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)

\[
\mathcal{L}_{MS}^{SL}(\Phi, \Theta) = \mathcal{L}_{SL}(\Phi, \Theta) + \lambda_{MS} \mathbb{E}_{s \sim p_\Phi} [\mathcal{L}_{MS}(s)]
\]

[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 $\lambda_{MS}$

![Graph showing test classification error and number of parameters vs. $\log_{10} \lambda_{MS}$]
Model-Size-Aware TAN Structure Learning - Experiments

- Pareto frontier

![Graph showing Pareto frontier for usps dataset]

- Test classification error [%]
  - usps

- #parameters

Wolfgang Roth et al.
January 15, 2020
Quantization in Bayesian Network Classifiers

- Quantize log-probabilities $\theta$ to negative fixed-point values

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

- Apply the straight-through gradient estimator

![Diagram]

$$f(x) = x$$
Quantization Experiments

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

Test classification error [%]

satimage

FC NN ReLU
FC NN sign
BNC NB STE (ours)
BNC NB B&B [2]
BNC NB float32

Bayesian Networks vs. Neural Networks

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

![Graph showing test classification error vs. #operations for different models (usps dataset)]
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/