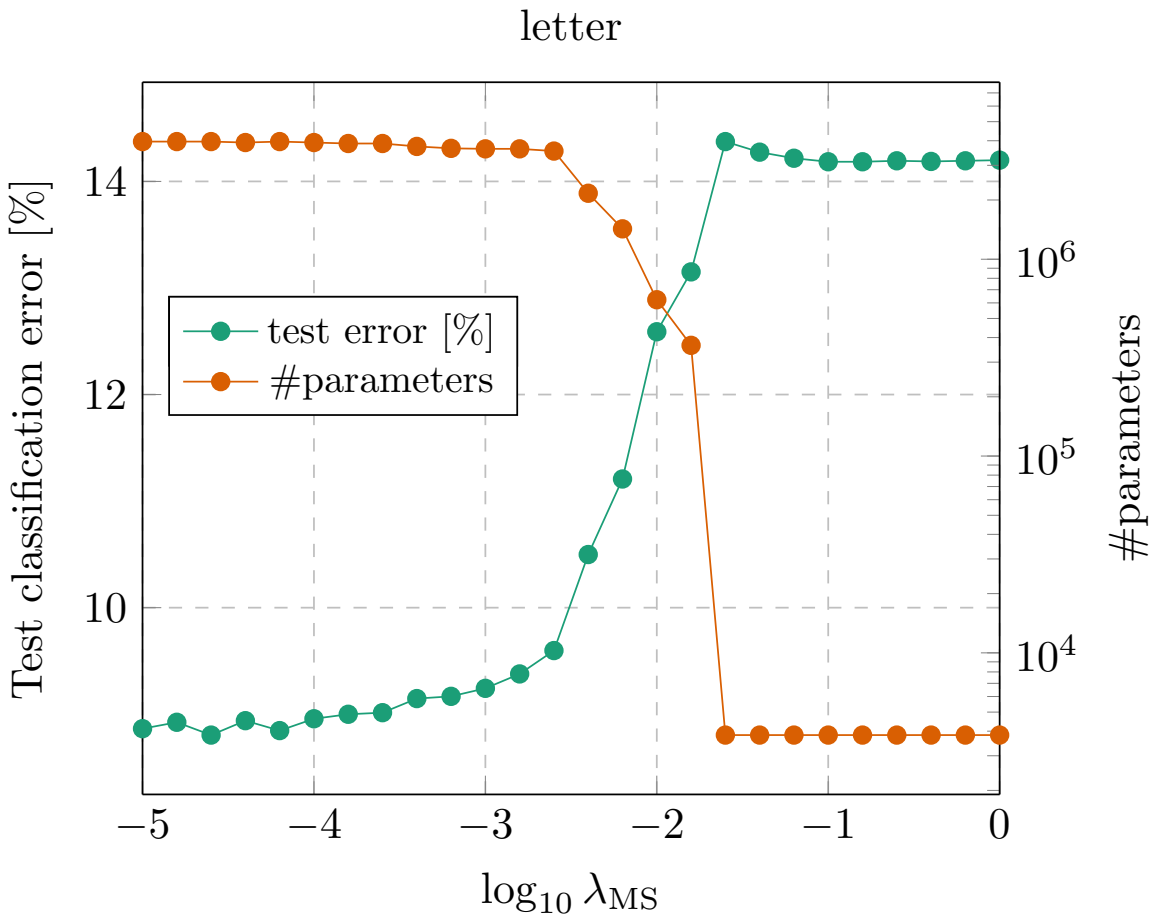
- We extend the differentiable TAN structure loss from [1]
  - ✓ Structure learning using backpropagation
- New term penalizes number of parameters (=model size)

$$\mathcal{L}_{\mathrm{SL}}^{\mathrm{MS}}(\Phi, \boldsymbol{\Theta}) = \underbrace{\mathcal{L}_{\mathrm{SL}}(\Phi, \boldsymbol{\Theta})}_{\text{proposed in [1]}} + \overbrace{\lambda_{\mathrm{MS}}}^{\text{trade-off}} \underbrace{\mathbb{E}_{\mathbf{S} \sim p_{\Phi}} \left[\mathcal{L}_{\mathrm{MS}}(\mathbf{S}) \right]}_{\#\text{model parameters}}$$

