Computationally efficient probabilistic inference with noisy threshold models based on a CP tensor decomposition

Jirka Vomlel and Petr Tichavský

Institute of Information Theory and Automation (ÚTIA)
Academy of Sciences of the Czech Republic

Contents

- Motivation

Contents

Contents

- Motivation
- Noisy threshold models
- CP-decomposition of conditional probability tables

Contents

- Motivation
- Noisy threshold models
- CP-decomposition of conditional probability tables
- Experiments

Contents

Quick Medical Reference - Decision Theoretic (QMR-DT) Miller et al. (1986) and Shwe et al. (1991).

- 570 diseases in the first level

Quick Medical Reference - Decision Theoretic (QMR-DT) Miller et al. (1986) and Shwe et al. (1991).

- 570 diseases in the first level
- 4075 observations in the second level

Quick Medical Reference - Decision Theoretic (QMR-DT) Miller et al. (1986) and Shwe et al. (1991).

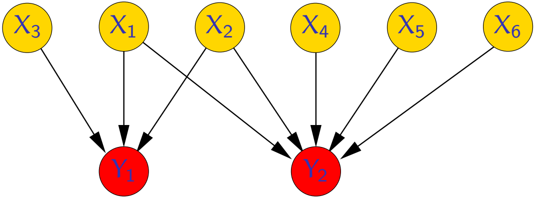
- 570 diseases in the first level
- 4075 observations in the second level
- all variables are binary

Quick Medical Reference - Decision Theoretic (QMR-DT) Miller et al. (1986) and Shwe et al. (1991).

- 570 diseases in the first level
- 4075 observations in the second level
- all variables are binary
- conditional probability tables are noisy-or models

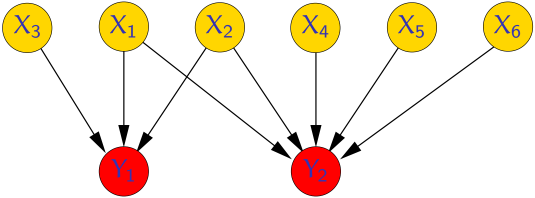
Quick Medical Reference - Decision Theoretic (QMR-DT) Miller et al. (1986) and Shwe et al. (1991).

- 570 diseases in the first level
- 4075 observations in the second level
- all variables are binary
- conditional probability tables are noisy-or models



Quick Medical Reference - Decision Theoretic (QMR-DT) Miller et al. (1986) and Shwe et al. (1991).

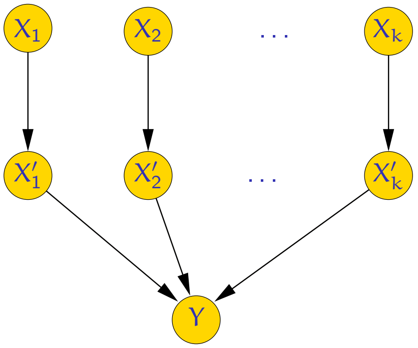
- 570 diseases in the first level
- 4075 observations in the second level
- all variables are binary
- conditional probability tables are noisy-or models



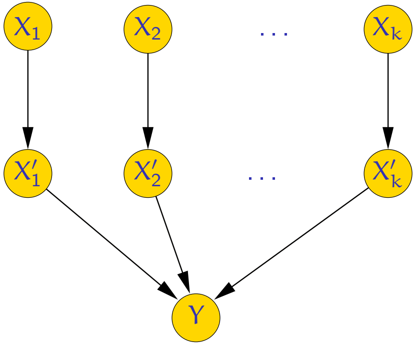
Definition (The inference task)

Given a subset of observations (e.g. Y_1 and Y_2) compute probabilities of diseases (e.g. $P(X_i|Y_1 = y_1, Y_2 = y_2)$, $i = 1, \ldots, 6$.

