An Analysis of Bayesian Network Model-Approximation Techniques

Adamo Santana[1] and **Gregory Provan**[2]
Department of Computer Science, University College Cork, Cork, Ireland

Abstract. Two approaches have been used to perform approximate inference in Bayesian networks for which exact inference is infeasible: employing an approximation algorithm, or approximating the structure. In this article we compare two structure-approximation techniques, edge-deletion and approximate structure learning based on sub-sampling, in terms of relative accuracy and computational efficiency. Our empirical results indicate that edge-deletion techniques dominate the subsampling/induction strategy, in both accuracy and performance of generating the approximate network. We show, for several large Bayesian networks, how edge-deletion can create approximate networks with order-of-magnitude inference speedups and relatively little loss of accuracy.

1 Introduction

Bayesian networks (BNs) have become an important tool for modeling and probabilistic inference. As the size and complexity of BN models increase, so too do the demands of performing inference. In cases where exact inference is intractable, it is important to use approximation techniques to enable inference to take place. Such approximation may apply to the inference algorithms (e.g., stochastic sampling algorithms [2], or other approaches [3, 8]), or to the BN model $\mathcal{B}$ (e.g., edge-reduction [1, 10], probability-table/state-space approximation approaches [5, 7]). In this article, we focus on generating a space-bounded, approximate model, in cases where we have limitations on the space for embedding a BN model.

Our objective is the examine the tradeoff between space and performance of different approximations, i.e., given an approximate model $\mathcal{B}'$, what kinds of inference speedups do we obtain for what levels of inference accuracy, with respect to $\mathcal{B}$? This goal contrasts with the objectives of previous network-approximation (e.g,. edge-deletion) approaches [1, 10], where the primary interest was deleting edges while remaining within a certain error bound.

Our contributions are the following. First, we compare two BN-approximation approaches, one using BN threshold-based sub-sampling and network induction, and the other using threshold-based edge deletion. We show that the sampling/induction approach is limited by the accuracy of the induction algorithm, and produces networks which are inferior to the edge-deletion approach, due to the network-induction. We also show that, on a range of networks, the edge-deletion approach can produce several orders-of-magnitude speedups in inference with small penalties in inference accuracy.

[1] adamo@ufpa.br; Supported by CAPES CBE/PDEE 0005/2007.
[2] g.provan@cs.ucc.ie; Supported by SFI grant 04/IN3/I524.

2 Technical Preliminaries

A BN model $\mathcal{B}$ is defined as a tuple $(\mathcal{G}, \mathcal{P})$, where $\mathcal{G}$ is a directed acyclic graph (DAG), and $\mathcal{P}$ is a set of *probability distributions* constructed from vertices $\mathcal{V} = \{V_i\}$ in $\mathcal{G}$ such that $Pr\{\mathcal{V}\} = \prod_{i=1}^{n} Pr\{V_i|pa(V_i)\}$, where $pa(V_i)$ are the parents of V_i in $\mathcal{G}$. We compare two approaches, sub-sampling plus machine learning (SSML), and edge deletion (ED).

SSML Approach: In SSML, we generate from $\mathcal{B}$ a training dataset T composed of 10,000 random samples, using the GeNIe tool [4]. We then used a sampling threshold ϕ to prune from T all cases for which $Pr(\mathcal{B}) < \phi$, to create T_ϕ. For each value of ϕ examined, we induced an approximate network $\mathcal{B}_\phi$ from T_ϕ using the constrained based PC-algorithm [9].

ED Approach: In ED, we generate from $\mathcal{B}$ an approximate network $\mathcal{B}_\kappa$ by pruning from $\mathcal{B}$ all those edges whose Kulback-Leibler (KL) divergence is below a threshold κ. The KL divergence [6] was chosen as the metric for indicating the importance of the dependence related to each edge of the network since it is one of the most widely used methods for measuring the distance between distributions.

