International Journal of Approximate Reasoning 61 (2015) 1–15

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Propagation effects of model-calculated probability values in Bayesian networks

S.P.D. Woudenberg *, L.C. van der Gaag

Utrecht University, PO Box 80.089, 3508 TB Utrecht, The Netherlands

ARTICLE INFO

Article history:
Received 24 November 2014
Received in revised form 13 March 2015
Accepted 13 March 2015
Available online 21 March 2015

Keywords:
Bayesian network engineering
Probabilistic causal interaction model
(Leaky) noisy-OR model
Noisy-MAX model
Sensitivity analysis

ABSTRACT

Probabilistic causal interaction models have become quite popular among Bayesian-network engineers as elicitation of all probabilities required often proves the main bottleneck in building a real-world network with domain experts. The best-known interaction models are the noisy-OR model and its generalisations. These models in essence are parameterised conditional probability tables for which just a limited number of parameter probabilities are required. The models assume specific properties of intercausal interaction and cannot be applied uncritically. Given their clear engineering advantages however, they are subject to ill-considered use. This paper demonstrates that such ill-considered use can result in poorly calibrated output probabilities from a Bayesian network. By studying, in an analytical way, the propagation effects of noisy-OR calculated probability values, we identify conditions under which use of the model can be harmful for a network's performance. These conditions demonstrate that use of the noisy-OR model for mere pragmatic reasons is sometimes warranted, even when the model's underlying assumptions are not met in reality.

1. Introduction

When building a Bayesian network with the help of domain experts, the process of eliciting all probabilities required for its quantification is generally considered the most daunting and time consuming among the engineering tasks involved [10,26]. The task is often impeded by the experts feeling uncomfortable with providing concrete numbers to describe their knowledge and experience [12]. In order to reduce the amount of time spent on probability elicitation and to alleviate the burden for the experts involved, often probabilistic causal interaction models are used for the quantification task [13,20,21,25]. These models can in general be looked upon as parameterised conditional probability tables for the effect variable of a causal mechanism with multiple cause variables. The models require a limited number of parameters, from which the values for all remaining probabilities in the table of the effect variable are readily calculated. The rules provided for this purpose are derived from properties of probabilistic interaction which are assumed to hold among the variables of a mechanism. Since only a limited number of parameters need be provided, use of a probabilistic causal interaction model often implies a substantial reduction of the number of probabilities to be assessed explicitly by experts.

The most popular among the causal interaction models are the noisy-OR model and its generalisations [8,14,25]. Various empirical studies have been conducted with the noisy-OR model specifically, to investigate the occurrence of its underlying pattern of causal interaction in reality [32], and to study the effect of its use on the process of knowledge elicitation [14,31];

* Corresponding author.
E-mail addresses: S.P.D.Woudenberg@uu.nl (S.P.D. Woudenberg), L.C.vanderGaag@uu.nl (L.C. van der Gaag).

http://dx.doi.org/10.1016/j.ijar.2015.03.005
0888-613X/

empirical studies have further been conducted to gain insight in the possible effects of use of the model on the performance of Bayesian networks developed for practical applications [1,2,23]. The overall conclusion from these studies is that use of the noisy-OR model serves to considerably reduce the burden of probability elicitation, without severely hampering the performance of a network.

Despite its clear advantages for Bayesian-network engineering, the noisy-OR model cannot be applied uncritically. The probability values calculated from the model's rules can only be considered appropriate approximations of the true probabilities if the properties underlying the noisy-OR model actually hold in the domain of application. In practise unfortunately, Bayesian-network engineers are not all aware of these properties, which makes the noisy-OR model subject to ill-considered use. The consistently positive results of using the model in practical applications moreover, have led to the suggestion that Bayesian networks are quite robust against the inaccuracies that are induced in their conditional probability tables by using the noisy-OR model.

In this paper, we show that ill-considered use of the noisy-OR model *can* result in poorly calibrated output probabilities. For this purpose, we study the propagation effects of noisy-OR calculated probability values that deviate from the true probability values. More specifically, we employ sensitivity-analysis techniques for determining when use of the model may harm a network's output probabilities, and hence its overall performance. The conclusions of our investigations do not contradict the findings from earlier experimental studies which have led to the suggestion that the noisy-OR model can be used without severely hampering the performance of the network, even when the underlying assumptions of the model are not met in reality. Our investigations serve to provide a formal underpinning of these findings and show that use of causal interaction models such as the noisy-OR model for mere pragmatic reasons may often be warranted.

The paper is organised as follows. In Section 2 we introduce our notational conventions, and briefly introduce Bayesian networks and the noisy-OR model. We further review the technique of sensitivity analysis of Bayesian networks which will be used in Section 3 for studying in an analytical way, the propagation effects of noisy-OR calculated values on output probabilities computed from a basic causal mechanism. Our results are extended in Section 4 to apply to more involved Bayesian networks. The results of our analyses of various generalisations of the noisy-OR model are addressed in Section 5. The paper ends with our concluding observations in Section 6.

2. Preliminaries

In this section, we begin by introducing our notational conventions and briefly review Bayesian networks in general; more elaborate introductions are found in various textbooks [7,15,16]. We then describe the noisy-OR model in detail. Since in this paper we will exploit insights from sensitivity analysis of Bayesian networks, preliminaries from this field are provided as well.

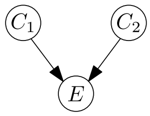
2.1. Bayesian networks

We consider (discrete) random variables, which are denoted by capital letters and whose values are denoted by indexed small letters. We assume that a random variable E adopts a value e^i from its possible values $e^0, \dots, e^m$, $m \geq 1$; for a binary variable E, we will write $\bar{e}$ and e to denote its two values e^0 and e^1, representing the absence and the presence of the modelled concept respectively. Sets of random variables are denoted by bold-face capital letters and joint value combinations for these sets are written in bold-face small letters. We further consider joint probability distributions Pr over our sets of random variables.

A Bayesian network is a graphical representation of a joint probability distribution over a set of variables. The network includes a directed acyclic graph in which the nodes represent the variables and in which the arcs describe the (in)dependency relation among the variables. We will use the term variable to refer to a random variable itself and to the node representing it, interchangeably. The arcs of the graph of a Bayesian network are often looked upon as capturing a causal relationship between the connected variables; although we will not make any claims with respect to a causal interpretation, we will adopt the terminology involved. If the graph includes an arc pointing from a variable C to a variable E, we say that C is a cause variable for E, and E is an effect variable of C. The phrase causal mechanism is used to refer to a single effect variable with its associated n cause variables, $n \geq 0$, in the graph of a Bayesian network. In this paper we will focus mostly on causal mechanisms with two or more cause variables $C_1, \dots, C_n$, $n \geq 2$, unless stated otherwise.

In addition to the graphical structure, a Bayesian network includes a conditional probability table, or CPT, for each variable. For a variable E, this table specifies the conditional probability distributions $\Pr(E \mid \mathbf{c})$ over E given all possible joint value combinations $\mathbf{c}$ for its cause variables. Fig. 1 shows an example causal mechanism; the figure depicts the graphical structure of the mechanism, with the binary effect variable E and the two binary cause variables C_1 and C_2, and shows the conditional probability table for E. Note that the table specifies four conditional probability distributions for the effect variable, which number is exponential in the number of cause variables discerned.

A Bayesian network describes a unique joint probability distribution and hence provides for computing any prior or posterior probability of interest over its variables. The problem of establishing probabilities from a Bayesian network was proven to be NP-hard in general by Cooper [4]; Roth later showed #P-hardness [27]. For networks of which the graphical structure has bounded treewidth, the problem can be solved in polynomial time, that is, polynomial in the network's size.



C_1	C_2	E	
		Present	Absent
present	present	0.99	0.01
present	absent	**0.80**	0.20
absent	present	**0.95**	0.05
absent	absent	0.00	1.00

Fig. 1. The graph of an example causal mechanism (*left*), along with the conditional probability table for its effect variable E (*right*).

Various algorithms have been designed for this purpose, among which the junction-tree propagation algorithm is the most efficient to date [15,18].

2.2. The noisy-OR model

When constructing a Bayesian network for a real-world application, the first step is to configure a directed acyclic graph to describe the probabilistic independencies between the relevant variables. Subsequently, conditional probability tables need be specified for all variables, that is, for the effect variables of all causal mechanisms in the graph. For an effect variable E, such a table details an exponential number of conditional probability distributions $\Pr(E \mid \mathbf{c})$ over E, given all possible value combinations $\mathbf{c}$ for the cause variables $C_1, \ldots, C_n$, $n \geq 0$, of E. Experience shows that obtaining all required numbers is a daunting and time consuming task, especially if these numbers have to be elicited from domain experts. The noisy-OR model has been designed to alleviate this burden for mechanisms involving binary variables only [25].

The noisy-OR model in essence is a parameterised conditional probability table for the effect variable E of a causal mechanism with two or more cause variables. The model takes for its parameters the conditional probabilities $\Pr(e \mid \bar{c}_1, \ldots, \bar{c}_{j-1}, c_j, \bar{c}_{j+1}, \ldots, \bar{c}_n)$, $j = 1, \ldots, n$, of the effect e to arise in the presence of the single cause c_j; it thus takes as many parameters as the number of cause variables in the mechanism to which it is applied. The remaining probabilities need not be provided explicitly, but are generated by the model: it sets the conditional probability $\Pr(e \mid \bar{c}_1, \ldots, \bar{c}_n)$ of the effect e to arise in the absence of all known causes to zero, and defines the remaining probabilities through

$$\Pr(e \mid \mathbf{c}) = 1 - \prod_{j \in J} (1 - \Pr(e \mid \bar{c}_1, \ldots, \bar{c}_{j-1}, c_j, \bar{c}_{j+1}, \ldots, \bar{c}_n))$$

where J is the set of indices of the cause variables C_j that are marked as being present in the joint value combination $\mathbf{c}$. The probability of the effect e arising in the presence of multiple simultaneous causes thus is calculated from the probabilities of e given each of these causes separately. We demonstrate the use of the noisy-OR model with an example.

