Probabilistic Independence of Causal Influences

Adam Zagorecki
Engineering Systems Department
Defence Academy of the UK
Cranfield University
Shrivenham, UK

Marek Druzdzel
Decision Systems Laboratory
School of Information Sciences
University of Pittsburgh
Pittsburgh, PA 15260, USA

Abstract

One practical problem with building large scale Bayesian network models is an exponential growth of the number of numerical parameters in conditional probability tables. Obtaining large number of probabilities from domain experts is too expensive and too time demanding in practice. A widely accepted solution to this problem is the assumption of independence of causal influences (ICI) which allows for parametric models that define conditional probability distributions using only a number of parameters that is linear in the number of causes. ICI models, such as the noisy-OR and the noisy-AND gates, have been widely used by practitioners. In this paper we propose PICI, probabilistic ICI, an extension of the ICI assumption that leads to more expressive parametric models. We provide examples of three PICI models and demonstrate how they can cope with a combination of positive and negative influences, something that is hard for noisy-OR and noisy-AND gates.

1 INTRODUCTION

Bayesian networks (Pearl, 1988) have proved their value as a modeling tool in many distinct domains. One of the most serious bottlenecks in applying this modeling tool is the costly process of creating the model. Although practically applicable algorithms for learning BN from data are available (Heckerman, 1999), still there are many applications that lack quality data and require using an expert's knowledge to build models.

One of the most serious problems related to building practical models is the large number of conditional probability distributions needed to be specified when a node in the graph has large number of parent nodes. In the case of discrete variables, which we assume in this paper, conditional probabilities are encoded in the form of conditional probability tables (CPTs) that are indexed by all possible combinations of parent states. For example, 15 binary parent variables result in over 32,000 parameters, a number that is impossible to specify in a direct manner. There are two qualitatively different approaches to avoiding the problem of specifying large CPTs. The first is to exploit internal structure within a CPT – which basically means to encode efficiently symmetries in CPTs (Boutilier et al., 1996; Boutilier et al., 1995). The other is to assume some model of interaction among causes (parent influences) that defines the effect's (child node's) CPT. The most popular class of model in this category is based on the concept known as *causal independence* or *independence of causal influences* (ICI) (Heckerman and Breese, 1994) which we describe in Section 2. These two approaches should be viewed as complementary. It is because they are able to capture two distinct types of interactions between causes.

In practical applications, the noisy-OR (Good, 1961; Pearl, 1988) model together with its extension capable of handling multi-valued variables, the noisy-MAX (Henrion, 1989), and the complementary models the noisy-AND/MIN (Díez and Druzdzel, 2002) are the most often applied ICI models. One of the obvious limitations of these models is that they capture only a small set of patterns of interactions among causes, in particular they do not allow for combining both positive and negative influences. In this paper we introduce an extension to the independence of causal influences which allows us to define models that capture both positive and negative influences. We believe that the new models are of practical importance as practitioners with whom we have had contact often express a need for conditional distribution models that allow for a combination of promoting and inhibiting causes.

The problem of insufficient expressive power of the ICI models has been recognized by practitioners and we are aware of at least two attempts to propose models that are based on the ICI idea, but are not strictly ICI models. The first of them is the *recursive noisy-OR* (Lemmer and Gossink, 2004) which allows the specification of interactions among parent causes, but still is limited to only positive influences. The other interesting proposal is the CAST logic (Chang et al., 1994; Rosen and Smith, 1996) which allows for combining both positive and negative influences in a single model, however detaches from a probabilistic interpretation of the parameters, and consequently leads to difficulties with their interpretation and it can not be exploited to speed-up inference.

The remainder of this paper is organized as follows. In Section 2, we briefly describe independence of causal influences. In Section 3 we discuss the amechanistic property of the ICI models, while in Section 4 we introduce an extension of ICI – a new family of models – probabilistic independence of causal influences. In the following two Sections 5 and 6, we introduce two examples of models for local probability distributions, that belong to the new family. Finally, we conclude our paper with discussion of our proposal in Section 7.

2 INDEPENDENCE OF CAUSAL INFLUENCES (ICI)

In this section we briefly introduce the concept of independence of causal influences (ICI). First

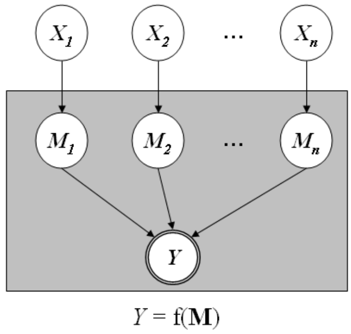

Figure 1: General form of independence of causal interactions

however we introduce notation used throughout the paper. We denote an effect (child) variable as Y and its n parent variables as $\mathbf{X} = \{X_1, \ldots, X_n\}$. We denote the states of a variable with lower case, e.g. $X = x$. When it is unambiguous we use x instead of $X = x$.

In the ICI model, the interaction between variables X_i and Y is defined by means of (1) the *mechanism* variables M_i, introduced to quantify the influence of each cause on the effect separately, and (2) the deterministic function f that maps the outputs of M_i into Y. Formally, the causal independence model is a model for which two independence assertions hold: (1) for any two mechanism variables M_i and M_j $(i \neq j)$ M_i is independent of M_j given $X_1, \ldots, X_n$, and (2) M_i and any other variable in the network that does not belong to the causal mechanism are independent given $X_1, \ldots, X_n$ and Y. An ICI model is shown in Figure 1.

