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

Jiří Vomlel
Institute of Information Theory and Automation of the AS CR
Prague, Czech Republic
vomlel@utia.cas.cz

Petr Tichavský
Institute of Information Theory and Automation of the AS CR
Prague, Czech Republic
tichavsk@utia.cas.cz

Abstract

Conditional probability tables (CPTs) of threshold functions represent a generalization of two popular models – noisy-or and noisy-and. They constitute an alternative to these two models in case they are too rough. When using the standard inference techniques the inference complexity is exponential with respect to the number of parents of a variable. In case the CPTs take a special form (in this paper it is the noisy-threshold model) more efficient inference techniques could be employed. Each CPT defined for variables with finite number of states can be viewed as a tensor (a multilinear array). Tensors can be decomposed as linear combinations of rank-one tensors, where a rank one tensor is an outer product of vectors. Such decomposition is referred to as Canonical Polyadic (CP) or CANDECOMP-PARAFAC (CP) decomposition. The tensor decomposition offers a compact representation of CPTs which can be efficiently utilized in probabilistic inference. In this paper we propose a CP decomposition of tensors corresponding to CPTs of threshold functions and their noisy counterparts. We performed experiments on subnetworks of the well-known QMR-DT network generalized by replacing noisy-or by noisy-threshold models. Each generated subnetwork contained more then one hundred variables. The results of our experiments reveal that by using the suggested decomposition of CPTs we can get computational savings in several orders of magnitude.

1 Introduction

In many applications of Bayesian networks (Jensen and Nielsen, 2007), conditional probability tables (CPTs) have a certain local structure. Canonical models (Díez and Druzdzel, 2006) form a commonly used class of CPTs with the local structure being defined as a combination of a deterministic part with independent probabilistic influence of each parent variable, see Figure 1.

*This work was supported by Czech Science Foundation through projects 201/08/0539 and 102/09/1278.

The joint probability distribution of the Bayesian network in Figure 1 is

$$P(Y|X_1', \ldots, X_k') \prod_{i=1}^{k} P(X_i'|X_i) P(X_i) \ ,$$

where the first term $P(Y|X_1', \ldots, X_k')$ corresponds to a deterministic function and terms $P(X_i'|X_i)$ to the probabilistic part (often called noise). We can replace the Bayesian network of Figure 1 by a model without auxiliary variables $X_1', \ldots, X_k'$ by marginalizing them out from the Bayesian network. The values of $P(Y|X_1, \ldots, X_k)$ can be computed from the

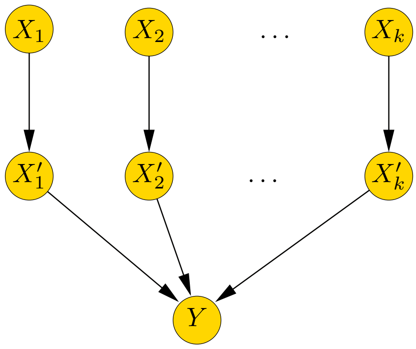

Figure 1: A Bayesian network with a canonical model with explicit deterministic part $P(Y|X'_1, \dots, X'_k)$ and probabilistic parts $P(X'_i|X_i), i = 1, \dots, k$.

original model by

$$P(Y|X_1, \dots, X_k) = \sum_{X'_1} \dots \sum_{X'_k} P(Y, X'_1, \dots, X'_k) \cdot \prod_{i=1}^{k} P(X'_i|X_i) \ .$$

Assume CPT with the state y of variable Y being observed. As it was suggested in (Savicky and Vomlel, 2007) we can rewrite each CPT as a product of two-dimensional potentials $\psi_i, i = 1, \dots, k$

$$P(y|X_1, \dots, X_k) = \sum_{B} \prod_{i=1}^{k} \psi(B, X_i) \ , \quad (1)$$

where B is an auxiliary variable. This transformation can be visualized by the undirected graph given in Figure 2.

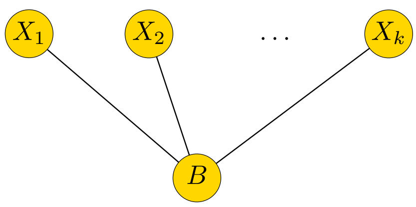

Figure 2: Model of $P(y|X_1, \dots, X_k)$ after the transformation using auxiliary variable B.