Example 1. We consider the causal mechanism from Fig. 1 on the left, and take the cause variables C_1 and C_2 to represent the intake of alcohol and of the GHB party drug respectively. Either substance can cause a stimulating effect, which is represented by the effect variable E. Suppose that a domain expert is able to assess the probability of a stimulating effect arising from either substance, yet feels uncomfortable attaching a concrete number to the joint effect of concurrent intake of both substances. The conditional probability table from Fig. 1 on the right then results from applying the noisy-OR model to the effect variable E of the mechanism. The probabilities printed in bold are the parameter probabilities provided by the expert; the probability $\Pr(e \mid \bar{c}_1, \bar{c}_2)$ is set to zero and the probability $\Pr(e \mid c_1, c_2)$ is calculated by the noisy-OR model from the two provided parameter probabilities. We note that the probability of the stimulating effect arising with the concurrent intake of both substances is calculated to be higher than the probability of the effect to arise upon the intake of either one of these stimuli.

Underlying the noisy-OR model are two basic assumptions concerning the probabilistic interactions among the variables of a causal mechanism. The assumption of *accountability* states that the effect e of a mechanism cannot arise as long as none of its causes are present; this assumption thus sets $\Pr(e \mid \bar{c}_1, \ldots, \bar{c}_n) = 0$. The assumption of *exception independence* pertains to the exception mechanisms that may inhibit the effect to arise in the presence of a cause. Each arc $C_j \rightarrow E$ in a causal mechanism can be viewed as an essentially deterministic causal relation between the variables C_j and E which sometimes is perturbed, that is, on exception, the effect does not show even though the cause is present. The exception mechanism involved can be made explicit by a so-called inhibitor variable I_j with a probability distribution capturing the probability of the exception occurring. The property of exception independence now states that the inhibitor variables I_j for a causal mechanism are mutually independent. For further details of the noisy-OR model and its underlying properties of intercausal interaction, we refer to [25].

Since its introduction, the noisy-OR model has given rise to several variants and generalisations, the most prominent among these being the leaky noisy-OR model [14] and the noisy-MAX model [8,14]. With the leaky noisy-OR model the probability of the effect occurring in the absence of all explicitly modelled causes is assumed to be non-zero and is captured explicitly by an additional parameter. The noisy-MAX model generalises the noisy-OR model to discrete variables with arbitrary numbers of values. We will elaborate on these generalisations in Section 5.

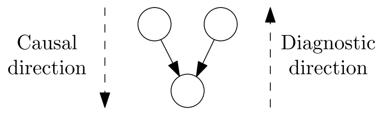


Fig. 2. A causal mechanism with dashed arrowing indicating the directions of causal and diagnostic propagation.

2.3. Sensitivity analysis of Bayesian networks

The probability values specified in the conditional probability tables of a Bayesian network in general are either estimated from suitable data or assessed by domain experts. Either way, the probability values obtained are likely to include at least some degree of inaccuracy. To study the possible effects of these inaccuracies on its output, a Bayesian network can be subjected to a sensitivity analysis [3,5,22]. Informally spoken, such an analysis amounts to systematically varying one of the network's probability values and computing the effect of the variation on an output probability of interest. The result of a sensitivity analysis is a sensitivity function $f(x)$ that expresses the network's output probability in the probability value x being varied [6].

A sensitivity function $f(x)$ cannot be arbitrarily shaped, but is either a linear function or a (rectangular) hyperbolic function in the probability value x under study [6]. More specifically, for an output probability pertaining to a variable without any observed descendants, the sensitivity function is linear in the probability value being varied and hence takes the form $f(x) = \alpha \cdot (x + \beta)$ where the constants α and β are built from the network's non-varied probability values. The linear function arises basically from studying the effects of inaccuracies in the probability values for the (possibly indirect) causes of the variable of interest, and hence is said to capture the effects of causal propagation; Fig. 2 visualises this direction of propagation.

Hyperbolic sensitivity functions arise from studying the effects of inaccuracies in a network's probability values on output probabilities for variables with observed successors; we say that these functions capture the effects of diagnostic propagation. More formally, a hyperbolic sensitivity function $f(x)$ takes the following form:

$$f(x) = \frac{\alpha' \cdot x + \beta'}{\alpha \cdot x + \beta}$$

where the constants $\alpha, \alpha', \beta, \beta'$ are again built from the non-varied probability values of the network. We note that both x and $f(x)$ represent probabilities and hence range from 0 to 1. For studying the effects of variation of a probability value x on a network's output therefore, only a fragment of one of the branches of the hyperbola $f(x)$ is relevant. The window defined by the range $[0, 1]$ for both $f(x)$ and x, is called the unit window for the sensitivity function [11].

The hyperbolic function $f(x)$ can be written in the following more suitable form [11]:

$$f(x) = \frac{r}{x - s} + t$$

with

$$s = -\frac{\beta}{\alpha} \quad , \quad t = \frac{\alpha'}{\alpha} \quad \text{and} \quad r = \frac{\beta' \cdot \alpha - \alpha' \cdot \beta}{\alpha^2}$$

In this form, the constant s denotes the vertical asymptote of the hyperbola and t denotes its horizontal asymptote. The constants s and t determine the general shape of the hyperbola, and the quadrants in which its two branches lie more specifically. The constant r defines the locations of the vertices of the two branches of a hyperbola in general: the vertex of a hyperbola branch is the point where the absolute value of its first derivative equals 1. For a sensitivity function, the vertex is one of the four points $(s \pm \sqrt{|r|}, t \pm \sqrt{|r|})$, dependent of the quadrant of the branch under study; we note that the vertex need not lie within the unit window to which the sensitivity function is restricted.

While recent research in sensitivity analysis of Bayesian networks has focused mostly on varying a single probability value, it is possible also to perform a sensitivity analysis in which multiple probability values are varied simultaneously; such an analysis serves to reveal the joint combined effect of variation of these values on the output probability of interest. For a higher-order sensitivity analysis in essence similar observations hold as reviewed above for a one-way analysis in which a single probability value is varied. For a two-way sensitivity analysis for example, the sensitivity function $f(x, y)$ takes the form of a quotient of two functions that are bi-linear in the two probability values x and y under study [5]:

$$f(x, y) = \frac{\alpha'_1 \cdot x \cdot y + \alpha'_2 \cdot x + \alpha'_3 \cdot y + \beta'}{\alpha_1 \cdot x \cdot y + \alpha_2 \cdot x + \alpha_3 \cdot y + \beta}$$

where the constants involved again are built from the non-varied probability values of the network at hand; the function $f(x, y)$ reduces to a bi-linear function upon causal propagation. The cross-product terms of the function, involving both probability values x and y, capture the interaction effects of the two values on the output probability of interest. We note that this information cannot be revealed by one-way analyses for each of the two probability values separately. A sensitivity function which results from varying two probability values from the same conditional probability table will lack such cross-product terms whenever the two probabilities relate to logically incompatible conditions.

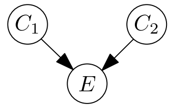


Fig. 3. A basic causal mechanism with the effect variable E and cause variables C_1, C_2.

3. Basic propagation effects of noisy-OR calculated probability values

The noisy-OR model is highly popular among Bayesian-network engineers, because its use serves to substantially reduce the number of conditional probabilities for which values need actually be obtained. Experiences with the model have further suggested that even if the noisy-OR calculated probability values deviate from the true probabilities, the overall performance of the network at hand does not substantially degrade. In view of these experiences, we consider again Example 1, in which we modelled the stimulating effect from alcohol use and from taking the GHB party drug. We recall that the noisy-OR model calculated a probability value of 0.99 for a stimulating effect arising when both substances are taken. In reality however, this probability is much smaller, as concurrent intake of alcohol and GHB is known to have a depressing rather than a stimulating effect. We would like to note that, in general, experts may be quite reluctant to provide a concrete probability value for a combined effect, mostly because they have little experience with observing the joint effect of the simultaneous presence of the causes involved. They will often be able, however, to indicate, based upon their domain knowledge, whether the combined effect will be stronger or weaker than the effect of each cause separately. In this section, we will study in an analytical way what the consequences can be on the output of a Bayesian network, when a noisy-OR calculated value deviates from the true probability. We will focus to this end on a basic causal mechanism for which we investigate the effects upon causal propagation and upon diagnostic propagation separately; in Section 4, we extend our results to more involved network snippets.

