On tensor rank of conditional probability tables in Bayesian networks

Jirka Vomlel and Petr Tichavský

ÚTIA, Academy of Sciences of the Czech Republic, Prague

Tensor (an object of multilinear algebra)

Tensor is a mapping $\mathcal{A} : \mathbb{I} \to \mathbb{R}$, where

- $\mathbb{I} = I_1 \times \ldots \times I_k$,
- k is a natural number called the order of tensor $\mathcal{A}$,
- $I_j, j = 1, \ldots, k$ are index sets (typically, sets of integers),
- cardinalities of $I_j, j = 1, \ldots, k$ are called dimensions of tensor $\mathcal{A}$.

Example

Example of a tensor of order $k = 4$ and dimensions $n_1 = n_2 = n_3 = n_4 = 2$:

$$\mathcal{A} = \begin{pmatrix} \begin{pmatrix} 0\,1 \\ 1\,1 \end{pmatrix} & \begin{pmatrix} 1\,1 \\ 1\,1 \end{pmatrix} \\ \begin{pmatrix} 1\,1 \\ 1\,1 \end{pmatrix} & \begin{pmatrix} 1\,1 \\ 1\,1 \end{pmatrix} \end{pmatrix}$$

Tensor rank

Tensor $\mathcal{A}$ has rank one if it can be written as an outer product of vectors $\boldsymbol{a}_j = (a_{j,i})_{i \in I_j}$, $j = 1, \ldots, k$:

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

with the outer product $\otimes$ being defined as

$$\mathcal{A}_{i_1,\ldots,i_k} = a_{1,i_1} \cdot \ldots \cdot a_{k,i_k} \ ,$$

for all $(i_1, \ldots, i_k) \in I_1 \times \ldots \times I_k$.

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

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

The rank of a tensor $\mathcal{A}$, denoted $rank(\mathcal{A})$, is the minimal r over all such decompositions.

The decomposition of a tensor $\mathcal{A}$ to tensors of rank one that sum up to $\mathcal{A}$ is called CP tensor decomposition.

Example

The tensor $\mathcal{A}$ from the above example can be written as:

$$\begin{aligned} \mathcal{A} = \ & (1,1) \otimes (1,1) \otimes (1,1) \otimes (1,1) \\ & -(1,0) \otimes (1,0) \otimes (1,0) \otimes (1,0) \ . \end{aligned}$$

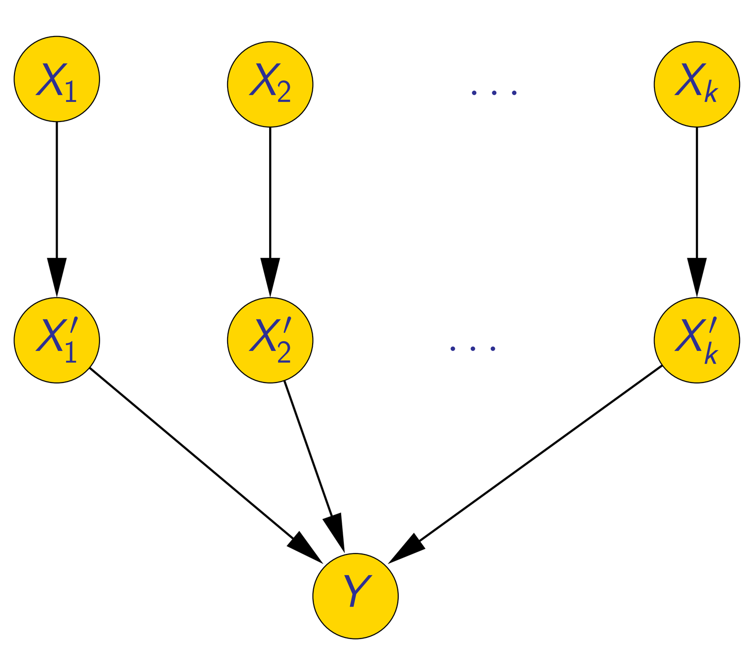
This implies that its rank is at most 2.

Conditional probability tables (CPTs) as tensors

- Let $X_i, i = 1, \ldots, k$ be discrete random variables taking states from $I_j, j = 1, \ldots, k$ and
- let Y be a discrete random variable with an observed state y.

Then each CPT $P(Y = y | X_1, \ldots, X_k)$ can be viewed as a tensor.

