NP-hardness of MAP in Ternary Tree Bayesian Networks

Cassio P. de Campos

Technical Report No. IDSIA-06-13
December 2013

IDSIA / USI-SUPSI
Dalle Molle Institute for Artificial Intelligence
Galleria 2, 6928 Manno, Switzerland

NP-hardness of MAP in Ternary Tree Bayesian Networks

Cassio P. de Campos

December 2013

Abstract

This paper strengthens the NP-hardness result for the (partial) maximum a posteriori (MAP) problem in Bayesian networks with topology of trees (every variable has at most one parent) and variable cardinality at most three. MAP is the problem of querying the most probable state configuration of some (not necessarily all) of the network variables given evidence. It is demonstrated that the problem remains hard even in such simplistic networks.

1 Introduction

A Bayesian network (BN) is a probabilistic graphical model that relies on a structured dependency among random variables to represent a joint probability distribution in a compact and efficient manner [Pearl 1988]. One of the hardest inference problems in BNs is the *maximum a posteriori* (or MAP) problem, where one looks for states of some variables that maximize their joint probability, given some other variables as evidence (there may exist variables that are neither queried nor part of the evidence). The reader is assumed to have read the previous work on which this text is based [de Campos 2011]. The same notation as there will be used here.

Definition 1 *A Bayesian network (BN) $\mathcal{N}$ is a triple $(\mathcal{G}, \mathcal{X}, \mathcal{P})$, where $\mathcal{G} = (\mathbf{V}_\mathcal{G}, \mathbf{E}_\mathcal{G})$ is a directed acyclic graph with nodes $\mathbf{V}_\mathcal{G}$ associated (in a one-to-one mapping) to random variables $\mathcal{X} = \{X_1, \ldots, X_n\}$ over discrete domains $\{\Omega_{X_1}, \ldots, \Omega_{X_n}\}$ and $\mathcal{P}$ is a collection of probability values $p(x_i|\pi_{X_i}) \in \mathcal{Q}$,[1] with $\sum_{x_i \in \Omega_{X_i}} p(x_i|\pi_{X_i}) = 1$, where $x_i \in \Omega_{X_i}$ is a category or state of X_i and $\pi_{X_i} \in \times_{X \in PA_{X_i}} \Omega_X$ a complete instantiation for the parents PA_{X_i} of X_i in $\mathcal{G}$. Furthermore, every variable is conditionally independent of its non-descendant non-parents given its parents.*

The joint probability distribution represented by a BN $(\mathcal{G}, \mathcal{X}, \mathcal{P})$ is obtained by $p(\mathbf{x}) = \prod_i p(x_i|\pi_{X_i})$, where $\mathbf{x} \in \Omega_\mathcal{X}$ and all states x_i, π_{X_i} (for every i) agree with $\mathbf{x}$. Now we introduce some notation. Singletons $\{X_i\}$ and $\{x_i\}$ are respectively denoted as X_i and x_i. Nodes of the graph and their associated random variables are used interchangeably. Uppercase letters are used for random variables and lowercase letters for their corresponding states. Bold letters are employed for vectors/sets. The input size of a BN, called simply b here, is given by the length of the bit string to specify all the local conditional probability distributions and the structure to describe the graph.

The MAP problem is to find an instantiation $\mathbf{x}_{\text{opt}} \in \Omega_{\mathcal{X}^{\text{map}}}$, with $\mathcal{X}^{\text{map}} \subseteq \mathcal{X} \setminus \mathbf{E}$, such that its probability is maximized:

$$\mathbf{x}_{\text{opt}} = \operatorname*{argmax}_{\mathbf{x} \in \Omega_\mathbf{X}} p(\mathbf{x}|\mathbf{e}) = \operatorname*{argmax}_{\mathbf{x} \in \Omega_\mathbf{X}} p(\mathbf{x}, \mathbf{e}), \tag{1}$$

because $p(\mathbf{e})$ (assumed to be non-zero[2]) is a constant with respect to the maximization.

[1] $\mathcal{Q}$ denotes the non-negative rational numbers defined by fractions of integers.

[2] If $p(\mathbf{e})$ is zero, then so is $p(\mathbf{x}, \mathbf{e})$, and the problem vanishes, as any $\mathbf{x}$ will be a maximizer for the problem.