3.1. Effects upon causal propagation

We consider the conditional probability tables for the three variables of the basic mechanism from Fig. 3. We assume that the prior probability distributions for the cause variables C_1 and C_2 are non-degenerate, that is, we assume that the probabilities $\Pr(c_i)$ and $\Pr(\bar{c}_i)$, $i = 1, 2$, are non-zero. We further assume that the conditional probability $\Pr(e \mid \bar{c}_1, \bar{c}_2)$ for the effect variable E equals 0 and that values for the probabilities $\Pr(e \mid \bar{c}_1, c_2)$ and $\Pr(e \mid c_1, \bar{c}_2)$ have been obtained from data or from experts. We now focus on the value of the fourth probability $\Pr(e \mid c_1, c_2)$ of the conditional probability table $\Pr(E \mid C_1, C_2)$ for E. We begin by investigating the effects of possible deviations from the true probability value of $\Pr(e \mid c_1, c_2)$ on the prior distribution of E. We have that the probability $\Pr(e)$ equals:

$$\Pr(e) = \Pr(e \mid c_1, c_2) \cdot \Pr(c_1) \cdot \Pr(c_2) + \Pr(e \mid \bar{c}_1, c_2) \cdot \Pr(\bar{c}_1) \cdot \Pr(c_2) + \Pr(e \mid c_1, \bar{c}_2) \cdot \Pr(c_1) \cdot \Pr(\bar{c}_2)$$

where the cause variables C_1 and C_2 have been taken to be mutually independent a priori, as read from the mechanism under study. We now write $\Pr(e)$ as a function of the value x of the probability $\Pr(e \mid c_1, c_2)$. The result is a linear function in x:

$$\begin{aligned}\Pr(e)(x) &= x \cdot \Pr(c_1) \cdot \Pr(c_2) + \Pr(e \mid \bar{c}_1, c_2) \cdot \Pr(\bar{c}_1) \cdot \Pr(c_2) + \Pr(e \mid c_1, \bar{c}_2) \cdot \Pr(c_1) \cdot \Pr(\bar{c}_2) \\ &= \alpha \cdot (x + \beta)\end{aligned}$$

where

$$\begin{aligned}\alpha &= \Pr(c_1) \cdot \Pr(c_2) \\ \beta &= \Pr(e \mid \bar{c}_1, c_2) \cdot \lambda_{C_1}^{-1} + \Pr(e \mid c_1, \bar{c}_2) \cdot \lambda_{C_2}^{-1}\end{aligned}$$

with $\lambda_{C_i}^{-1} = \Pr(\bar{c}_i)/\Pr(c_i)$, $i = 1, 2$. The reciprocal likelihood ratio $\lambda_{C_i}^{-1}$ expresses the degree to which the cause C_i is more likely to be absent than to be present. A large ratio $\lambda_{C_i}^{-1} \in (1, \infty)$ reflects a high probability of C_i being absent; a small ratio $\lambda_{C_i}^{-1} \in (0, 1)$ on the other hand indicates that C_i is more likely to be present.

The gradient α of the function $\Pr(e)(x)$ describes the effect that a deviation from the true value of $\Pr(e \mid c_1, c_2)$ can have on the prior probability of the effect e arising. This gradient is restricted by $0 \leq \alpha \leq 1$ and depends solely on the prior probabilities of each of the two causes being present. We note that large values for α are found only with high probabilities of the presence of these causes. The constant β in the offset of the function equally depends on the prior probabilities $\Pr(c_1)$ and $\Pr(c_2)$, yet is also dependent of the two probabilities $\Pr(e \mid \bar{c}_1, c_2)$ and $\Pr(e \mid c_1, \bar{c}_2)$ of the effect arising in the presence of a single cause. We note that since the overall offset $\alpha \cdot \beta$ is necessarily restricted to the interval [0, 1], the range of possible values for the constant β is constrained by the value of the gradient α.

To illustrate the above considerations, Fig. 4 depicts, for the basic mechanism under study, two example functions expressing the prior probability of interest $\Pr(e)$ in terms of the value x for the conditional probability $\Pr(e \mid c_1, c_2)$. For both

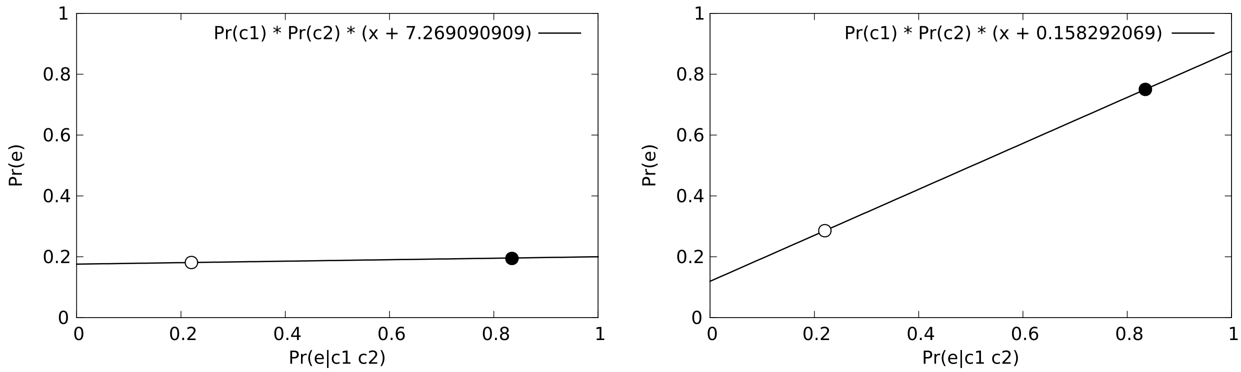


Fig. 4. The effects on $\Pr(e)$ of varying the value of $\Pr(e \mid c_1, c_2)$, given small values (*left*) and given large values (*right*) for $\Pr(c_1)$ and $\Pr(c_2)$; the values for $\Pr(e)$ found with the true probability $\Pr(e \mid c_1, c_2)$ are indicated by the open circles and the output probabilities found with the noisy-OR calculated value for $\Pr(e \mid c_1, c_2)$ are marked by the black dots.

functions, the two probabilities $\Pr(e \mid \bar{c}_1, c_2)$ and $\Pr(e \mid c_1, \bar{c}_2)$ were set to 0.43 and 0.71, respectively. For the function on the left, $\Pr(c_1)$ and $\Pr(c_2)$ were assigned the small values 0.22 and 0.11 respectively, resulting in a small gradient $\alpha = 0.024$. This small gradient conveys the information that even a substantial deviation from the true value of $\Pr(e \mid c_1, c_2)$ will have just a minor effect on the probability of interest. Given the larger values 0.83 and 0.91 for $\Pr(c_1)$ and $\Pr(c_2)$, the effect on $\Pr(e)$ of a deviation from the true probability $\Pr(e \mid c_1, c_2)$ can be more substantial, as is demonstrated by the function on the right; this function has a gradient equal to $\alpha = 0.755$. We now observe that since the two functions describe the effect of any arbitrary deviation from the true value of the probability $\Pr(e \mid c_1, c_2)$, they also capture the effect of using the noisy-OR calculated value for $\Pr(e \mid c_1, c_2)$. Fig. 4 indicates, with each depicted function, the two output probabilities which result from using the true probability value of 0.22 (represented by the open circle) and using the noisy-OR calculated value of 0.83 (represented by the black dot) respectively. The figure thereby illustrates that using the noisy-OR value will have a larger effect on the output probability in view of large prior probabilities for the separate causes than in view of small prior probabilities.

The effects of causal propagation through the basic mechanism are further investigated by assuming the presence of one of the causes, that is, by considering $\Pr(e \mid c_1)$ or $\Pr(e \mid c_2)$ for the probability of interest. We note that assuming absence of one of the causes would be irrelevant for our purposes since the probabilities $\Pr(e \mid \bar{c}_1)$ and $\Pr(e \mid \bar{c}_2)$ are not (algebraically) dependent of the value of the conditional probability $\Pr(e \mid c_1, c_2)$ under study, that is, varying $\Pr(e \mid c_1, c_2)$ cannot influence $\Pr(e \mid \bar{c}_1)$ or $\Pr(e \mid \bar{c}_2)$. As an example we now express the probability of interest $\Pr(e \mid c_1)$ as a function of the probability value $x = \Pr(e \mid c_1, c_2)$, where we again take the two cause variables C_1 and C_2 to be mutually independent a priori:

$$\Pr(e \mid c_1)(x) = \Pr(c_2) \cdot (x + \Pr(e \mid c_1, \bar{c}_2) \cdot \lambda_{C_2}^{-1})$$

Upon comparison with the function $\Pr(e)(x)$ derived above, we note that the two functions differ in the role of the prior probability $\Pr(c_1)$ of the presence of the observed cause: where the gradient of $\Pr(e)(x)$ includes both $\Pr(c_1)$ and $\Pr(c_2)$, the gradient of $\Pr(e \mid c_1)(x)$ no longer reveals a dependency of $\Pr(c_1)$. Analogous observations hold for the function derived for the probability of interest $\Pr(e \mid c_2)$.

We have considered the consequences of deviating noisy-OR probability values upon causal propagation through the basic mechanism; we note that by a deviating noisy-OR value, we refer to a noisy-OR calculated probability value which deviates from the associated true probability value. We have shown that a large deviation from the true value of the probability $\Pr(e \mid c_1, c_2)$ can give a large shift in a prior or posterior probability of the effect only if the yet unobserved causes have large probabilities of being present. The larger these probabilities, the larger the gradient of the associated function will be and the larger the effect on the output probability of interest can become. In view of relatively small probabilities of the (yet unobserved) causes of a mechanism being present therefore, a network engineer may safely apply the noisy-OR model for the conditional probability table of the effect variable. If these probabilities may become substantially larger upon inference however, caution is advised when considering application of the model. Since strong effects may then arise upon causal propagation, the network engineer should verify that the true probability value does not deviate significantly from the noisy-OR calculated one. We would like to emphasise that the extent to which a large shift in the output probability of a mechanism will actually affect the overall performance of the Bayesian network at hand is strongly dependent of the network's graphical structure and (other) parameter probabilities. We will return to this observation in Section 4.

