Estimating the Impact of Variables in Bayesian Belief Networks

Sicco Pier van Gosliga

PhD candidate – University of Amsterdam

September 27, 2013

Introduction

This talk will be about how to assess the relevance of observing missing evidence when we want to compute the probability of a certain hypothesis.

I will discuss the following topics:

- Bayesian Belief Networks
  (their structure, use and methods of inference)
- Relevance of variables
  (real impact, approximation method)
- Experimental results

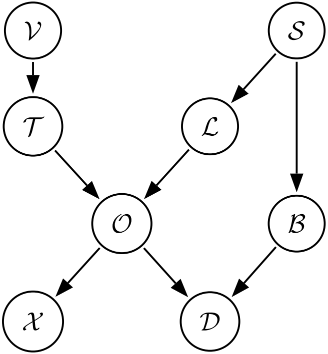
Bayesian Belief Networks

BBNs are probabilistic models for calculating the posterior probability of an event given current observations and prior knowledge.

$P(\mathcal{V} = v) = 0.01$

$P(\mathcal{T}\|\mathcal{V})$	$\mathcal{T} = t$
$\mathcal{V} = v$	0.99
$\mathcal{V} = \neg v$	0.01

$P(\mathcal{X}\|\mathcal{O})$	$\mathcal{O} = o$
$\mathcal{X} = x$	0.95
$\mathcal{X} = \neg x$	0.02



$P(\mathcal{S} = s) = 0.50$

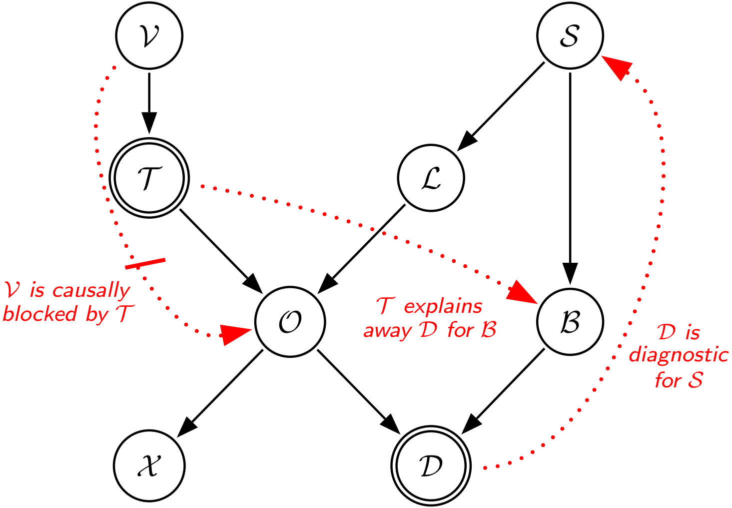
$P(\mathcal{L}\|\mathcal{S})$	$\mathcal{L} = l$
$\mathcal{S} = s$	0.99
$\mathcal{S} = \neg s$	0.90

$P(\mathcal{B}\|\mathcal{S})$	$\mathcal{B} = b$
$\mathcal{S} = s$	0.70
$\mathcal{S} = \neg s$	0.40

	$\neg o$		o	
$P(\mathcal{D}\|\mathcal{O},\mathcal{B})$	$\neg b$	b	$\neg b$	b
$D = d$	0.10	0.80	0.70	0.90

Bayesian Belief Networks

Things to consider while calculating: $P(\mathcal{L}=l \mid \mathcal{T}=\neg t, \mathcal{D}=d)$



Bayesian Belief Networks

Applications:

- Decision Support Systems, e.g. investigations and sensor networks
- Risk Assessments, e.g. safety of CO^2 storage
- Naive Bayesian Classifier, e.g. spam filters

Methods for inference:

- Joint Probability Table
- Pearl Message Passing, Cut-Set Conditioning (Pearl, 1982)
- Junction Tree algorithm (Lauritzen and Spiegelhalter, 1988)
- Approximate Methods, e.g. Loopy Belief Propagation

Inference for arbitrary BBNs is for all methods NP-hard (Cooper, 1990)
Reasons to measure relevance: pruning and evidence collection.

Relevance of Variables: real maximum impact

How to measure relevance?