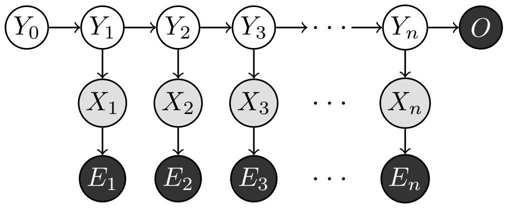

Figure 1: Tree used to prove Theorem 3.

Table 1: Probability values used in the proof of Theorem 3.

$p(Y_i \vert Y_{i-1})$	y_{i-1}^T	y_{i-1}^F	y_{i-1}^*
y_i^T	$\frac{1+t_i}{2}$	0	0
y_i^F	$\frac{1-t_i}{2}$	0	0
y_i^*	0	1	1

$p(E_i \vert X_i)$	x_i^T	x_i^F
e_i^T	t_i	1
e_i^F	$1-t_i$	0

$p(X_i \vert Y_i)$	y_i^T	y_i^F	y_i^*
x_i^T	$\frac{t_i}{1+t_i}$	1	$\frac{1}{2}$
x_i^F	$\frac{1}{1+t_i}$	0	$\frac{1}{2}$

Definition 2 *Given a BN $\mathcal{N} = (\mathcal{G}, \mathcal{X}, \mathcal{P})$ such that the maximum cardinality of any variable is at most z and the minimum treewidth is at most w, $\mathbf{X} \subseteq \mathcal{X} \setminus \mathbf{E}$, a rational r and an instantiation $\mathbf{e} \in \Omega_{\mathbf{E}}$, Decision-MAP-$z$-$w$ is the problem of deciding if there is $\mathbf{x} \in \Omega_{\mathbf{X}}$ such that $p(\mathbf{x}, \mathbf{e}) > r$.*

2 Hardness of MAP-3-1

MAP has been proven to be hard even in trees with cardinality five [de Campos 2011]. The next theorem strengthens that result.

Theorem 3 *Decision-MAP-z-w is NP-hard even if $z = 3$, $w = 1$.*

Proof We use a reduction from the partition problem [Garey and Johnson 1979], which is the problem of deciding whether a list of positive integer numbers $s_1, \ldots, s_n$ can be partitioned in a way such that $\sum_{i \in I} s_i = \sum_{i \notin I} s_i$, where $I \subseteq N = \{1, \ldots, n\}$ (the notation $i \notin I$ means that $i \in N \setminus I$). We say that the partition problem is a *yes-instance* if there is such a I, otherwise we call it a *no-instance*. Instead of working with the partition problem in that form, we define $S := \sum_i s_i/2$ and $v_i := s_i/S$, $i = 1, \ldots, n$, and work with the partition problem using v_i instead of s_i (note that S is an integer, otherwise it would be trivially a no-instance, and that the hardness of the problem remains the same, as there is a polynomial-time computable bijection between solutions and decisions of the two variants).

We build a tree over variables $X_1, \ldots, X_n, Y_0, \ldots, Y_n, O, E_1, \ldots, E_n$ with graph as in Figure 1. The root node is associated to the ternary variable Y_0 taking values in $\{y_0^T, y_0^F, y_0^*\}$ such that $p(y_0^T) = 1$. For variable O, we have $p(o^T|y_n^F) = p(o^T|y_n^*) = 1$ and $p(o^T|y_n^T) = 0$. The remaining variables are defined in Table 1.

Consider the computation of MAP where the variables of interest are $\mathbf{X} = \{X_1, \ldots, X_n\}$ (the gray ones in Figure 1) and evidence $(\mathbf{e}^T, o^T) = \{\forall i : E_i = e_i^T\} \cup \{O = o^T\}$ (dark nodes in the figure). It follows that

$$p(\mathbf{x}, \mathbf{e}^T, o^T) = \sum_{y_n} p(o^T|y_n)p(y_n, \mathbf{x}, \mathbf{e}^T) = p(y_n^F, \mathbf{x}, \mathbf{e}^T) + p(y_n^*, \mathbf{x}, \mathbf{e}^T) =$$

$$= p(\mathbf{x}, \mathbf{e}^T) - p(y_n^T, \mathbf{x}, \mathbf{e}^T) = p(\mathbf{e}^T|\mathbf{x}) \left(p(\mathbf{x}) - p(y_n^T, \mathbf{x})\right).$$

