

Bayesian Network Inference With NIN-AND Tree Models

Yang Xiang, University of Guelph, Canada

Abstract

Non-impeding noisy-AND (NIN-AND) tree models were developed to improve efficiency and expressiveness in acquisition of conditional probability tables (CPTs) when constructing Bayesian networks (BNs). To take advantage of these models in the BN inference, we propose a multiplicative factorization of these models and a compilation of NIN-AND tree modeled BNs for lazy propagation (LP). Soundness of the method and its efficiency improvement are shown.

1 Introduction

BNs allow uncertain knowledge to be expressed with a linear number of CPTs. Each CPT, often about an effect conditioned on its n causes, is exponential in size in terms of n . Noisy-OR (Pearl, 1988) and several causal independence models (CIMs), e.g., (Heckerman and Breese, 1996; Galan and Diez, 2000; Lemmer and Gossink, 2004), reduce the number of parameters to linear, but are limited to reinforcing interactions (Xiang and Jia, 2007). This is overcome in NIN-AND tree models that allow undermining and recursive mixture of both.

Besides making acquisition easier, CIMs are explored for the inference efficiency by different approaches. These include parent divorcing (Olesen et al., 1989), temporal belief nets (Heckerman, 1993), heterogeneous factorization (Zhang and Poole, 1996), multiplicative factorization (MF) (Takikawa and D’Ambrosio, 1999), tensor rank-one decomposition (Savicky and Vomlel, 2007), among others. Most explore reinforcing CIM noisy-OR and noisy-MAX.

This work takes the approach by (Madsen and D’Ambrosio, 2000), who apply a MF to noisy-MAX for LP in junction trees (JTs). We propose a MF of binary NIN-AND tree models and a compilation of NIN-AND tree modeled BNs into JTs for LP. This approach has the following properties: (a) It is based on NIN-AND tree CIMs, and expresses mixture of reinforcing and undermining interactions. (b) It is based on MFs, and does not impose constraints to

variable elimination ordering (as, e.g., heterogeneous factorization does (Zhang and Poole, 1996)). (c) It is based on LP in JTs, uses the smallest runtime factors (unlike standard belief propagation in JTs, e.g., (Jensen et al., 1990), where each runtime factor is a product of several CPTs), and provides exact posterior marginals for all variables. (d) Moralization, as commonly performed, is not needed, which is similar to rank-one decomposition (Savicky and Vomlel, 2007).

Sec. 2 reviews NIN-AND tree models. MFs of single NIN-AND gate models and their soundness are presented in Secs. 3 and 4. The MF of multi-gate NIN-AND tree models is covered in Sec. 5, with soundness proven in Sec. 6. How to compile factorized models for LP is described in Sec. 7. Sec. 8 demonstrates the efficiency gain.

2 NIN-AND Tree Causal Models

This section introduces briefly binary NIN-AND tree models, which this work focuses on. More details on these models can be found in (Xiang and Jia, 2007). An uncertain cause can produce an effect but does not always do so. Denote the effect by e with domain $D_e = \{e^0, e^1\}$, where e^1 denotes $e = true$, and the set of all causes (including a leaky variable if any) of e by $C = \{c_1, \dots, c_n\}$, where each cause has domain $D_{c_i} = \{c_i^0, c_i^1\}$. At times, we write these domains exchangeably as $D_e = D_{c_i} = \{0, 1\}$.

A *single-causal success* is an event where c_i causes e to occur when other causes are false.

Denote the event by $e^1 \leftarrow c_i^1$ and its probability by $P(e^1 \leftarrow c_i^1)$. Smoking causing lung cancer is denoted $lc^1 \leftarrow smk^1$. A *single-causal failure*, where e is false when c_i is true and other causes are false, is denoted by $e^0 \leftarrow c_i^1$. A *multi-causal success* is an event where a set $X = \{c_1, \dots, c_k\}$ of causes ($k \leq n$) cause e , and is denoted by $e^1 \leftarrow c_1^1, \dots, c_k^1$ or $e^1 \leftarrow \underline{x}^1$.

A CPT $P(e|C)$ relates to causal probabilities as follows: If $C = \{c_1, c_2, c_3\}$, then $P(e^1 | c_1^1, c_2^0, c_3^1) = P(e^1 \leftarrow c_1^1, c_3^1)$.