Noisy threshold - a generalization of noisy-or



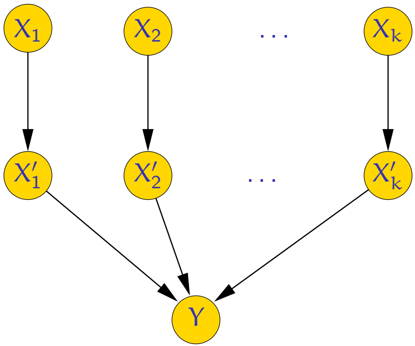
Noisy threshold - a generalization of noisy-or



Y takes value 1 if at least ℓ out of k parents take value 1:

$$P(Y = 1 | X'_1 = x'_1, \ldots, X'_k = x'_k) = \begin{cases} 1 & \text{if } x'_1 + \ldots + x'_k \geqslant \ell \\ 0 & \text{otherwise.} \end{cases}$$

Noisy threshold - a generalization of noisy-or



Y takes value 1 if at least ℓ out of k parents take value 1:

$$P(Y = 1 | X'_1 = x'_1, \ldots, X'_k = x'_k) = \begin{cases} 1 & \text{if } x'_1 + \ldots + x'_k \geqslant \ell \\ 0 & \text{otherwise.} \end{cases}$$

Noise: for $i = 1, \ldots, k$

$$P(X'_i = 1 | X_i = x_i) = \begin{cases} 0 & \text{if } x_i = 0 \\ \pi_i & \text{otherwise.} \end{cases}$$

An example for $k = 4, \ell = 1$, and $\pi_i = 1, i = 1, \ldots, k$ - i.e., for deterministic OR function

$P(Y = 1 | X_1 = x_1, \ldots, X_4 = x_4)$

An example for $k = 4, \ell = 1,$ and $\pi_i = 1, i = 1, \ldots, k$ - i.e., for deterministic OR function

$$P(Y = 1 | X_1 = x_1, \ldots, X_4 = x_4)$$

$$= \begin{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \\ \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \end{pmatrix}$$

An example for $k = 4, \ell = 1,$ and $\pi_i = 1, i = 1, \ldots, k$ - i.e., for deterministic OR function

$$
\begin{aligned}
& P(Y = 1 | X_1 = x_1, \ldots, X_4 = x_4) \\
= & \begin{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \\ \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \end{pmatrix} \\
= & \begin{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \\ \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \end{pmatrix} - \begin{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \end{pmatrix}
\end{aligned}
$$

An example for $k = 4, \ell = 1,$ and $\pi_i = 1, i = 1, \ldots, k$ - i.e., for deterministic OR function

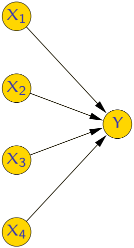
$$
\begin{aligned}
& P(Y = 1 | X_1 = x_1, \ldots, X_4 = x_4) \\
= & \begin{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \\ \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \end{pmatrix} \\
= & \begin{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \\ \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \end{pmatrix} - \begin{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \end{pmatrix} \\
= & (1,1) \otimes (1,1) \otimes (1,1) \otimes (1,1) \; - \; (1,0) \otimes (1,0) \otimes (1,0) \otimes (1,0)
\end{aligned}
$$

An example for $k = 4, \ell = 1,$ and $\pi_i = 1, i = 1, \ldots, k$ - i.e., for deterministic OR function

$$
\begin{aligned}
& P(Y = 1 | X_1 = x_1, \ldots, X_4 = x_4) \\
= & \begin{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \\ \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \end{pmatrix} \\
= & \begin{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \\ \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \end{pmatrix} - \begin{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \end{pmatrix} \\
= & (1,1) \otimes (1,1) \otimes (1,1) \otimes (1,1) \; - \; (1,0) \otimes (1,0) \otimes (1,0) \otimes (1,0) \\
= & (1,1)^{\otimes k} - (1,0)^{\otimes k}
\end{aligned}
$$

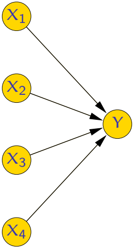
Compilation of the threshold model for $\ell = 1$ - the standard approach

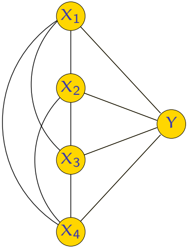
Lauritzen and Spiegelhalter (1988), Jensen et al. (1990), Shafer and Shenoy (1990)



Compilation of the threshold model for $\ell = 1$ - the standard approach

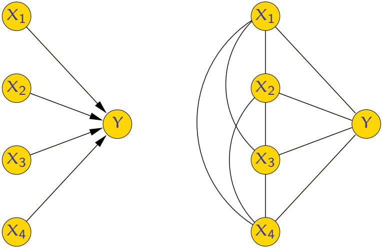
Lauritzen and Spiegelhalter (1988), Jensen et al. (1990), Shafer and Shenoy (1990)





Compilation of the threshold model for $\ell = 1$ - the standard approach

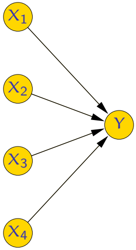
Lauritzen and Spiegelhalter (1988), Jensen et al. (1990), Shafer and Shenoy (1990)



The total table size is $2^5 = 32$.

