Addressing the Practical Limitations of Noisy-OR using Conditional Inter-causal Anti-Correlation with Ranked Nodes

Takao Noguchi, Norman Fenton and Martin Neil

Abstract—Numerous methods have been proposed to simplify the problem of eliciting complex conditional probability tables in Bayesian networks. One of the most popular methods - “Noisy-OR” - approximates the required relationship in many real-world situations between a set of variables that are potential causes of an effect variable. However, the Noisy-OR function has the conditional inter-causal independence (CII) property which means that ‘explaining away’ behavior – one of the most powerful benefits of BN inference – is not present when the effect variable is observed as false. Hence, for many real-world problems where the Noisy-OR has been used or proposed it may be deficient as an approximation of the required relationship. However, there is a very simple alternative solution, namely to define the variables as ranked nodes and to use the ranked node weighted average function. This does not have the CII property – instead we prove it has the conditional anti-correlation property required to ensure that explaining away works in all cases. Moreover, ranked node variables are not restricted to binary states, and hence provide a more comprehensive and general solution to Noisy-OR in all cases.

Index Terms—Knowledge Representation Formalisms and Methods, Probabilistic algorithms, Bayesian networks

I. INTRODUCTION

Bayesian Networks (BNs) have been applied in many areas to provide practical solutions to risk assessment problems [1], [2]. A BN is a directed acyclic graph, with nodes representing variables of interest. The relationships between the variables are expressed as conditional probability tables. In the absence of relevant data a conditional probability table is typically elicited from human experts. However, such elicitation can be difficult because the number of required inputs can grow exponentially with the number of variables. To reduce this burden a number of functions have been proposed to approximate practical situations.

This work was supported in part by: the European Research Council under project ERC-2013-AdG3139182 (BAYES-KNOWLEDGE); the Leverhulme Trust under Grant RPG-2016-118 CAUSAL-DYNAMICS; the EPSRC under project EP/P009964/1: PAMBAYESIAN: Patient Managed decision-support using Bayes Networks. We also acknowledge Agena Ltd for software support.

T. Noguchi is with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London, E1 4NS. E-mail: t.noguchi@qmul.ac.uk

N. E. Fenton is with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London, E1 4NS, and Agena Ltd. E-mail: n.fenton@qmul.ac.uk

M. Neil is with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London, E1 4NS, and Agena Ltd. E-mail: m.neil@qmul.ac.uk

One of the most widely-used such functions is Noisy-OR [3], which formalizes probabilistic relations between multiple causes of a single effect. Noisy-OR represents a distribution of causes and an effect, when all the causes are independent and binary. Consider the example where the ‘causes’ are different types of known security threats and the ‘effect’ is a security breach. Noisy-OR provides a useful approximation in such practical applications, because if the effect is observed to be present then the known absence of some of the causes makes the remaining causes more likely to have caused the effect. This is a classic example of explaining-away behavior which is an important property of inference with BNs. Specifically, with Noisy-OR causes are dependent on each other when conditioned on the presence of the effect. However, when the causes are instead conditioned on the absence of an effect, these same causes become independent. This property of the Noisy-OR is referred to as the conditional inter-causal independence (CII) property [4].

Thus, with Noisy-OR, when the effect is observed as false, there is no explaining away behavior: if one of the causes is subsequently observed to be true, there is no increase in the probability that the remaining causes are false. Similarly, if one of the causes is observed to be false, there is no decrease in the probability that the remaining causes are false. This can be a problem in practice since the underlying relationship often requires this explaining away behavior even when the effect is observed as false. To see this consider again the above security breach example: if we know that a breach has not occurred, but also discover that one of the most important types of threats has occurred, then the probability that the remaining threats have occurred should clearly decrease. But, this does not happen with a Noisy-OR model of the problem, even though a Noisy-OR model for this type of problem works well for both forward inferences and for backward inference when a breach has been observed.

A previous paper [5] has proposed an extension to the Noisy-OR function to partly overcome this deficiency in the explaining-away behavior. However, in order to achieve the full explaining-away behavior we require conditional anti-correlation between causes. The method proposed in this paper – ranked nodes [6], [7] – produces the necessary conditional inter-causal anti-correlation when conditioned on any state (present or absent) of an effect. This anti-correlation property enables the explaining-away behavior, regardless of an observed effect.

