Tensor rank-one decomposition of noisy-or models

Petr Savický and Jiří Vomlel

ÚI a ÚTIA AV ČR

Alšovice 2008, Rakovice

Additive decomposition of a probability table

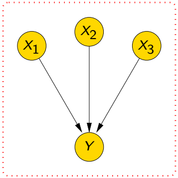
We say that a conditional probability table $P(Y|X_1, \ldots X_n)$ can be decomposed by use of an auxiliary variable B if there exist tables $\xi(B, Y)$ and $\varphi_i(B, X_i)$, for $i = 1, \ldots, n$ such that

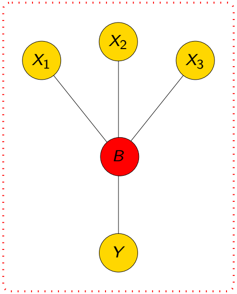
$$P(Y|X_1, \ldots, X_n) = \sum_B \xi(Y, B) \cdot \prod_{i=1}^{n} \varphi_i(B, X_i)$$

Additive decomposition of a probability table

We say that a conditional probability table $P(Y|X_1, \ldots X_n)$ can be decomposed by use of an auxiliary variable B if there exist tables $\xi(B, Y)$ and $\varphi_i(B, X_i)$, for $i = 1, \ldots, n$ such that

$$P(Y|X_1, \ldots, X_n) = \sum_B \xi(Y, B) \cdot \prod_{i=1}^{n} \varphi_i(B, X_i)$$





Correspondence to tensor rank-one decomposition

A decomposition using the auxiliary variable B

$$P(Y|X_1,\ldots,X_n) \;=\; \sum_B \xi(Y,B)\cdot\prod_{i=1}^{n}\varphi_i(B,X_i)$$

that has the minimal number of states of B

Correspondence to tensor rank-one decomposition

A decomposition using the auxiliary variable B

$$P(Y|X_1,\ldots,X_n) \;=\; \sum_B \xi(Y,B)\cdot\prod_{i=1}^{n}\varphi_i(B,X_i)$$

that has the minimal number of states of B

is in fact a (minimal) tensor rank-one decomposition of $P(Y|X_1,\ldots,X_n)$.

Tensor rank

Definition (Tensor)

Let $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_n$, where $\mathcal{X}_i \subset \mathbb{N}$. $\psi : \mathcal{X} \mapsto \mathbb{R}$ is tensor of order n.

Tensor rank

Definition (Tensor)

Let $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_n$, where $\mathcal{X}_i \subset \mathbb{N}$. $\psi : \mathcal{X} \mapsto \mathbb{R}$ is tensor of order n.

Definition (Tensor of rank one)

A tensor ψ of order n has rank 1 if it is the outer product of n vectors

Tensor rank

Definition (Tensor)

Let $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_n$, where $\mathcal{X}_i \subset \mathbb{N}$. $\psi : \mathcal{X} \mapsto \mathbb{R}$ is tensor of order n.

Definition (Tensor of rank one)

A tensor ψ of order n has rank 1 if it is the outer product of n vectors , i.e.

$$\psi(x_1, \ldots, x_n) \quad = \quad \psi_1(x_1) \cdot \ldots \cdot \psi_n(x_n) \text{for all } x \in \mathcal{X}.$$

Tensor rank

Definition (Tensor)

Let $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_n$, where $\mathcal{X}_i \subset \mathbb{N}$. $\psi : \mathcal{X} \mapsto \mathbb{R}$ is tensor of order n.

Definition (Tensor of rank one)

A tensor ψ of order n has rank 1 if it is the outer product of n vectors , i.e.

$$\psi(x_1, \ldots, x_n) \quad = \quad \psi_1(x_1) \cdot \ldots \cdot \psi_n(x_n) \text{for all } x \in \mathcal{X}.$$

Example (A 3rd-order tensor that has rank 1)

$$\begin{pmatrix} (1,2) & (2,4) \\ (2,4) & (4,8) \end{pmatrix}$$

Tensor rank

Definition (Tensor)

Let $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_n$, where $\mathcal{X}_i \subset \mathbb{N}$. $\psi : \mathcal{X} \mapsto \mathbb{R}$ is tensor of order n.

