An Empirical Study of the Use of the Noisy-Or Model in a Real-Life Bayesian Network

Janneke H. Bolt and Linda C. van der Gaag

Department of Information and Computing Sciences, Utrecht University,
P.O. Box 80.089, 3508 TB Utrecht, The Netherlands
{janneke,linda}@cs.uu.nl

Abstract. The use of the noisy-OR model is advocated throughout the literature as an approach to lightening the task of obtaining all probabilities required for a Bayesian network. Little evidence is available, however, as to the effects of using the model on a network's performance. In this paper, we construct a noisy-OR version of a real-life hand-built Bayesian network of moderate size, and compare the performance of the original network with that of the constructed noisy-OR version. Empirical results from using the two networks on real-life data show that the performance of the original network does not degrade by using the noisy-OR model.

1 Introduction

When building a Bayesian network, the task of obtaining all probabilities required is generally acknowledged to be the most daunting among the engineering tasks involved [1]. A Bayesian network of realistic size easily requires hundreds and sometimes even thousands of probabilities for its conditional probability tables. While for some application domains these probabilities are readily available or can be estimated from data, for other domains they need to be provided by domain experts. Over the last decades, researchers have taken two different approaches to lightening this quantification task. On the one hand, researchers have focused on methods and tools to support the elicitation of many probabilities from domain experts in little time [2]. On the other hand, researchers have designed causal independency models to support the quantification [3,4].

A causal independency model is a parameterised specification of a conditional probability table for a Bayesian network. It requires the assessment of just a small number of conditional probabilities for a table under construction; these probabilities are the model's parameters. The other probabilities required to arrive at a fully specified table then are computed from the model's parameters through simple mathematical functions. Well-known examples of causal independency models are the noisy-OR model which was originally devised by J. Pearl [5], and its generalisation, the noisy-MAX model [6].

To provide for parameterisation, causal independency models assume a specific type of interaction of the causes of a common effect. Before using such a model for a network under construction, therefore, a network engineer has to

E. Hüllermeier, R. Kruse, and F. Hoffmann (Eds.): IPMU 2010, Part I, CCIS 80, pp. 11–20, 2010.

verify whether the assumed interaction actually holds in reality. Yet, the assumption's validity is not always easily checked; engineers, moreover, may not always be acquainted with the full details of the interaction assumption underlying the model. As a consequence, causal independency models are used in practice where they may not be entirely appropriate. Little is known about the effects of simply using these models on the performance of a Bayesian network in a real-life setting, however. Some researchers studied how using noisy-MAX approximations influence posterior probabilities computed from three different Bayesian networks [7]. Using randomly generated evidence, they found that the posterior probabilities computed from the noisy-MAX versions quite closely approximated the ones computed from the original networks. In another study, the noisy-OR model was used upon learning conditional probability tables from data [8]. The researchers found that the performance of the original network was improved upon by using noisy-OR approximations instead of uniform distributions for the tables for which sufficient data were lacking.

In this paper, we present the results from an empirical study of the effects of applying the noisy-OR model without verification of its underlying assumption. For the study, we used our CSF network for the early detection of Classical Swine Fever (CSF) in pigs, which was constructed and quantified in close collaboration with two domain experts. This network includes 32 stochastic variables, for which over 1100 conditional probabilities are specified; 470 of these probabilities were estimated directly by one of the experts. The basic idea of the study was to substitute new probability tables for the expert-provided ones, where these new tables were obtained from applying the leaky noisy-OR model, and to compare the performance of the original network with that of its leaky noisy-OR version. More specifically, the conditional probability tables of 11 of the 32 variables in total were replaced by noisy-OR approximations, which resulted in 122 of the 470 directly estimated probabilities being replaced by computed ones. Since the performance of a Bayesian network for diagnostic reasoning in biomedical applications is commonly expressed in terms of its sensitivity and specificity, we studied the performances of the two versions of the CSF network in terms of these characteristics, using real data collected from veterinary practice and from experimental infections. We found that use of the noisy-OR model had little effect on the network's performance, even though some of the substituted probability tables differed substantially from the original expert-provided ones.