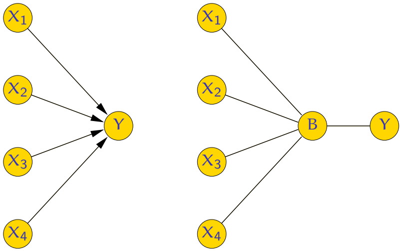
Compilation of the threshold model for $\ell = 1$ - after the suggested decomposition

Díez and Galán (2002), Vomlel (2002), Savický and Vomlel (2007)



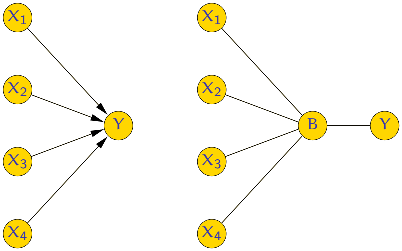
Compilation of the threshold model for $\ell = 1$ - after the suggested decomposition

Díez and Galán (2002), Vomlel (2002), Savický and Vomlel (2007)



Compilation of the threshold model for $\ell = 1$ - after the suggested decomposition

Díez and Galán (2002), Vomlel (2002), Savický and Vomlel (2007)



The total table size is $5 \cdot 2^2 = 20$.

Decomposition of $\mathbf{T}(\ell, k)$ into sum of tensor products

- $P(Y = 1|\mathbf{X} = x)$ can be viewed as a **tensor** $\mathbf{T}(\ell, k)$.

Decomposition of $\mathbf{T}(\ell, k)$ into sum of tensor products

- $P(Y = 1|\mathbf{X} = x)$ can be viewed as a **tensor** $\mathbf{T}(\ell, k)$.
- All dimensions of $\mathbf{T}(\ell, k)$ are equal to 2.

Decomposition of $\mathbf{T}(\ell, k)$ into sum of tensor products

- $P(Y = 1|\mathbf{X} = x)$ can be viewed as a **tensor** $\mathbf{T}(\ell, k)$.
- All dimensions of $\mathbf{T}(\ell, k)$ are equal to 2.
- $\mathbf{T}(\ell, k)$ is symmetric.

Decomposition of $\mathbf{T}(\ell, k)$ into sum of tensor products

- $P(Y = 1 | \mathbf{X} = x)$ can be viewed as a **tensor** $\mathbf{T}(\ell, k)$.
- All dimensions of $\mathbf{T}(\ell, k)$ are equal to 2.
- $\mathbf{T}(\ell, k)$ is symmetric.

Definition (Symmetric rank)

Symmetric rank (srank) is the minimum number r such that

$$\mathbf{T}(\ell, k) = \sum_{i=1}^{r} b_i \cdot \mathbf{a}_i^{\otimes k}$$

where for $i = 1, \ldots, k$:

- $b_i \in \mathbb{R}$ and
- $\mathbf{a}_i$ are real-valued vectors of length 2.

Decomposition of $\mathbf{T}(\ell, k)$ into sum of tensor products

- $P(Y = 1|\mathbf{X} = x)$ can be viewed as a **tensor** $\mathbf{T}(\ell, k)$.
- All dimensions of $\mathbf{T}(\ell, k)$ are equal to 2.
- $\mathbf{T}(\ell, k)$ is symmetric.

Definition (Symmetric rank)

Symmetric rank (srank) is the minimum number r such that

$$\mathbf{T}(\ell, k) = \sum_{i=1}^{r} b_i \cdot \mathbf{a}_i^{\otimes k}$$

where for $i = 1, \ldots, k$:

- $b_i \in \mathbb{R}$ and
- $\mathbf{a}_i$ are real-valued vectors of length 2.

