Research Task A - The Paper:
A Differential Approach to Inference in Bayesian Networks

André E. dos Santos   dossantos@cs.uregina.ca
Department of Computer Science
University of Regina
Regina, Canada

Abstract

Bayesian networks (BNs) is a probabilistic graphical model for dealing with uncertainty in artificial intelligence. However, BNs are known for requiring a high amount of processing during its running time of execution. This strongly limits the applications of BNs to embedded systems, which are characterized for their primitive computational resources. Darwiche (2000) proposes a new approach for inference in Bayesian networks based on partial differentiation. Once the BN is processed, one can compute in constant-time answers to a large class of probabilistic queries. One of the key contributions is the possible cost-effective implementation on a variety of software and hardware platforms.

The paper has a mathematical heavy notation and it is not a easy reading. A reason is the weak background on probabilistic graphical models. Thus, it is not self contained, which means one might have to look for second sources. Throughout the paper, several advantages are presented All claims have proof and are sound. However, some points are not discussed in the paper, such the generalization of the proposed approach for other networks.

1. Introduction

We present a critique for the paper "A Differential Approach to Inference in Bayesian Networks", by Darwiche (2000). The paper is the first of many published articles to present the idea of compiled BNs(Darwiche, 2003; Chavira et al., 2006; Park and Darwiche, 2004; Chavira and Darwiche, 2007). It is also has been sum up as an important part of his book "Modeling and Reasoning with Bayesian Networks" (2009). The paper was published on the Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI). UAI is yet today the leading conference of the field and is annually hold in the USA. Some of its notorious sponsors are Google, Microsoft Research and Facebook.

Bayesian networks (BNs) (Pearl, 1988) is a probabilistic graphical model for dealing with uncertainty in artificial intelligence. BNs overcame the acquisition intractability problem when reasoning with the a join probability distribution. However, BNs are known for requiring a high amount of processing during its running time of execution (Koller and Friedman, 2009). This strongly limits the applications of BNs to embedded systems, which are characterized for their primitive computational resources. Due those limitations, consumer electronics often utilizes semantic weaker frameworks, such as Fuzzy logic or ruled-based systems.

Darwiche (2000) proposes a new approach for inference in Bayesian networks based on partial differentiation. Its first step is compiling the BN into a polynomial. Then, it computes the partial derivatives of this polynomial with respect to each variable on the BN. Once such derivatives are made available, one can compute in constant-time answers to a large class of probabilistic queries. The differential approach presents two key contributions. First, it enphasizes the role of partial differentiation on probabilistic reasoning, giving a new utility to central tasks of BNs. Second, it helps the migration of BN applications to embedded systems, which are known for constraint in computational resources.

This paper is organized as follows. In Section 2, Bayesian networks is reviewed. Section 3 presents ideas proposed in the paper. The critique is drawn in Section 4. Conclusions are given in Section 5.

2. Background

A *joint probability distribution* (JPD) is function P on the Cartesian product V of the variable domains such that the following two conditions hold:

$$0 \leq P(v) \leq 1.0 \text{ for each configuration } v \in V; \text{ and} \tag{1}$$

$$\sum_{v \in V} P(v) = 1.0 \tag{2}$$

Unlike *fuzzy logic* and *rule-based systems*, probability theory provides a rigorous mathematical foundation for the representation of uncertainty and for reasoning with uncertainty. However, acquisition intractability is one of the problems of using only the jpd to model the problem domain. For instance, specifying one probability value would be difficult with 50 variables, let alone 2^{50} probability values.

By exploiting the notion of *probabilistic conditional independence*, Pearl (1988) solved the jpd acquisition problem by introducing Bayesian networks. A *Bayesian network* (BN) Pearl (1988) is a directed acyclic graph (DAG) on finite set of discrete variables U together with *conditional probability tables* (CPTs) $P(v_1|Pa(v_1))$, $P(v_2|Pa(v_2))$, ..., $P(v_n|Pa(v_n))$, where $Pa(v_i)$ denotes the parents of the variable v_i in the DAG. The product of the CPTs for the DAG on U is a *joint probability distribution* $P(U)$ (Pearl, 1988). For example, Figure 1 shows a Bayesian network with CPTs $P(a)$ and $P(b|a)$ in Table 1.
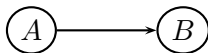

Figure 1: A BN from (Darwiche, 2000).

Unfortunately, both exact and approximate inference in Bayesian networks are, in general, *NP-hard* (Koller and Friedman, 2009). Thus, the use of exponential complexity algorithms is justified. That implies a limitation for BN applications to embedded systems, which are known for constraint in computational resources.