Causal probabilities satisfy the following:

$$P(e^1 \leftarrow \emptyset) = 0 \quad (1)$$

$$P(e^0 \leftarrow \underline{x}^1) = 1 - P(e^1 \leftarrow \underline{x}^1) \quad (2)$$

Eqn. (1) denotes that the effect never occurs if all causes are false. Eqn. (2) relates causal success to failure.

Causes reinforce each other if collectively they are at least as effective as when some are active. Radiotherapy and chemotherapy are reinforcing causes for curing cancer. If collectively causes are less effective, they undermine each other. Taking either one of two desirable jobs leads to happiness for Dave. When taking both, chance of his happiness is reduced due to overstress. For $C = \{c_1, c_2\}$, if c_1, c_2 undermine each other, we have $P(e^1 | c_1^1, c_2^0) > 0$, $P(e^1 | c_1^0, c_2^1) > 0$, and

$$P(e^1 | c_1^1, c_2^1) < \min(P(e^1 | c_1^1, c_2^0), P(e^1 | c_1^0, c_2^1)).$$

Def. 1 defines both causal interactions.

Definition 1 Let $R = \{W_1, W_2, \dots\}$ be a partition of a set X of causes, $R' \subset R$ be any proper subset of R , and $Y = \cup_{W_i \in R'} W_i$. Sets of causes in R **reinforce** each other, iff

$$\forall R' P(e^1 \leftarrow y^1) \leq P(e^1 \leftarrow \underline{x}^1).$$

Sets of causes in R **undermine** each other, iff $\forall R' P(e^1 \leftarrow y^1) > P(e^1 \leftarrow \underline{x}^1)$.

Reinforcement and undermining occur between individual causes as well as sets of them. When it is between individuals, each W_i is a singleton. Otherwise, W_i can be a generic set. Consider $X = \{c_1, c_2, c_3, c_4\}$, $W_1 = \{c_1, c_2\}$, $W_2 = \{c_3, c_4\}$, $R = \{W_1, W_2\}$, where c_1, c_2 reinforce each other, and so do c_3 and c_4 . But W_1 and W_2 can undermine each other.

Disjoint sets of causes W_1, \dots, W_m satisfy *failure conjunction* iff

$$(e^0 \leftarrow \underline{w}_1^1, \dots, \underline{w}_m^1) = (e^0 \leftarrow \underline{w}_1^1) \wedge \dots \wedge (e^0 \leftarrow \underline{w}_m^1).$$

That is, when causes collectively fail to produce the effect, each must have failed to do so. They also satisfy *failure independence* iff

$$\begin{aligned} & P((e^0 \leftarrow \underline{w}_1^1) \wedge \dots \wedge (e^0 \leftarrow \underline{w}_m^1)) \\ &= P(e^0 \leftarrow \underline{w}_1^1) \times \dots \times P(e^0 \leftarrow \underline{w}_m^1). \end{aligned} \quad (3)$$

Disjoint sets of causes W_1, \dots, W_m satisfy *success conjunction* iff

$$(e^1 \leftarrow \underline{w}_1^1, \dots, \underline{w}_m^1) = (e^1 \leftarrow \underline{w}_1^1) \wedge \dots \wedge (e^1 \leftarrow \underline{w}_m^1).$$

That is, collective success requires individual effectiveness. They also satisfy *success independence* iff

$$\begin{aligned} & P((e^1 \leftarrow \underline{w}_1^1) \wedge \dots \wedge (e^1 \leftarrow \underline{w}_m^1)) \\ &= P(e^1 \leftarrow \underline{w}_1^1) \times \dots \times P(e^1 \leftarrow \underline{w}_m^1). \end{aligned} \quad (4)$$

Causes are undermining when they satisfy success conjunction and independence. Hence, undermining can be modeled by a direct NIN-AND gate (Fig. 1 (a)). Its root nodes (top) are

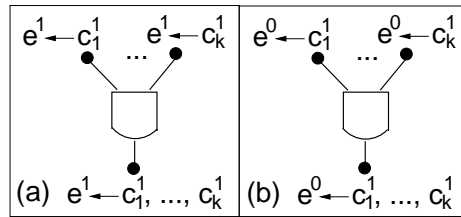


Figure 1: (a) A direct NIN-AND gate. (b) A dual NIN-AND gate.

single-causal successes, and its leaf (bottom) is the multi-causal success in question. Success conjunction is expressed by NIN-AND gate, and success independence by disconnection of roots other than through the gate. Probability of leaf event is computed by Eqn. (4). Similarly, causes are reinforcing when they satisfy failure conjunction and independence. Hence, reinforcement can be modeled by a dual gate (Fig. 1 (b)). Leaf event probability is obtained by Eqn. (3).