In order to guarantee the above equality, variable B has to have certain number of states. However, the equality can be always satisfied if the number of states of B is the product of the number of states of variables $X_1, \dots, X_k$. The transformation becomes computationally advantageous if the number of states is low, which is the case of CPTs of canonical models. It was observed in (Savicky and Vomlel, 2007) that since each CPT can be understood as a tensor[1] the minimum number of states of B equals the rank of tensor A whose values are defined as

$$\boldsymbol{A}_{i_1,\dots,i_k} = P(y|X_1 = x_{i_1}, \dots, X_k = x_{i_k}),$$

for all combinations of states $(x_{i_1}, \dots, x_{i_k})$ of variables $X_1, \dots, X_k$. The decomposition of tensors into the form corresponding to the right hand side of formula (1) has been studied for more than forty years (Carroll and Chang, 1970; Harshman, 1970) and it is known now as Canonical Polyadic (CP) or CANDECOMP-PARAFAC (CP) decomposition. In (Comon et al., 2008) it is called an outer-product decomposition.

In this paper we deal with conditional probability tables representing one specific type of canonical models – deterministic threshold functions and their noisy counterparts. An (ℓ, k) threshold function is a function of k binary arguments that takes the value one if at least ℓ out of its k arguments take value one – otherwise the function value is zero. The noisy version allows noise at the inputs of the function. The noisy threshold models represent a generalization of two popular models - noisy-or and noisy-and. They constitute an alternative to noisy-or and noisy-and in case they are too rough. The conditional probability tables of the threshold functions appear, for example, in medical applications of Bayesian networks (Visscher et al., 2009; van Gerven et al., 2007; Jurgelenaite et al., 2006; Jurgelenaite and Heskes, 2006).

For CP tensor decompositions of other canonical models see Vomlel (2011) where the tensors of ℓ-out-of-k functions are studied and Savicky and Vomlel (2007), where CP decompositions of

[1] The formal definition of a tensor can be found in the next section.

tensors of several other canonical models (noisy-max, noisy-min, noisy-add, noisy-xor) are described.

The rest of this paper is organized as follows. In Section 2 we introduce the necessary tensor notation, define tensors of the threshold functions, and present their basic properties. Section 3 represents the main original contribution of this paper. We propose an algorithm for the CP decomposition of tensors of the threshold functions based on the upper bound of the symmetric rank of these tensors. In Section 4 we conclude the paper by computational comparisons performed on a generalized version of the QMR-DT network.

2 Preliminaries

Tensor is a mapping[2] $\boldsymbol{A} : \mathbb{I} \rightarrow \mathbb{R}$, where $\mathbb{I} = I_1 \times \ldots \times I_k$, k is a natural number called the order of tensor $\boldsymbol{A}$, and $I_j, j = 1, \ldots, k$ are index sets. Typically, I_j are sets of integers of cardinality n_j. Then we can say that tensor $\boldsymbol{A}$ has dimensions $n_1, \ldots, n_k$. In this paper all index sets will be $I_j = \{0, 1\}, j = 1, \ldots, k$.

Tensor $\boldsymbol{A}$ has rank one if it can be written as an outer product of vectors, i. e.,

$$\boldsymbol{A} = \boldsymbol{a}_1 \otimes \ldots \otimes \boldsymbol{a}_k \ ,$$

where $\boldsymbol{a}_j, j = 1, \ldots, k$ are real valued vectors of length $|I_j|$.

Each tensor can be decomposed as a linear combination of rank-one tensors:

$$\boldsymbol{A} = \sum_{i=1}^{r} b_i \cdot \boldsymbol{a}_{i,1} \otimes \ldots \otimes \boldsymbol{a}_{i,k} \ , \qquad (2)$$

The rank of a tensor $\boldsymbol{A}$, denoted $rank(\boldsymbol{A})$, is the minimal r over all such decompositions. The decomposition of a tensor $\boldsymbol{A}$ to tensors of rank one that sum up to $\boldsymbol{A}$ is called CP tensor decomposition.

A special class of tensors that appear in the problems that motivated our research in this area (Savicky and Vomlel, 2007; Vomlel, 2002) are tensors representing functions. In this paper we will pay special attention to tensors representing the threshold function, i.e. a Boolean function taking value 1 if and only if ℓ of more of its k inputs have value 1.

