Most Frugal Explanations in Bayesian networks

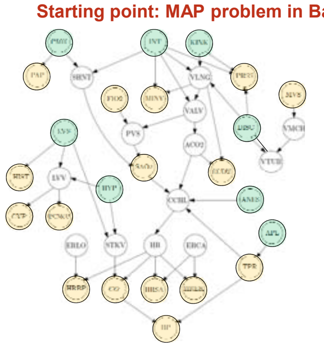
Johan Kwisthout

Donders Institute for Brain, Cognition and Behaviour
Radboud University Nijmegen

Overview

- Most Frugal Explanation: why, what, & how
- Computational complexity results + their interpretation
- Simulation results
- Conclusion + future work

Starting point: MAP problem in Bayesian Networks



Hypothesized causes

Observable findings

Intermediate variables

MAP problem: given ◯, what is the most probable joint value assignment to ◯ ? Note that we need to marginalize over ◯ !

Beinlich et al. (1989). *The ALARM monitoring system: a case study with two probabilistic inference techniques for belief networks.* 2nd European Conference on AI and Medicine.

Known complexity results for MAP

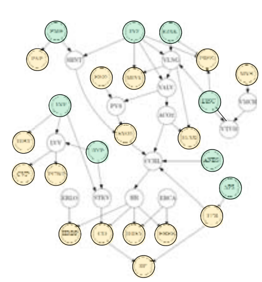
- MAP is NP^{PP}-hard to compute exactly (Park and Darwiche, 2004), and is thus *strictly harder* than MPE and INFERENCE
- MAP remains NP-hard in general to compute exactly, even in graphs with bounded treewidth and cardinality (De Campos, 2011)
- It is NP-hard to approximate MAP, even on polytrees (Park and Darwiche, 2004)

Bottom line: MAP is intractable to compute either exactly or approximately, even under rather strong constraints on topology

Observations

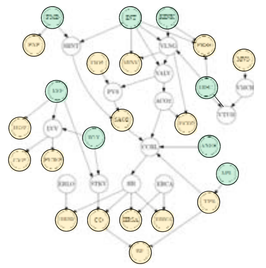
- MAP *is* tractable when either:
  - Both treewidth, cardinality, and number of intermediate variables are bounded (Kwisthout, 2011)
  - Both treewidth and cardinality are bounded and the MAP explanation has a high probability (Bodlaender et al., 2002; Kwisthout, 2011)
  - Are these constraints reasonable in real-world large networks?
- Druzdzel and Suermondt (1994): In general, for any particular inference query in a Bayesian network, most (intermediate) variables are not relevant: they do not influence the output
- **Can we utilize this property of probability distributions?**

All Intermediate variables are equal…



- To compute MAP, we marginalize over all intermediate variables
- Some of them have little or no effect on the outcome of the MAP explanation
- That seems like a waste of computational resources…

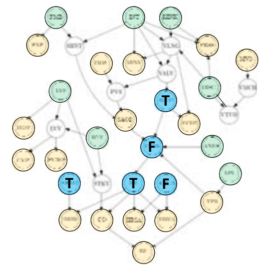
…but some are more equal than others!



Example:

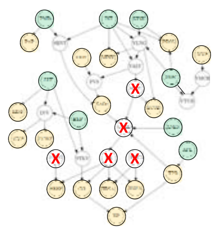
- We have that PCWP and BP take their "alarming" values and the other observable variables take default values
- MAP explanation for the diagnostic variables: HYP = true, all others = false
- We marginalized over 13 intermediate variables to arrive to that decision

…but some are more equal than others!