Table 1: Table (a) provides the prior probability of variable A and Table (b) provides the conditional probability of B given A.

(a)

A	$P(A)$
1	0.3
0	0.7

(b)

A	B	$P(B\|A)$
1	1	0.1
1	0	0.9
0	1	0.8
0	0	0.2

3. A Differential Approach to Inference in Bayesian Networks

In this section we show the mains ideas proposed by the paper "A Differential Approach to Inference in Bayesian Networks" (2000). We give an overview on how to represent a BN as a polynomial, and then compiling and evaluating the polynomial representation.

3.1 Polynomial Representation of Bayesian networks

We now show how to represent a BN as a polynomial $\mathcal{F}$ that includes two type os variables: evidence indicator λ and network parameter θ (probability values).

We can parameterize a CPT as follows. A *parameterize* table receives a evidence indicator λ multiplied to the probability value. For instance, Table 2 shows the parameterized CPTs for the BN of Table 1. Notice that $\bar{x}$ denotes the instantiation $x = 0$.

Table 2: Table (a) provides the parameterized CPT for $P(A)$ and Table (b) provides the parameterized CPT for $P(B|A)$.

(a)

A	$P(A)$
1	$0.3\lambda_A$
0	$0.7\lambda_{\bar{A}}$

(b)

A	B	$P(B\|A)$
1	1	$0.1\lambda_B$
1	0	$0.9\lambda_{\bar{B}}$
0	1	$0.8\lambda_B$
0	0	$0.2\lambda_{\bar{B}}$

The polynomial $\mathcal{F}$ of a BN is represented by the formula:

$$\mathcal{F}(\lambda, \theta) = \sum_U \prod \theta_i \prod \lambda_i \tag{3}$$

As an example, the polynomial of the BN in Figure 1 is

$$\mathcal{F} = \theta_A \theta_{B|A} \lambda_A \lambda_B + \theta_A \theta_{\bar{B}|A} \lambda_A \lambda_{\bar{B}} + \theta_{\bar{A}} \theta_{B|\bar{A}} \lambda_{\bar{A}} \lambda_B + \theta_{\bar{A}} \theta_{\bar{B}|\bar{A}} \lambda_{\bar{A}} \lambda_{\bar{B}} \tag{4}$$

The size of the polynomial is exponential to the size of the network.

Now, with a evidence we can set a indicator λ to either 0 or 1. If is consistent with the evidence is 1, and is 0 otherwise. For example, if the evidence is $A\bar{B}$ that is, A is true,

3

B is false, then $\mathcal{F}(\lambda_A = 1, \lambda_{\bar{A}} = 0, \lambda_B = 0, \lambda_{\bar{B}} = 1, \theta_A = 0.3, \theta_{\bar{A}} = 0.7, \theta_{B|A} = 0.1, \theta_{\bar{B}|A} = 0.9, \theta_{B|\bar{A}} = 0.8, \theta_{\bar{B}|\bar{A}} = 0.2,)$, which equal to 0.27 in this case. The probability values are the correspondents in the respectively CPTs. Therefore, the polynomial is a linear function of many variables, where each one corresponds to either an evidence indicator λ or a probability value θ. Each parameter has degree one. For each instantiation given an evidence, the polynomial can be evaluated to compute the probability of the variable given the evidence.

The author highlights that BNs represented by polynomial function has been already exploited by other authors (Castillo et al., 1996, 1997; Russell et al., 1995). However, the evidence indicators were fixed to a particular value. Hence, they could not be used to answer queries with respect to varying evidence.

3.2 Compiling a Bayesian networks

The paper presents a algorithm, called VE_COMPLETE, to compile a polynomial representation of a BN. VE_COMPLETE receives as input a BN and an elimination ordering of the variables. The result is a more compact polynomial following the elimination ordering used to construct this new presentation. Thereby, a rooted DAG can be drawn for the compiled polynomial. For instance, the compiled polynomial in (4) following the elimination ordering $\{B, A\}$ is

$$\mathcal{F} = \theta_A \lambda_A (\theta_{B|A} \lambda_B + \theta_{\bar{B}|A} \lambda_{\bar{B}}) + \theta_{\bar{A}} \lambda_{\bar{A}} (\theta_{B|\bar{A}} \lambda_B + \theta_{\bar{B}|\bar{A}} \lambda_{\bar{B}}) \tag{5}$$