Even though little is known about the effects of their inadvertent application, the use of causal independency models, and of the noisy-OR and noisy-MAX models more specifically, is advocated throughout the literature as an approach to substantially lighten the task of obtaining all probabilities required for a Bayesian network. Although more fundamental research is required to corroborate our findings, the results from our empirical study warrant the cautious conclusion that the noisy-OR model can indeed be applied without extensive knowledge elicitation efforts for Bayesian networks for diagnostic applications.

The paper is organised as follows. In Section 2, we briefly describe our CSF network. Section 3 illustrates the noisy-OR model and its use in Bayesian networks

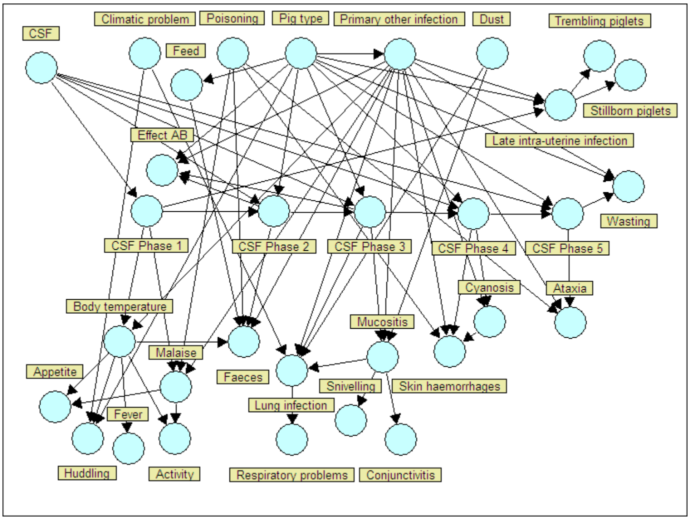

Fig. 1. The graphical structure of the CSF network for the early detection of Classical Swine Fever in individual pigs

in general. In Section 4, we describe the construction of a leaky noisy-OR version of the CSF network. The performances of the two networks are compared in Section 5. The paper ends in Section 6 with our concluding observations.

2 A Bayesian Network for Classical Swine Fever

A Bayesian network describes a joint probability distribution over a collection of stochastic variables. The variables and the qualitative (in)dependency relationships between them are modelled by nodes and arcs respectively, in a graphical structure. The strengths of the dependencies between the variables are expressed through probabilities. More specifically, for each variable X in the graphical structure, a conditional probability table is specified which describes the probability distributions $\Pr(X \mid \mathbf{pa}(X))$ over the values of X for each possible combination of values for the parents $\mathbf{pa}(X)$ of X in the structure; note that a conditional probability table thus specifies an exponential number of probability distributions, that is, exponential in the number of parents involved.

As an example, Figure 1 depicts the graphical structure of our CSF network for the early detection of Classical Swine Fever in pigs. Classical Swine Fever is a highly infectious viral disease of pigs which has a tendency to spread rapidly, both within a herd and between herds, and which can develop quite aggressively with large proportions of affected animals dying. Because of the major socio-economical consequences that an outbreak may have, reducing the time between first infection of a herd and first detection is of major importance. In close collaboration with experts from the Central Veterinary Institute in the

Table 1. The conditional probability table of the variable *Appetite* in the CSF network; the boldfaced probabilities were provided directly by our domain expert

Body temperature	*Malaise*	*Appetite* reduced	normal
elevated	yes	**0.9000**	0.1000
	no	**0.2500**	0.7500
normal	yes	**0.8500**	0.1500
	no	**0.0050**	0.9950

Netherlands, we developed the depicted network to support veterinarians in the early detection of the disease. The network takes for its input the clinical signs observed in an individual pig and returns the posterior probability of these signs being caused by a CSF infection.

The CSF network currently includes 32 stochastic variables. To describe the strengths of the dependencies among these variables, the network's graphical structure is supplemented with over 1100 probabilities, organised in conditional probability tables. As an example, Table 1 shows the conditional probability table for the variable *Appetite*, which models whether or not a pig has a reduced appetite. Since the probabilities from a single distribution sum to 1, only some of the probabilities from the table were estimated explicitly by our domain expert; these probabilities are shown in boldface. All in all, 470 of the network's more than 1100 probabilities were estimated directly by one of the experts. We would like to note that some of the network's dependency relationships are deterministic, which accounts for the observation that fewer than half of the probabilities were expert provided. We further note that we will specify the probabilities in all tables throughout the paper with a precision of four decimals, since the domain expert provided his probability assessments up to this high precision.