By organizing multiple direct and dual NIN-AND gates in a tree, mixture of reinforcement and undermining at multiple levels can be expressed in an NIN-AND tree model. Consider $C = \{c_1, c_2, c_3\}$, where c_1 and c_3 undermine each other, but collectively they reinforce c_2 . Assuming event conjunction and independence, their interaction relative to event $e^1 \leftarrow c_1^1, c_2^1, c_3^1$ can be expressed by NIN-AND tree in Fig. 2. Top gate is direct and bottom gate (leaf gate) is

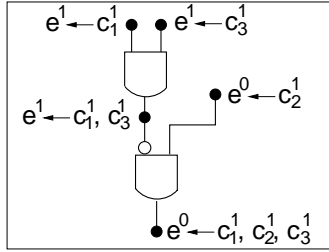


Figure 2: An NIN-AND tree

dual. Link downward from node $e^1 \leftarrow c_1^1, c_3^1$ has a white oval end (a negation link) and negates the event into $e^0 \leftarrow c_1^1, c_3^1$. All other links are forward links. Probability of leaf event is computed by Eqns. (3) and (4). For instance, from single-causal probabilities for root events, $P(e^1 \leftarrow c_1^1) = 0.85$, $P(e^1 \leftarrow c_2^1) = 0.8$, $P(e^1 \leftarrow c_3^1) = 0.7$, derive $P(e^0 \leftarrow c_1^1, c_2^1, c_3^1)$ as follows:

$$\begin{aligned} P(e^1 \leftarrow c_1^1, c_3^1) &= 0.85 \times 0.7 = 0.595 \\ P(e^0 \leftarrow c_1^1, c_2^1, c_3^1) &= P(e^0 \leftarrow c_1^1, c_3^1)P(e^0 \leftarrow c_2^1) \\ &= 0.405 \times 0.2 = 0.081 \end{aligned}$$

Hence, $P(e^1 \leftarrow c_1^1, c_2^1, c_3^1) = 0.919$ by Eqn. (2).

An NIN-AND tree is *minimal* if consecutive gates are opposite in type, e.g., Fig. 2. Every NIN-AND tree model is equivalent to a unique minimal NIN-AND tree model (Xiang et al., 2009). In this work, we assume minimal models.

3 MF of Direct Gate Local Models

Consider the BN family in Fig. 3 (a), where causal interaction is modeled by direct NIN-AND gate in Fig. 1 (a) with $k = n$ and with single-causal probabilities $P(e^1 \leftarrow c_i^1)$ ($i = 1, \dots, n$) specified.

A MF of a direct NIN-AND gate model consists of a Markov network (MN) (Pearl, 1988)

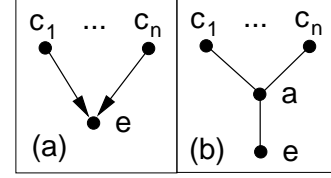


Figure 3: (a) A BN family. (b) A MN segment.

segment and a set of generalized potentials. The MN segment contains the BN family variables plus auxiliary variable a of domain $\{a^0, a^1, a^2\}$ or simply $\{0, 1, 2\}$. They are connected into an *undirected* graph as Fig. 3 (b).

For each link $\langle x, y \rangle$ in the MN segment, a *generalized potential* made of reals, possibly negative, is assigned. Table 1 shows potentials $f(x, y)$ for links $\langle c_i, a \rangle$ and $\langle a, e \rangle$. Let

Table 1: $f(a, c_i)$ (left) and $f(e, a)$ (right) of a direct NIN-AND gate model

a	c_i	$f(a, c_i)$	e	a	$f(e, a)$
0	0	1	0	0	-1
0	1	$P(e^1 \leftarrow c_i^1)$	0	1	1
1	0	1	0	2	1
1	1	1	1	0	1
2	0	1	1	1	0
2	1	0	1	2	-1

$g(e, c_1, \dots, c_n)$ denote product of link potentials with variable a marginalized out,

$$g(e, c_1, \dots, c_n) = \sum_a f(e, a) \prod_{i=1}^n f(a, c_i).$$

We refer to $g(e, c_1, \dots, c_n)$ as *marginalized product*. Theorem 1 shows $g(e, c_1, \dots, c_n)$ to be the CPT defined by the direct NIN-AND gate model.