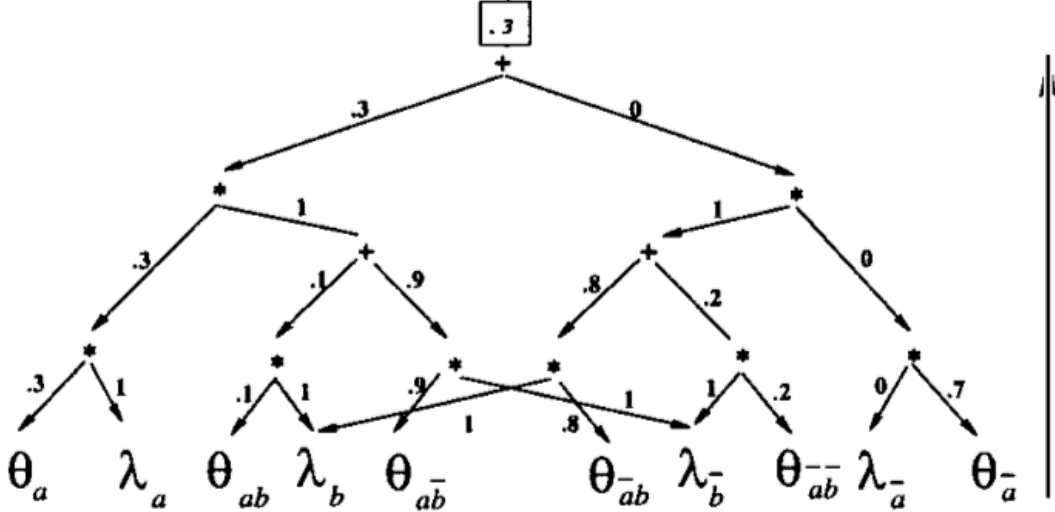and depicted in Figure 2.

Figure 2: Rooted DAG representing Equation (5).

3.3 Evaluating and Differentiating a Polynomial Representation

The evaluation of the complied polynomial and computing the partial derivatives is a two-phase message passing scheme in which each message is simply a number. The first phase, massages are sent from nodes to their parents in the rooted DAG following the operations of each node. Figure 2 shows this process, where it leads to assigning the value 0.3 to the root, indicating that the probability of evidence E is $P(E) = 0.3$. Second phase, messages are sent from nodes to children in the same rooted DAG, leading to computation of all partial derivatives. This processes is illustrated in Figure 3.
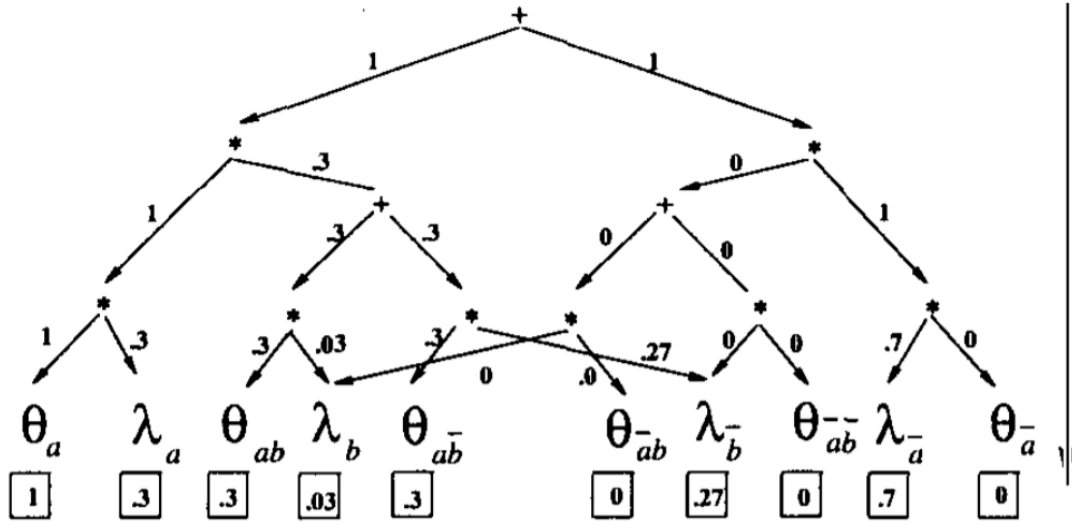

Figure 3: Second phase of the evaluation of rooted DAG in Figure 5, under evidence $E = A$.

The number of messages passed in both phases is twice the number of edges in the compilation of the polynomial. Since each message can be computed in constant time, the whole computation is linear in the size of the polynomial.

After the two-phase steps we have the capacity to answer, in constant time, a very large class of probabilistic queries, relating to classical inference, parameter estimation, model validation and sensitivity analysis. Some of those queries are (i) in the root node the probability of the evidence, and (ii) the posterior marginal of any network variable.

4. Criticism

In this section we present a critique for the paper "A Differential Approach to Inference in Bayesian Networks" (Darwiche, 2000).