3 The Noisy-OR Model

For a Bayesian network of realistic size, hundreds or even thousands of conditional probabilities are required. One of the approaches to lightening the task of obtaining all these probabilities is the use of causal independency models. A causal independency model is a parameterised specification of a conditional probability table. The model requires the assessment of just a small number of conditional probabilities to arrive at a fully specified probability table; these probabilities are the model's parameters. The other probabilities for the table are computed from these parameters through simple functions. Several researchers have studied different types of causal independency model along with their properties, and by now an entire family of models have been described [3,4].

The best-known causal independency model is the noisy-OR model for binary variables [5]. This model pertains to a causal mechanism composed of an effect variable E and parent variables C_i, $i = 1, \ldots, n$, $n \geq 2$, which model possible

Table 2. The conditional probability tables that result from application of the noisy-OR model (*left*) and the leaky noisy-OR model (*right*), respectively, for the variable *Appetite* in the CSF network; the parameters for the models are shown in boldface

Body temp.	*Malaise*	*Appetite* reduced	normal
elevated	yes	0.8875	0.1125
	no	**0.2500**	0.7500
normal	yes	**0.8500**	0.1500
	no	0.000	1.000

Body temp.	*Malaise*	*Appetite* reduced	normal
elevated	yes	0.8881	0.1119
	no	**0.2500**	0.7500
normal	yes	**0.8500**	0.1500
	no	**0.0050**	0.9950

causes of the effect. The effect variable has values e for the effect being present and $\bar{e}$ for the effect being absent; each cause variable C_i has the values c_i and $\bar{c}_i$ for the presence and absence of the cause, respectively. The variable *Appetite* in the CSF network constitutes an example of such a causal mechanism, along with its parent variables *Body temperature*, which models an elevated body temperature as a possible cause for a reduced appetite, and *Malaise*, modelling a sense of malaise as another cause for such a finding.

The noisy-OR model now provides a parameterised probability table for the effect variable E given its parent variables in the causal mechanism. The parameters of the model are the probabilities $\Pr(e \mid \bar{c}_1, \ldots, \bar{c}_{j-1}, c_j, \bar{c}_{j+1}, \ldots, \bar{c}_n)$ of the effect occurring in the presence of a single cause c_j. Note that the model thus has n parameters, that is, the number of parameters to be assessed explicitly is linear in the number of possible causes of the modelled effect. The probability $\Pr(e \mid \bar{c}_1, \ldots, \bar{c}_n)$ of the effect occurring spontaneously in the absence of all causes, is taken to be zero by the noisy-OR model. To arrive at a fully specified probability table for the variable E, the conditional probabilities of the effect e occurring given all possible combinations of values $\mathbf{c}$ involving multiple causes are taken to be

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

where $\mathcal{J}$ is the set of indices of the cause variables C_j which have the value c_j in the combination of values $\mathbf{c}$ under consideration. As an example, Table 2 shows, on the left, the conditional probability table that would result from application of the noisy-OR model for the variable *Appetite* in the CSF network.

The noisy-OR model assumes a disjunctive interaction of the causes of a common effect, which implies that the properties of exception independence and accountability are assumed to hold. The property of exception independence states that the presence of any single cause essentially suffices to produce the effect and that the hidden processes that may inhibit the occurrence of the effect are mutually independent. The property of accountability states that the effect of the causal mechanism is absent if none of the possible causes is present.