- However, ACO2, CCHL, ERLO, HR, and ERCA do not really contribute to the explanation: for **every** joint value assignment to these variables, HYP = true would be the MAP explanation
- We use this knowledge and assign them an **arbitrary value** and again compute MAP, now marginalizing over just 8 variables (was 13)

Most Frugal Explanation (MFE)



Evidence variables
Hypothesis variables
Irrelevant Intermediate variables
Relevant Intermediate variables

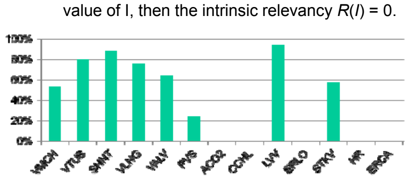
- **Instance**: Bayesian network, partitioned into (observed) **evidence** variables E, **hypothesis** variables H, **irrelevant** intermediate variables I^-, **relevant** intermediate variables I^+
- **Output**: the joint value assignment to the hypothesis variables that is most probable for the **maximal number** of joint value assignments to the irrelevant intermediate variables
- We discuss shortly what **relevant** means in this context

Most Frugal Explanation (MFE)

- Note different notion of "best explanation" of MFE versus MAP:
  - MAP: best explanation is explanation that is most probable
  - MFE: best explanation is explanation that is most probable for the maximum number of joint value assignments to the irrelevant intermediate variables
- MFE explanation **=** MAP explanation if all intermediate variables are designated as relevant
- MFE explanation ≈ MAP explanation if all intermediate variables that ***actually are*** relevant are ***designated*** as relevant
- If only a **subset** of the *actual* relevant variables are designated as relevant, MFE and MAP may diverge

What is relevant?

- An intermediate variable $I \in Int$ has an intrinsic relevancy $R(I)$, defined as the fraction of joint value assignments to $Int \setminus \{ I \}$ for which the MFE is not identical for all values of I
- Example: if the MFE is **always** dependent on the value of I, then the intrinsic relevancy $R(I) = 1$. If it is **never** dependent on the value of I, then the intrinsic relevancy $R(I) = 0$.



Intrinsic relevancy of the intermediate variables in the ALARM network, **PCWP** = high; **BP** = high; other evidence variables take non-alarming values

How to determine what is relevant?

- Intrinsic relevancy is formal parameter of an intermediate variable (relative to evidence), MFE uses subjective partition between relevant and irrelevant variables
- Partition can be made based on previous experience, heuristics, domain knowledge, or by approximating intrinsic relevancy using a few samples [intrinsic relevancy is intractable to compute exactly]
- The more (intrinsically) relevant variables are included, the more the distribution of MFE explanations tends to be skewed to a single explanation

Overview

- Most Frugal Explanation: why, what, & how
- **Computational complexity results + their interpretation**
- Simulation results
- Conclusion + future work

Computational Complexity

- Computing MFE is **intractable**: it has MAP as a special case and it can be shown to be strictly harder than MAP under usual complexity-theoretic assumptions
- Computing MFE remains intractable when treewidth, cardinality, are the number of relevant intermediate variables are bounded
- However, we can efficiently **approximate** MFE (with a poly-time randomized sampling algorithm) when (in addition) the distribution of MFE explanations is "skewed" towards one explanation – we then need only few samples to decide with low possibility of error

Computational Complexity

- Computing MFE is **intractable**: it has MAP as a special case and it can be shown to be strictly harder than MAP under usual complexity-theoretic assumptions
- To be precise, computing MFE is $NP^{PP^{PP}}$-complete
- Three "sources of complexity" working on top of each other:
  1. ("NP") selecting a joint value assignment out of potentially exponentially many candidate assignments to the explanation set;
  2. ("PP") solving an inference problem over the variables in the set $\mathbf{I}^+$;
  3. ("PP") deciding upon a threshold of the joint value assignments to the set $\mathbf{I}^-$.

Computational Complexity

- The reduction itself is rather straightforward; the most difficult part (I guess) is in understanding the SAT variant that is the canonical complete problem for this class:

  E-MAJMAJSAT

  Instance: A Boolean formula φ whose n variables $x_1 \ldots x_n$ are partitioned into three sets $\mathbf{E} = x_1 \ldots x_k$, $\mathbf{M_1} = x_{k+1} \ldots x_l$, and $\mathbf{M_2} = x_{l+1} \ldots x_n$ for some numbers k, l with $1 \leq k \leq l \leq n$.