- Entropy based distance measures, e.g. Kullback–Leibler divergence:

$$D_{KL}(P \parallel Q) = \sum_i \ln\left(\frac{p_i}{q_i}\right) p_i \tag{1}$$

- Absolute disance, i.e. Chebyshev distance measure:

$$D_{Ch}(\ P,\ Q\ ) = \max_i \left(\ |\ p_i\ -\ q_i\ |\ \right) \tag{2}$$

The maximum impact of X on Y in the context of $\mathbf{e}$:

$$\delta_{\mathbf{e}}(Y|X) = \max_{x_i \in X; x_j \in X} D_{Ch}\Big(P(Y|x_i, \mathbf{e}), P(Y|x_j, \mathbf{e})\Big) \tag{3}$$

The real maximum impact of X on Y:

$$\delta(Y|X) = \max_{\mathbf{e} \in \mathrm{E}} \Big(\delta_{\mathbf{e}}(Y|X)\Big) \tag{4}$$

Relevance of Variables: real maximum impact

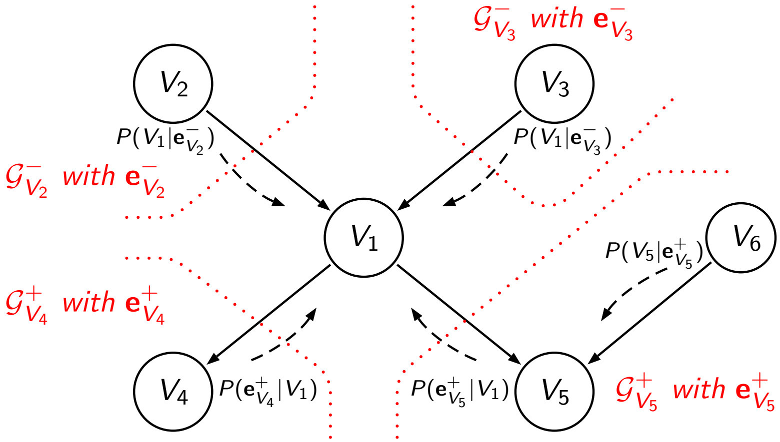
The real impact between variables in the Asia BBN:

	$\mathcal{A}$	$\mathcal{T}$	$\mathcal{S}$	$\mathcal{L}$	$\mathcal{O}$	$\mathcal{X}$	$\mathcal{B}$	$\mathcal{D}$
$\delta(\mathcal{A},\ldots)$ **e**	-	0.390 $\{d,\neg l,\neg s,x\}$	0.176 $\{\neg b,o\}$	0.345 $\{\neg b,o\}$	0.390 $\{d,\neg l,\neg s,x\}$	0.189 $\{\neg b,d,\neg l\}$	0.163 $\{d,\neg l,\neg s,x\}$	0.206 $\{\neg b,\neg l,x\}$
$\delta(\mathcal{T},\ldots)$ **e**	0.038 $\emptyset$	-	0.487 $\{\neg b,o\}$	0.990 $\{o,\neg s\}$	1.000 $\{\neg l\}$	0.930 $\{\neg l\}$	0.419 $\{d,\neg l,\neg s\}$	0.600 $\{\neg b,\neg l\}$
$\delta(\mathcal{S},\ldots)$ **e**	0.016 $\{o\}$	0.496 $\{a,o\}$	-	0.528 $\{b,d,\neg t,x\}$	0.528 $\{b,d,\neg t,x\}$	0.345 $\{\neg b,d,\neg t\}$	0.303 $\{d,t\}$	0.341 $\{\neg t,x\}$
$\delta(\mathcal{L},\ldots)$ **e**	0.038 $\{o\}$	0.990 $\{\neg a,o\}$	0.509 $\{\neg b\}$	-	1.000 $\{\neg t\}$	0.930 $\{\neg t\}$	0.419 $\{d,\neg s,\neg t\}$	0.600 $\{\neg b,\neg t\}$
$\delta(\mathcal{O},\ldots)$ **e**	0.038 $\{\neg l\}$	1.000 $\{\neg l\}$	0.509 $\{\neg b,\neg t\}$	1.000 $\{\neg t\}$	-	0.930 $\emptyset$	0.419 $\{d,\neg s\}$	0.600 $\{\neg b\}$
$\delta(\mathcal{X},\ldots)$ **e**	0.023 $\{\neg b,d,\neg l\}$	0.871 $\{a,\neg b,d,\neg l\}$	0.434 $\{\neg b,d,\neg t\}$	0.922 $\{\neg b,d,s,\neg t\}$	0.934 $\{a,\neg b,d,s\}$	-	0.314 $\{a,d,\neg s\}$	0.459 $\{a,\neg b,s\}$
$\delta(\mathcal{B},\ldots)$ **e**	0.015 $\{d,\neg l,x\}$	0.402 $\{d,\neg l,x\}$	0.319 $\{\neg d,\neg t,x\}$	0.399 $\{d,\neg s,\neg t,x\}$	0.427 $\{\neg a,d,\neg s,x\}$	0.356 $\{a,d,s\}$	-	0.700 $\{\neg l,\neg t\}$
$\delta(\mathcal{D},\ldots)$ **e**	0.020 $\{\neg b,\neg l,x\}$	0.622 $\{a,\neg b,\neg l,x\}$	0.323 $\{\neg b,\neg t,x\}$	0.634 $\{\neg b,\neg t,x\}$	0.634 $\{\neg b,\neg t,x\}$	0.455 $\{a,\neg b,s\}$	0.714 $\{\neg l,\neg t\}$	-

