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

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:

$$D_{Ch}(P,Q) = \max_i \Big(\big|p_i - q_i\big|\Big) \tag{1}$$

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 $\mathcal{G}$ 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 $\mathbf{e}$. 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 $\mathbf{e}$ for other variables in $\mathcal{G}$, where set E holds all evidence configurations for $\mathcal{G}$.

$$\delta_{\mathbf{e}}(V_1|V_2) = \max_{v_{2i}\in V_2; v_{2j}\in V_2} D_{Ch}\Big(P(V_1|v_{2i},\mathbf{e}), P(V_1|v_{2j},\mathbf{e})\Big) \tag{2}$$

$$\delta(V_1|V_2) = \max_{\mathbf{e}\in\mathsf{E}} \Big(\delta_{\mathbf{e}}(V_1|V_2)\Big) \tag{3}$$

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 $\mathcal{G}$ that contains V_1's parents, children and parents of children. Suppose $\mathbf{e}$ is a set of evidence, and we want to compute the posteriors for V_1 given this evidence set: $P(V_1|\mathbf{e})$. Each edge to V_1 can be considered to partition $\mathcal{G}$ in subgraphs: edge $V_2 \rightarrow V_1$ divides $\mathcal{G}$ in an upper subgraph $\mathcal{G}^-_{V_2}$ and a lower subgraph $\mathcal{G}^+_{V_2}$. Let $\mathbf{e}^-_{V_2}$ be the subset of $\mathbf{e}$ that concerns the variables in $\mathcal{G}^-_{V_2}$. Likewise, edges $V_3 \rightarrow V_1$, $V_1 \rightarrow V_4$ and $V_1 \rightarrow V_5$ create the following subgraphs end evidence sets: $\mathcal{G}^-_{V_3}$ with $\mathbf{e}^-_{V_3}$, $\mathcal{G}^+_{V_4}$ with $\mathbf{e}^+_{V_4}$ and $\mathcal{G}^+_{V_5}$ with $\mathbf{e}^+_{V_5}$. Applying Bayes' rule, the posterior probability distribution for V_1 given $\mathbf{e}$ can then be computed as follows:

$$\begin{aligned} P(V_1|\mathbf{e}) &= P(V_1|\mathbf{e}^-_{V_2},\mathbf{e}^-_{V_3},\mathbf{e}^+_{V_4},\mathbf{e}^+_{V_5}) && (4)\\ &= \eta\, P(V_1|\mathbf{e}^-_{V_2},\mathbf{e}^-_{V_3})\, P(\mathbf{e}^+_{V_4}|V_1)\, P(\mathbf{e}^+_{V_5}|V_1) \\ &= \eta\big(\sum_{V_2 V_3} P(V_1|V_2,V_3)P(V_2|\mathbf{e}^-_{V_2})P(V_3|\mathbf{e}^-_{V_3})\big)\lambda(V_1) && (5)\end{aligned}$$

In Eq. 4 and 5, η is a normalizing constant and $\lambda(V_1) = P(\mathbf{e}^+_{V_4}|V_1)P(\mathbf{e}^+_{V_5}|V_1)$. The posteriors of V_1 are computed by combining prior information $P(V_1|V_2,V_3)$ and current observations in evidence set $\mathbf{e}$. Each parent and child variable of V_1 contributes a parameter conditioned by a subset of $\mathbf{e}$. 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 $p = P(v_{1k}|v_{2i},v_{3c})$ and $q = P(v_{1k}|v_{2j},v_{3c})$. For the diagnostic parameters of V_1's children, we set $\lambda = \lambda(v_{1k})$. The λ 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.

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

$$\text{unless } p \text{ or } q \text{ is 0, or } p \text{ or } q \text{ is 1, then } \Delta_{V_1}(V_1|V_2) = 1 \tag{6}$$

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

$$\text{unless } \; p+q-1=0 \text{ , then: } \; \lambda = \frac{1}{2} \tag{7}$$

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.

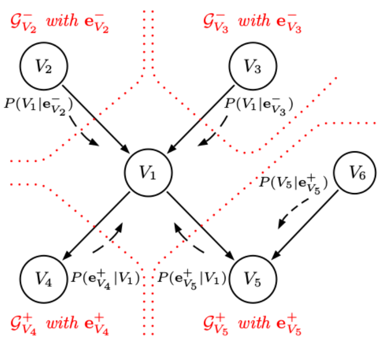

Fig. 1. The markov blanket of variable V_1.

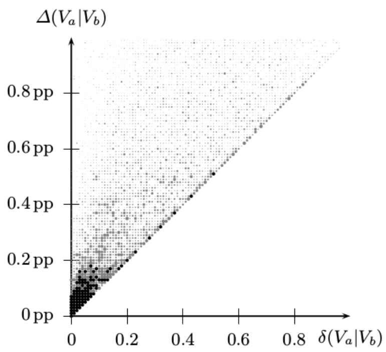

Fig. 2. Estimated impact as function of the real impact.

References

[1] H. Chan and A. Darwiche, *A distance measure for bounding probabilistic belief change*, Int. Journal of Approximate Reasoning, vol. 38, no. 2, 149–174 (2005)

[2] G. F. Cooper, *Probabilistic Inference Using Belief Networks is NP-hard*, Technical Report KSL-87-27, Medical Computer Science Group, Stanford University (1987)

[3] V. M. H. Coupé and L. C. van der Gaag, *Properties of sensitivity analysis of bayesian belief networks*, Annals of Mathematics and A.I., vol. 36, no. 4, 323–356 (2002)

[4] J. S. Ide and F. G. Cozman, *Random generation of bayesian networks*, Proc. of the 16th Brazilian Symp. on Artificial Intelligence, Springer-Verlag, 366–375 (2002)

[5] F.V. Jensen, S.H. Aldenryd, and K.B. Jensen, *Sensitivity analysis in bayesian networks*, Symbolic and Quantitative Approaches to Reasoning and Uncertainty, Lecture Notes in Computer Science, vol. 946, Springer, 243–250 (1995)

[6] S. Kullback and R. A. Leibler, *On Information and Sufficiency*, The Annals of Mathematical Statistics, vol. 22, no. 1, 79–86 (1951)

[7] J. Lin, *Divergence measures based on the shannon entropy*, Information Theory, IEEE Transactions on Information Theory, vol. 37, no. 1, 145–151 (1991)

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

[9] R. A. van Engelen, *Approximating Bayesian belief networks by arc removal*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, 916–920 (1997)