By calculations with this specification of the network, one obtains $p(\mathbf{x}) = 2^{-n}$ (no matter the actual states of $\mathbf{x}$), $p(y_n^T, \mathbf{x}) = 2^{-n}p(\mathbf{e}^T|\mathbf{x}) = 2^{-n}\prod_{i \in I} t_i$, where $I \subseteq N$ is the set of indices of the elements such that X_i is at the state x_i^T. Denote $\mathbf{t} = \prod_{i \in I} t_i$. Then

$$p(\mathbf{x}, \mathbf{e}^T, o^T) = 2^{-n}\mathbf{t}(1 - \mathbf{t}).$$

This is a concave quadratic function on $0 \le \mathbf{t} \le 1$ with maximum at 2^{-1} such that $\mathbf{t}(1-\mathbf{t})$ monotonically increases when $\mathbf{t}$ approaches one half (from both sides). If we could set t_i to be exactly 2^{-v_i}, then $\frac{1}{2^n}\mathbf{t}(1-\mathbf{t}) = \frac{1}{2^n}2^{-\sum_{i \in I} v_i}(1 - 2^{-\sum_{i \in I} v_i})$, which achieves the maximum of $\frac{1}{2^n}2^{-1}(1 - 2^{-1})$ if and only if $\sum_{i \in I} v_i = 1$, that is, if and only if there is an even partition.

It remains to show that we can specify values t_i using only polynomially many bits in b such that they are very close to 2^{-v_i}, and hence yes-instances of partition are separated from no-instances. For that, one just needs to follow the results in [Maua et al. 2013, de Campos et al. 2013] regarding errors introduced by using rationals in place of real numbers, or the very same approach as in Theorem 3 of [de Campos 2011], which we copy here for completeness.

Compute each t_i to be equal to 2^{-v_i} with $4b + 3$ bits of precision and by rounding it up (if necessary), that is, $t_i = 2^{-v_i} + \text{error}_i$, where $0 \le \text{error}_i < 2^{-(4b+3)}$. Clearly t_i can be computed in polynomial time and space in b (this ensures that the specification of the Bayesian network, which requires rational numbers, is polynomial in b). Note that $2^{-v_i} \le t_i \le 2^{-v_i} + \text{error}_i < 2^{-v_i} + 2^{-(4b+3)} \le 2^{-v_i + 2^{-4b}}$ (in short, this holds because 2^{-4b} in the exponent makes the value grow faster than the linear addition of $2^{-(4b+3)}$).

If I is not an even partition, then we know that one of the two conditions hold: (i) $\sum_{i \in I} s_i \le S - 1 \Rightarrow \sum_{i \in I} v_i \le 1 - \frac{1}{S}$, or (ii) $\sum_{i \in I} s_i \ge S + 1 \Rightarrow \sum_{i \in I} v_i \ge 1 + \frac{1}{S}$, because the original numbers s_i are integers. Consider these two cases.

If $\sum_{i \in I} s_i \ge S + 1$, then $\mathbf{t} < \prod_{i \in I} 2^{-v_i + 2^{-4b}}$ equals to

$$2^{\sum_{i \in I}(-v_i + 2^{-4b})} \le 2^{\frac{n}{2^{4b}} - (1 + \frac{1}{S})} \le 2^{-1 - (\frac{1}{2^b} - \frac{1}{2^{3b}})} = l,$$

by using $S \le 2^b$ and $n \le b < 2^b$. On the other hand, if $\sum_{i \in I} s_i \le S - 1$, then $\mathbf{t} \ge \prod_{i \in I} 2^{-v_i}$ equals to

$$2^{-\sum_{i \in I} v_i} \ge 2^{-(1 - \frac{1}{S})} = 2^{-1 + \frac{1}{S}} \ge 2^{-1 + \frac{1}{2^b}} = u.$$

Now suppose I' is an even partition. Then we know that the corresponding $\mathbf{t}'$ satisfies $2^{-1} \le \mathbf{t}'$ and

$$\mathbf{t}' < \prod_{i \in I'} 2^{-v_i + 2^{-4b}} = 2^{\sum_{i \in I'}(-v_i + 2^{-4b})} \le 2^{-1 + \frac{1}{2^{3b}}} = a.$$