We adopt several metrics for the "quality" of an approximate network $\mathcal{B}'$ with respect to $\mathcal{B}$: the *error* ϵ on a test set is the difference in posterior probability averaged over the set V_t of target nodes; the *complexity reduction factor*, $\frac{CT(\mathcal{B}')}{CT(\mathcal{B})}$, is the relative network complexity, based on using the maximum clique table size of $\mathcal{B}'$, $CT(\mathcal{B}')$, as an inference complexity measure; and the *network reduction factor*, $\mathcal{S}(\mathcal{B}', \mathcal{B})$, is a measure of the degree of isomorphism between $\mathcal{B}'$ and $\mathcal{B}$.

3 Experimental Analysis

We empirically compared the SSML and ED approaches to BN approximation using 7 benchmark networks: C17, Alarm, Hailfinder, Pignet, Barley, Munin and C250 (a circuit with 250 nodes and 500 arcs). In our experiments, we created networks based on sub-sampling thresholds of $\phi = e^{-10},\ e^{-5},\ 5e^{-10}$, and KL thresholds $\kappa = 0.1,\ 0.15$ and 0.2. To test the error ϵ of each approximate network, we sampled to create a testing data set of 500 cases, such that we chose a set of "target" nodes whose posterior distributions we computed during testing. We computed several comparative measures, including the error rate ϵ for classification, the KL-divergence between

the distributions and the maximum clique table size CT of the networks. Figure 1 shows that, whereas for ED the CT values never increase with increasing threshold κ (meaning that the network gets no computationally harder to evaluate), with SSML the CT values *can increase* with ϕ. This anomalous performance is due to the induction process, in which we cannot guarantee that the network structure learned will monotonically decrease in size and CT values with ϕ; in contrast, with ED, this is guaranteed as edge pruning occurs. This failure to guarantee that approximate networks will be computationally simpler with SSML means that it may not be possible to use this approach unless structure-based constraints can be applied during the induction phase.

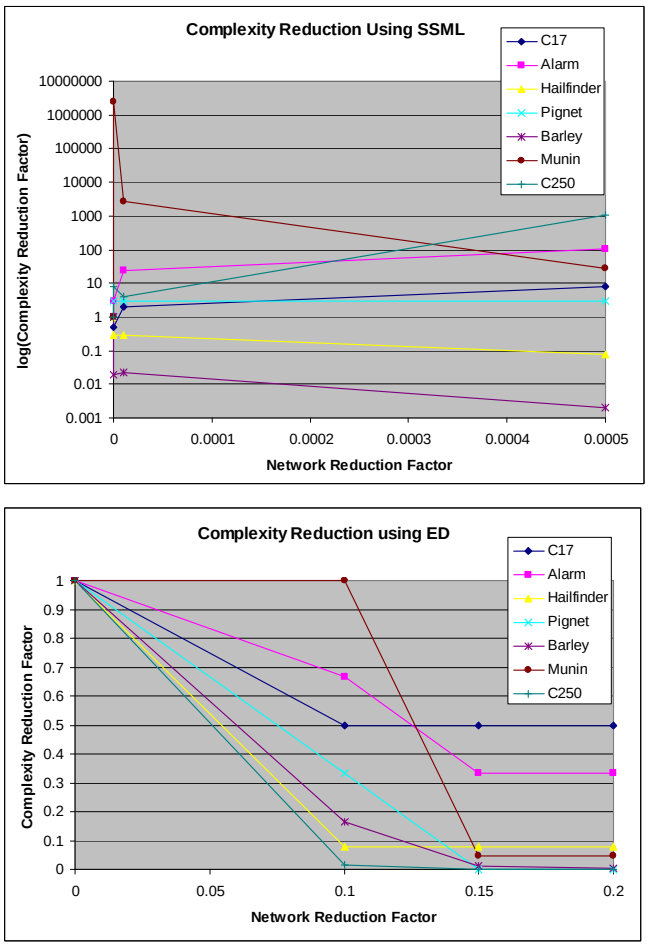

Figure 1. Comparison of SSML and ED for approximation of Network Complexity

The other major difference is the computational cost for the structure-approximation. The ED approach, since it uses the computing of divergences and pruning of edges with KL values, is proved very efficient. In contrast, the SSML has high computational cost, since it involves computing posteriors for the original network and inducing an approximate network for a training set; both expensive tasks for complex BNs.