Theorem 1 *Given a MF of a direct NIN-AND gate model over e and $C = \{c_1, \dots, c_n\}$, then the marginalized product $g(e, c_1, \dots, c_n)$ is the CPT of the direct gate model, i.e., $g(e, c_1, \dots, c_n) = P(e|c_1, \dots, c_n)$.*

Proof: Start from Fig. 4, and then consider Eqn. (5) for $e = e^1$,

$$\prod_{i=1}^n f(a^0, c_i) - \prod_{i=1}^n f(a^2, c_i).$$

$$\begin{aligned}
g(e, c_1, \dots, c_n) &= f(e, a^0) \prod_{i=1}^n f(a^0, c_i) + f(e, a^1) \prod_{i=1}^n f(a^1, c_i) + f(e, a^2) \prod_{i=1}^n f(a^2, c_i) \\
&= \begin{cases} f(e^0, a^0) \prod_{i=1}^n f(a^0, c_i) + f(e^0, a^1) \prod_{i=1}^n f(a^1, c_i) + f(e^0, a^2) \prod_{i=1}^n f(a^2, c_i) & \text{if } e = e^0 \\ f(e^1, a^0) \prod_{i=1}^n f(a^0, c_i) + f(e^1, a^1) \prod_{i=1}^n f(a^1, c_i) + f(e^1, a^2) \prod_{i=1}^n f(a^2, c_i) & \text{if } e = e^1 \end{cases} \\
&= \begin{cases} -\prod_{i=1}^n f(a^0, c_i) + \prod_{i=1}^n f(a^1, c_i) + \prod_{i=1}^n f(a^2, c_i) & \text{if } e = e^0 \\ \prod_{i=1}^n f(a^0, c_i) - \prod_{i=1}^n f(a^2, c_i) & \text{if } e = e^1 \end{cases} \\
&= \begin{cases} 1 - (\prod_{i=1}^n f(a^0, c_i) - \prod_{i=1}^n f(a^2, c_i)) & \text{if } e = e^0 \\ \prod_{i=1}^n f(a^0, c_i) - \prod_{i=1}^n f(a^2, c_i) & \text{if } e = e^1 \end{cases} \tag{5}
\end{aligned}$$

Figure 4: Proof of Prop. 1 (see Table 1 for definition of factors)

If for one or more i , $c_i = c_i^1$, it becomes $\prod_{c_i=c_i^1} P(e^1 \leftarrow c_i^1)$, and satisfies $P(e^1 \leftarrow c_1, \dots, c_n)$ from Eqn. (4). If for every i , $c_i = c_i^0$, it becomes $1 - 1 = 0$, and satisfies Eqn. (1). Therefore, $g(e^1, c_1, \dots, c_n)$ corresponds to $P(e^1|c_1, \dots, c_n)$ as defined by direct gate model.

From Eqn. (5) for $e = e^0$, it follows that $g(e, c_1, \dots, c_n)$ defines the CPT $P(e|c_1, \dots, c_n)$ (cf. Eqn. (2)) and the theorem follows. \square

4 MF of Dual Gate Local Models

Next, consider the BN family in Fig. 3 (a), modeled by the dual NIN-AND gate in Fig. 1 (b) with $k = n$ and with single-causal probabilities specified.

A MF of a dual NIN-AND gate model consists of a MN segment (Fig. 3 (b)), and a set of generalized potentials, one per link. Table 2 shows potentials for links $\langle c_i, a \rangle$ and $\langle a, e \rangle$.

Table 2: $f(a, c_i)$ (left) and $f(e, a)$ (right) of a dual NIN-AND gate model

a	c_i	$f(a, c_i)$	e	a	$f(e, a)$
0	0	1	0	0	1
0	1	$P(e^0 \leftarrow c_i^1)$	0	1	0
1	0	1	0	2	0
1	1	1	1	0	-1
2	0	1	1	1	1
2	1	0	1	2	0

Theorem 2 shows that the marginalized product of link potentials is the CPT defined by the dual NIN-AND gate model. Its proof is omitted due to space.

Theorem 2 *Given a MF of a dual NIN-AND gate model over e and $C = \{c_1, \dots, c_n\}$, the marginalized product is the CPT of dual gate model, i.e., $g(e, c_1, \dots, c_n) = P(e|c_1, \dots, c_n)$.*

5 MF of Tree Local Models

Secs. 3 and 4 deal with MFs of single-gate NIN-AND tree models. We now consider general tree models. Let a BN family over e and $C = \{c_1, \dots, c_n\}$ be modeled by an NIN-AND tree T and single-causal probabilities.