[1] Roth and Pernkopf, Differentiable TAN Structure Learning for Bayesian Network Classifiers, PGM 2020

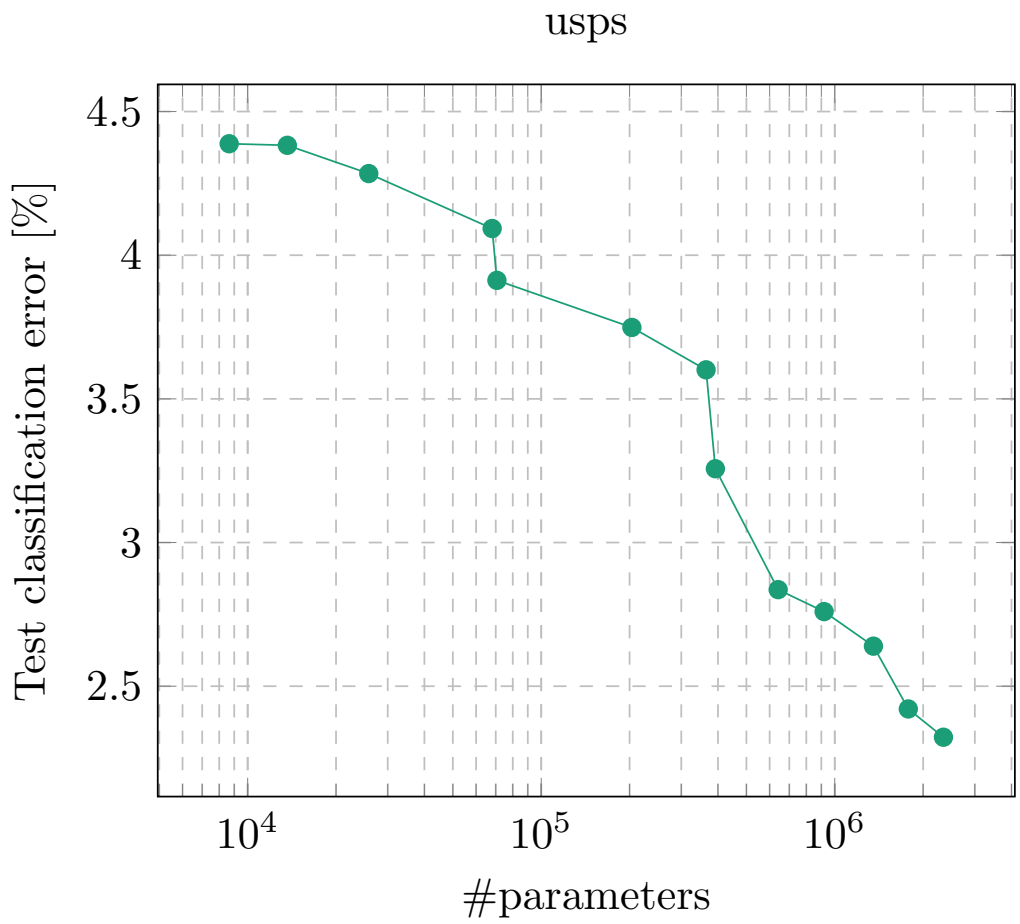
Model-Size-Aware TAN Structure Learning - Experiments