Definition 1. Tensor $\boldsymbol{T}(\ell, k) : \{0,1\}^k \rightarrow \{0,1\}$ represents an (ℓ, k)-threshold function if it holds for $(i_1, \ldots, i_k) \in \{0,1\}^k$:

$$\begin{aligned} \boldsymbol{T}_{i_1,\ldots,i_k}(\ell, k) &= \delta(i_1 + \ldots + i_k \geq \ell) \\ \delta(i \geq \ell) &= \begin{cases} 1 & \text{if } i \geq \ell \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Example 1.

$$\boldsymbol{T}(2,4) = \begin{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} & \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \\ \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \end{pmatrix} .$$

Tensors representing the threshold function have certain nice properties, e. g., they are symmetric.

Definition 2. Tensor $\boldsymbol{A} : \{0,1\}^k \rightarrow \mathbb{R}$ is symmetric if for $(i_1, \ldots, i_k) \in \{0,1\}^k$ it holds that

$$\boldsymbol{A}_{i_1,\ldots,i_k} = \boldsymbol{A}_{i_{\sigma(1)},\ldots,i_{\sigma(k)}} \ ,$$

for any permutation σ of $\{1, \ldots, k\}$.

For symmetric tensors it is possible to define a symmetric rank as follows.

Definition 3. The symmetric rank $srank(\boldsymbol{A})$ of a tensor $\boldsymbol{A}$ is the minimum number of symmetric rank-one tensors such that their linear combination equals $\boldsymbol{A}$

$$\begin{aligned} \boldsymbol{A} &= \sum_{i=1}^{r} b_i \cdot \underbrace{\boldsymbol{a}_i \otimes \ldots \otimes \boldsymbol{a}_i}_{k \text{ copies}} \\ &= \sum_{i=1}^{r} b_i \cdot \boldsymbol{a}_i^{\otimes k} \ , \qquad (3) \end{aligned}$$

where we adopt the notation of (Comon et al., 2008).

Remark. It is not known whether it holds for symmetric tensors $\boldsymbol{A}$ that $rank(\boldsymbol{A}) = srank(\boldsymbol{A})$.

[2] Often tensor values are from $\mathbb{C}$. However in this paper we will restrict them to be from $\mathbb{R}$.

Each symmetric tensor $\boldsymbol{A} : \{0,1\}^k \rightarrow \mathbb{R}$ of rank one can be written as

$$\boldsymbol{A} = \begin{cases} (0,a)^{\otimes k} & \text{if } \boldsymbol{A}_{0,\dots,0} = 0 \\ b \cdot (1,a)^{\otimes k} & \text{otherwise,} \end{cases} \quad (4)$$

where $a, b \in \mathbb{R}$.

In the following lemma we treat the border cases with symmetric rank one.

Lemma 1. The symmetric rank of tensors $\boldsymbol{T}(\ell,k)$ representing the respective (ℓ,k)-threshold function for $\ell \in \{0,k\}$ is one.

Proof.

$$\begin{aligned} \boldsymbol{T}(k,k) &= (0,1)^{\otimes k} \\ \boldsymbol{T}(0,k) &= (1,1)^{\otimes k} . \end{aligned}$$

□

In the next lemma another we present a case with a low symmetric rank equal to two.

Lemma 2. The symmetric rank of tensors $\boldsymbol{T}(1,k)$ representing the respective $(1,k)$-threshold function is two.

Proof.

$$\boldsymbol{T}(1,k) = (1,1)^{\otimes k} - (1,0)^{\otimes k}$$

and there does not exist any vector $\boldsymbol{a}$ such that

$$\boldsymbol{T}(1,k) = \boldsymbol{a}^{\otimes k} .$$

To see this note that $\boldsymbol{T}(1,k)_{0,\dots,0} = 0$. This requires $\boldsymbol{a} = (0,a), a \in \mathbb{R}$. But tensor $(0,a)^{\otimes k}$ has all its values but the one at $(1,\dots,1)$ equal to zero and thus cannot be equal to $\boldsymbol{T}(1,k)$. □

3 A CP tensor decomposition

In this section we will construct a CP decomposition of the tensor $\boldsymbol{T}(\ell,k)$ representing an (ℓ,k)-threshold function. In the previous section we have already treated the border cases for $\ell \in \{0,1,k\}$.