Definition (Tensor of rank one)

A tensor ψ of order n has rank 1 if it is the outer product of n vectors , i.e.

$$\psi(x_1, \ldots, x_n) \;=\; \psi_1(x_1) \cdot \ldots \cdot \psi_n(x_n) \text{for all } x \in \mathcal{X}.$$

Example (A 3rd-order tensor that has rank 1)

$$\begin{pmatrix} (1,2) & (2,4) \\ (2,4) & (4,8) \end{pmatrix} \;=\; (1,2) \otimes (1,2) \otimes (1,2)$$

Tensor rank

Definition (Tensor rank)

The minimal number of rank-one tensors necessary to add in order to yield a tensor ψ is called rank of the tensor ψ.

Example (A rank-2 tensor)

$$\begin{pmatrix} (1,2) & (2,4) \\ (2,4) & (4,9) \end{pmatrix}$$

Tensor rank

Definition (Tensor rank)

The minimal number of rank-one tensors necessary to add in order to yield a tensor ψ is called rank of the tensor ψ.

Example (A rank-2 tensor)

$$\begin{pmatrix} (1,2) & (2,4) \\ (2,4) & (4,9) \end{pmatrix} = \begin{pmatrix} (1,2) & (2,4) \\ (2,4) & (4,8) \end{pmatrix} + \begin{pmatrix} (0,0) & (0,0) \\ (0,0) & (0,1) \end{pmatrix}$$

Tensor rank

Definition (Tensor rank)

The minimal number of rank-one tensors necessary to add in order to yield a tensor ψ is called rank of the tensor ψ.

Example (A rank-2 tensor)

$$\begin{aligned}\begin{pmatrix} (1,2) & (2,4) \\ (2,4) & (4,9) \end{pmatrix} &= \begin{pmatrix} (1,2) & (2,4) \\ (2,4) & (4,8) \end{pmatrix} + \begin{pmatrix} (0,0) & (0,0) \\ (0,0) & (0,1) \end{pmatrix} \\ &= (1,2) \otimes (1,2) \otimes (1,2) \; + \; (0,1) \otimes (0,1) \otimes (0,1)\end{aligned}$$

Tensor rank

Definition (Tensor rank)

The minimal number of rank-one tensors necessary to add in order to yield a tensor ψ is called rank of the tensor ψ.

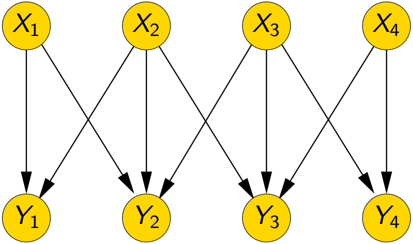
Example (A rank-2 tensor)

$$\begin{pmatrix} (1,2) & (2,4) \\ (2,4) & (4,9) \end{pmatrix} = \begin{pmatrix} (1,2) & (2,4) \\ (2,4) & (4,8) \end{pmatrix} + \begin{pmatrix} (0,0) & (0,0) \\ (0,0) & (0,1) \end{pmatrix}$$

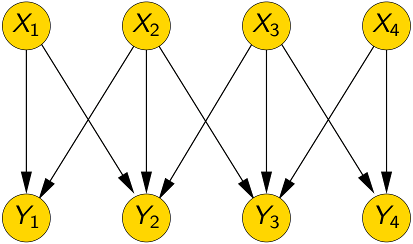
$$= (1,2) \otimes (1,2) \otimes (1,2) + (0,1) \otimes (0,1) \otimes (0,1)$$

$$\neq (a,b) \otimes (c,d) \otimes (e,f) \text{ for any } a,b,c,d,e,f \in \mathbb{R}.$$

BN2O model

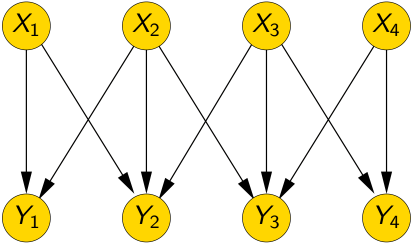


BN2O model



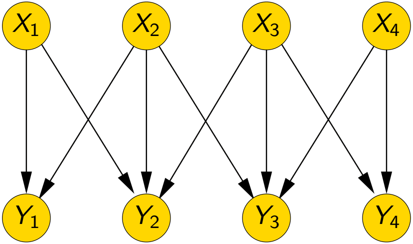
- Y_i is false only if all its parents with value true are inhibited.

