Estimating the Impact of Variables in Bayesian Belief Networks
van Gosliga, S.P.; Groen, F.C.A.

Published in:
Tenth Tbilisi Symposium on Language, Logic and Computation: Gudauri, Georgia, 23-27 September 2013

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Tenth Tbilisi Symposium on Language, Logic and Computation

Gudauri, Georgia
23–27 September 2013

Centre for Language, Logic and Speech
and
Razmadze Mathematical Institute at the Tbilisi State University
Institute for Logic, Language and Computation
Estimating the Impact of Variables in Bayesian Belief Networks
Sicco Pier van Gosliga and Frans Groen

Introduction
Bayesian belief networks (BBNs) are often designed to aid decision making. A BBN models uncertainty and enables to compute posterior probabilities given prior information and current observations [8]. In this paper we focus on a solution to two practical problems that arise with the application of BBNs: First, in real world applications observations are associated with costs. To keep these costs within acceptable limits we would like to prioritize observations most relevant to our decision. Second, models can grow too large for feasible computations [2]. Rather than restricting the design of a BBN for decision making, we pursue an ad-hoc sub-model tailored to its relevance for the decision maker. For these reasons, we propose an efficient approximation algorithm to compute the maximum impact of observing a variable in respect to the posteriors of other variables in a BBN. The algorithm is guaranteed to never underestimate the real impact. First the impact of the variables within the markov blanket of a variable are calculated, followed by a message passing algorithm to include other variables in the network. The method is closely related the field of sensitivity analysis [3][5], which mostly focuses on aiding the design of a BBN.

Methodology
Distance measure. To quantify belief changes in posterior probability distributions we will use the maximum absolute distance as defined by the Chebyshev distance function $D_{Ch}(P, Q)$ which takes the largest difference between pairs of probability values in two discrete probability distributions P and Q:

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

(1)

Distance functions to quantify differences in probability distributions are often based on entropy [1][6][7]. Entropy based distance functions are relative distance measures. As a result, small absolute differences can be valued equally important as a large difference when both span an equal order of magnitude. Also, entropy based distance functions evaluate the general difference between two discrete probability distributions, while the Chebyshev distance focuses on the maximum difference. Since decisions are based upon the absolute posterior probability of a specific outcome, the Chebyshev distance is the measure that directly relates to the decision making. Van Engelen [9] introduced a method to compute an absolute upper bound for the maximum absolute error based on the K-L divergence. However, its reliance on computing prior marginals in advance limits its applicability for pruning BBNs. We base our method directly on the Chebyshev distance and local prior conditional distributions rather than prior marginals.
Aim. We aim to get a safe estimate of the real maximum impact. Given a BBN $G$ containing variables $V_1$ and $V_2$ we first define, in Eq. 2, the real maximum impact of $V_2$ on $V_1$ in the context of evidence $e$. We then define, in Eq. 3, the generalized real maximum impact of $V_2$ on $V_1$ as the maximum absolute difference that two different instantiations of $V_2$ can cause in the posterior probability for any state of $V_1$ given any possible combination of evidence $e$ for other variables in $G$, where set $E$ holds all evidence configurations for $G$.

$$\delta_e(V_1|V_2) = \max_{e_2 \in e_{V_2}: e_2 \in V_2} D_{CHS} \left( P(V_1|V_2,e), P(V_1|V_2,\emptyset) \right)$$ (2)

$$\hat{\delta}(V_1|V_2) = \max_{e \in E} \left( \delta_e(V_1|V_2) \right)$$ (3)

Algorithm

First Phase. We first calculate the maximum impact a variable may potentially have on other variables within its markov blanket, if the BBN beyond the markov blanket could take any form. Figure 1 shows $V_1$’s markov blanket, a subgraph of $G$ that contains $V_1$’s parents, children and parents of children. Suppose $e$ is a set of evidence, and we want to compute the posteriors for $V_1$ given this evidence set: $P(V_1|e)$. Each edge to $V_1$ can be considered to partition $G$ in subgraphs: edge $V_2 \rightarrow V_1$ divides $G$ in an upper subgraph $G_{V_2}^+$ and a lower subgraph $G_{V_2}^-$. Let $e_{V_2}$ be the subset of $e$ that concerns the variables in $G_{V_2}^-$. Likewise, edges $V_3 \rightarrow V_1$, $V_2 \rightarrow V_4$ and $V_1 \rightarrow V_3$ create the following subgraphs end evidence sets: $G_{V_3}$ with $e_{V_3}$, $G_{V_4}$ with $e_{V_4}$, and $G_{V_3}$ with $e_{V_3}^+$. Applying Bayes’ rule, the posterior probability distribution for $V_1$ given $e$ can then be computed as follows:

$$P(V_1|e) = P(V_1|V_2, e_{V_2}, e_{V_4}, e_{V_3}^+) \left( \frac{1}{\eta} \right)$$ (4)

$$= \frac{\eta}{\left( \sum_{e_{V_2}} P(V_1|V_2, V_3) P(V_2|e_{V_2}) P(V_3|e_{V_3}) \lambda(V_1) \right)} \lambda(V_1)$$ (5)