3.2. Effects upon diagnostic propagation

Thus far we examined the consequences that a deviating noisy-OR calculated value can have on a probability of interest which is established by causal propagation through the basic mechanism under study. In this section we investigate the consequences of such a value upon propagation in the diagnostic direction, that is, upon propagating evidence for the effect

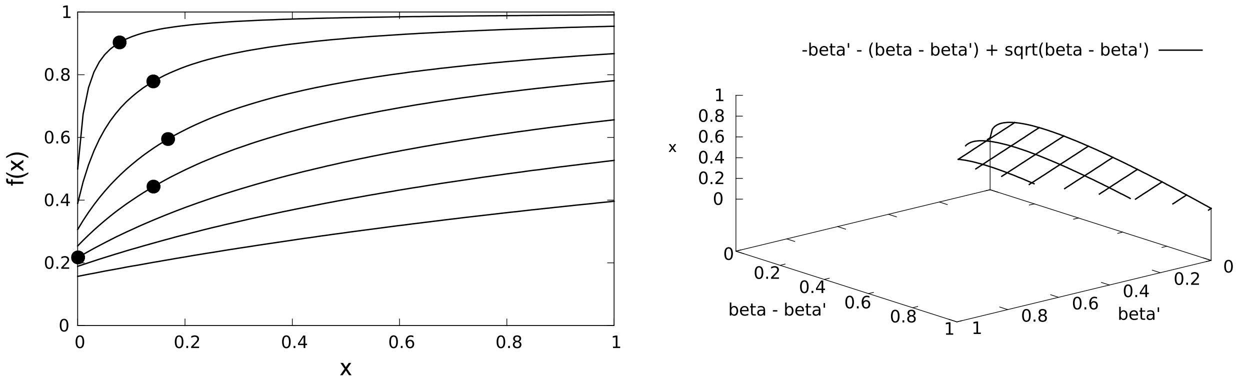


Fig. 5. Several example fourth-quadrant hyperbola branches restricted to the unit window with their vertices indicated by a dot (*left*), and the *x*-coordinate of the vertex, plotted as a function of $\beta - \beta'$ and β' (*right*).

variable to an unobserved cause variable. For this purpose, we consider again the basic causal mechanism from Fig. 3 and express the example output probability $\Pr(c_1 \mid e)$ as a function of the value x of the probability $\Pr(e \mid c_1, c_2)$:

$$\Pr(c_1 \mid e)(x) = \frac{x \cdot \Pr(c_1) \cdot \Pr(c_2) + \Pr(e \mid c_1, \bar{c}_2) \cdot \Pr(c_1) \cdot \Pr(\bar{c}_2)}{\Pr(e)(x)} = \frac{x + \beta'}{x + \beta}$$

where

$$\beta' = \Pr(e \mid c_1, \bar{c}_2) \cdot \lambda_{C_2}^{-1}$$
$$\beta = \Pr(e \mid \bar{c}_1, c_2) \cdot \lambda_{C_1}^{-1} + \Pr(e \mid c_1, \bar{c}_2) \cdot \lambda_{C_2}^{-1}$$

with $\lambda_{C_i}^{-1}$, $i = 1, 2$, as before. To acquire more detailed insight in the values which the constants β and β' can attain, we begin by writing the denominator $\Pr(e)(x)$ as

$$\Pr(e)(x) = \Pr(c_1, e)(x) + \Pr(\bar{c}_1, e)(x)$$

Writing the functions $\Pr(c_1, e)(x)$ and $\Pr(\bar{c}_1, e)(x)$ in terms of the parameter probabilities $\Pr(e \mid c_1, \bar{c}_2)$ and $\Pr(e \mid \bar{c}_1, c_2)$ now gives

$$\Pr(e)(x) = \Pr(c_1) \cdot \Pr(c_2) \cdot \left(x + \Pr(e \mid c_1, \bar{c}_2) \cdot \lambda_{C_2}^{-1} + \Pr(e \mid \bar{c}_1, c_2) \cdot \lambda_{C_1}^{-1}\right)$$

In the function $\Pr(c_1 \mid e)(x)$ therefore, both the numerator and the denominator include the multiplicative term $\Pr(c_1) \cdot \Pr(c_2)$, which can be divided out of the equation. As a result, we find that the function $\Pr(e)(x)$ is proportional to

$$\Pr(e)(x) \propto x + \Pr(e \mid c_1, \bar{c}_2) \cdot \lambda_{C_2}^{-1} + \Pr(e \mid \bar{c}_1, c_2) \cdot \lambda_{C_1}^{-1} \propto (x + \beta') + (\beta - \beta')$$

where

$$\Pr(c_1, e)(x) \propto x + \beta' \quad \text{and} \quad \Pr(\bar{c}_1, e)(x) \propto \beta - \beta'$$

Since the denominator $\Pr(e)(x)$ of the function $\Pr(c_1 \mid e)(x)$ is required to be larger than zero for all values x, we have that $\beta > 0$. From $\Pr(c_1, e)(x)$ expressing a probability for all x we further find that $\beta' \geq 0$. From also $\Pr(\bar{c}_1, e)(x)$ being a probability for all x, we conclude that $\beta \geq \beta'$.

Now, the function $\Pr(c_1 \mid e)(x)$ derived above is hyperbolic in the probability value x, as expected from insights in sensitivity analysis of Bayesian networks in general. Building upon the properties of hyperbolic functions reviewed in Section 2.3, we find that the vertical asymptote s of the function equals $s = -\beta$. As $\beta > 0$, we have that this asymptote lies to the left of the unit window. The function further has its horizontal asymptote at $t = 1$. From these findings we conclude that the function $\Pr(c_1 \mid e)(x)$ is a fragment of a fourth-quadrant hyperbola branch. For illustrative purposes Fig. 5 depicts, on the left, seven example fourth-quadrant branches, restricted to the unit window. The figure shows that while a relatively small deviation in the x-value may have a considerable effect on the output value with some branches, with other branches the output is hardly affected by even large deviations in x.

From studies of sensitivity functions from Bayesian networks in general, we know that the effect of deviations from the true x-value on the output probability is largely dependent of the location of the vertex of the hyperbola branch at hand. In the plot of Fig. 5 on the left, the vertices of the example branches are indicated by a dot if lying within the unit window. In general, the closer the vertex of the hyperbola branch lies to the upper-left corner of the unit window, that is, the closer it is to the point $(0, 1)$, the larger the effect of a deviation from x to smaller values can be. Our function $\Pr(c_1 \mid e)(x)$ has its vertex at

$$(s + \sqrt{|r|}, 1 - \sqrt{|r|}) = \left(-\beta + \sqrt{(\beta - \beta')}, 1 - \sqrt{(\beta - \beta')}\right)$$

The vertex lies within the unit window for values of β, β' with $\beta < \sqrt{(\beta - \beta')} < 1$. From $\beta < \sqrt{(\beta - \beta')}$ and $\beta \geq \beta'$, we find that only relatively small values of β' can result in a vertex with an x-coordinate in the unit range. This observation is supported by the plot of Fig. 5 on the right, which depicts the x-coordinate of the vertex, if within the unit range, as a function of β' and $\beta - \beta'$. The plot on the left in addition provides some examples. The function at the bottom of the plot has the values $\beta' = 0.395$ and $\beta = 2.521$; this function has a vertex to the left of the window as a result of the asymptote $s = -2.521$ and in fact is almost linear within the unit window. In contrast, the function depicted as the third from below has its vertex within the window; it has the smaller value $\beta' = 0.170$ and the value $\beta = 0.782$. Although a relatively small value of the constant β' will give a vertex with a positive x-coordinate, the vertex will approach the upper-left corner of the unit window only if in addition the difference $\beta - \beta'$ is quite small. We note that with a small value of $\beta - \beta'$, the function $\Pr(c_1 \mid e)(x) = (x + \beta')/(x + \beta)$ indeed approaches 1. Once more considering the plot of Fig. 5 on the left, we observe that the vertex of the top function lies close to the upper-left corner of the unit window; this function has $\beta' = 0.009$ and $\beta = 0.019$. From these considerations, we conclude that deviations from the true value x of the probability $\Pr(e \mid c_1, c_2)$ can have large effects on the output probability $\Pr(c_1 \mid e)$ only if both constants β' and β are small. Small values of $\beta' = \Pr(e \mid c_1, \bar{c}_2) \cdot \lambda_{C_2}^{-1}$ are found when at least one of the probabilities $\Pr(e \mid c_1, \bar{c}_2)$ and $\Pr(\bar{c}_2)$ is small; small values of β are found if, in addition, at least one of $\Pr(e \mid \bar{c}_1, c_2)$ and $\Pr(\bar{c}_1)$ is small.

Since the function $\Pr(c_1 \mid e)(x)$ derived above describes the effect of any arbitrary deviation from the true value of the probability $\Pr(e \mid c_1, c_2)$, it also captures the effect of using the noisy-OR calculated value for $\Pr(e \mid c_1, c_2)$ upon inference. The above analysis shows that strong effects on the output probability $\Pr(c_1 \mid e)$ can be expected only if both β' and β have rather small values, that is, if at least one of the following conditions holds:

- both $\Pr(e \mid c_1, \bar{c}_2)$ and $\Pr(e \mid \bar{c}_1, c_2)$ have small values;
- both $\Pr(e \mid c_1, \bar{c}_2)$ and $\Pr(\bar{c}_1)$ are small;
- both $\Pr(e \mid \bar{c}_1, c_2)$ and $\Pr(\bar{c}_2)$ are small;
- both $\Pr(\bar{c}_1)$ and $\Pr(\bar{c}_2)$ are small, that is, both causes are quite likely to be present.