- Varying trade-off parameter λ_{MS}



Model-Size-Aware TAN Structure Learning - Experiments

- Pareto frontier

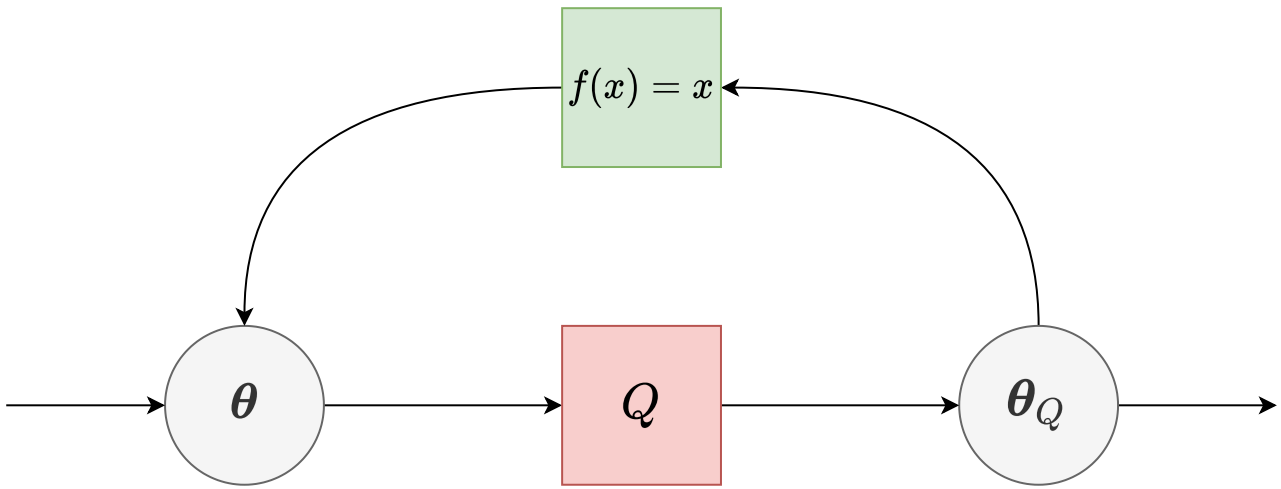


Quantization in Bayesian Network Classifiers

- Quantize log-probabilities θ to negative fixed-point values

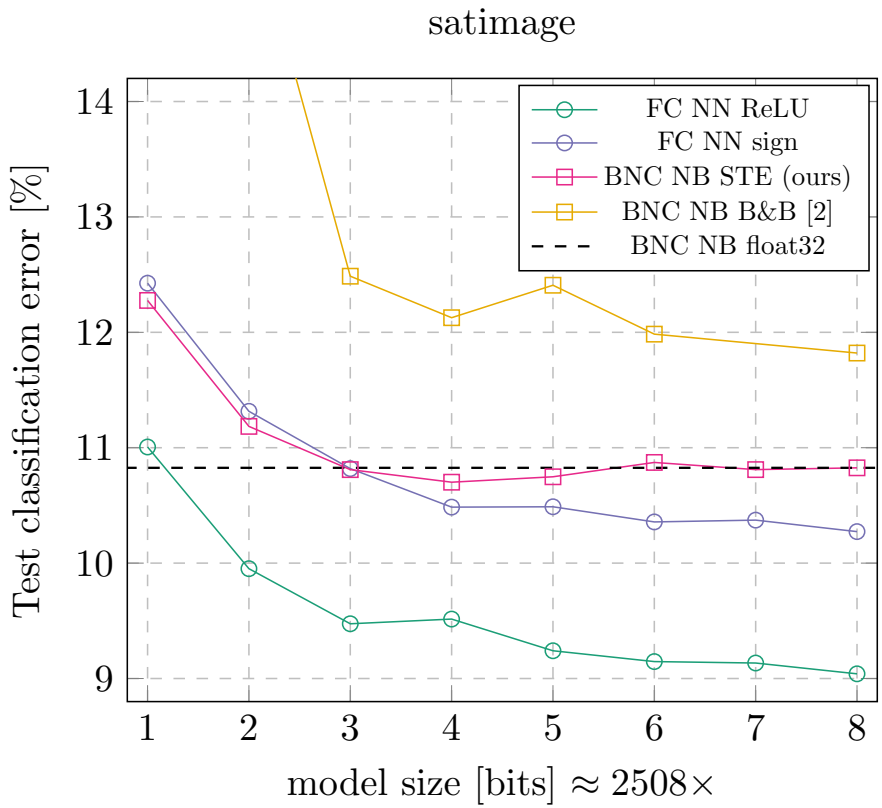
$$Q(\theta) = \text{clip}\left(\text{round}\left(\theta \cdot 2^{B_F}\right) \cdot 2^{-B_F}, -U, 0\right)$$