The ranked node solution also directly and easily resolves
another major limitation of the Noisy-OR function, namely that it is applicable only when all of the variables are binary. Although this limitation has been addressed with various extensions to the Noisy-OR function (typically to allow multiple-valued causes, for example noisy-MAX [8], [9]) these extensions retain the CII property of the Noisy-OR and hence still suffer from a lack of explaining away behavior when the effect is observed as false.

In this paper we show that the use of ranked-nodes overcomes both of the above limitations of the Noisy-OR: it produces explaining-away behavior regardless of values observed for an effect, and it is applicable to multiple-valued variables. In Section II, we introduce notation and define the ranked-node function. In Section III, we show that with the ranked-node function, causes are conditionally anti-correlated. In Section IV, we provide examples showing how the ranked node function can be used as a simple and better alternative approximation of Noisy-OR in practice. Section V provides our conclusions and recommendations.

II. RANKED-NODE FUNCTION

A ranked node in a BN is a variable with a finite ordered set of labelled states (such as {Very Low, Low, Medium, High, Very High}) mapped onto equal intervals over an underlying numerical scale from 0 to 1. The notion of ranked node functions was introduced in [6] as a practical and efficient way of defining the conditional probability tables for BN fragments in which, like the Noisy-OR, there are multiple causes and a single effect, but for which the variables are an ordinal scale rather than Boolean. Suppose, for example, there are three known independent factors (x1, x2, and x3) that can cause a disease y, then in practice it may be more useful and accurate to define the x1 and even y as ranked nodes rather than simply Boolean. Ranked nodes and their associated functions have been implemented in the BN software package AgenaRisk [10] and the method has been widely used and validated [1], [7]. The ranked node function exploits the underlying numerical scale to reduce the number of parameters to be elicited from experts. For example, one key ranked node function that approximates the required relationship between the causes xi and the effect y is a truncated Normal distribution over the interval (0, 1) whose mean is a weighted average of the xi and whose variance is a simple measure of the uncertainty. Hence, in the above example, we only need to elicit 4 parameters (3 weight parameters and a variance) to generate a conditional probability table which would have 625 entries if all 4 variables had 5 states.

While ranked nodes are now widely used, what has never been demonstrated formally is that the (weighted mean) ranked node function provides a simple and powerful alternative to Noisy-OR in all cases – not just where we need to use multiple-state variables. For a start ranked nodes can have just two states – and in such situations it is possible to achieve similar approximations to Noisy-OR on forward propagation (all cases) and backward propagation (when the effect node is ‘true’). But, crucially, because we will prove that the ranked node function has the conditional anti-correlation property it produces different (but ‘more appropriate’) results for backward propagation when the effect node is ‘false’ and specifically, unlike Noisy-OR, we get the explaining away behavior in all cases.

In what follows, we use a Greek character to denote a variable which needs to be elicited from human experts. Multiple causes are denoted with a column vector x, and the ith element in x is denoted as xi.

**Figure 1** An example BN fragment with ranked nodes. This fragment represent a relationship between n = 3 causes (x1, x2, and x3) and an effect (y). The effect has α = 2 possible states, and each cause has βi = 3 possible states.

Formally, the ranked-node function models a causal relation such as that represented in Figure 1. In this causal relation, a single effect y can be caused by any of n causes (x1, x2, ..., xn). An effect can take α possible states (s1, s2, ..., sα). Typically, when α = 3, s1 is “low”, s2 is “medium” and s3 is “high”. Similarly, the ith cause xi can take βi possible states (s1, s2, ..., sβi).