Whether the anticipated effects will actually arise upon propagation depends on the true value of $\Pr(e \mid c_1, c_2)$ and on the noisy-OR calculated one: only if at least one of these values is smaller than the x-coordinate of the vertex of the function can using the noisy-OR value strongly affect the output. In addition to the above considerations, we note that effects on the output probability, albeit weaker ones, may also be found with larger β values in view of a small value of β'.

We have considered the consequences of deviating noisy-OR probability values upon diagnostic propagation through the basic mechanism. We have found, as with causal propagation, that a large deviation from the true probability value can give a large shift in an output probability of interest if the yet unobserved causes have large probabilities of being present. In combination with small values for the noisy-OR parameter probabilities can the effects become especially strong. A network engineer is strongly advised against applying the noisy-OR model for the effect variable if the true conditional probability of the effect arising or the noisy-OR calculated value for this probability is quite small. The noisy-OR model can be more or less safely applied if none of the yet unobserved causes is likely to occur. Once again we would like to emphasise that the extent to which a large shift in the output probability of a mechanism can actually affect the overall performance of a network as a whole, is strongly dependent of the network's graphical structure and (other) parameter probabilities.

4. Propagation effects of noisy-OR calculated values in general

In the previous section, we studied the possible consequences of deviating noisy-OR calculated probability values upon propagation through a causal mechanism with two mutually independent cause variables. We now extend our study to more involved network snippets. We first consider causal mechanisms which involve a direct dependency between their pair of cause variables, and then turn to mechanisms with more than two cause variables. We end by briefly reviewing the possible consequences of deviating noisy-OR values in larger Bayesian networks. Throughout this section we will focus on just the differences in effects between the more involved network snippets and the basic causal mechanism considered in Section 3.

4.1. Causal mechanisms with dependent cause variables

The basic mechanism studied in Section 3 included the independent cause variables C_1 and C_2, and the effect variable E. We now pursue our investigations by no longer assuming a priori independence of the two cause variables. As an example, we study the mechanism with the added extra arc $C_1 \to C_2$. We assume that the prior probability distribution over C_1 and the conditional distributions over C_2 given C_1 are non-degenerate. We further assume that $\Pr(e \mid \bar{c}_1, \bar{c}_2)$ equals zero, and that values for the probabilities $\Pr(e \mid \bar{c}_1, c_2)$ and $\Pr(e \mid c_1, \bar{c}_2)$ have been estimated from data or assessed by experts. We focus again on the fourth probability $\Pr(e \mid c_1, c_2)$ in the conditional probability table for E, and begin again by addressing the consequences for the prior probability $\Pr(e)$, of propagating a value x for $\Pr(e \mid c_1, c_2)$ which deviates from the true value. As with the basic mechanism, we find a linear function expressing the probability $\Pr(e)$ in this value:

$$\Pr(e)(x) = \alpha \cdot (x + \beta)$$

The main difference with the function obtained with the basic mechanism is that the gradient now equals $\alpha = \Pr(c_1) \cdot \Pr(c_2 \mid c_1)$. Since it now takes the dependency between the two cause variables into consideration, α is no longer dependent of the prior probabilities of each of the two causes separately, but of the prior probability of their joint presence instead. The gradient may now attain a large value with a moderately likely cause c_2 which becomes quite likely in the presence of c_1; a dependency capturing a negative probabilistic influence between two likely causes on the other hand, may forestall a large gradient. The propagation effects in the causal direction are otherwise in line with the effects with the basic mechanism: a large deviation from the true value of the probability $\Pr(e \mid c_1, c_2)$ can give a large shift in the prior probability of the effect only if the two causes have a large probability of being present simultaneously.

For investigating the effects upon diagnostic propagation through the extended mechanism, we consider again the example posterior probability of interest $\Pr(c_1 \mid e)$, and express it as a function of the value x of $\Pr(e \mid c_1, c_2)$. The following hyperbolic function is found:

$$\Pr(c_1 \mid e) = \frac{x + \beta'}{x + \beta}$$

with $0 \leq \beta' \leq \beta$, $\beta > 0$, as before. While with the basic mechanism, the constants β and β' included the reciprocal likelihood ratios for the two cause variables separately, they now involve the reciprocal *conditional* likelihood ratios

$$\lambda^{-1}_{C_i|C_j} = \frac{\Pr(\bar{c}_i \mid c_j)}{\Pr(c_i \mid c_j)}$$

for C_i given c_j, $i, j = 1, 2$, $i \neq j$. In view of these conditional ratios however, similar observations hold as with the basic mechanism. Deviations from the true value x of the probability $\Pr(e \mid c_1, c_2)$ can have a large effect on the output probability $\Pr(c_1 \mid e)$ only if both constants β and β' are small, that is, if at least one of the following conditions hold:

- both $\Pr(e \mid c_1, \bar{c}_2)$ and $\Pr(e \mid \bar{c}_1, c_2)$ have small values;
- both $\Pr(e \mid c_1, \bar{c}_2)$ and $\Pr(\bar{c}_1 \mid c_2)$ are small;
- both $\Pr(e \mid \bar{c}_1, c_2)$ and $\Pr(\bar{c}_2 \mid c_1)$ are small;
- both $\Pr(c_1 \mid c_2)$ and $\Pr(c_2 \mid c_1)$ are quite large, that is, there is a strong positive probabilistic influence between the two cause variables.

From the above considerations, we conclude that the presence of a direct dependency between the two cause variables of a basic causal mechanism does not result in significantly different consequences of propagating deviating noisy-OR calculated values. Large effects on a probability of interest are expected in fact under essentially the same conditions as derived in Section 3 with the basic mechanism. When considering application of the noisy-OR model for the effect variable of a mechanism with an explicit intercausal dependency, extra caution is advised however, if the mutual dependency between the causes is positive and quite strong. A network engineer should then carefully consider the direction and strength of the qualitative probabilistic influence associated with the explicit dependency; for further information on qualitative concepts of probability, we refer to [24].

4.2. Causal mechanisms with multiple cause variables

Having focused on two-cause mechanisms so far, we now turn to mechanisms involving more than two cause variables. We focus more specifically on a mechanism with the three cause variables C_1, C_2, C_3, and the effect variable E. In view of the three cause variables, the conditional probability table for E includes eight probabilities. When employing the noisy-OR model for the variable E, values for three of these probabilities are to be specified explicitly: these are the conditional probabilities of the effect e arising in the presence of just one of the three causes. The conditional probability of the effect arising spontaneously again is set to zero. We now focus on the remaining four probabilities of the table, which are calculated from the model. While we supposed in our investigations of the two-cause mechanism that the single noisy-OR calculated value under study deviated from the true probability, we cannot now reasonably assume for our three-cause mechanism that only one of the calculated probability values is deviant. Any of the four noisy-OR calculated values may deviate from its true probability in reality. We note that if we could assume that just the calculated value for the effect arising in the presence of all causes differed from the true probability value, our analysis would be analogous to the one presented in Section 3. For studying the joint effects of four deviating probability values however, we need to perform a higher-order analysis.

We begin again by addressing the consequences for the prior probability of interest $\Pr(e)$, of propagating possibly deviating values w, x, y and z for the four probabilities $\Pr(e \mid c_1, c_2, c_3)$, $\Pr(e \mid \bar{c}_1, c_2, c_3)$, $\Pr(e \mid c_1, \bar{c}_2, c_3)$ and $\Pr(e \mid c_1, c_2, \bar{c}_3)$, respectively. The following multi-dimensional function expresses $\Pr(e)$ in these values:

$$\Pr(e)(w, x, y, z) = \alpha_1 \cdot w + \alpha_2 \cdot x + \alpha_3 \cdot y + \alpha_4 \cdot z + \beta$$

where the constants α_j, $j = 1, \ldots, 4$, are built from the prior probabilities of the three causes c_i, $i = 1, 2, 3$. The constant α_4 for example equals $\Pr(c_1) \cdot \Pr(c_2) \cdot \Pr(\bar{c}_3)$, and captures the function's partial gradient associated with the dimension $z = \Pr(e \mid c_1, c_2, \bar{c}_3)$. The constant β in the offset of the function also involves the prior probabilities of the three causes and in addition includes the values of the parameter probabilities $\Pr(e \mid \bar{c}_1, \bar{c}_2, c_3)$, $\Pr(e \mid \bar{c}_1, c_2, \bar{c}_3)$ and $\Pr(e \mid c_1, \bar{c}_2, \bar{c}_3)$. Focusing again on the partial gradients of the function, we observe that α_4 for example can attain a large value only with large prior probabilities of the causes c_1 and c_2 being present and a large probability of the absence of c_3; analogous observations hold for the other partial gradients of the function. A large partial gradient can thus be found only with quite skewed prior probability distributions over the cause variables. Since the four dimensions of the function correspond to different combinations of causes being present, a large partial gradient can be found in a single dimension only. We conclude that the causal propagation effects with the three-cause mechanism are much in line with the effects found with the basic mechanism: large deviations from the true values of at least one of the probabilities $\Pr(e \mid c_1, c_2, c_3)$, $\Pr(e \mid \bar{c}_1, c_2, c_3)$, $\Pr(e \mid c_1, \bar{c}_2, c_3)$ and $\Pr(e \mid c_1, c_2, \bar{c}_3)$ can give a large shift in the prior probability of the effect only with highly skewed prior probability distributions over the three cause variables.