In Eq. 4 and 5, $\eta$ is a normalizing constant and $\lambda(V_1) = P(e_{V_1}^+|V_1) P(e_{V_3}^-|V_1)$. The posteriors of $V_1$ are computed by combining prior information $P(V_1|V_2, V_3)$ and current observations in evidence set $e$. Each parent and child variable of $V_1$ contributes a parameter conditioned by a subset of $e$. The maximizing causal parameters for $V_1$’s parents can be derived from the conditional probability table $P(V_1|V_2, V_3)$. For computing the local potential maximum impact of $V_2$ on $V_1$, $\Delta_{V_1}(V_1|V_2)$, we set these as $p = P(v_{1k}|v_{2i}, v_3e)$ and $q = P(v_{1k}|v_{2j}, v_3e)$. For the diagnostic parameters of $V_1$’s children, we set $\lambda = \lambda(v_{1k})$. The value that maximizes $\Delta_{V_1}(V_1|V_2)$ can be calculated in closed form with Eq. 7. The value $\Delta_{V_1}(V_1|V_2)$ can then be computed with Eq. 6.

$$\Delta_{V_1}(V_1|V_2) = \max_{i,j,k,e} \left( \frac{1}{2\lambda p - \lambda - p + 1} \lambda p - \frac{1}{2\lambda q - \lambda - q + 1} \lambda q \right)$$ (6)

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

$$\lambda = \frac{(p-1)p(q^2-q) + p(-q) + p + q - 1}{p + q - 1}$$ (7)

unless $p + q - 1 = 0$, then: $\lambda = \frac{1}{2}$
Second Phase. The second phase assesses the maximum potential impact between nodes further apart. A propagation algorithm is used to investigate all d-connecting paths between each pair of nodes. Multiple d-connecting paths may exist between a pair of variables, and are likely to partly overlap. For a single path segment, we take the product of all local potential impact values along intermediate edges as a safe overestimate of the real maximum impact. When parallel path segments converge we take their sum to approximate their joint influence.

Experiments
To assess the quality of our approximation of the real maximum impact values, the algorithm was tested on randomly generated BBNs. The BBNs were generated by using an algorithm designed and implemented by [4]. The $\Delta(V_a|V_b)$ values were compared to the corresponding $\delta(V_a|V_b)$ values, based on simulated evidence. In each BBN, all possible constellations of evidence were taken into account to discover the real maximum impact. In total 48000 pairs of variables were evaluated in singly connected graphs, and 90000 pairs of variables in multiply connected graphs. The experiments show that accuracy of the $\Delta(V_a|V_b)$ values decreases with the number of states a variable may have, the maximum degree of the graph, and the number of edges between $V_a$ and $V_b$. The algorithm never returned an underestimate of the real maximum impact. Figure 2 gives the estimated impact as function of the real impact in these experiments.

Discussion
The method was successfully used to estimate the impact of variables in a BBN. Instead of getting an optimally accurate assessment of the real maximum impact, we have chosen an approach that is guaranteed never to return an underestimate of the real impact. Not underestimating a variable’s impact in the context of decision making can be crucial. To improve the method’s accuracy and applicability, it could be extended to respect existing evidence and to get the joint impact of multiple variables to a single goal variable.

References
A syntactic characterisation of the Gabbay–de Jongh logics

Jeroen Goudsmit

[11] introduced refutation systems as a formal means of reasoning non-derivability. One can think of a refutation system as a proof system for non-derivability. The “proofs” in this system are formal inferences of the refutability of a given formula \( \phi \), written \( \neg \vdash \phi \), and these inferences are called refutations. When every refutable formula is non-derivable the refutation system is sound, and when the converse holds it is said to be complete.

[12] proposed a refutation system for intuitionistic propositional logic (IPC), which consisted of the rules listed below for \( L = IPC \). In stating that this system is sound and complete for IPC he, in essence, conjectured that IPC is the sole intermediate logic with the disjunction property. The situation however turned out more subtle, see [2] for a survey.

\[
\begin{align*}
\text{Ax} & : \quad \neg \vdash p \\
\text{Subs} & : \quad \neg \vdash \sigma(\phi) \quad \neg \vdash \psi \\
\text{MT} & : \quad \phi \vdash \psi \rightarrow \phi \\
\text{DP} & : \quad \neg \vdash \phi \quad \neg \vdash \psi
\end{align*}
\]

The axiom Ax states that no propositional variable is derivable, which is a valid statement for any intermediate logic. The reverse substitution rule Subs is also sound for any intermediate logic, and the same goes for modes tollens, MT. Finally, the rule DP is basically the counter-positive of the disjunction property, which states that if \( \phi \lor \psi \) is derivable, then one amongst \( \phi \) and \( \psi \) must be derivable. To summarise, Ax, Subs and MT form a sounds refutation system for each intermediate logic \( L \). When we let \( L \) be CPC, classical propositional logic, it is complete. The latter is clear when one realises that a counter model of a formula \( \phi \) gives rise to a substitution \( \sigma \) such that \( \sigma(\phi) \equiv \bot \). Note that, although each refutation system can be sound for many logics, there is but one for which it is both sound and complete. It is in this sense that a refutation system syntactically characterises a logic.

[13] proved that replacing the rule DP by the rules GDP\(_n\) for all \( n \geq 2 \) as below suffices to refute all non-theorems of IPC. That is to say, the refutation system containing Ax, Subs, MT and