BN2O model



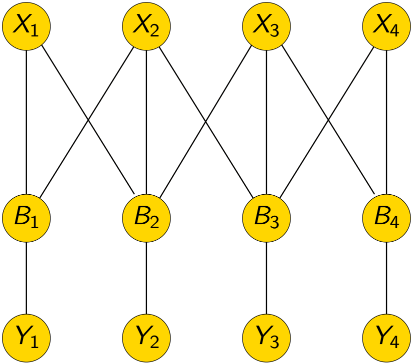
- Y_i is false only if all its parents with value true are inhibited.
- Let $p_{i,j}$ be the inhibition probability for the parent node X_j of Y_i.

BN2O model



- Y_i is false only if all its parents with value true are inhibited.
- Let $p_{i,j}$ be the inhibition probability for the parent node X_j of Y_i.
- $P(Y_i = 0 | X_{pa(Y_i)} = x_{pa(Y_i)}) \quad = \quad \prod_{j \in pa(Y_i)}^{n} p_{i,j}^{x_j}$

BN2O model after rank-one decomposition



BN2O model after rank-one decomposition (Díez and Galán, 2003)

$$P(Y_i = y_i | X_1 = x_1, \ldots, X_n = x_n) \quad = \quad \sum_{b_i=0}^{1} \xi(b_i, y_i) \cdot \prod_{j=1}^{n} \varphi_j(b_i, x_j)$$

BN2O model after rank-one decomposition (Díez and Galán, 2003)

$$P(Y_i = y_i | X_1 = x_1, \ldots, X_n = x_n) = \sum_{b_i=0}^{1} \xi(b_i, y_i) \cdot \prod_{j=1}^{n} \varphi_j(b_i, x_j)$$

$$\xi(b_i, y_i) = \begin{cases} +1 & b_i = y_i \\ -1 & b_i = y_i - 1 \\ 0 & \text{otherwise.} \end{cases}$$

BN2O model after rank-one decomposition (Díez and Galán, 2003)

$$
\begin{aligned}
P(Y_i = y_i | X_1 = x_1, \ldots, X_n = x_n) &= \sum_{b_i=0}^{1} \xi(b_i, y_i) \cdot \prod_{j=1}^{n} \varphi_j(b_i, x_j) \\
\xi(b_i, y_i) &= \begin{cases} +1 & b_i = y_i \\ -1 & b_i = y_i - 1 \\ 0 & \text{otherwise.} \end{cases} \\
\varphi_j(b_i, x_j) &= \begin{cases} p_{i,j} & \text{if } b_i = x_j - 1 \\ 1 & \text{otherwise.} \end{cases}
\end{aligned}
$$

Arithmetic circuits

Circuit inputs:

Arithmetic circuits

Circuit inputs: 7

- *BN parameters*

$$\theta_{x|\mathbf{u}} \quad = \quad P(X = x|\mathbf{U} = \mathbf{u})$$

Arithmetic circuits

Circuit inputs: 7

- *BN parameters*

$$\theta_{x|\mathbf{u}} = P(X = x|\mathbf{U} = \mathbf{u})$$

- *evidence indicators*

$$\lambda_x = \begin{cases} 1 & \text{if state } x \text{ of } X \text{ is consistent with evidence } \mathbf{e} \\ 0 & \text{otherwise.} \end{cases}$$

If there is no evidence for X, then $\lambda_x = 1$ for all states x of X.

Arithmetic circuits

Circuit inputs: 7

- *BN parameters*

$$\theta_{x|\mathbf{u}} \quad = \quad P(X = x|\mathbf{U} = \mathbf{u})$$

- *evidence indicators*

$$\lambda_x \quad = \quad \begin{cases} 1 & \text{if state } x \text{ of } X \text{ is consistent with evidence } \mathbf{e} \\ 0 & \text{otherwise.} \end{cases}$$

If there is no evidence for X, then $\lambda_x = 1$ for all states x of X.

Circuit output:

- probability of evidence $P(\mathbf{e})$.