To study the effects of deviating noisy-OR values upon diagnostic propagation, we now express the example posterior probability of interest $\Pr(c_3 \mid e)$ as a function of the values w, x, y and z for the four probabilities under study, to find that

$$\Pr(c_3 \mid e)(w, x, y, z) = \frac{\alpha_1 \cdot w + \alpha_2 \cdot x + \alpha_3 \cdot y + \beta'}{\alpha_1 \cdot w + \alpha_2 \cdot x + \alpha_3 \cdot y + \alpha_4 \cdot z + \beta}$$

where the constants α_j, $j = 1, \ldots, 4$, again are built from the prior probabilities of the cause variables C_i, $i = 1, 2, 3$, and the constants β and β' in addition involve the parameter probabilities of the noisy-OR model. We note that while the value z of the probability $\Pr(e \mid c_1, c_2, \bar{c}_3)$ no longer affects the numerator of the probability of interest $\Pr(c_3 \mid e)$, it does still influence the denominator. Focusing on the partial gradient α_1 as an example, we observe that, by taking the value zero for x, y, z, the multi-dimensional function above reduces to:

$$\Pr(c_3 \mid e)(w, 0, 0, 0) = \frac{w + \beta'/\alpha_1}{w + \beta/\alpha_1}$$

which again is a fragment of a fourth-quadrant hyperbola branch, to which the observations from Section 3.2 apply, that is, from the function we find that deviations of w from the true probability $\Pr(e \mid c_1, c_2, c_3)$ may give large effects on the output probability of interest only if both β'/α_1 and β/α_1 are quite small. We note that to this end at least the value of α_1 need to be quite large, that is, each of the three causes needs to have a large prior probability of being present. In view of larger (fixed) values for x, y and z, the constants involved in the hyperbolic function expressing $\Pr(c_3 \mid e)$ in w will become larger, and the propagation effect will diminish. We conclude that large effects on the example output probability $\Pr(c_3 \mid e)$ can be found only with a large value for α_1, and small values for $\alpha_2, \alpha_3, \alpha_4$; such effects are then found for the smaller value range of x, y, z.

From the above considerations, we conclude that the inclusion of additional cause variables in a causal mechanism does not give rise to significantly different consequences of propagating deviating noisy-OR calculated values: in essence, large propagation effects are expected under essentially the same conditions as derived in Section 3. Although large propagation effects can still occur, they will become less likely as the number of cause variables increases.

4.3. On ill-considered use of the noisy-OR model in larger networks

In our analysis of the possible effects of ill-considered use of the noisy-OR model, we investigated the results from locally propagating deviating noisy-OR calculated values through a causal mechanism, that is, we focused on output probabilities pertaining to one of the variables involved in the mechanism under study. In real-world applications however, the conditional probability tables resulting from application of the noisy-OR model are used upon propagating information throughout a larger Bayesian network. A large deviation of a locally computed probability may then be reduced by further propagation; alternatively, small deviations in locally computed probabilities may jointly induce a large deviation in the overall output probability of interest. Such effects are dependent of the topology of the graphical structure of the network at hand and of the skewness properties of the other conditional probability tables used in the propagation. Although over the years, quite some insights have been gained in the sensitivities exhibited by a network; we refer to [11] for such insights. For a specific Bayesian network under study the effects of deviating noisy-OR calculated values should be investigated by experimental evaluation. The conclusions from the present paper provide directions for focusing such an evaluation on the relevant model-computed values and output probabilities of interest.

5. Propagation effects with generalisations of the noisy-OR model

Despite its clear engineering advantages, the noisy-OR model has only restricted applicability because it assumes the properties of accountability and exception independence to hold for a causal mechanism and in addition assumes all variables involved to be binary. To enhance its applicability, researchers have developed various generalisations of the basic model. A well-known example is the leaky noisy-OR model, which does not assume the property of accountability for a mechanism [9,14]; in order to allow application with non-binary discrete variables moreover, the noisy-MAX model has been suggested [8]. Similar to the basic noisy-OR model, these generalisations can be viewed as parameterised conditional probability tables for which just a limited number of parameter probabilities are to be specified. In this section we briefly address the possible effects upon propagation of deviating probability values calculated from the leaky noisy-OR and noisy-MAX models. Upon doing so, we will focus once more on the differences in effects between the generalised models and the basic noisy-OR model studied in the previous sections.

5.1. Propagation effects with the leaky noisy-OR model

The property of accountability assumed by the noisy-OR model presupposes that all causes of an effect have been identified and explicitly modelled in a causal mechanism. In real-world applications however, incompleteness of information is inherent to any model, and causes may escape explicit representation. For mechanisms in which causes are left implicit, the leaky noisy-OR model may be employed. This model closely resembles the noisy-OR model, yet does not assume accountability. Instead, it provides for capturing the influence of all yet unmodelled causes on a common effect by a so-called leak probability [14]. For a causal mechanism with an effect variable E and the cause variables $C_1, \ldots, C_n$, $n \geq 2$, the leaky noisy-OR model requires the same parameter probabilities as the noisy-OR model, and further takes the extra parameter $\Pr(e \mid \bar{c}_1, \ldots, \bar{c}_n)$ to capture the probability of the effect e occurring spontaneously in the absence of any of its modelled causes. This leak probability $\Pr(e \mid \bar{c}_1, \ldots, \bar{c}_n)$ typically attains small values in practise. For calculating the probabilities of the effect arising in the presence of multiple causes, a computation rule has been defined just as with the standard noisy-OR model [9,14]. This rule assumes that the leak probability is not embedded in the parameter probabilities obtained, but is straightforwardly rewritten for parameter probabilities in which the leak probability is implicitly comprised; for further details, we refer to [9]. Since our observations below hold for both parameter probabilities with and without an implicit leak probability, we study the propagation effects of deviating model-calculated probabilities in the conditional probability table for the effect variable E through the original formulation of the computation rule:

$$\Pr(e \mid \mathbf{c}) = 1 - (1 - \Pr(e \mid \bar{c}_1, \ldots, \bar{c}_n)) \cdot \prod_{j \in J} \frac{1 - \Pr(e \mid \bar{c}_1, \ldots, c_j, \ldots, \bar{c}_n)}{1 - \Pr(e \mid \bar{c}_1, \ldots, \bar{c}_n)}$$

where J again is the set of indices of the cause variables C_j that are marked as being present in the cause combination $\mathbf{c}$. We note that if there are no unmodelled causes for the effect e, that is, if $\Pr(e \mid \bar{c}_1, \ldots, \bar{c}_n) = 0$, then the leaky noisy-OR model results in the same calculated values as the noisy-OR model. In the remainder of this section, we investigate the consequences of the extra parameter of the leaky noisy-OR model and hence assume a non-zero leak probability.

For studying the possible propagation effects of deviating model-calculated probability values, we consider again a basic two-cause mechanism, with the effect variable E and the cause variables C_1, C_2. We assume that values have been obtained for the parameter probabilities $\Pr(e \mid c_1, \bar{c}_2), \Pr(e \mid \bar{c}_1, c_2)$ and $\Pr(e \mid \bar{c}_1, \bar{c}_2)$, and address the fourth probability $\Pr(e \mid c_1, c_2)$ from the conditional probability table of E. We begin again by considering propagation in the causal direction, and focus on the probability of interest $\Pr(e)$. We note that this probability $\Pr(e)$ is algebraically dependent of the leak probability $\Pr(e \mid \bar{c}_1, \bar{c}_2)$. When expressing $\Pr(e)$ as a function of the value x of the probability $\Pr(e \mid c_1, c_2)$, this dependency is reflected in the offset of the function: we find that

$$\Pr(e)(x) = \alpha \cdot (x + \beta)$$

with

$$\alpha = \Pr(c_1) \cdot \Pr(c_2)$$
$$\beta = \Pr(e \mid \bar{c}_1, c_2) \cdot \lambda_{C_1}^{-1} + \Pr(e \mid c_1, \bar{c}_2) \cdot \lambda_{C_2}^{-1} + \Pr(e \mid \bar{c}_1, \bar{c}_2) \cdot \lambda_{C_1}^{-1} \cdot \lambda_{C_2}^{-1}$$

When compared to the function obtained with the noisy-OR model, we find that the gradient α is the same in both functions. The noisy-OR and leaky noisy-OR models therefore share the property that large effects of a deviating model-calculated value upon causal propagation can be found only with large prior probabilities of the causes being present. The constant β in the offsets of the two functions differ however: with the leaky noisy-OR model, this constant includes the extra term $\Pr(e \mid \bar{c}_1, \bar{c}_2) \cdot \lambda_{C_1}^{-1} \cdot \lambda_{C_2}^{-1}$ involving the leak probability. Since this probability is quite small in general, the additional term will be quite small with large prior probabilities of the causes being present. With small probabilities of the causes being present on the other hand, the extra term in the constant β may become quite large. The larger value of β will then cause the range of values for the gradient α to be more restricted, and smaller propagation effects will be found.

We would like to note that the dependency of $\Pr(e \mid c_1, c_2)$ on the leak probability $\Pr(e \mid \bar{c}_1, \ldots, \bar{c}_2)$ appears to be non-linear in the computation rule of the leaky noisy-OR model. The apparent non-linearity arises from the leak probability being comprised in the parameter probabilities and the model taking the comprised leak into consideration only once. The leak thus is cancelled out from each of the parameter probabilities, thereby effectively removing the non-linearity. Our observations therefore are not influenced by the apparent non-linear dependency of the model-calculated value for $\Pr(e \mid c_1, c_2)$ on the leak probability.