Practical problem: infeasable to assess for larger BBNs.

Relevance of Variables: localized maximum impact

Computing the posterior probability distribution of variable V_1:



$$P(V_1|\mathbf{e}) = P(V_1|\mathbf{e}^-_{V_2}, \mathbf{e}^-_{V_3}, \mathbf{e}^+_{V_4}, \mathbf{e}^+_{V_5}) = \eta\, P(V_1|\mathbf{e}^-_{V_2}, \mathbf{e}^-_{V_3})\, P(\mathbf{e}^+_{V_4}|V_1)\, P(\mathbf{e}^+_{V_5}|V_1)$$

Relevance of Variables: localized maximum impact

The posterior probability distribution of V_1:

$$P(V_1|\mathbf{e}) = \eta\, P(V_1|\mathbf{e}^-_{V_2}, \mathbf{e}^-_{V_3})\, P(\mathbf{e}^+_{V_4}|V_1)\, P(\mathbf{e}^+_{V_5}|V_1)$$
$$= \eta\, \Big(\sum_{V_2 V_3} P(V_1|V_2, V_3)\, P(V_2|\mathbf{e}^-_{V_2})\, P(V_3|\mathbf{e}^-_{V_3}) \Big)\ \lambda(V_1)$$

The localized maximum potential impact of V_2 on V_1, using D_{Ch}:

$$\Delta_{V_1}(V_1|V_2) = \max_{i,j,k,c} \Big| \eta_a\, P(v_{1k}|v_{2i}, v_{3c})\, \lambda(v_{1k}) - \eta_b\, P(v_{1k}|v_{2j}, v_{3c})\, \lambda(v_{1k}) \Big|$$

Setting: $p = P(v_{1k}|v_{2i}, v_{3c})$, $q = P(v_{1k}|v_{2j}, v_{3c})$ and $\lambda = \lambda(v_{1k})$

$$\lambda = \frac{\sqrt{(p-1)\ p\ (q^2 - q)}\ +\ p\ (-q) + p + q - 1}{p + q - 1}$$

unless $p + q - 1 = 0$, then: $\lambda = \frac{1}{2}$

Relevance of Variables: localized maximum impact

Now we can directly compute the localized maximum impact:

$$\Delta_{V_1}(V_1|V_2) = \max_{i,j,k,c} \left| \eta_a P(v_{1k}|v_{2i}, v_{3c}) \lambda(v_{1k}) - \eta_b P(v_{1k}|v_{2j}, v_{3c}) \lambda(v_{1k}) \right|$$

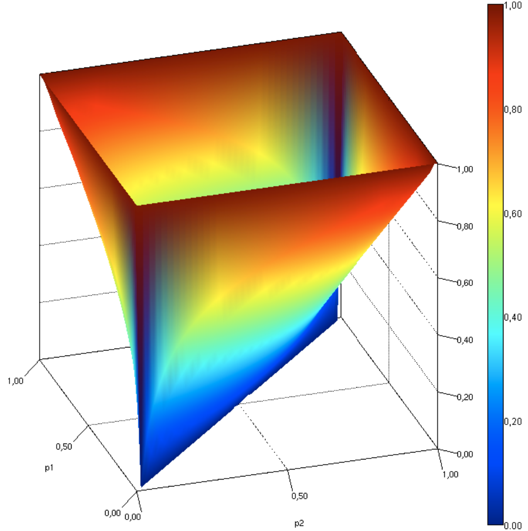
$$= \max_{i,j,k,c} \left| \frac{1}{2\lambda p - \lambda - p + 1} \lambda p - \frac{1}{2\lambda q - \lambda - q + 1} \lambda q \right|$$

unless p or q is 0, or p or q is 1, then $\Delta_{V_1}(V_1|V_2) = 1$

$\geq \delta_{\mathbf{e}}(V_1|V_2)$ (for polytrees)