- Apply the straight-through gradient estimator



Quantization Experiments

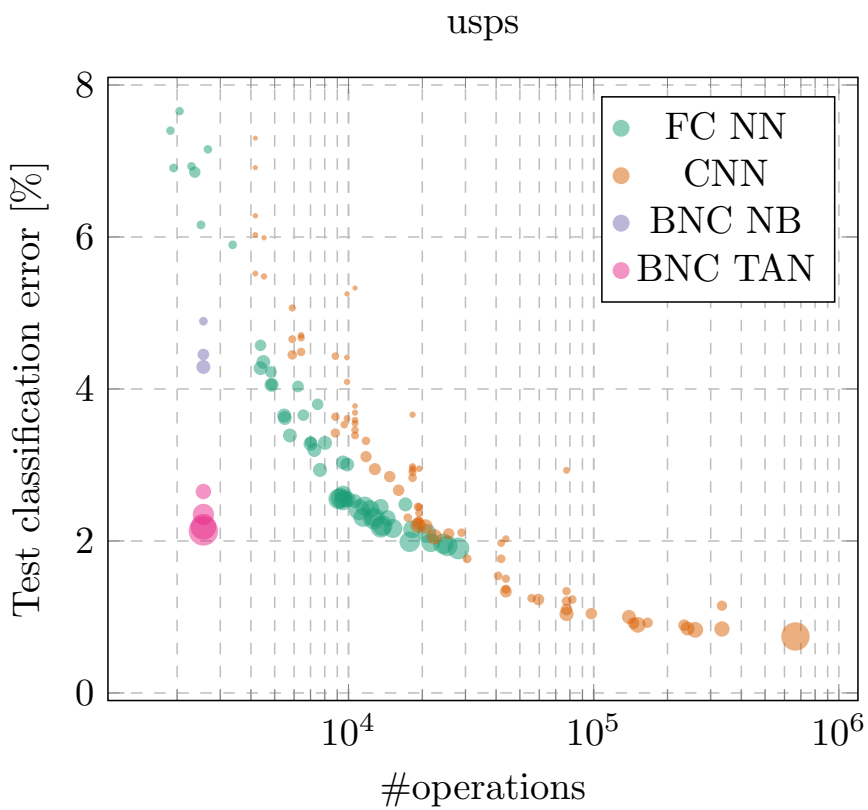
- Fully-connected NNs (FC NNs) vs. Bayesian Networks (BNCs)
- Matched model sizes on x-axis



[2] Tschiatschek et al., Integer Bayesian Network Classifiers, ECML 2014

Bayesian Networks vs. Neural Networks

- Pareto optimal models with respect to
  - classification error [%]
  - model size [#bits for model parameters]
  - #operations to compute predictions



Conclusion

- Transfer ideas from deep learning to Bayesian networks
- Differentiable model-size-aware TAN structure learning
  - Jointly train conditional probability tables and structure
  - Trade off between accuracy and model size
  - Easy to implement using automatic differentiation frameworks
- Quantization-aware training for Bayesian networks
  - Simple and effective quantization using straight-through estimator
- Bayesian networks can be a viable alternative to DNNs in the small-scale setting

Code available @ `https://github.com/wroth8/bnc/`