A MF of an NIN-AND tree model consists of a MN segment and a set of generalized potentials, one per link. The MN segment G is obtained from T as follows:

1. For each root in T , labeled $e \leftarrow c_i^1$, relabel it by the cause variable c_i .
2. For each gate, relabel its output node with an auxiliary variable, $a \in \{a^0, a^1, a^2\}$, connect its input nodes to a , and delete the gate. If the gate deleted is direct or dual, refer to variable a as direct or dual.
3. For leaf gate and corresponding auxiliary variable a , create a new node, label it by variable e , and connect a to e .

Consider the BN family in Fig. 5 (a), modeled by NIN-AND tree in Fig. 2. The MN segment is shown in Fig. 5 (b), where a, b are auxiliary.

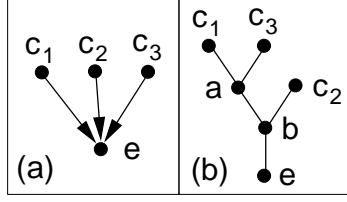


Figure 5: (a) A BN family. (b) A MN segment.

Link potentials are assigned as follows, where a, b are auxiliary variables.

1. For each link $\langle c_i, a \rangle$, where a is direct, assign the potential as Table 1 (left).
2. For each link $\langle c_i, a \rangle$, where a is dual, assign the potential as Table 2 (left).
3. For each link $\langle a, b \rangle$, where b is closer to e than a , assign the potential as Table 3.

Table 3: $f(a, b)$ for an NIN-AND tree model

b	a	$f(a, b)$	b	a	$f()$	b	a	$f()$
0	0	-1	1	0	0	2	0	0
0	1	1	1	1	1	2	1	0
0	2	1	1	2	0	2	2	1

4. For each link $\langle a, e \rangle$, where a is direct, assign the potential as Table 1 (right).
5. For each link $\langle a, e \rangle$, where a is dual, assign the potential as Table 2 (right).

We denote an NIN-AND tree model as $TM = (e, C, T, SP)$, where T is the NIN-AND tree and SP is the set of single-causal probabilities one per cause in C . We denote the MF of TM as $\phi = (e, C, G, F)$, where G is the MN segment and F is the set of potentials over links of G . Prop. 1 characterizes the MN segment, whose proof is straightforward.

Proposition 1 *Let $TM = (e, C, T, SP)$ be an NIN-AND tree model, where T contains $m \geq 1$ gates. Let $\phi = (e, C, G, F)$ be the MF of TM . The following hold for G :*

1. G is a tree of m internal nodes, all of which are auxiliary variables, one per gate of T .
2. Terminal nodes (degree = 1) are c_1, \dots, c_n and e only.

Although G is undirected, we refer to groups of nodes through their causal relation. For each internal node, nodes along causal direction to e are *downstream* and nodes against the direction to a_i are *upstream*. Links of G and link potentials are referred to accordingly as upstream or downstream. In Fig. 5 (b), node c_1 , link $\langle c_1, a \rangle$, and potential $f(a, c_1)$ are upstream to a , while b is downstream to a .

6 Soundness of MF of Tree Models

We show that ϕ represents the CPT of TM exactly. As it has been shown in Theorems 1 and 2 for single-gate TM s, we focus on multi-gate TM s below. First, we define a condition on potentials of ϕ in a subtree rooted at a link $\langle x, y \rangle$.

Definition 2 *Let $\phi = (e, C, G, F)$ be the MF of an NIN-AND tree model $TM = (e, C, T, SP)$ of two or more gates. Let y be an internal node of G , x be an upstream neighbor of y , c_1, \dots, c_k ($k \geq 1$) be upstream causes of y via link $\langle x, y \rangle$, and R be the set of internal nodes upstream of y via link $\langle x, y \rangle$. Let $g(b, R, c_1, \dots, c_k)$ be the product of potentials assigned to links upstream of y . Then*

$$g(y, c_1, \dots, c_k) = \sum_R g(y, R, c_1, \dots, c_k)$$

is a **valid** marginalized product (VMP) iff the following hold:

$$g(y^0, c_1^0, \dots, c_k^0) = 1; \quad (6)$$

$$g(y^0, c_1, \dots, c_k) = P(e \leftarrow c_u^1, \dots, c_v^1), \quad (7)$$

where one or more $c_i = c_i^1$, denoted c_u, \dots, c_v , and $e = e^1$ (e^0) if y is direct (dual);

$$g(y^1, c_1, \dots, c_k) = 1; \quad (8)$$

$$g(y^2, c_1^0, \dots, c_k^0) = 1; \quad (9)$$

$$g(y^2, c_1, \dots, c_k) = 0, \quad (10)$$

where one or more $c_i = c_i^1$.