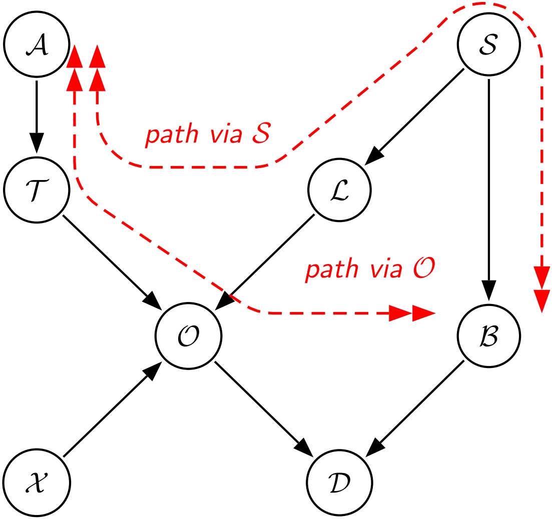
N.B. Here i, j, k, c refer to indices used for setting p, q and λ.

$\Delta_{V1}(V_1|V_2)$ given p and q:



Relevance of Variables: propagated maximum impact

How to assess the maximum impact of variables further apart?



Relevance of Variables: propagated maximum impact

How to compute the propagated maximum impact $\Delta(\mathcal{B}|\mathcal{A})$:

We consider these paths between $\mathcal{A}$ and $\mathcal{B}$:

- $\{\mathcal{A}\rightarrow\mathcal{T},\ \mathcal{T}\rightarrow\mathcal{L},\ \mathcal{L}\rightarrow\mathcal{S},\ \mathcal{S}\rightarrow\mathcal{B}\}$
- $\{\mathcal{A}\rightarrow\mathcal{T},\ \mathcal{T}\rightarrow\mathcal{O},\ \mathcal{O}\rightarrow\mathcal{B}\}$

The paths diverge at $\mathcal{T}$, converge at $\mathcal{B}$:

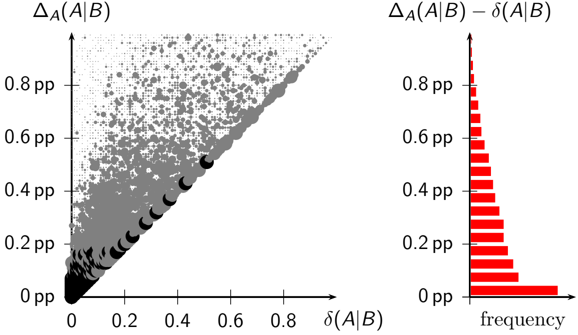
- The product of local impact values for each path segment.
- Where parallel path segments converge we take their sum.

$$\begin{aligned}\Delta(\mathcal{B}|\mathcal{A}) &= \Delta_{\mathcal{T}}(\mathcal{T}|\mathcal{A})\cdot\Delta(\mathcal{B}|\mathcal{T})\\ &= \Delta_{\mathcal{A}}(\mathcal{T}|\mathcal{A})\cdot\big(\Delta_{\mathcal{L}}(\mathcal{L}|\mathcal{T})\cdot\Delta_{\mathcal{S}}(\mathcal{S}|\mathcal{L})\cdot\Delta_{\mathcal{B}}(\mathcal{B}|\mathcal{S})+\\ &\qquad\qquad\Delta_{\mathcal{O}}(\mathcal{O}|\mathcal{T})\cdot\Delta_{\mathcal{B}}(\mathcal{B}|\mathcal{O})\big)\end{aligned}$$

Propagation is feasible for BBNs that are 'too large' to compute.

Experimental Results

Localized maximum impact as function of the real impact, in polytrees:



Conclusions:

- For polytrees, $\Delta_A(B|A)$ never underestimates $\delta(B|A)$.
- The 'quality' of $\Delta(B|A)$ very much depends on the structure.
- For multiply connected graphs $\Delta(B|A)$ never underestimates $\delta(B|A)$.

Thank you.

siccopier.vangosliga@gmail.com