Since exact inference was computationally infeasible for these larger networks, we used several sampling-based inference algorithms [2], all of which generate 10,000 samples to ensure close convergence of the results to the exact value.

Figure 2 displays the results of tradeoffs made over a range of KL threshold values using ED, showing that a significant reduction in relative inference complexity occurs, with little loss of accuracy. For example, our data indicate that for the C250 network, we have $O(10^6)$ faster inference with $> 90\%$ accuracy; for Munin, we have $O(10^5)$ faster inference with $\sim 80\%$ accuracy.

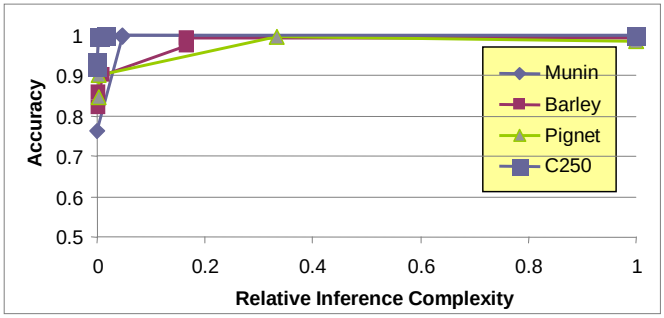

Figure 2. Tradeoff curves for four larger networks using ED

4 Conclusions

This paper compared two models for BN structure approximation, based on sub-sampling with network induction (SSML) and edge deletion (ED), to identify the types of tradeoff of inference-speedup and loss of accuracy possible with each approach. We showed that SSML cannot guarantee monotonically faster inference with increasing network approximations; this arises because the network structure induced from approximate data (sampled from the original network) has high variance. In contrast, with ED, the tradeoffs of accuracy for faster inference are guaranteed to be monotonic. We have showed, for several large BNs, how ED can create approximate networks with order-of-magnitude inference speedups with relatively little loss of accuracy.

REFERENCES

[1] A. Choi and A. Darwiche, 'An Edge Deletion Semantics for Belief Propagation and Its Practical Impact on Approximation Quality', *Proceedings AAAI*, **21**(2), 1107, (2006).

[2] B. D'Ambrosio, 'Inference in Bayesian networks', *AI Magazine*, **20**(2), 21–36, (1999).

[3] R. Dechter and I. Rish, 'Mini-buckets: A general scheme for bounded inference', *J. of the ACM*, **50**(2), 107–153, (2003).

[4] M.J. Druzdzel, 'GeNIe: A development environment for graphical decision-analytic models'', *Proc. AMIA*, (1999).

[5] C.U. Kjaerulff, 'Reduction of computation complexity in bayesian networks through removal of weak dependencies', *Proc. of 10th Conf. on UAI*, (1994).

[6] S. Kullback and R.A. Leibler, 'On Information and Sufficiency', *Annals of Math. Stat.*, **22**(1), 79–86, (1951).

[7] C.L. Liu and M.P. Wellman, 'Bounding probabilistic relationships in Bayesian networks using qualitative influences: methods and applications', *International Journal of Approximate Reasoning*, **36**(1), 31–73, (2004).

[8] F.T. Ramos and F.G. Cozman, 'Anytime anyspace probabilistic inference', *Int. J. of Approx. Reasoning*, **38**, 53–80, (2005).

[9] P. Spirtes, C.N. Glymour, and R. Scheines, *Causation, Prediction, and Search*, MIT Press, 2000.

[10] R.A. van Engelen, 'Approximating Bayesian belief networks by arc removal', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(8), 916–920, (1997).