If we restrict the rank-one tensors in the decomposition as defined by formula (3) to tensors with a non-zero value at the $(0,\dots,0)$ position then we can rewrite formula (3) as

$$\boldsymbol{A} = \sum_{i=1}^{r} b_i \cdot (1,a_i)^{\otimes k} , \quad (5)$$

where $a_i, b_i \in \mathbb{R}, i = 1,\dots,k$. Note that, generally, this restricted decomposition need not be a minimal decomposition of $\boldsymbol{A}$ (with respect to r). On the other hand, it can be used to provide an upper bound on the symmetric rank of a tensor $\boldsymbol{A}$.

It follows from formula (5) that a sufficient condition for a symmetric decomposition of a tensor of an (ℓ,k)-threshold function to have rank r is the following system of equations:

$$\begin{aligned} a_1^0 \cdot b_1 + \ldots + a_r^0 \cdot b_r &= \delta(0 \geq \ell) \\ a_1^1 \cdot b_1 + \ldots + a_r^1 \cdot b_r &= \delta(1 \geq \ell) \\ &\vdots \\ a_1^k \cdot b_1 + \ldots + a_r^k \cdot b_r &= \delta(k \geq \ell) , \end{aligned} \quad (6)$$

which is a system of $k+1$ equations with $2r$ variables.

Let $m \in \mathbb{N}^+$, $\boldsymbol{a} = (a_1,\dots,a_m), a_j \in \mathbb{R}, j \in \{1,\dots,m\}$, and $V(\boldsymbol{a})$ be a Vandermonde matrix $m \times m$ defined as

$$V(\boldsymbol{a}) = \begin{pmatrix} 1 & \ldots & 1 \\ a_1 & \ldots & a_m \\ \\ a_1^{m-1} & \ldots & a_m^{m-1} \end{pmatrix} .$$

Further let $\boldsymbol{e}(\ell)$ be the vector of length k having its values $\boldsymbol{e}_i(\ell)$ at positions $i = 1,\dots,\ell$ equal to zero and for $i = \ell+1,\dots,k$ equal to one i.e.,

$$\boldsymbol{e}(\ell) = (\delta(0 \geq \ell),\dots,\delta(k-1 \geq \ell))^T . \quad (7)$$

A sufficient condition for the solution of the system (6) for $r = k$ is to solve the following system of $3k$ equations. Let $\boldsymbol{a} = (a_1,\dots,a_k)$, $\boldsymbol{b} = (b_1,\dots,b_k)$, $\boldsymbol{c} = (c_1,\dots,c_k)$, and $\boldsymbol{a} * \boldsymbol{b}$ denote elementwise multiplication (Hadamard product) of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. All but the last equation of system (6) correspond to system (8), all but the first equation of system (6) correspond to systems (9) and (10):

$$\begin{aligned} V(\boldsymbol{a}) \cdot \boldsymbol{b} &= \boldsymbol{e}(\ell) && (8) \\ V(\boldsymbol{a}) \cdot \boldsymbol{c} &= \boldsymbol{e}(\ell-1) && (9) \\ \boldsymbol{c} &= \boldsymbol{a} * \boldsymbol{b} . && (10) \end{aligned}$$

If values of $a_i, i = 1,\dots,k$ are distinct then

$$\begin{aligned} \boldsymbol{b} &= V(\boldsymbol{a})^{-1} \cdot \boldsymbol{e}(\ell) && (11) \\ \boldsymbol{c} &= V(\boldsymbol{a})^{-1} \cdot \boldsymbol{e}(\ell-1) . && (12) \end{aligned}$$