- This decomposition is called **Canonical Polyadic** (CP) or **CANDECOMP-PARAFAC** (CP) or **tensor rank-one**.

Theoretical results

Results in the proceedings:

- $\mathit{srank}(\mathbf{T}(0, k)) = 1$.

Theoretical results

Results in the proceedings:

- $\mathrm{srank}(\mathbf{T}(0,k)) = 1$.
- $\mathrm{srank}(\mathbf{T}(k,k)) = 1$.

Theoretical results

Results in the proceedings:

- $srank(\mathbf{T}(0, k)) = 1$.
- $srank(\mathbf{T}(k, k)) = 1$.
- $srank(\mathbf{T}(1, k)) = 2$.

Theoretical results

Results in the proceedings:

- $srank(\mathbf{T}(0, k)) = 1$.
- $srank(\mathbf{T}(k, k)) = 1$.
- $srank(\mathbf{T}(1, k)) = 2$.
- $srank(\mathbf{T}(k-1, k)) = k$.

Theoretical results

Results in the proceedings:

- $\mathrm{srank}(\mathbf{T}(0,k)) = 1$.
- $\mathrm{srank}(\mathbf{T}(k,k)) = 1$.
- $\mathrm{srank}(\mathbf{T}(1,k)) = 2$.
- $\mathrm{srank}(\mathbf{T}(k-1,k)) = k$.
- $\mathrm{srank}(\mathbf{T}(\ell,k)) \leqslant k$ for $\ell = 3,\ldots,k-2$.

Theoretical results

Results in the proceedings:

- $\mathrm{srank}(\mathbf{T}(0, k)) = 1$.
- $\mathrm{srank}(\mathbf{T}(k, k)) = 1$.
- $\mathrm{srank}(\mathbf{T}(1, k)) = 2$.
- $\mathrm{srank}(\mathbf{T}(k-1, k)) = k$.
- $\mathrm{srank}(\mathbf{T}(\ell, k)) \leqslant k$ for $\ell = 3, \dots, k-2$.
- An algorithm for CP-decomposition to k factors.

Theoretical results

Results in the proceedings:

- $\mathrm{srank}(\mathbf{T}(0, k)) = 1$.
- $\mathrm{srank}(\mathbf{T}(k, k)) = 1$.
- $\mathrm{srank}(\mathbf{T}(1, k)) = 2$.
- $\mathrm{srank}(\mathbf{T}(k - 1, k)) = k$.
- $\mathrm{srank}(\mathbf{T}(\ell, k)) \leqslant k$ for $\ell = 3, \ldots, k - 2$.
- An algorithm for CP-decomposition to k factors.
- For the noisy threshold the above values represent upper bounds.

Theoretical results

Results in the proceedings:

- $\mathrm{srank}(\mathbf{T}(0, k)) = 1$.
- $\mathrm{srank}(\mathbf{T}(k, k)) = 1$.
- $\mathrm{srank}(\mathbf{T}(1, k)) = 2$.
- $\mathrm{srank}(\mathbf{T}(k-1, k)) = k$.
- $\mathrm{srank}(\mathbf{T}(\ell, k)) \leqslant k$ for $\ell = 3, \dots, k-2$.
- An algorithm for CP-decomposition to k factors.
- For the noisy threshold the above values represent upper bounds.

New results (not in the proceedings):

Theoretical results

Results in the proceedings:

- $\mathrm{srank}(\mathbf{T}(0, k)) = 1$.
- $\mathrm{srank}(\mathbf{T}(k, k)) = 1$.
- $\mathrm{srank}(\mathbf{T}(1, k)) = 2$.
- $\mathrm{srank}(\mathbf{T}(k-1, k)) = k$.
- $\mathrm{srank}(\mathbf{T}(\ell, k)) \leqslant k$ for $\ell = 3, \ldots, k-2$.
- An algorithm for CP-decomposition to k factors.
- For the noisy threshold the above values represent upper bounds.

New results (not in the proceedings):

- $\mathrm{srank}(\mathbf{T}(\ell, k)) \geqslant k-1$ for $\ell = 3, \ldots, k-2$.