The possible effects from diagnostic propagation with the leaky noisy-OR model are also largely similar to those found with the standard noisy-OR model. When expressing the output probability $\Pr(c_1 \mid e)$ for example, as a function of the value x for the probability $\Pr(e \mid c_1, c_2)$, we find that the only difference from the function obtained with the noisy-OR model is in the function's denominator. The following hyperbolic function is found:

$$\Pr(c_1 \mid e) = \frac{x + \beta'}{x + \beta}$$

where the constant β includes the extra term $\Pr(e \mid \bar{c}_1, \bar{c}_2) \cdot \lambda_{C_1}^{-1} \cdot \lambda_{C_2}^{-1}$ when compared with the function resulting from the noisy-OR model. The constant β may thus attain larger values with the leaky noisy-OR model than with the noisy-OR model. We recall that this constant determines the position of the vertical asymptote $s = -\beta$ as well as the location of the vertex of the hyperbola branch under study. The larger β now implies that the vertical asymptote of the hyperbola branch lies further to the left of the unit window, which in turn causes the vertex to move away from the upper-left corner of the window. The effect of a deviating model-calculated value may therefore be smaller with the leaky noisy-OR model than with the noisy-OR model.

Based upon the above considerations, we conclude that application of the leaky noisy-OR model will result in similar propagation effects as the basic noisy-OR model. The extra parameter of the leaky noisy-OR model will tend to have a weakening influence on the effect however. In fact, the larger the leak probability for a causal mechanism is, the weaker the propagation effects will be.

5.2. Propagation effects of the noisy-MAX model

All causal interaction models discussed above, pertain to mechanisms with binary variables only. As real-world Bayesian networks often include non-binary discrete variables to describe their domain knowledge, researchers have generalised the noisy-OR model to provide for mechanisms involving such variables [8,14]. The most commonly used among these generalisations is the noisy-MAX model. Underlying this model are various assumptions concerning the variables of the causal mechanism at hand. The value domain of each cause variable C_i is assumed to include a designated value modelling absence of the cause, denoted as c_i^0; the other values of the variable then capture different levels of manifestation of the cause. The value domain of the effect variable E is also supposed to include a designated value e^0 modelling absence, and in addition is assumed to allow a total ordering; in the sequel, we will take $e^i < e^j$ whenever $i < j$. The noisy-MAX model now builds upon the properties of accountability and exception independence just like the noisy-OR model. With the noisy-MAX model, the assumption of accountability states that $\Pr(e^0 \mid c_1^0, \ldots, c_n^0) = 1$ and, hence, $\Pr(e^i \mid c_1^0, \ldots, c_n^0) = 0$ for all values e^i with $i > 0$. The parameter probabilities for the noisy-MAX model describe for each cause variable separately, the influence of its different manifestation levels on the possible values of the effect variable, that is, the model takes the parameter probabilities $\Pr(e^i \mid c_1^0, \ldots, c_{j-1}^0, c_j^k, c_{j+1}^0, \ldots, c_n^0)$ for all values c_j^k, $k > 0$, of the cause variable C_j and all values e^i, $i \geq 0$, of the effect variable E. The remaining probabilities for the conditional probability table for E are defined through

$$\Pr(e^i \mid \mathbf{c}) = \begin{cases} \Pr(E \leq e^i \mid \mathbf{c}) - \Pr(E \leq e^{i-1} \mid \mathbf{c}) & \text{for } i > 0 \\ \Pr(E \leq e^0 \mid \mathbf{c}) & \text{for } i = 0 \end{cases}$$

with

$$\Pr(E \leq e^i \mid \mathbf{c}) = \prod_{j \in J} \sum_{\ell = 0, \ldots, i} \Pr(e^\ell \mid c_1^0, \ldots, c_{j-1}^0, c_j^k, c_{j+1}^0, \ldots, c_n^0)$$

where J is the set of indices of the cause variables C_j that are marked as having a value c_j^k with $k > 0$ in the joint value combination $\mathbf{c}$.

To study the possible effects of propagating deviating model-calculated probability values, we consider again a basic two-cause mechanism, with the effect variable E and the cause variables C_1, C_2. For ease of exposition, we assume that the cause variable C_1 is ternary and that the other variables are binary. For the binary variables, we will adhere to our former notations; the values of the ternary variable C_1 are written c_1^0, c_1^1, c_1^2, where c_1^0 denotes absence of the cause at hand. We assume that the probability $\Pr(e \mid c_1^0, \bar{c}_2)$ is set to zero, and that values have been obtained for the parameter probabilities $\Pr(e \mid c_1^1, \bar{c}_2)$, $\Pr(e \mid c_1^2, \bar{c}_2)$ and $\Pr(e \mid c_1^0, c_2)$. We now focus on the two remaining probabilities $\Pr(e \mid c_1^1, c_2)$ and $\Pr(e \mid c_1^2, c_2)$ from the probability table for E, and write the prior probability of interest $\Pr(e)$ in terms of the values x and y for these probabilities. The following bi-linear function is found:

$$\begin{aligned}\Pr(e)(x, y) &= x \cdot \Pr(c_1^1) \cdot \Pr(c_2) + \Pr(e \mid c_1^1, \bar{c}_2) \cdot \Pr(c_1^1) \cdot \Pr(\bar{c}_2) + \ y \ \cdot \Pr(c_1^2) \cdot \Pr(c_2) + \Pr(e \mid c_1^2, \bar{c}_2) \cdot \Pr(c_1^2) \cdot \Pr(\bar{c}_2) \\ &\quad + \Pr(e \mid c_1^0, c_2) \cdot \Pr(c_1^0) \cdot \Pr(c_2) \\ &= \alpha_1 \cdot x + \alpha_2 \cdot y + \beta\end{aligned}$$

where the constants α_i, $i = 1, 2$, again are built from prior probabilities of the two cause variables; the constant α_1 for example, equals $\Pr(c_1^1) \cdot \Pr(c_2)$ and captures the function's partial gradient associated with the dimension $x = \Pr(e \mid c_1^1, c_2)$. The constant β also involves these prior probabilities and in addition includes the parameter probabilities $\Pr(e \mid c_1^j, \bar{c}_2)$, $j = 1, 2$, and $\Pr(e \mid c_1^0, c_2)$. Focusing again on the partial gradients of the function, we observe that α_1 can attain a large value only with large probabilities of c_2 being present and of C_1 having the value c_1^1. Analogously, α_2 attains a large value only with large prior probabilities of c_2 being present and of C_1 being equal to c_1^2. We thus once more find propagation effects in line with those found with the noisy-OR model in Section 3: large deviations from the true values of at least one of the probabilities $\Pr(e \mid c_1^1, c_2)$ and $\Pr(e \mid c_1^2, c_2)$ can give a large shift in the prior probability of the effect only with highly skewed prior probability distributions over the cause variables. We note moreover that such a large partial gradient can be found in just a single dimension.

For investigating the effects of deviating noisy-MAX calculated values upon diagnostic propagation, we express the example probability of interest $\Pr(c_1^1 \mid e)$ as a function of the value x for the probability $\Pr(e \mid c_1^1, c_2)$ and the value y for the probability $\Pr(e \mid c_1^2, c_2)$, to find that

$$\Pr(c_1^1 \mid e)(x, y) = \frac{x + \beta'}{x + \alpha \cdot y + \beta}$$

where

$$\begin{aligned}\alpha &= \frac{\Pr(c_1^2)}{\Pr(c_1^1)} \\ \beta' &= \Pr(e \mid c_1^1, \bar{c}_2) \cdot \lambda_{C_2}^{-1} \\ \beta &= \Pr(e \mid c_1^0, c_2) \cdot \frac{\Pr(c_1^0)}{\Pr(c_1^1)} + \Pr(e \mid c_1^1, \bar{c}_2) \cdot \lambda_{C_2}^{-1} + \Pr(e \mid c_1^2, \bar{c}_2) \cdot \frac{\Pr(c_1^2)}{\Pr(c_1^1)} \cdot \lambda_{C_2}^{-1}\end{aligned}$$

We observe that the function above once again is a quotient of two multi-linear functions. By a similar analysis as used in Section 4.2, we start by taking the value zero for y, as a result of which the function reduces to a one-dimensional hyperbolic function. The constants β' and β involved in this function again depend to a large extent on the skewness of the prior probability distributions over the two cause variables. With large prior probabilities $\Pr(c_2)$ and $\Pr(c_1^1)$, deviations of x from the true probability $\Pr(e \mid c_1^1, c_2)$ may again give large effects on the output probability of interest. In view of a larger (fixed) value for y however, the constants involved in the function expressing $\Pr(c_1^1 \mid e)$ in x will become larger, and the propagation effect will diminish. Large effects can thus only be found for the smaller value range of y.

We conclude our analysis of the noisy-MAX model by mentioning that similar results hold for use of the model with causal mechanisms involving a non-binary effect variable and/or multiple non-binary cause variables.

5.3. *On other generalisations of the noisy-OR model*

Over the years, many other generalisations and extensions of the noisy-OR model have been developed, which range from fairly simple models to more involved ones such as the (inhibited) recursive noisy-OR [17,19] and the NIN-AND tree [29,30] models. The techniques and approach presented in this paper can be applied rather straightforwardly for studying the effects of propagating deviating values by the simpler models. For the more involved models, our techniques may not suffice. The NIN-AND tree model, for example, employs a tree structure to describe a causal mechanism. As this tree structure is inherent to the model, it should also be taken into consideration upon investigating propagation effects. In essence, each of the building block of an NIN-AND tree could be studied separately to gain at least some basic insights; we note that one of the building blocks is the noisy-OR model. Further analysis then depends heavily on the causal mechanism considered and on the actual NIN-AND tree corresponding with that mechanism. As such an analysis is beyond the scope of the current paper, we leave it for future research.