Note that the explicit formula for the inverse of the Vandermonde matrix is known. Let

$$p(x) = \prod_{j=1}^{k}(x - a_j) \text{ and} \quad (13)$$

$$p_i(x) = \prod_{j \neq i}(x - a_j) \quad (14)$$

be polynomials in variable x. Further let $p[j], j = 1, \ldots, k+1$ denote the coefficient in the term with x^{j-1} of polynomial $p(x)$ so that

$$p(x) = \sum_{j=1}^{k+1} p[j] \cdot x^{j-1} \ .$$

Finally, let $p|_{x=a_i}$ denote the substitution of a_i in place of x in polynomial $p(x)$. We can write equations (11) and (12) for $i = 1, \ldots, k$ as

$$b_i = \frac{\sum_{j=\ell+1}^{k} p_i[j]}{p_i|_{x=a_i}} \quad (15)$$

$$c_i = \frac{\sum_{j=\ell}^{k} p_i[j]}{p_i|_{x=a_i}} \ . \quad (16)$$

Substituting (15) and (16) into (10) we get for $i = 1, \ldots, k$:

$$a_i = \frac{c_i}{b_i} = \frac{\sum_{j=\ell}^{k} p_i[j]}{\sum_{j=\ell+1}^{k} p_i[j]} \ , \quad (17)$$

where the right hand side depends on $a_j, j \neq i$ only. Due to symmetry, if one equation of (17) holds then all equations hold and the system can be reduced to one equation only, e.g., to

$$a_k = \frac{\sum_{j=\ell}^{k} p_k[j]}{\sum_{j=\ell+1}^{k} p_k[j]} \ , \quad (18)$$

where the right hand side depends on $a_1, \ldots, a_{k-1}$. If we set $a_1, \ldots, a_k$ so that they are pairwise distinct, a_k satisfies (18), and from (11) we compute $b_1, \ldots, b_k$, which is under the above constraints always possible, then we have a solution of system (6). As a consequence we have following lemma.

Lemma 3. *The symmetric rank of a tensor $\boldsymbol{T}(\ell, k)$ representing (ℓ, k)-threshold function for $\ell = 2, \ldots, k-1$ is at most k.*

Table 1: An algorithm for the CP tensor decomposition of tensors $\boldsymbol{T}(\ell, k)$ defined by formula (5) for $r = k$.

Input: k, ℓ
Output: $(a_1, \ldots, a_k)$ and $(b_1, \ldots, b_k)$

Find $\boldsymbol{a}_0 = (a_1, \ldots, a_k)$ at random such that:
 $a_i \neq a_j, i \neq j$
 $\sum_{j=\ell+1}^{k} p_k[j] \neq 0$
 $a_k = \frac{\sum_{j=\ell}^{k} p_k[j]}{\sum_{j=\ell+1}^{k} p_k[j]}$
Find $\boldsymbol{a} = (a_1, \ldots, a_k)$ minimizing $\kappa(V(\boldsymbol{a}))$
 subject to $a_k = \frac{\sum_{j=\ell}^{k} p_k[j]}{\sum_{j=\ell+1}^{k} p_k[j]}$
 starting at initial point $\boldsymbol{a}_0$
For $i \in \{1, \ldots, k\}$ compute:
 $b_i \leftarrow \frac{\sum_{j=\ell+1}^{k} p_i[j]}{p_i|_{x=a_i}}$

Remark. Lemma 3 slightly lowers the general upper bound for symmetric tensors from (Comon et al., 2008, Section 4.1) for tensors of (ℓ, k)-threshold function. With complex numbers being allowed in the decomposition their upper bound for symmetric tensors with all dimensions being two is $k+1$.

The main contribution of this paper is the construction of a CP tensor decomposition of any tensor $\boldsymbol{T}(\ell, k)$ representing the respective (ℓ, k)-threshold function for $\ell = 2, \ldots, k-1$ to the sum of k symmetric tensors of rank-one. In Table 1 we summarize the algorithm for the CP decomposition defined by formula (5) for $r = k$.

Initial values of $a_1, \ldots, a_{k-1}$ are drawn at random from Gaussian distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$. If $(a_1, \ldots, a_{k-1})$ are such that either $(a_1, \ldots, a_k)$ are not pairwise distinct or $\sum_{j=\ell+1}^{k} p_k[j] = 0$ then a new configuration is generated. In our experiments, we never had to generate new values of $(a_1, \ldots, a_{k-1})$. But since there is no guarantee that it cannot happened therefore the loop is needed. From the computational point of view it is advantageous to have $\boldsymbol{a}$ defined so that the condition number $\kappa(V(\boldsymbol{a}))$ of matrix $V(\boldsymbol{a})$ is minimized.