Lemma 1 shows that the VMP condition holds when x is a root.

Lemma 1 *If x is a cause c_i , then $g(y, c_i)$ is a VMP.*

Proof: Here, $R = \emptyset$, $k = 1$, and $g(y, c_i) = f(y, c_i)$. If y is direct, $f(y, c_i)$ is as Table 1 (left), If y is dual, $f(y, c_i)$ is as Table 2 (left). In either case, Eqns. (6) through (10) hold. \square

Prop. 2 shows that the VMP condition holds for an arbitrary y node.

Proposition 2 *For any y , $g(y, c_1, \dots, c_k)$ is a VMP.*

Proof: We prove by induction in the length L of longest path from y to an upstream c_i . Case for $L = 1$ is shown in Lemma 1. We assume that the statement holds for $L \leq w$, where $w \geq 1$, and consider $L = w + 1$.

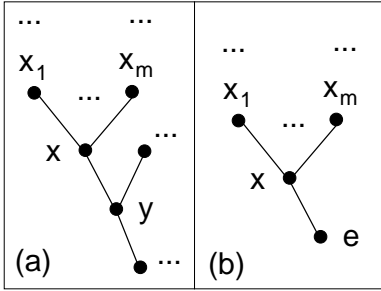


Figure 6: Illustration for proofs of Prop. 2 (a) and Theorem 3 (b)

Suppose x has upstream neighbors x_1, \dots, x_m ($m \geq 2$), (see Fig. 6 (a)). Denote corresponding upstream cause sets as S_1, \dots, S_m that form a partition of c_1, \dots, c_k . Denote marginalized product of $f(x, x_i)$ and potentials upstream of x_i by $g_i(x, S_i)$, where $i = 1, \dots, m$. By inductive assumption, each $g_i(x, S_i)$ is a VMP. We have

$$g(y, c_1, \dots, c_k) = \sum_x f(y, x) \prod_{i=1}^m g_i(x, S_i).$$

From Table 3 for $f(y, x)$, we have

$$g(y, c_1, \dots, c_k)$$

$$= \begin{cases} - \prod_{i=1}^m g_i(x^0, S_i) \\ + \prod_{i=1}^m g_i(x^1, S_i) \\ + \prod_{i=1}^m g_i(x^2, S_i) & (y = y^0) \\ \prod_{i=1}^m g_i(x^1, S_i) & (y = y^1) \\ \prod_{i=1}^m g_i(x^2, S_i) & (y = y^2). \end{cases}$$

From Eqns. (6), (8), (9) for $g_i(x, S_i)$, we have

$$- \prod_{i=1}^m g_i(x^0, \underline{s}_i^0) + \prod_{i=1}^m g_i(x^1, \underline{s}_i^0) + \prod_{i=1}^m g_i(x^2, \underline{s}_i^0) = -1 + 1 + 1.$$

Hence, Eqn. (6) holds for $g(y, c_1, \dots, c_k)$.

Next, consider Eqn. (7). Since T is minimal, if y is direct, x is dual. From Eqns. (6), (7) for $g_i(x, S_i)$,

$$- \prod_{i=1}^m g_i(x^0, S_i) = - \prod_{c_i \in S_i, c_i = c_i^1} P(e^0 \leftarrow c_i^1).$$

From Eqn. (8), $\prod_{i=1}^m g_i(x^1, S_i) = 1$. From Eqn. (10), $\prod_{i=1}^m g_i(x^2, S_i) = 0$. Hence, we have

$$- \prod_{i=1}^m g_i(x^0, S_i) + \prod_{i=1}^m g_i(x^1, S_i) + \prod_{i=1}^m g_i(x^2, S_i) = 1 - \prod_{c_i \in S_i, c_i = c_i^1} P(e^0 \leftarrow c_i^1) = P(e^1 \leftarrow c_u^1, \dots, c_v^1),$$

by Eqn. (3), where c_u, \dots, c_v are those in c_1, \dots, c_k with $c_i = c_i^1$.