For many causal mechanisms in practice, the assumption of accountability underlying the noisy-OR model is not met: the effect of these mechanisms can occur even if all modelled causes are absent, that is, $\Pr(e \mid \bar{c}_1, \ldots, \bar{c}_n) \neq 0$. If the property of exception independence does hold for such a mechanism, the leaky noisy-OR model can be employed for constructing a conditional probability table for the effect variable. This model is closely related to the noisy-OR model yet includes the probability $\Pr(e \mid \bar{c}_1, \ldots, \bar{c}_n)$ as a leak probability. In the literature, two different views of the leaky noisy-OR model have been described, resulting in two variants of the model. The first view assumes that all estimated parameter probabilities, given each cause separately, implicitly include the leak probability [9,10]. The second view assumes that the leak probability needs to be explicitly assessed and incorporated into all other parameter probabilities [6]. It has been argued that the first view applies to probabilities derived from data and that the second view is more appropriate for probabilities provided by experts [3]. Since the CSF network used in our current study was constructed by hand and all probabilities were provided by an expert, we focus here on the second view. With this view, the leaky noisy-OR model has $n+1$ parameters; these are the n conditional probabilities $\Pr(e \mid \bar{c}_1, \ldots, \bar{c}_{j-1}, c_j, \bar{c}_{j+1}, \ldots, \bar{c}_n)$ from the noisy-OR model and the leak probability $\Pr(e \mid \bar{c}_1, \ldots, \bar{c}_n)$. To arrive at a fully specified probability table for the effect variable E, the conditional probabilities of e for all combinations of values $\mathbf{c}$ involving multiple causes are taken to be

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

where $\mathcal{J}$ again denotes the set of indices of the cause variables C_j which have the value c_j in the combination of values $\mathbf{c}$ under consideration.

The noisy-OR model being specified for binary variables only, researchers have introduced the noisy-MAX model as a generalisation to multiple-valued variables [6,9]. A general introduction to this model is beyond the scope of the present paper. In the sequel, however, we exploit the property that for a causal mechanism involving a binary effect variable and multiple-valued cause variables, the computation rule of the (leaky) noisy-MAX model is a simple generalisation of the function of the (leaky) noisy-OR model. As an example from the CSF network, we consider the causal mechanism involving the effect variable *Body temperature* (B), modelling whether or not a pig has an increased body temperature, and the cause variables *Primary other infection* (P), modelling the presence of an infection other than a CSF infection, and *CSF Phase 1* (C), modelling whether or not a pig has entered the first phase of an infection with the CSF virus. The variables *Body temperature* and *CSF Phase 1* are binary; the variable *Primary other infection* has the four values none ($\bar{p}$), respiratory infection (p_1), intestinal infection (p_2), and respiratory+intestinal infection (p_3). The leaky noisy-MAX model for this mechanism has for its parameters the conditional probabilities $\Pr(b \mid \bar{p}c)$, $\Pr(b \mid p_i\bar{c})$, $i = 1, 2, 3$, and the leak probability $\Pr(b \mid \bar{p}\bar{c})$. The probability $\Pr(b \mid p_1 c)$ of an increased body temperature given both a respiratory and a first-phase CSF infection now is computed to be

$$\Pr(b \mid p_1 c) = 1 - (1 - \Pr(b \mid \bar{p}\bar{c})) \cdot (1 - \Pr(a \mid p_1\bar{c})) \cdot (1 - \Pr(a \mid \bar{p}c))$$

In our empirical study of the effects of using the leaky noisy-OR model in the CSF network, we exploited this simple function for multiple-valued cause variables.

4 Constructing a Leaky Noisy-OR Version

As described above, the leaky noisy-OR model assumes the property of exception independence to hold in an application domain, which may be hard to verify in practice. In this paper, we empirically study the effect of simply applying the leaky noisy-OR model, without proper verification of its underlying assumption, on a network's performance. For this purpose, we constructed a leaky noisy-OR version of the CSF network and compared the performances of the two networks.

From the original network, we selected all variables that could be considered effect variables in a causal mechanism. With this criterion, a total of 11 variables were selected, which are listed in Table 3. For the selected variables, we substituted new conditional probability tables which were constructed using the leaky noisy-OR model, for the original ones. For the model's parameters, we used the probabilities as originally assessed by our expert; we thus did not construct best-fitting approximations of the original tables. While for most of the selected variables very small Kullback-Leibler distances were found between the originally specified probability distributions and the leaky noisy-OR tables, for some of the variables these distances proved to be of considerable size; the fourth column of Table 3 records the maximum Kullback-Leibler distance found for each of the selected variables. The second and third columns of the table specify, for each variable separately, the numbers of probabilities which have to be assessed directly for the full probability table and for the leaky noisy-OR table, respectively. Table 3 shows that by using the leaky noisy-OR model, the number of probabilities to be specified explicitly by the domain expert, would be reduced from 220 to 98. Since for the entire network a total of 470 parameters