Since the values of $\boldsymbol{a}$ and $\boldsymbol{b}$ can be precomputed one can spend some time with an optimization algorithm searching for values $\boldsymbol{a}$ minimizing $\kappa(V(\boldsymbol{a}))$. We experimented with Nelder-Mead method restarted from different starting points. Note that if $a_1, \ldots, a_k$ are distinct then for $i = 1, \ldots, k$ $p_i|_{x=a_i} \neq 0$ and b_i are well-defined. Around $k = 25$ Vandermonde matrices get easily badly conditioned. However, in the numerical experiments we have not observed any significant errors[3] even in computations performed with CPTs for (ℓ, k)-threshold function with $k = 25, 26, 27$[4].

In the next example we will show that, generally, Lemma 3 does not provide a tight upper bound even for $\ell \in 2, \ldots, k-1$.

Example 2.

$$\begin{aligned}\boldsymbol{T}(2,5) &= (1, -\frac{1}{2})^{\otimes 5} - 3 \cdot (1, \frac{1}{2})^{\otimes 5} \\ &\quad + (1 - \frac{\sqrt{3}}{2}) \cdot (1, -\frac{\sqrt{3}}{2})^{\otimes 5} \\ &\quad + (1 + \frac{\sqrt{3}}{2}) \cdot (1, \frac{\sqrt{3}}{2})^{\otimes 5}\end{aligned}$$

which implies that for $k = 5$

$$srank(\boldsymbol{T}(2,5)) \leq 4 = k - 1 < k .$$

On the other hand there exist tensors for which Lemma 3 provides a tight upper bound.

Lemma 4. *The symmetric rank of tensors $\boldsymbol{T}(k-1,k)$ representing $(k-1,k)$-threshold function is k.*

Proof. We will prove that $srank(\boldsymbol{T}(k-1,k)) > k-1$ by contradiction. Assume $r = k-1$. In this case the system (6) corresponds to

$$\begin{aligned} a_1^0 \cdot b_1 + \ldots + a_{k-1}^0 \cdot b_{k-1} &= 0 \\ &\vdots \\ a_1^{k-2} \cdot b_1 + \ldots + a_{k-1}^{k-2} \cdot b_{k-1} &= 0 \\ a_1^{k-1} \cdot b_1 + \ldots + a_{k-1}^{k-1} \cdot b_{k-1} &= 1 \\ a_1^{k} \cdot b_1 + \ldots + a_{k-1}^{k} \cdot b_{k-1} &= 1 . \quad (19)\end{aligned}$$

Let $\boldsymbol{a} = (a_1, \ldots, a_{k-1})$. We can write the first $k-1$ equalities of (19) using a Vandermonde matrix as

$$\boldsymbol{V}(\boldsymbol{a}) \cdot \boldsymbol{b} = \boldsymbol{0} ,$$

which cannot hold unless the Vandermonde matrix is singular, which is not the case. Note that even if rank-one tensor of the form $(0,a)^{\otimes k}$ (which we excluded from formula (5)) were allowed to be part of the decomposition then it would not add any value to the left hand side of the first k equations of system (19) since all its values except $\boldsymbol{A}_{1,\ldots,1}$ are zero. Therefore $srank(\boldsymbol{T}(k-1,k)) > k-1$. This together with Lemma 3 implies $srank(\boldsymbol{T}(k-1,k)) = k$. □

Due to Theorem 6 from (Savicky and Vomlel, 2007) the results derived for deterministic part of canonical models can be easily combined with the probabilistic part representing the noise. See Section 3.5 in (Savicky and Vomlel, 2007) for details. A consequence is that the upper bound of the symmetric rank (Lemma 3) for tensors $\boldsymbol{T}(\ell,k)$ representing (ℓ,k)-threshold functions is also an upper bound on the rank of their noisy counterparts.

In this section all results were derived for $P(y|X_1, \ldots, X_k)$ with $y = 1$. Corresponding results for $y = 0$ can be achieved after flipping the values on each coordinate in tensors and in vectors that generate the tensors of a CP decomposition.

4 Experiments

We performed experiments with the Quick Medical Reference - Decision Theoretic version (QMR-DT) derived from the original QMR (Miller et al., 1986) by (Shwe et al., 1991). The Bayesian network of QMR-DT contains 570 diseases (variables X_i) and 4075 observations (variables Y_j). The conditional probability tables for observations given related diseases are noisy-or models. We generalized the QMR-DT by replacing noisy-or with noisy-threshold models. The experiments were performed with subnetworks of QMR-DT. In the first test, we randomly selected 14 observations. We included all their parents in the generated subnetwork.