Similarly, if y is dual, we have

$$g(y^0, c_1, \dots, c_k) = P(e^0 \leftarrow c_u^1, \dots, c_v^1).$$

Hence, Eqn. (7) holds for $g(y, c_1, \dots, c_k)$.

From Eqn. (8) for $g_i(x, S_i)$, we have Eqn. (8) for $g(y, c_1, \dots, c_k)$. From Eqn. (9) for $g_i(x, S_i)$, it holds for $g(y, c_1, \dots, c_k)$. Finally, from Eqn. (10) for $g_i(x, S_i)$, we have Eqn. (10) for $g(y, c_1, \dots, c_k)$. \square

Theorem 3 concludes soundness of ϕ .

Theorem 3 *Let $\phi = (e, C, G, F)$ be the MF of an NIN-AND tree model $TM = (e, C, T, SP)$ of two or more gates. Then the marginalized product of all potentials satisfies $g(e, c_1, \dots, c_n) = P(e|c_1, \dots, c_n)$.*

Proof sketch: Due to space limit, a proof sketch is given below. Let x be the (only) neighbor of e in G , shown in Fig. 6 (b), upstream neighbors of x be x_1, \dots, x_m ($m \geq 2$), and corresponding upstream cause sets be S_1, \dots, S_m that partition c_1, \dots, c_n . Denote marginalized product of $f(x, x_i)$ and potentials upstream of x_i by $g_i(x, S_i)$, where $i = 1, \dots, m$.

Consider the case where x is direct. From Table 1, Prop. 2, and Eqn. (8), it can be shown that

$$g(e^0, c_1, \dots, c_n) = 1 - g(e^1, c_1, \dots, c_n).$$

Applying Eqns. (1), (4), (6), (7), (9) and (10) to $g(e^1, c_1, \dots, c_n)$, it can be shown that $g(e, c_1, \dots, c_n)$ corresponds to $P(e|c_1, \dots, c_n)$ defined by TM.

The case for dual x can be similarly shown, and the theorem follows. \square

7 Compiling NIN-AND Tree Modeled BNs for LP

Consider a binary BN over variable set V with DAG D . Each root of D is assigned a prior, collected in set PR . Each single-parent non-root is assigned a CPT, collected in set PS . Family of each multi-parent non-root is expressed as an NIN-AND tree model, collected in set Ψ . Then, $\Gamma = (V, D, PR, PS, \Psi)$ is an *NIN-AND tree modeled BN* (NATBN).

For effective inference, we first convert Γ into a Markov network as follows:

1. Each root x of D retains prior $P(x) \in PR$.
2. For each non-root x with a single parent y , drop direction in link $y \rightarrow x$, and assign $P(x|y) \in PS$ to link $\langle y, x \rangle$.
3. For each multi-parent non-root x in D with parent set $\pi(x)$, and the NIN-AND tree model $TM \in \Psi$ over family $\{x\} \cup \pi(x)$, derive the MF ϕ of TM .

Replace subgraph of D over $\{x\} \cup \pi(x)$ (e.g., Fig. 5 (a)) by the MN segment of ϕ (e.g., Fig. 5 (b)). Assign each link of the MN segment a potential according to ϕ .

We denote the result $\Delta = (V^+, D^+, PR, LT)$, where V^+ is the set V of variables plus all auxiliary variables, D^+ is the resultant undirected graph over node set V^+ , PR is the set of potentials associated with nodes in a subset of V^+ , and LT is the set of link potentials. We refer to Δ as the *multiplicatively factorized MN* (MFMN) of Γ .

Next, we compile Δ for LP as follows:

1. Triangulate D^+ and construct a JT T^+ .
2. For each potential in PR and LT , assign to a cluster in T^+ that contains corresponding variables.

T^+ can then be used for LP. The above does not involve moralization. Note that each potential in T^+ is over at most two variables.

8 Efficiency Improvement

A collection of 120 NATBNs are simulated, divided into 4 sets of 30 each. Each NATBN contains 100 variables. For NATBNs in the same set, the number of causes per NIN-AND tree model, i.e., n , is identically upper-bounded. The bounds are 5, 7, 9 and 11, respectively. All NATBNs have the same density (5% more links than singly-connected). Each is compiled into a MFMN for LP.

For comparison, from each NATBN, a *peer BN* is derived by assigning each multi-parent variable the CPT determined from its NIN-AND tree model. Peer BNs are compiled for LP normally.