6. Conclusions

When building a Bayesian network with the help of domain experts, the elicitation of all probabilities required often proves the main bottleneck in the engineering process. In order to reduce the amount of time spent on probability elicitation and to alleviate the burden for the experts involved, researchers have developed various probabilistic causal interaction models. These models basically are parameterised conditional probability tables which require just a limited number of parameter probabilities. The remaining probabilities then are calculated using model-specific rules, which are derived from properties of probabilistic interaction among the variables involved. Since not all network engineers are fully aware of these properties, causal interaction models are inherently subject to ill-considered use.

In this paper, we addressed the extent to which ill-considered use of a probabilistic causal interaction model can be harmful for the overall performance of a Bayesian network; upon doing so, we focused more specifically on the commonly employed noisy-OR model and its variants. Using techniques from sensitivity analysis, we expressed various output probabilities of interest as functions of the values of one or more conditional probabilities from a causal mechanism. These functions served to reveal the propagation effects of any deviation from the true probability values on the output, and hence provided for studying the effects of deviating model-calculated values.

We demonstrated that ill-considered use of a causal interaction model can result in poorly calibrated output probabilities, and identified conditions under which large propagation effects on the output can be expected. We found for example that a deviating model-calculated value may have a large effect on an output probability upon propagation, only if the yet unobserved cause variables in the mechanism have quite skewed probability distributions and/or the obtained parameter probabilities have small values. Throughout the paper, we stated the conditions under which large propagation effects can be expected, in further detail. We would like to emphasise that these conditions pertain to output probabilities within a causal mechanism under study. Strong local effects may yet be subdued upon further propagation throughout a larger network.

While one of the results from this paper is the observation that ill-considered use of the noisy-OR model and its variants can lead to poorly calibrated network output, this result does not contradict the findings from earlier experimental studies which have led to the suggestion that Bayesian networks are quite robust against the inaccuracies induced in their conditional probability tables by the use of these models. In fact, our investigations have provided a formal underpinning of these findings and show that use of a causal interaction model for mere pragmatic reasons may be warranted, even when the model's underlying assumptions are not met in reality. Network engineers are advised however, to verify whether large propagation effects may be expected before applying the noisy-OR model, using the insights from the paper.

Our investigations of the effects of deviating noisy-OR calculated probability values have revealed that the presence of cancellation effects among the causes in a causal mechanism may induce quite strong propagation effects on probabilities of interest. If a modelled cause serves to annihilate to some extent the effect of another cause, the noisy-OR calculated value will be considerably higher than the true probability value. Large prior probabilities of the causes being present will then induce strong effects on an output probability upon propagation in the diagnostic direction. Motivated by this observation, we are now focusing our further investigations on causal mechanisms which embed such cancellation effects among their causes. As it will be highly advantageous from a network-engineering point of view to have available a causal interaction model for describing cancellation effects, we are now in the process of designing parameterised conditional probability tables for this type of interaction.

Acknowledgements

This paper is an extended version of the ECSQARU 2011 paper *Using the noisy-OR model can be harmful ... but it often is not* [28]. We would like to thank all reviewers associated with the ECSQARU 2011 conference and with IJAR, for their helpful suggestions for extending and improving the paper. The research is partly funded with financial support from the Prevention of and Fight against Crime Programme of the EU European Commission – Directorate-General Home Affairs.

References

[1] V. Anand, S.M. Downs, Probabilistic asthma case finding: a noisy-OR reformulation, in: AMIA Annual Symposium Proceedings, 2008, pp. 6–10.
[2] J. Bolt, L.C. van der Gaag, An empirical study of the use of the noisy-OR model in a real-life Bayesian network, in: E. Hüllermeier, R. Kruse, F. Hoffmann (Eds.), Information Processing and Management of Uncertainty in Knowledge-Based Systems, in: Theory and Methods, vol. 80, Springer, Berlin, 2010, pp. 11–20.
[3] E. Castillo, J.M. Gutiérrez, A.S. Gadi, Sensitivity analysis in discrete Bayesian networks, IEEE Trans. Syst. Man Cybern. 27 (1997) 412–423.
[4] G.F. Cooper, The computational complexity of probabilistic inference using Bayesian belief networks, Artif. Intell. 42 (1990) 393–405.
[5] V.M.H. Coupé, L.C. van der Gaag, J.D.F. Habbema, Sensitivity analysis: an aid for belief-network quantification, Knowl. Eng. Rev. 15 (2000) 215–232.
[6] V.M.H. Coupé, L.C. van der Gaag, Properties of sensitivity analysis of Bayesian belief networks, Ann. Math. Artif. Intell. 36 (2002) 323–356.
[7] A. Darwiche, Modeling and Reasoning with Bayesian Networks, Cambridge University Press, Cambridge, 2009.
[8] F.J. Díez, Parameter adjustment in Bayes networks. The generalized noisy-OR gate, in: D. Heckerman (Ed.), Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence (UAI-01), Morgan Kaufmann Publishers, 2001, pp. 165–182.
[9] F.J. Díez, M.J. Druzdzel, Canonical probabilistic models for knowledge engineering, Technical Report CISIAD-06-01, UNED, Madrid, Spain, 2006.
[10] M.J. Druzdzel, L.C. van der Gaag, Building probabilistic networks: where do the numbers come from? Guest editors introduction, IEEE Trans. Knowl. Data Eng. 12 (2000) 481–486.
[11] L.C. van der Gaag, S. Renooij, V.M.H. Coupé, Sensitivity analysis of probabilistic networks, in: P. Lucas, J.A. Gámez, A. Salmeron (Eds.), Advances in Probabilistic Graphical Models, in: Studies in Fuzziness and Soft Computing, vol. 213, Springer, Berlin, 2007, pp. 103–124.
[12] L.C. van der Gaag, S. Renooij, C.L.M. Witteman, B.M.P. Aleman, B.G. Taal, Probabilities for a probabilistic network: a case study in oesophageal cancer, Artif. Intell. Med. 25 (2002) 123–148.
[13] D. Heckerman, J. Breese, Causal independence for probabilistic assessment and inference using Bayesian networks, IEEE Trans. Syst. Man Cybern. 26 (1996) 826–831.
[14] M. Henrion, Some practical issues in constructing belief networks, in: J. Lemmer, T. Levitt, L. Kanal (Eds.), Proceedings of the Third Annual Conference on Uncertainty in Artificial Intelligence (UAI-87), Elsevier, 1987, pp. 161–174.
[15] F.V. Jensen, Th.D. Nielsen, Bayesian Networks and Decision Graphs, second edition, Springer Verlag, New York, 2007.

[16] D. Koller, N. Friedman, Probabilistic Graphical Models. Principles and Techniques, The MIT Press, Cambridge, 2009.
[17] U. Kuter, D. Nau, D. Gossink, J.F. Lemmer, Interactive course-of-action planning using causal models, in: M. Pechoucek, A. Tate (Eds.), Proceedings of the Third International Conference on Knowledge Systems for Coalition Operations (KSCO-2004), Prague, 2004, pp. 37–51.
[18] S.L. Lauritzen, D.J. Spiegelhalter, Local computations with probabilities on graphical structures and their application to expert systems, J. R. Stat. Soc. B 50 (1988) 157–224.
[19] J.F. Lemmer, D.E. Gossink, Recursive noisy OR – a rule for estimating complex probabilistic interactions, IEEE Trans. Syst. Man Cybern., Part B, Cybern. 34 (2004) 2252–2261.
[20] P.J.F. Lucas, Bayesian network modelling through qualitative patterns, Artif. Intell. 163 (2005) 233–263.
[21] J. Ma, Qualitative approach to Bayesian networks with multiple causes, IEEE Trans. Syst. Man Cybern. 42 (2012) 382–391.
[22] M.G. Morgan, M. Henrion, Uncertainty, a Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis, Cambridge University Press, Cambridge, 1990.
[23] A. Oniśko, M.J. Druzdzel, H. Wasyluk, Learning Bayesian network parameters from small data sets: application of noisy-OR gates, Int. J. Approx. Reason. 27 (2001) 165–182.
[24] S. Parsons, Qualitative Methods for Reasoning under Uncertainty, MIT Press, 2001.
[25] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann Publishers, San Francisco, 1988.
[26] S. Renooij, Probability elicitation for belief networks: issues to consider, Knowl. Eng. Rev. 16 (2001) 255–269.
[27] D. Roth, On the hardness of approximate reasoning, Artif. Intell. 82 (1996) 273–302.
[28] S.P.D. Woudenberg, L.C. van der Gaag, Using the noisy-OR model can be harmful ... but it often is not, in: W. Liu (Ed.), Proceedings of the 11th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Springer-Verlag, Berlin, 2011, pp. 122–133.
[29] Y. Xiang, N. Jia, Modeling causal reinforcement and undermining for efficient CPT elicitation, IEEE Trans. Knowl. Data Eng. 19 (2007) 1708–1718.
[30] Y. Xiang, Non-impeding noisy-AND tree causal models over multi-valued variables, Int. J. Approx. Reason. 53 (2012) 988–1002.
[31] A. Zagorecki, M. Druzdzel, An empirical study of probability elicitation under noisy-OR assumption, in: Proceedings of the 17th International Florida Artificial Intelligence Research Symposium, 2004, pp. 880–885.
[32] A. Zagorecki, M.J. Druzdzel, Knowledge engineering for Bayesian networks: how common are noisy-MAX distributions in practice? IEEE Trans. Syst. Man Cybern. 43 (2013) 186–195.