[3] We compared one dimensional marginal probabilities computed from the full CPTs with marginal probabilities computed in models after CP decomposition.

[4] Note that computational complexity with the standard method is exponential with respect to k. Therefore this method has difficulties with getting probabilities for higher k.

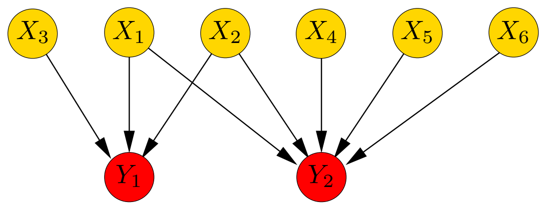

Figure 3: A part of the QMR-DT network.

In Figure 3 we give an example of a subnetwork of the QMR-DT network generated by two observations and their parents. In this way we generated forty different networks.

For each network we compared computational complexity of the junction tree method (Jensen et al., 1990) applied to models after two different transformations:

- moralization and triangulation (the standard method)
- the tensor CP decomposition applied to CPTs with number of parents higher than four[5] followed by triangulation.

We measured the computational complexity by the total table size of models computed by Hugin optimal triangulation[6].

In the second test, we repeated the same process with 28 observations instead of 14. In both tests we have got together eighty bipartite graphs with their size in the range from 46 to 585 nodes. The results of experiments are summarized in Figure 4. Note the logarithmic scales. If the total table size is larger than 2^{64} the models are intractable in Hugin. Numerical experiments reveal that we can get a gain in the order of several magnitudes and many intractable models became tractable.

A different approach exploiting a local structure in CPTs are arithmetic circuits (ACs) (Darwiche, 2003). In (Vomlel and Savicky, 2008) the CP tensor decomposition was used to preprocess Bayesian networks containing noisy-or models. The ACs of the preprocessed networks were compared with ACs created by Ace[7] from networks after parent divorcing. The CP tensor decomposition decreased the size of ACs for a majority of tested networks (about 88%). We conjecture we would get similar results for experiments reported in this section. However, we did not perform these experiments – they should be a topic of a our future research along with comparisons with other methods exploiting local structure of CPTs.

[5]For CPTs with less than four parents we used moralization instead.

[6]Hugin Expert A/S, `http://www.hugin.com`

[7]Ace, A Bayesian Network Compiler, 2008, `http://reasoning.cs.ucla.edu/ace/`

5 Conclusions

We proposed a CP decomposition of tensors corresponding to threshold functions. We applied this decomposition to probabilistic inference in Bayesian networks containing conditional probability tables representing noisy threshold functions. We performed computational experiments with a generalized version of QMR-DT where the noisy-or models were replaced by noisy threshold models. The CP tensor decomposition led to a computational gain in the order of several magnitudes and made many intractable models manageable.

Acknowledgments

We would like to thank Frank Jensen for the Hugin optimal triangulation method and Gregory F. Cooper from University of Pittsburgh for the structural part of QMR-DT model.

References

J. D. Carroll and J. J. Chang. 1970. Analysis of individual differences in multidimensional scaling via an n-way generalization of Eckart-Young decomposition. *Psychometrika*, 35:283–319.

P. Comon, G. Golub, Lek-Heng Lim, and B. Mourrain. 2008. Symmetric tensors and symmetric tensor rank. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1254–1279.