For each NATBN and its peer BN, random observations over 5 variables are used in inference by LP. Table 4 summarizes results. Each row, indexed by n , contains results from one set of NATBNs. Each column shows sample mean (left) and standard deviation (right) of a given measure. Col. 2, *pjts*, indicates state space sizes of JTs compiled from peer BNs. Col. 3, *mjts*, shows state space sizes of JTs from NATBNs. Last two columns, *pinf* and *minf*, show LP time to compute posterior marginals for all variables using peer BNs and NATBNs.

As n grows, JT state space sizes for peer BNs grow rapidly, but only slightly for NATBNs.

Table 4: Experimental Results

n	pjts		mjts		pinf (ms)		minf (ms)	
5	1232.8	78.4	1796.8	150.6	53.0	28.7	41.1	11.2
7	1810.0	164.5	1899.8	189.0	68.0	13.5	44.9	15.0
9	2784.5	444.9	1916.5	205.6	146.8	43.4	53.3	18.4
11	5919.3	815.8	1846.5	166.0	566.4	269.4	49.9	20.0

Accordingly, computational savings of NATBNs grows with n . When $n = 5$, NATBNs use about 80% of inference time as peer BNs. When $n = 11$, inference time of NATBNs is one-order of magnitude less than peer BNs.

9 Conclusion

This contribution proposes a multiplicative factorization of NIN-AND tree modeled BNs and compilation of the resultant graphical model for LP. We have shown that the method allows exact inference for posterior marginals of all variables. Our experiments demonstrate a significant efficiency gain for sparse BNs with large family sizes. Research extending the method to multi-valued variables is ongoing.

Acknowledgments

I thank reviewers for their helpful comments, and apologize for not being able to address all of them fully due to space limit. Financial support from NSERC, Canada is acknowledged.

References

- S.F. Galan and F.J. Diez. 2000. Modeling dynamic causal interaction with Bayesian networks: temporal noisy gates. In *Proc. 2nd Inter. Workshop on Causal Networks*, pages 1–5.
- D. Heckerman and J.S. Breese. 1996. Causal independence for probabilistic assessment and inference using Bayesian networks. *IEEE Trans. on System, Man and Cybernetics*, 26(6):826–831.
- D. Heckerman. 1993. Causal independence for knowledge acquisition and inference. In D. Heckerman and A. Mamdani, editors, *Proc. 9th Conf. on Uncertainty in Artificial Intelligence*, pages 122–127. Morgan Kaufmann.
- F.V. Jensen, S.L. Lauritzen, and K.G. Olesen. 1990. Bayesian updating in causal probabilistic networks by local computations. *Computational Statistics Quarterly*, (4):269–282.
- J.F. Lemmer and D.E. Gossink. 2004. Recursive noisy OR - a rule for estimating complex probabilistic interactions. *IEEE Trans. on System, Man and Cybernetics, Part B*, 34(6):2252–2261.
- A.L. Madsen and B. D’Ambrosio. 2000. A factorized representation of independence of causal influence and lazy propagation. *Inter. J. Uncertainty, Fuzziness and Knowledge-Based Systems*, 8(2):151–166.
- K.G. Olesen, U. Kjrulff, F. Jensen, F.V. Jensen, B. Falck, S. Andreassen, and S.K. Andersen. 1989. A munin network for the median nerve—a case study on loops. *Applied Artificial Intelligence*, 3(2-3):385–403.
- J. Pearl. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.
- P. Savicky and J. Vomlel. 2007. Exploiting tensor rank-one decomposition in probabilistic inference. *Kybernetika*, 43(5):747–764.
- M. Takikawa and B. D’Ambrosio. 1999. Multiplicative factorization of noisy-max. In *Proc. 15th Conf. Uncertainty in Artificial Intelligence*, pages 622–630.
- Y. Xiang and N. Jia. 2007. Modeling causal reinforcement and undermining for efficient CPT elicitation. *IEEE Trans. Knowledge and Data Engineering*, 19(12):1708–1718.
- Y. Xiang, Y. Li, and J. Zhu. 2009. Towards effective elicitation of NIN-AND tree causal models. In L. Godo and A. Pugliese, editors, *Inter. Conf. on Scalable Uncertainty Management (SUM 2009)*, LNCS 5785, pages 282–296. Springer-Verlag Berlin Heidelberg.
- N. Zhang and D. Poole. 1996. Exploiting causal independence in bayesian network inference. *J. Artificial Intelligence Research*, 5:301–328.