Arithmetic circuits

Circuit inputs: 7

- *BN parameters*

$$\theta_{x|\mathbf{u}} = P(X = x|\mathbf{U} = \mathbf{u})$$

- *evidence indicators*

$$\lambda_x = \begin{cases} 1 & \text{if state } x \text{ of } X \text{ is consistent with evidence } \mathbf{e} \\ 0 & \text{otherwise.} \end{cases}$$

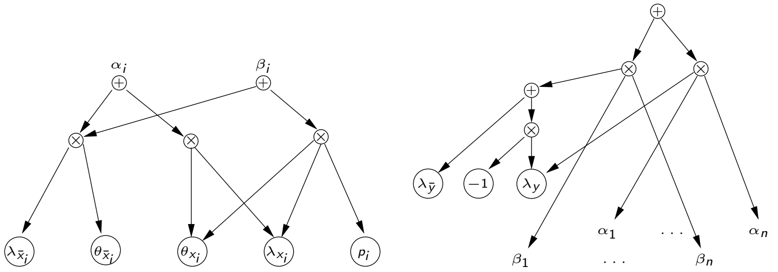
If there is no evidence for X, then $\lambda_x = 1$ for all states x of X.

Circuit output:

- probability of evidence $P(\mathbf{e})$.

Definition (Arithmetic circuit (AC))

Arithmetic circuit is a rooted, directed acyclic graph whose leaf nodes correspond to circuit inputs and whose other nodes are labeled with multiplication and addition operations. The root node corresponds to circuit output.



α_i
β_i
$\lambda_{\bar{x}_i}$
$\theta_{\bar{x}_i}$
θ_{x_i}
λ_{x_i}
p_i
$\lambda_{\bar{y}}$
-1
λ_y
α_1
$\dots$
α_n
β_1
$\dots$
β_n

Arithmetic circuits

- An AC may be used to represent the computations determined by a junction tree,

Arithmetic circuits

- An AC may be used to represent the computations determined by a junction tree,
- An AC may also represent more efficient computations, if they exist due to specific properties of the initial BN (e.g., determinism, context specific independence).

Construction of an AC

Darwiche (2002) proposed a method for constructing ACs of Bayesian networks.

Construction of an AC

Darwiche (2002) proposed a method for constructing ACs of Bayesian networks. It is based on

- encoding a Bayesian network using the logical conjunctive normal form (CNF),

Construction of an AC

Darwiche (2002) proposed a method for constructing ACs of Bayesian networks. It is based on

- encoding a Bayesian network using the logical conjunctive normal form (CNF),
- transforming the CNF into a special logical form called Deterministic, Decomposable Negation Normal Form (d-DNNF), and

Construction of an AC

Darwiche (2002) proposed a method for constructing ACs of Bayesian networks. It is based on

- encoding a Bayesian network using the logical conjunctive normal form (CNF),
- transforming the CNF into a special logical form called Deterministic, Decomposable Negation Normal Form (d-DNNF), and
- transforming the d-DNNF to the AC of the original Bayesian network.

Construction of an AC

Darwiche (2002) proposed a method for constructing ACs of Bayesian networks. It is based on

- encoding a Bayesian network using the logical conjunctive normal form (CNF),
- transforming the CNF into a special logical form called Deterministic, Decomposable Negation Normal Form (d-DNNF), and
- transforming the d-DNNF to the AC of the original Bayesian network.

We apply Darwiche's method to the BN2O transformed using the tensor rank-one decomposition.

Comparisons of AC sizes - created by Ace running on

`aligator.utia.cas.cz`: 8x AMD Opteron 8220, 64GB RAM

Median of the ratio o/t of the AC sizes for the original o and transformed model t computed from 10 models with randomly generated structure of the type x-y-e

x-y-e	o/t
6-40-160	1.49
10-15-60	1.24
10-20-80	1.19
20-30-150	1.03
20-40-200	1.01
30-15-150	474.86
50-50-200	1.11

x-y-e	o/t
10-1-10	1.47
10-15-90	1.21
10-30-90	1.13
20-30-300	0.98
30-10-100	303.48
50-20-80	1.90

Comparisons of AC sizes - created by Ace running on `aligator.utia.cas.cz`: 8x AMD Opteron 8220, 64GB RAM

Median of the ratio o/t of the AC sizes for the original o and transformed model t computed from 10 models with randomly generated structure of the type x-y-e

x-y-e	o/t
6-40-160	1.49
10-15-60	1.24
10-20-80	1.19
20-30-150	1.03
20-40-200	1.01
30-15-150	474.86
50-50-200	1.11

x-y-e	o/t
10-1-10	1.47
10-15-90	1.21
10-30-90	1.13
20-30-300	0.98
30-10-100	303.48
50-20-80	1.90

Models for which it was possible to construct an AC for the transformed model only (program run out of memory for the original model):

30-20-200 30-25-250 50-20-200

Transformed vs. original model AC size in the $\log_{10}$ scale