Theoretical results

Results in the proceedings:

- $\mathrm{srank}(\mathbf{T}(0, k)) = 1$.
- $\mathrm{srank}(\mathbf{T}(k, k)) = 1$.
- $\mathrm{srank}(\mathbf{T}(1, k)) = 2$.
- $\mathrm{srank}(\mathbf{T}(k-1, k)) = k$.
- $\mathrm{srank}(\mathbf{T}(\ell, k)) \leqslant k$ for $\ell = 3, \ldots, k-2$.
- An algorithm for CP-decomposition to k factors.
- For the noisy threshold the above values represent upper bounds.

New results (not in the proceedings):

- $\mathrm{srank}(\mathbf{T}(\ell, k)) \geqslant k-1$ for $\ell = 3, \ldots, k-2$.
- An algorithm for CP-decomposition to $k-1$ factors. But we don't know yet if we can avoid complex numbers for some values of ℓ.

Experimental results

Comparisons of the total table size:

- the standard junction tree method versus the CP tensor decomposition

Experimental results

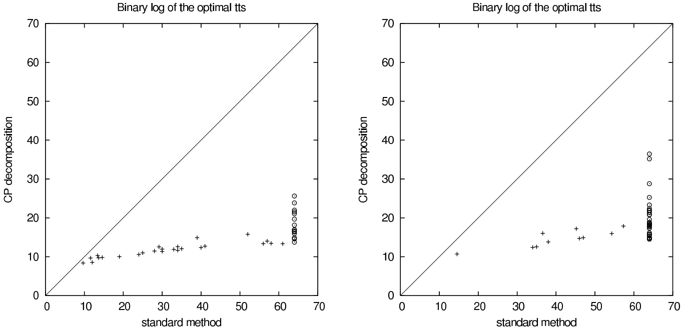
Comparisons of the total table size:

- the standard junction tree method versus the CP tensor decomposition
- using QMR subnetworks networks after 14 observations and after 28 observations.

Experimental results

Comparisons of the total table size:

- the standard junction tree method versus the CP tensor decomposition
- using QMR subnetworks networks after 14 observations and after 28 observations.



Relation to arithmetic circuits (ACs) of Darwiche et al.

- The model after the CP decomposition can be used as an input for Ace (Chavira and Darwiche).

Relation to arithmetic circuits (ACs) of Darwiche et al.

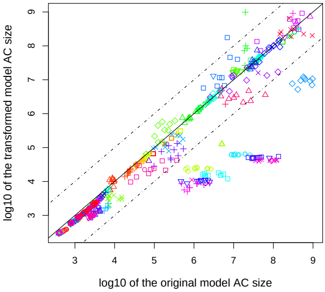
- The model after the CP decomposition can be used as an input for Ace (Chavira and Darwiche).
- Ace supports parent divorcing for noisy-or (i.e., $\ell = 1$).

Relation to arithmetic circuits (ACs) of Darwiche et al.

- The model after the CP decomposition can be used as an input for Ace (Chavira and Darwiche).
- Ace supports parent divorcing for noisy-or (i.e., $\ell = 1$).
- In our PGM'08 paper we reported comparisons of the ACs' size for random QMR-like networks:

Relation to arithmetic circuits (ACs) of Darwiche et al.

- The model after the CP decomposition can be used as an input for Ace (Chavira and Darwiche).
- Ace supports parent divorcing for noisy-or (i.e., $\ell = 1$).
- In our PGM'08 paper we reported comparisons of the ACs' size for random QMR-like networks:



Conclusions

- Theoretical results that give upper bounds for symmetric rank of tensors corresponding to threshold functions.

Conclusions

- Theoretical results that give upper bounds for symmetric rank of tensors corresponding to threshold functions.
- An algorithm for CP decomposition of these tensors.

Conclusions

- Theoretical results that give upper bounds for symmetric rank of tensors corresponding to threshold functions.
- An algorithm for CP decomposition of these tensors.
- The CP tensor decomposition lead to a computational gain in the order of several magnitudes and made many intractable models manageable.

Acknowledgments

Thanks to:

- Frank Jensen from Hugin for providing the Hugin optimal triangulation method and

Acknowledgments

Thanks to:

- Frank Jensen from Hugin for providing the Hugin optimal triangulation method and
- Gregory F. Cooper from University of Pittsburgh for the structural part of QMR-DT model.