A. Darwiche. 2003. A differential approach to inference in Bayesian networks. *Journal of the ACM*, 50:280–305.

F. J. Díez and M. J. Druzdzel. 2006. Canonical probabilistic models for knowledge engineering.

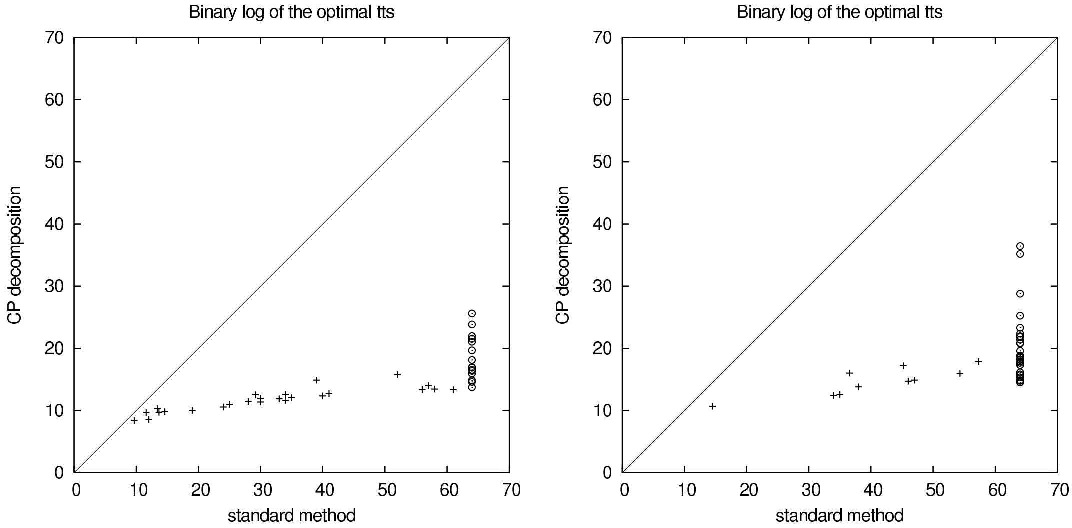

Figure 4: Comparison of total table size for QMR-DT subnetworks for the standard method and after the CP tensor decomposition. The left hand side graph is for networks after 14 observations, the right side graphs after 28 observations. The circle points positioned at the value of $tts = 2^{64}$ of the standard method correspond to networks where the standard method failed since $tts > 2^{64}$.

Technical Report CISIAD-06-01, UNED, Madrid, Spain.

R. A. Harshman. 1970. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-mode factor analysis. *UCLA Working Papers in Phonetics*, 16:1–84.

F. V. Jensen and T. D. Nielsen. 2007. *Bayesian Networks and Decision Graphs, 2nd ed.* Springer.

F. V. Jensen, S. L. Lauritzen, and K. G. Olesen. 1990. Bayesian updating in recursive graphical models by local computation. *Computational Statistics Quarterly*, 4:269–282.

R. Jurgelenaite and T. Heskes. 2006. EM algorithm for symmetric causal independence models. In *ECML'06*, volume 4212 of *Lecture Notes in Computer Science*, pages 234–245. Springer.

R. Jurgelenaite, P. Lucas, and T. Heskes. 2006. Exploring the noisy threshold function in designing Bayesian networks. In M. Bramer, F. Coenen, and T. Allen, editors, *Research and Development in Intelligent Systems XXII*, pages 133–146. Springer.

R. A. Miller, F. E. Fasarie, and J. D. Myers. 1986. Quick medical reference (QMR) for diagnostic assistance. *Medical Computing*, 3:34–48.

P. Savicky and J. Vomlel. 2007. Exploiting tensor rank-one decomposition in probabilistic inference. *Kybernetika*, 43(5):747–764.

M. Shwe, B. Middleton, D. Heckerman, M. Henrion, E. Horvitz, H. Lehmann, and G. Cooper. 1991. Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base. I. The probabilistic model and inference algorithms. *Methods of Information in Medicine*, 30:241–255.

M. A. J. van Gerven, R. Jurgelenaite, B. G. Taal, T. Heskes, and P. J. F. Lucas. 2007. Predicting carcinoid heart disease with the noisy-threshold classifier. *Artificial Intelligence in Medicine*, 40(1):45–55.

S. Visscher, P. J. F. Lucas, C. A. M. Schurink, and M. J. M. Bonten. 2009. Modelling treatment effects in a clinical bayesian network using boolean threshold functions. *Artificial Intelligence in Medicine*, 46(3):251–266.

J. Vomlel and P. Savicky. 2008. Arithmetic circuits of the noisy-or models. In Manfred Jaeger and Thomas D. Nielsen, editors, *Proceedings of PGM'08*, pages 297–304.

J. Vomlel. 2002. Exploiting functional dependence in Bayesian network inference. In *Proceedings of UAI'02*, pages 528–535. Morgan Kaufmann Publishers.

J. Vomlel. 2011. Rank of tensors of ℓ-out-of-k functions: An application in probabilistic inference. *Kybernetika*, 47(3):317–336.