Noisy-or: an example of a CPT of a low rank



- Let $X_i, i = 1, \ldots, k$, $X'_i, i = 1, \ldots, k$, and Y are binary variables taking states from set $\{0, 1\}$.
- Y takes value 1 if at least one of its parents take value 1:

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

- For $i = 1, \ldots, k$

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

Example

The CPT of $P(Y = 1 | X'_1 = x'_1, \ldots, X'_k = x'_k)$ corresponds to the tensor $\mathcal{A}$ from the above example.

Probabilistic inference

Compute $P(X_j | e)$ where e is the collected evidence.

Example

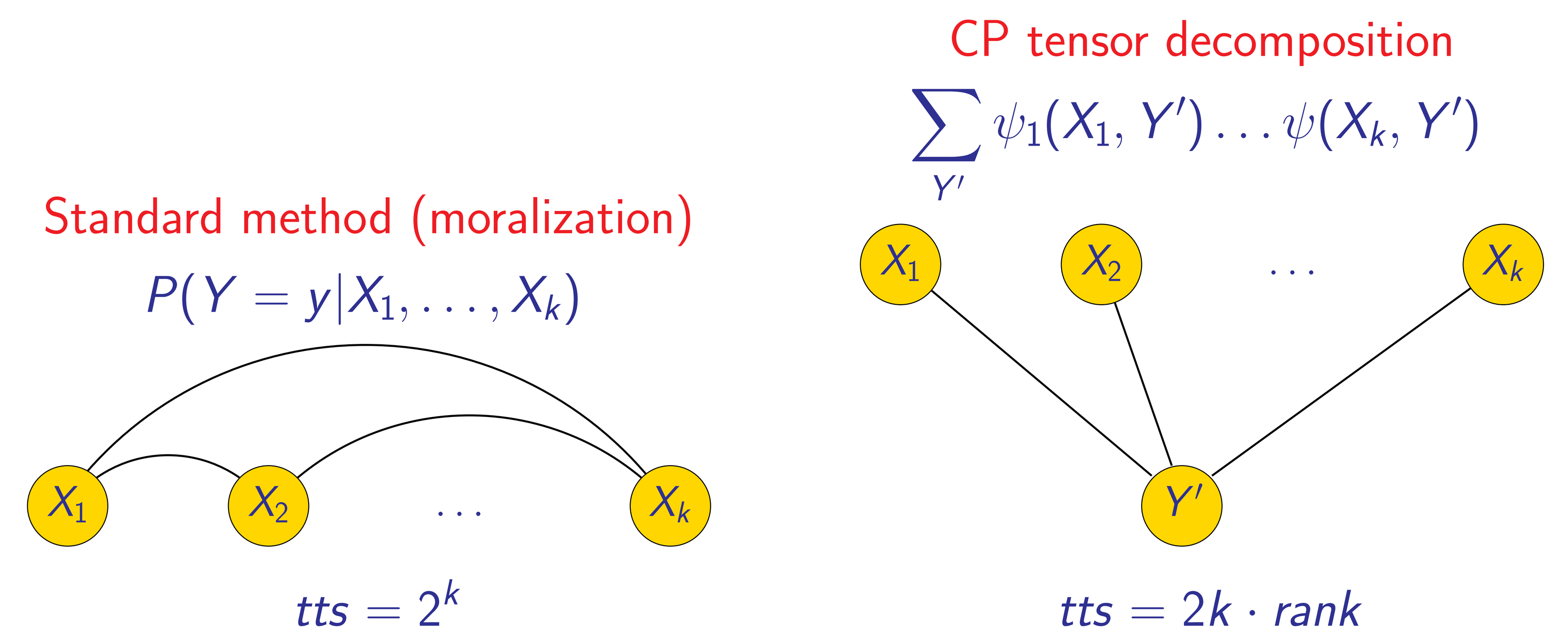
Compute $P(X_j | Y = y)$ for $j = 1, \ldots, 4$.

Low rank means efficient inference

A measure of the complexity of probabilistic inference is the total table size (tts). It is the total size of the tables used in a probabilistic inference method applied to a Bayesian network.

Example

Bayesian network with one noisy-or:

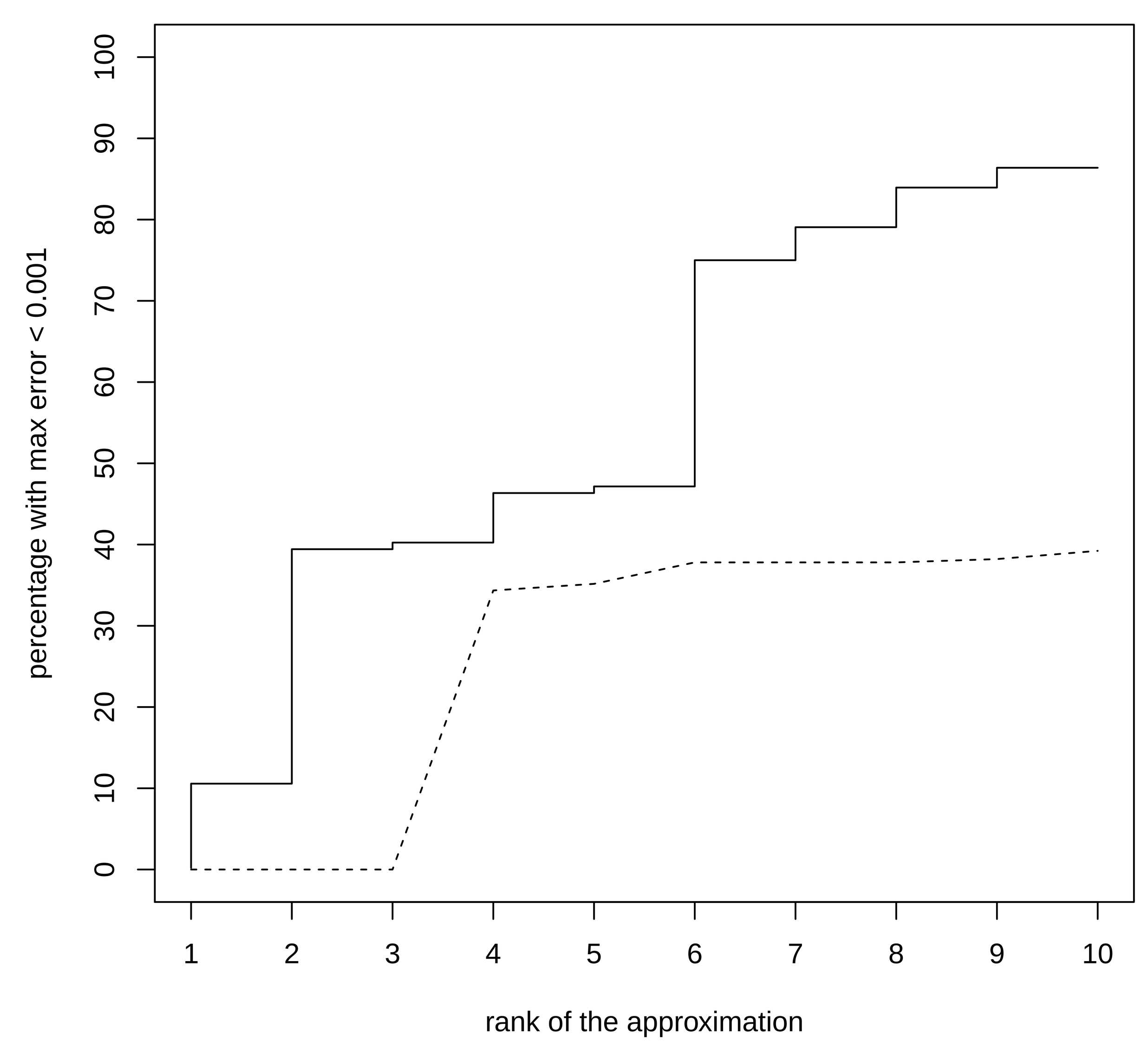


Real data and algorithms used

- 15 Bayesian networks from a Bayesian network repository: *alarm.net*, *hailfinder.net*, *barley2.net*, *pathfinder.net*, *munin1.net*, *munin2.net*, *munin3.net*, *munin4.net*, *mildew.net*, *hepar2.net*, *andes.net*, *win95pts.net*, *water.net*, *link.net*, and *insurance.net*.
- In the experiments we considered all CPTs for variables having at least three parents. It was 492 CPTs in total.
- For the CP tensor decomposition we used the Fast Damped Gauss-Newton (Levenberg-Marquard) algorithm implemented in the Matlab package TENSORBOX.

How common are low rank CPTs in real applications?

Percentage of CPTs that can be approximated with a maximum error smaller than 0.001 as a function of the rank of the approximation.



The full line corresponds to CPTs of real Bayesian networks, dashed line to CPTs of the same dimensions with their entries replaced by random numbers.

Conclusions

- Most of the conditional probability tables can be very well approximated by tables that have lower rank than one would expect from a general table of the same dimensions.
- The low rank approximation should be exploited (1) in model elicitation by using a compact parametrization of the internal structure of a CPT but also (2) during probabilistic inference.
- We conjecture that some low rank approximations of CPTs may actually correspond better to what a domain expert intended to model in the constructed CPT.

E-mail: vomlel@utia.cas.cz WWW: http://www.utia.cas.cz/vomlel