Probability that xi is si is modelled with a latent variable \( x_i \) and is defined as

\[
p(x_i = s_i^n) = \frac{1}{\beta_i} \leq \frac{j}{\beta_i}
\]

where \( x_i \sim N(\mu_i, \sigma_i^2) \)

Here, N indicates a normal distribution truncated within the range of (0,1). Similarly, probability that y is s_j is modelled with a latent variable \( y \) and is defined as

\[
p(y = s_j^n) = \frac{1}{\alpha} \leq \frac{j}{\alpha}
\]

where \( y \) can be modelled with one of several functions [6], and in this paper, we consider the weighted sum function:

\[
y|x \sim N(\omega^T \bar{x}, \tau)
\]

Here the weight vector \( \omega \) satisfies \( \omega_i > 0 \) ∀i, and \( \tau (\tau > 0) \) is the uncertainty parameter, elicited from human experts. When \( \sum_i \omega_i = 1 \), this weighted sum function is called the weighted mean function or WMEAN in [6]. The graphical representation of \( \bar{x} \) and \( y \) is provided in Figure 2.
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2018.2873314, IEEE Transactions on Knowledge and Data Engineering.

III. CONDITIONAL ANTI-CORRELATION

In this section, we show that given any value of \( \tilde{y} \), elements in \( \tilde{x} \) are anti-correlated. To this end, we first consider a non-truncated multivariate normal distribution and exploit the fact that sign of covariance in multivariate normal distribution does not change after truncation. First, let us introduce new variables \( \tilde{x} \) and \( \tilde{y} \):

\[
\tilde{x} \sim N(\mu, \Sigma) \\
\tilde{y} \sim N(\omega^T \tilde{x}, \tau)
\]

where \( N \) indicates a non-truncated normal distribution, and \( \Sigma \) is a covariance matrix. As we obtain \( \tilde{x} \) after truncating \( x \), and the covariance matrix is a diagonal matrix

\[
\Sigma = \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n^2 \end{bmatrix}
\]

Now to express \( [\tilde{x}, \tilde{y}]^T \) with multivariate normal distribution, we consider the cross-covariance between \( \tilde{x} \) and \( \tilde{y} \). The straightforward calculus shows that:

\[
\text{Cov} (\tilde{x}, \tilde{y}) = E[(\tilde{x} - E[\tilde{x}])(\tilde{y} - E[\tilde{y}])] = E[\tilde{x}\omega^T \tilde{x}] - \mu \omega^T \mu = \Sigma \omega
\]

Similarly, the variance of \( \tilde{y} \) is given by

\[
\text{Var}(\tilde{y}) = E[(\tilde{y} - E[\tilde{y}])^2] = \tau + \omega^T \Sigma \omega
\]

Thus,

\[
[\tilde{x}, \tilde{y}] \sim N \left( \begin{bmatrix} \mu \\
\omega^T \mu \end{bmatrix}, \begin{bmatrix} \Sigma & \Sigma \omega \\
\omega^T \Sigma & \tau + \omega^T \Sigma \omega \end{bmatrix} \right)
\]

Then, the conditional distribution of \( \tilde{x} \) given \( \tilde{y} = z \) follows multivariate normal distribution, whose mean is

\[
E[\tilde{x} | \tilde{y} = z] = \mu + \frac{z - \omega^T \mu}{\tau + \omega^T \Sigma \omega} \Sigma \omega
\]

And the covariance matrix is given by

\[
\text{Cov}(\tilde{x} | \tilde{y} = z) = \Sigma - \frac{\Sigma \omega \omega^T \Sigma}{\tau + \omega^T \Sigma \omega}
\]

(see Section 3.4 in [11] and Section 2.5 in [12], for more detailed derivations).

Recall the matrix \( \Sigma \) is a diagonal matrix, and the \( i \)th diagonal element is \( \sigma_i^2 \). Then, the covariance between \( \tilde{x}_i \) and \( \tilde{x}_j \) (\( i \neq j \)) is

\[
\frac{\sigma_i^2 \omega_i \omega_j}{\tau + \omega^T \Sigma \omega} < 0
\]

As \( \omega_i > 0 \) \( \forall i \) and \( \tau > 0 \), the above covariance is negative and also independent of the conditioned value of \( \tilde{y}, z \).

The sign of the covariance does not change after truncation [13], and thus, the above results indicate that \( \tilde{x}_i \) and \( \tilde{x}_j \) (\( i \neq j \)) are also anti-correlated once conditioned on \( \tilde{y} \).