To complete the proof, we show that the distance between $\mathbf{t}'$ and 2^{-1} is always less than the distance between $\mathbf{t}$ and 2^{-1} of a non-even partition plus a gap, that is,

$$\begin{aligned} |\mathbf{t}' - 2^{-1}| + 2^{-(3b+2)} &\le a - 2^{-1} + 2^{-(3b+2)} \\ &< \min\{u - 2^{-1}, 2^{-1} - l\} \le |\mathbf{t} - 2^{-1}|, \end{aligned} \tag{2}$$

which can be proved by analyzing the two elements of the min. The first term holds because

$$a+2^{-(3b+2)} - 2^{-1} < a \cdot 2^{\frac{1}{2^{2b}}} - 2^{-1}$$
$$= 2^{-1+\frac{2^{-b}+2^{-2b}}{2^b}} - 2^{-1} < 2^{-1+\frac{1}{2^b}} - 2^{-1} = u - 2^{-1}.$$

The second comes from the fact that the function $h(b) = a + l + 2^{-(3b+2)} = 2^{-1+\frac{1}{2^{3b}}} + 2^{-1-(\frac{1}{2^b}-\frac{1}{2^{3b}})} + 2^{-(3b+2)}$ is less than 1 for $b = 1, 2$ (by inspection), it is a monotonic increasing function for $b \geq 2$ (the derivative is always positive), and it has $\lim_{b\to\infty} h(b) = 1$. Hence, we conclude that $h(b) < 1$, which implies

$$a + l + 2^{-(3b+2)} < 1 \iff a - 2^{-1} + 2^{-(3b+2)} < 2^{-1} - l.$$

This concludes that there is a gap of at least $2^{-(3b+2)}$ between the worst value of $\mathbf{t}'$ (relative to an even partition) and the best value of $\mathbf{t}$ (relative to a non-even partition), which will be used next to specify the threshold of the MAP problem:

$$\max_{\mathbf{x}} p(\mathbf{x}, \mathbf{e}^T, o^T) > r = c \cdot \frac{1}{2^n}, \tag{3}$$

where c is defined as $a' \cdot (1 - a')$, with a' equals a evaluated up to $3b + 2$ bits and rounded up, which implies that $2^{-1} < a \leq a' < a + 2^{-(3b+2)}$. By Eq. (2), a' is closer to one half than any $\mathbf{t}$ of a non-even partition, so the value c is certainly greater than any value that would be obtained by a non-even partition. On the other hand, a' is farther from 2^{-1} than a, so we can conclude that c separates even and non-even partitions, that is, $\mathbf{t} \cdot (1-\mathbf{t}) < c \leq a \cdot (1-a) < \mathbf{t}' \cdot (1-\mathbf{t}')$ for any $\mathbf{t}$ corresponding to a non-even partition and any $\mathbf{t}'$ of an even partition. Thus, a solution of the MAP problem obtains $p(\mathbf{x}, \mathbf{e}^T, o^T) > r$ if and only there is an even partition.[3] □

3 Conclusion

This short paper strengthens the hardness results for the MAP problem in Bayesian networks, proving that it is NP-hard even for simple trees with binary and ternary variables. It remains open whether MAP is NP-hard in binary trees.

References

[de Campos 2011] C. P. de Campos. New Complexity Results for MAP in Bayesian Networks. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2100–2106, 2011. AAAI Press.

[de Campos et al. 2013] C. P. de Campos and F. G. Cozman. Complexity of Inferences in Polytree-shaped Semi-Qualitative Probabilistic Networks. In *AAAI Conference on Artificial Intelligence* pages 217–223, 2013.

[Garey and Johnson 1979] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1979.

[Maua et al. 2013] D. D. Maua, C. P. de Campos, A. Benavoli and A. Antonucci. On the Complexity of Strong and Epistemic Credal Networks. In *29th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 391–400, 2013.

[Pearl 1988] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[3]The conditional version of Decision-MAP could be used in the reduction too by including the term $\frac{1}{p(\mathbf{e}^T, o^T)}$ in r, which can be computed in polynomial time by a BN propagation in polytrees [Pearl 1988], and it does not depend on the choice $\mathbf{x}$.