Table 3. The numbers of parameters of the original conditional probability tables (CPT) and their leaky noisy-OR versions, and the maximum Kullback-Leibler distance found between the original and noisy-OR distributions for selected variables of the network

Variable	*Number of CPT parameters needed*	*Number of noisy-OR parameters needed*	*max. KL distance*
Lung infection	64	8	0.0168
Wasting	40	25	0.0006
Skin haemorrhages	32	17	0.0009
Huddling	20	15	0.0050
Malaise	16	6	0.0019
Mucositis	16	6	0.2206
Body temperature	8	5	0.0052
Cyanosis	8	5	0.0000
Late intra-uterine inf.	8	5	0.0005
Appetite	4	3	0.0007
Activity	4	3	0.0138
Total	220	98	

had to be estimated, this number is reduced by more than 25% by exploiting the leaky noisy-OR model for one-third of the variables.

5 Comparing the Performances of the CSF Networks

The performance of a Bayesian network for diagnostic reasoning in biomedical applications is commonly expressed in terms of its sensitivity and specificity characteristics. The network's specificity then is the percentage of cases without the disease under study whom the network identifies as not having the disease; its sensitivity is the percentage of cases with the disease whom the network singles out as having the disease. Since the CSF network chosen for our empirical study was designed for diagnostic reasoning, we use these two characteristics for comparing the performances of the original network and its noisy-OR version.

For establishing the specificity of the two networks, field data from 375 pigs without Classical Swine Fever were used. For each of these pigs, the posterior probability of the observed clinical signs being caused by a CSF infection was computed from both networks. The computed probabilities were subsequently compared against a threshold probability α. If a computed probability exceeded this threshold probability, a suspicion of CSF was issued. Based upon the numbers of generated suspicions, the specificities of both networks could be readily established. Table 4 records these specificities for various realistic values of α.

The sensitivities of the two networks were estimated using experimental data. These data were collected from small groups of pigs in which some individuals had been inoculated with the CSF virus. A total of 91 animals were followed over a period of up to 35 days. Data were recorded at least every two or three days; the recording days were expressed in terms of the number of days after infection of the inoculated animals. For each recording day, for each pig, the posterior probability of the observed clinical signs being caused by a CSF infection was computed from both CSF networks and compared against a threshold probability α as before. Figure 2 shows, as an example, the cumulative number of animals for which a CSF suspicion was issued by the original network as a function of the number of days post infection, using $\alpha = 0.001$; with the leaky noisy-OR version of the network, the exact same results were found. With the threshold values of 0.0005, 0.005, and 0.01, we also found no differences in sensitivity between the

Table 4. The specificities of the CSF network and its leaky noisy-OR version, respectively, for different values of the threshold probability α for issuing a CSF suspicion

threshold α	*specificity CSF network*	*specificity noisy-OR version*
0.05	99%	99%
0.01	98%	98%
0.005	96%	96%
0.001	89%	88%
0.0005	84%	84%

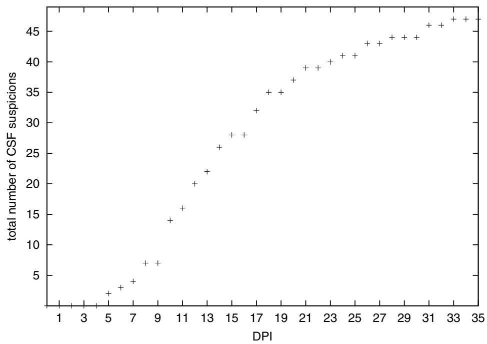

Fig. 2. The cumulative number of CSF suspicions issued by the original CSF network for 91 pigs, using $\alpha = 0.001$, as a function of the number of days post infection (DPI) of the inoculated animals

two networks. With $\alpha = 0.05$, the original network issued a CSF suspicion for one more animal than did the leaky noisy-OR version of the network.