  Question: Is there a truth assignment to the variables in $\mathbf{E}$ such that for the *majority* of truth assignments to the variables in $\mathbf{M_1}$ it holds, that the *majority* of truth assignments to the variables in $\mathbf{M_2}$ yield a satisfying truth instantiation to $\mathbf{E} \cup \mathbf{M_1} \cup \mathbf{M_2}$?

Computational Complexity

E-MAJMAJSAT

Instance: A Boolean formula φ whose n variables $x_1 \ldots x_n$ are partitioned into three sets $\mathbf{E} = x_1 \ldots x_k$, $\mathbf{M_1} = x_{k+1} \ldots x_l$, and $\mathbf{M_2} = x_{l+1} \ldots x_n$ for some numbers k, l with $1 \leq k \leq l \leq n$.

Question: Is there a truth assignment to the variables in $\mathbf{E}$ such that for the *majority* of truth assignments to the variables in $\mathbf{M_1}$ it holds, that the *majority* of truth assignments to the variables in $\mathbf{M_2}$ yield a satisfying truth instantiation to $\mathbf{E} \cup \mathbf{M_1} \cup \mathbf{M_2}$?

As an example, consider the formula $\varphi_{ex} = x_1 \wedge (x_2 \vee x_3) \wedge (x_4 \vee x_5)$ with $\mathbf{E} = \{x_1\}$, $\mathbf{M_1} = \{x_2, x_3\}$ and $\mathbf{M_2} = \{x_4, x_5\}$. This is a *yes*-example of E-MAJMAJSAT: for x_1=true, three out of four truth assignments to $\{x_2, x_3\}$ (all but $x_2 = x_3$=false) are such that the majority of truth assignments to $\{x_4, x_5\}$ satisfy φ_{ex}.

A randomized algorithm for MFE

- Computing MFE is **intractable** in general, even under strong constraints; however, have a look at this randomized algorithm:

Algorithm 1 Compute the Most Frugal Explanation.

```
Sampled-MFE(B, H, I+, I-, e, N)
for n = 1 to N do
    Choose i ∈ I- at random
    Determine h = argmax_h Pr(H = h, i, e)
    Collate the joint value assignments h
end for
Decide upon the joint value assignment h_maj that was picked most often
return h_maj
```

- This randomized algorithm repeatedly picks a joint value assignment $\mathbf{i} \in \mathbf{I}^-$ at random, determines the most probable explanation, and at the end decides upon which explanation was found most often

Fixed-parameter tractable approximation of MFE

- Approximating MFE is **intractable** in general but fixed-parameter tractable if the following parameters are bounded:

Parameter	Meaning
Treewidth (tw)	A measure on the network topology
Cardinality (c)	The maximum number of values each variable can take
#Relevants ($\|I^+\|$)	The number of relevant variables we marginalize over
Decisiveness (d)	A measure on the probability distribution, denoting the probability that for a given evidence set **E** with evidence **e** and an explanation set **H**, two random joint value assignments $\mathbf{i}_1$ and $\mathbf{i}_2$ to the irrelevant variables $\mathbf{I}^-$ would yield the same most probable explanations

Comparing MAP and MFE

- Computing MAP exactly can be done tractably when treewidth and cardinality are low, and either the total number of intermediate variables is low or the probability of the MAP explanation is high
- Approximating MFE can be done tractably when treewidth and cardinality are low, there are few relevant intermediate variables and the distribution of MFE explanations is skewed (not uniform)

➢ There can be ***many*** irrelevant intermediate variables
➢ Best explanation does not need to have (absolute) high probability
➢ Treewidth / cardinality are those of ***reduced*** junction tree

Overview

- Most Frugal Explanation: why, what, & how
- Computational complexity results + their interpretation
- **Simulation results**
- Conclusion + future work

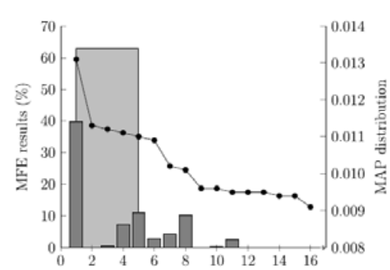
Simulations

- We ran MFE sampling algorithm on 100 randomly constructed networks to study its behavior under various parameters:
- Which variables are deemed relevant
  - All "objectively" relevant variables (R(I) > 0)
  - Only some of them (R(I) > some threshold)
  - Complete mismatch
- What is the 'decisiveness' of the probability distribution?
- When does MFE give a 'good' approximation of MAP, when does it not, how many samples are needed for a reliable decision?