This anti-correlation means that once conditioned on \( y, x_i \) and \( x_j \) (\( i \neq j \)) are less likely to be in the same state: for example, if \( x_j \) is \( s_j^1 \) (e.g., low), \( x_j \) is more likely to be \( s_j^1 \) (e.g., high). Therefore the above derivation shows that the causes are always anti-correlated once conditioned on the effect.

We have thus far considered only the weighted sum function, but other functions (e.g., weighted min) have also been proposed to model certain types of relations [6]. Unlike the weight sum function, however, the other functions do not produce the marginal distribution of \( \tilde{y} \) that follows a normal distribution. This non-normal distribution makes the derivation similar to the above difficult.

IV. IMPLICATIONS

As we discussed in the introduction, the conditional inter-causal anti-correlation in the ranked-node function produces the explaining-away behavior, whether an effect is observed to be true/present or false/absent. This conditional anti-correlation makes the ranked-node function an attractive choice in many applications. Here we provide an example to demonstrate that it is a natural (and superior) alternative to Noisy-OR even in the standard binary variable situation. In particular, we demonstrate that the ranked-node function preserves basic Noisy-OR behavior when the effect is true but breaks basic Noisy-OR behavior when the effect is false.

Figure 3 displays two networks, one with the Noisy-OR (the left panel) and the other with the ranked-node (the right panel). Figure 4 applies the models with these exact same parameters to the example of the security breach problem described in the Introduction. The figures shows that, with appropriately chosen parameters, the network with the ranked-node can closely match the network with the Noisy-OR, when all the causes and effect are binary-valued. These two networks are used in this section to contrast the Noisy-OR and the ranked-node functions. Below, we examine the two networks in Figure 3.

For the ranked-node function, we used the weighted sum function, as opposed to its more popular special case, weighted mean (i.e., WMEAN) function. This is because flexibility of the weighted sum function is required to produce the conditional probability table that is similar to the one in the Noisy-OR network. We note, however, that the weighted mean function is often sufficient to produce a useful conditional probability table in our practical experience [6].

The ranked-node function preserves the Noisy-OR behavior for forward inference: when a cause is observed to be true/high (false/low) the probability of effect being true/high (false/low) increases, and the increment is larger for a cause with a greater weight. In the network with the Noisy-OR in Figure 3, the probability of effect being true increases from .76 to .96, .89, and .84 when \( x_1, x_2 \) and \( x_3 \)

---

1 All BN calculations are performed using AgenaRisk (version 7 revision 4291). For ranked nodes this involves approximation of the exact solution.
are observed to be true respectively. Similarly with the ranked-node, the probability of effect being high increases from .76 to .94, .89 and .85 when \( x_1, x_2 \) and \( x_3 \) are observed to be high respectively. Also when all causes are observed to be false, the probability the effect is false can be non-zero. It is .10 with the Noisy-OR and .21 with the ranked-node function in Figure 3.

The ranked-node function also preserves the Noisy-OR behavior for backward inference when an effect is true/high: when a cause is subsequently observed as true/high the probability the remaining causes are true decrease. Instead when a cause is subsequently observed as false/low, the probability the remaining causes are true/high increase (see Table 1).

The ranked-node function, however, breaks the Noisy-OR behavior for backward inference when the effect is observed as false/low and hence provides the missing explaining away behavior: Specifically, when a cause is subsequently observed as true, the probability that the remaining causes are true decreases (instead of remaining unchanged as they do under Noisy-OR) and when a cause is subsequently observed as false, the probability that the remaining causes are true increases (instead of remaining unchanged as they do under Noisy-OR; see Table 2 for illustrations).

Further using the ranked-node with the weighted mean function, the above behavior is generalized to multi-valued causes and effect. When a cause is observed to be at a certain state (e.g., medium) the probability of effect being at the same state increases, and the increment is larger for a cause with a greater weight. When all causes are lowest, the probability of an effect being lowest is non-zero. Also when an effect is observed to be at a certain state, the probability of each cause being at the same state increases.