From the sensitivities and specificities reviewed above, we may conclude that the overall performance of the original CSF network does not degrade when the conditional probability tables of one-third of the variables are replaced by leaky noisy-OR tables. These characteristics, however, might hide differences in performance of the two networks for individual cases, which could indicate a fundamental problem of the leaky noisy-OR version. We therefore also studied, using the same threshold values, the networks' performance on all pig cases individually. With $\alpha = 0.005$, the leaky noisy-OR version of the network issued a false suspicion for one of the pigs which did not receive such a suspicion from the original network; it further did not issue a suspicion for one of the pigs which did receive a false suspicion from the original network. Using probability thresholds of 0.001 and 0.005, the leaky noisy-OR version of the CSF network issued an additional false suspicion on the one hand, yet also resulted in three fewer false suspicions on the other hand. With respect to the networks' sensitivities, we found that, with $\alpha = 0.05$, the leaky noisy-OR version failed to detect one of the cases that was detected by the original network. No further differences in performance were found. From these observations, we conclude that also in view of the 466 individual pig cases, the performance of the original CSF network is hardly affected by substituting leaky noisy-OR tables for its expert-provided conditional probability tables.

6 Concluding Observations

The use of causal independency models, and of the noisy-OR and noisy-MAX models more specifically, is advocated throughout the literature as an approach to substantially lightening the task of obtaining all probabilities required for a Bayesian network. Since these models assume a disjunctive interaction of the causes of a common effect, a network engineer has to verify, before using such a model, that the assumed interaction actually holds in reality. In practice, however, the assumption is hard to verify. Thus far, little evidence had been gathered about the effects of simply using these models without elaborate verification efforts, on the performance of a real-life network in practice. In this paper, we presented the results from an empirical study of these effects on a Bayesian network for the early detection of Classical Swine Fever in pigs. In this network, we substituted leaky noisy-OR tables for the expert-provided conditional probability tables for one-third of the variables. The performances of the original network and its leaky noisy-OR version on real-life data were investigated in terms of their sensitivity and specificity characteristics. The results of our study showed that using the leaky noisy-OR model had little effect on the performance of our CSF network. Although more research is required to corroborate our findings, a cautious conclusion from our study is that the (leaky) noisy-OR model can indeed be applied, without extensive knowledge elicitation efforts, to lighten the quantification task for Bayesian networks for diagnostic applications.

Acknowledgment. The research in this paper was supported by the Netherlands Organisation for Scientific Research.

References

1. Druzdzel, M.J., van der Gaag, L.C.: Building probabilistic networks: "Where do the numbers come from?" Guest editors' introduction. IEEE Transactions on Knowledge and Data Engineering 12, 481–486 (2000)
2. van der Gaag, L.C., Renooij, S., Witteman, C.L.M., Aleman, B., Taal, B.G.: How to elicit many probabilities. In: Laskey, K.B., Prade, H. (eds.) Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence, pp. 647–654 (1999)
3. Díez, F.J., Druzdzel, M.J.: Canonical probabilistic models for knowledge engineering. Technical Report CISIAD-06-01 (2007)
4. Lucas, P.J.F.: Bayesian network modelling through qualitative patterns. Artificial Intelligence 163, 233–263 (2005)
5. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers, Palo Alto (1988)
6. Díez, F.J.: Parameter adjustment in Bayes networks. The generalized noisy OR gate. In: Proceedings of the Nineth Annual Conference on Uncertainty in Artificial Intelligence, pp. 99–105 (1993)
7. Zagorecki, A., Druzdzel, M.J.: Knowledge engineering for Bayesian networks: How common are noisy-MAX distributions in practice? In: Brewka, G., Coradeschi, S., Perini, A., Traverso, P. (eds.) Proceedings of the Seventeenth European Conference on Artificial Intelligence, pp. 482–489 (2006)
8. Onisko, A., Druzdzel, M.J., Wasyluk, H.: Learning Bayesian network parameters from small data sets: Application of noisy-OR gates. International Journal of Approximate Reasoning 27(2), 165–182 (2001)
9. Henrion, M.: Some practical issues in constructing belief networks. In: Kanal, L.N., Levitt, T.S., Lemmer, J.F. (eds.) Uncertainty in Artificial Intelligence 3, pp. 161–173 (1989)
10. Zagorecki, A., Druzdzel, M.J.: An empirical study of probability elicitation under noisy-OR assumption. In: Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference, pp. 880–885 (2004)