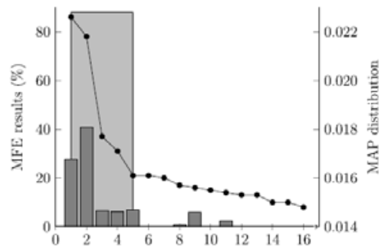
Simulations

- Each network had 40 variables, with two to four values, indegree at most five, random CPTs
- Five arbitrary variables were denoted as evidence, five as explanation variables
- R(I) was computed for all 30 intermediate variables
- We computed MAP and MFE under various settings of $\mathbf{I}^+$ / $\mathbf{I}^-$ and computed deviation using three measures
  - Ratio between MAP and MFE explanation
  - Rank (k, were MFE happens to be the k-th MAP)
  - Hamming distance between MAP and MFE explanation

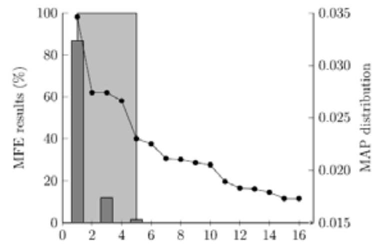
Illustrations



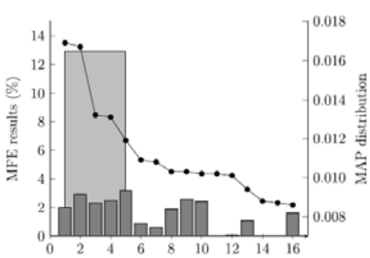
Illustrations



Illustrations



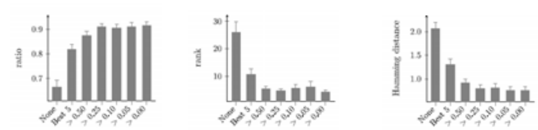
Illustrations



Simulation results

Table 2
Overview of simulation results. In this simulation the partition between relevant and irrelevant variables was varied and ranged from 'none' (all variables are irrelevant), 'best 5' (the five variables with the highest relevance are deemed relevant) to a relevance threshold between 0.50 and 0.00, yielding an average I^+ size between 11.32 and 16.35.

Cond.	I^+ size	Ratio	Rank	Dist.	% MFE	% MAP	% 5-MAP
None	0.00	0.66	25.90	2.05	0.08	0.03	0.14
Best 5	5.00	0.82	10.73	1.30	0.13	0.08	0.27
>0.50	11.32	0.87	5.36	0.87	0.25	0.17	0.46
>0.25	14.93	0.91	4.59	0.79	0.38	0.25	0.58
>0.10	15.79	0.91	5.56	0.81	0.39	0.25	0.60
>0.05	15.99	0.91	6.09	0.75	0.41	0.27	0.60
>0.00	16.35	0.92	4.12	0.75	0.41	0.26	0.61



Conclusion and further work

- We provided a **heuristic to MAP** – denoted MFE – based on the property that in general, only few intermediate variables are really relevant in establishing the most probable explanation
- The heuristic is such that we **partition** intermediate variables into relevant set (marginalized over) and irrelevant set (sampled over)
- Does not buy tractability per se, but has **efficient approximation algorithm** under situational constraints that are more **favorable** than constraints needed to render MAP tractable
- Illustrated by simulations with random networks

If you want to read more…

Artificial Intelligence 218 (2015) 56–71
Contents lists available at ScienceDirect
Artificial Intelligence
www.elsevier.com/locate/artint

Most frugal explanations in Bayesian networks

Johan Kwisthout *
Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, The Netherlands

- See references, preprint, supplementary material on my website: **http://www.dcc.ru.nl/~johank/**