V. CONCLUSION

While the ranked nodes and their associated functions are now widely known and used, this paper is the first to prove that the weighted sum function for ranked nodes satisfies the conditional anti-correlation property. The weighted sum function is a generalization of the more popular weighted mean function. Given that ranked nodes are used to approximate the conditional probability tables for relations that would otherwise be impossible to elicit completely, this property is crucial. That is because in many situations full ‘explaining away’ behavior is required for the relation.

In contrast, the Noisy-OR (and its various generalizations to multiple-state variables) does NOT have the conditional anti-correlation property. Instead, the Noisy-OR has the conditional inter-causal independence property, which means that there is no explaining away behavior when the effect variable is observed as false.

As we demonstrated in Section IV, the ranked node function breaks the behavior of Noisy-OR only when an effect is observed as false and generalizes the behavior to situations where causes and/or effects are measured on an ordinal, rather than Boolean, scale. Thus, with the ranked-node function, a practitioner can expect the explaining-away behavior, whether an effect is observed to be true or false or whether causes and an effect are binary or ordinal multiple-valued. Hence, although there are situations where Noisy-OR is an adequate practical approximation for modelling the effect of independent causal variables [3], ranked-nodes are more generally applicable and are implemented in a widely available BN software package [10]

REFERENCES

Figure 3 Two closely matching Bayesian networks. The left panel shows the network constructed with the Noisy-OR, and the right panel shows the network constructed with the ranked-node function. The parameters for the Noisy-OR is as follows: $x_1$, $x_2$, and $x_3$ cause $y$ to be true with .9, .7, and .5 probabilities and the leak parameter is set at .1. Thus, the probability of $y$ being true when all the three causes are true is given by $1 - (1 - 0.1) \times (1 - 0.9) \times (1 - 0.7) \times (1 - 0.5)$. The parameter for the ranked-node is as follows: the weight vector $\omega$ is [9/15, 7/15, 5/15] and the uncertainty parameter $\tau$ is set at 0.01. The same parameters are used for the practical example in Fig. 4.

<table>
<thead>
<tr>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>False</td>
<td>True</td>
<td>False</td>
</tr>
<tr>
<td>50%</td>
<td>50%</td>
<td>24.13%</td>
</tr>
<tr>
<td>False</td>
<td>True</td>
<td>False</td>
</tr>
<tr>
<td>50%</td>
<td>50%</td>
<td>24.02%</td>
</tr>
</tbody>
</table>

Table 1: Backward Inference with the Presence of Effect.

<table>
<thead>
<tr>
<th>Prior</th>
<th>Noisy-OR</th>
<th>Ranked-Node</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(x_1)$</td>
<td>.50</td>
<td>.63 (0.0)</td>
</tr>
<tr>
<td>$p(x_2)$</td>
<td>.50</td>
<td>.59 (.71)</td>
</tr>
<tr>
<td>$p(x_3)$</td>
<td>.50</td>
<td>.55 (.63)</td>
</tr>
</tbody>
</table>

Each row shows the probability of each cause to be true/high when the effect is true/high. The digits in brackets represent the probability after observing $x_1$ to be false/low.

Table 2: Backward Inference with the Absence of Effect.

<table>
<thead>
<tr>
<th>Prior</th>
<th>Noisy-OR</th>
<th>Ranked-Node</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(x_1)$</td>
<td>.50</td>
<td>.09 (0.0)</td>
</tr>
<tr>
<td>$p(x_2)$</td>
<td>.50</td>
<td>.23 (.23)</td>
</tr>
<tr>
<td>$p(x_3)$</td>
<td>.50</td>
<td>.33 (.33)</td>
</tr>
</tbody>
</table>

Each row shows the probability of each cause to be true/high when the effect is true/high. The digits in brackets represent the probability after observing $x_3$ to be false/low.
Figure 4 There are three known different types of threats that can cause a security breach. Using the same model parameter values as those defined in Figure 3 (so the Noisy-OR weights for the 3 threats are respectively 0.9, 0.7 and 0.5), this example demonstrates how ‘explaining away’ when the effect is observed as false fails with the Noisy-OR model but works in the similar Ranked nodes model.