The paper is well written and presents a good flow of ideas. Even so, the paper has a mathematical heavy notation and it is not a easy reading. The background on probabilistic graphical models is not provided. For instance, Section 2 of this paper was drawn from my

personal knowledge of the field. Thus, the paper is not self contained, which can lead some readers to search for second sources.

The paper presents several advantages. Proofs are provided and state the soundness of the mathematical claims. However, some points are not discussed For instance, the generalization of the proposed approach for other networks and how to deal with intractable networks are not discussed. Even thought the author being known for several winnings with his application on the Probabilistic Inference Competition of the proceedings, the paper also does not present experimental results to endure its claims. There is a typo on the Section 1, first result of the paper: "Given a variable elimination order of *with w* [...]", the correct word is *width*. On the polynomial representation example, the the third network parameter $\theta_{B|A}$ should be $\theta_{B|\bar{A}}$.

The main advantage of the paper is to simplify inference systems to the point where they could be implemented cost-effectively on a variety of software and hardware platforms. That would allow emmbeded systems that normally uses Fuzzy logic or rule-based system to starts using BNs, a more powerful semantic tool for dealing with uncertanty. And with the advantage not requiring a lot of processing during running time of execution This is a direct drawn from the three main results of the paper: (i) compiling the polynomial in linear in the size of the polynomial, (ii) answering a large number of queries by (i), and (iii) computing the partial derivatives of λ and θ simultaneously in time linear in the size of the polynomial.

There is no doubt the paper presents a breakthrough on BNs field. The author even states that the he is not aware of any computational framework for BNs, which combines the simplicity, comprehensiveness and computational complexity of the presents framework based on partial derivatives. However, after network polynomials (Darwiche, 2000), Poon and Domingos (2011) generalized network polynomials to unnormalized distributions and has been the "state of the art" on probabilistic graphical models.

5. Conclusion

We have presented an overview and critique of the paper "A Differential Approach to Inference in Bayesian Networks" by Darwiche (2000). The main advantage of the paper is to simplify inference systems to the point where they could be implemented cost-effectively on a variety of software and hardware platforms. The differential approach presents two key contributions. First, it enphasizes the role of partial differentiation on probabilistic reasoning, giving a new utility to central tasks of BNs. Second, it helps the migration of BN applications to embedded systems, which are known for constraint in computational resources.

Darwiche (2000) proposes a new approach for inference in Bayesian networks based on partial differentiation. Its first step is compiling the BN into a polynomial. Then, it computes the partial derivatives of this polynomial with respect to each variables. The partial derivatives are important when estimating BN parameters and when performing sensitivity analysis. Moreover, with the derivatives it is possible to compute in constant-time answers to a large class of probabilistic queries.

The paper presents a breakthrough on BNs field, which let to new (more comprehensive) networks (Poon and Domingos, 2011). However, the differential approach to inference

in BNs was the first computational framework to combine partial derivatives for such a purpose. That has been leading to a great advancements on the field and made more viable BN applications to embedded systems.

References

Enrique Castillo, José Manuel Gutiérrez, and Ali S Hadi. Goal oriented symbolic propagation in bayesian networks. In *PROCEEDINGS OF THE NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE*, pages 1263–1268. Citeseer, 1996.

Enrique Castillo, José Manuel Gutiérrez, and Ali S Hadi. Sensitivity analysis in discrete bayesian networks. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 27(4):412–423, 1997.

Mark Chavira and Adnan Darwiche. Compiling bayesian networks using variable elimination. In *IJCAI*, pages 2443–2449, 2007.

Mark Chavira, Adnan Darwiche, and Manfred Jaeger. Compiling relational bayesian networks for exact inference. *International Journal of Approximate Reasoning*, 42(1):4–20, 2006.

Adnan Darwiche. A differential approach to inference in bayesian networks. *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence (UAI2000)*, 2000.

Adnan Darwiche. A differential approach to inference in bayesian networks. *Journal of the ACM (JACM)*, 50(3):280–305, 2003.

Adnan Darwiche. *Modeling and reasoning with Bayesian networks*. Cambridge University Press, 2009.

D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.

James D Park and Adnan Darwiche. A differential semantics for jointree algorithms. *Artificial Intelligence*, 156(2):197–216, 2004.

J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

Hoifung Poon and Pedro Domingos. Sum-product networks: A new deep architecture. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 689–690. IEEE, 2011.

Stuart Russell, John Binder, Daphne Koller, and Keiji Kanazawa. Local learning in probabilistic networks with hidden variables. In *IJCAI*, volume 95, pages 1146–1152. Citeseer, 1995.