The most popular example of an ICI model is the noisy-OR model. The noisy-OR model assumes that all variables involved in the interaction are binary. The mechanism variables in the context of the noisy-OR are often referred to as *inhibitors*. The inhibitors have the same range as Y and their CPTs are defined as follows:

$$\begin{aligned} P(M_i = y | X_i = x_i) &= p_i \\ P(M_i = y | X_i = \overline{x}_i) &= 0\,. \end{aligned} \quad (1)$$

Function f that combines the individual influences is the deterministic OR. It is important to note that the domain of the function defining the individual influences are the outcomes (states) of Y (each mechanism variable maps $Range(X_i)$ to $Range(Y)$). This means that f is of the form $Y = f(M_1, \ldots, M_n)$, where typically all variables M_i and Y take values from the same set. In the case of the noisy-OR model it is $\{y, \overline{y}\}$. The noisy-MAX model is an extension of the noisy-OR model to multi-valued variables where the combination function is the deterministic MAX defined over Y's outcomes.

3 AMECHANISTIC PROPERTY

The amechanistic property of the causal interaction models was explicated by Breese and Heckerman, originally under the name of *atemporal* (Heckerman, 1993), although later the authors changed the name to *amechanistic* (Heckerman and Breese, 1996). The amechanistic property of ICI models relates to a major problem of this proposal — namely the problem of determining semantic meaning of causal mechanisms. In practice, it is often impossible to say anything about the nature of causal mechanisms (as they can often be simply artificial constructs for the sake of modeling) and therefore they, or their parameters, can not be specified explicitly. Even though Heckerman and Breese proposed a strict definition of the amechanistic property, in this paper we will broaden this definition and assume that an ICI model is an amechanistic model, when its parameterization can be defined exclusively by means of (a subset) of conditional probabilities $P(Y|\mathbf{X})$ without mentioning M_is explicitly. This removes the burden of defining mechanisms directly.

To achieve this goal it is assumed that one of the states of each cause X_i is a special state (also referred to as as the *distinguished* state). Usually such a state is the normal state, like *ok* in hardware diagnostic systems or *absent* for a disease in a medical system, but such association depends on the modeled domain. We will use * symbol to denote the distinguished state. Given that all causes X_i are in their distinguished states, the effect variable Y is guaranteed to be in its distinguished state. The idea is to allow for easy elicitation of parameters of the intermediate nodes M_i, even though these can not be observed directly. This is achieved through a particular way of setting (controlling) the causes X_i. Assuming that all causes except for a single cause X_i are in their distinguished states and X_i is in some state (not distinguished), it is easy to determine the probability distribution for the hidden variable M_i.

Not surprisingly, the noisy-OR model is an example of an amechanistic model. In this case, the distinguished states are usually *false* or *absent* states, because the effect variable is guaranteed to be in the distinguished state given that all the causes are in their distinguished states $X_i = \overline{x}_i$. Equation 1 reflects the amechanistic assumption and results in the fact that the parameters of the mechanisms (inhibitors) can be obtained directly as the conditional probabilities: $P(Y = y|\overline{x}_1, \ldots, \overline{x}_{i-1}, x_i, \overline{x}_{i+1}, \ldots, \overline{x}_n)$. Similarly, the noisy-MAX is an amechanistic model – it assumes that each parent variable has a distinguished state (arbitrarily selected) and the effect variable has a distinguished state. In the case of the effect variable, the distinguished state is assumed to be the lowest state (with respect to the ordering relation imposed on the effect variable's states). We strongly believe that the amechanistic property is highly significant from the point of view of knowledge acquisition. Even though we introduce a parametric model instead of a traditional CPT, the amechanistic property causes the parametric model to be defined in terms of a conditional probability distribution $P(Y|\mathbf{X})$ and, therefore, is conceptually consistent with the BN framework. We believe that the specification of a parametric model in terms of probabilities has contributed to the great popularity of the noisy-OR model.

4 PROBABILISTIC ICI

The combination function in the ICI models is defined as a mapping of mechanisms' states into the states of the effect variable Y. Therefore, it can be written as $Y = f(\mathbf{M})$, where $\mathbf{M}$ is a vector of mechanism variables. Let Q_i be a set of parameters of CPT of node M_i, and

$\mathbf{Q} = \{Q_1, \ldots, Q_n\}$ be a set of all parameters of all mechanism variables. Now we define the new family *probabilistic independence of causal interactions* (PICI) for local probability distributions. A PICI model for the variable Y consists of (1) a set of n mechanism variables M_i, where each variable M_i corresponds to exactly one parent X_i and has the same range as Y, and (2) a combination function f that transforms a set of probability distributions Q_i into a single probability distribution over Y. The mechanisms M_i in the PICI obey the same independence assumptions as in the ICI. The PICI family is defined in a way similar to the ICI family, with the exception of the combination function, that is defined in the form $P(Y) = f(\mathbf{Q}, \mathbf{M})$. The PICI family includes both ICI models, which can be easily seen from its definition, as $f(\mathbf{M})$ is a subset of $f(\mathbf{Q}, \mathbf{M})$, assuming $\mathbf{Q}$ is the empty set.

In other words, in the case of ICI for a given instantiation of the states of the mechanism variables, the state of Y is a function of the states of the mechanism variables, while for the PICI the distribution over Y is a function of the states of the mechanism variables and some parameters $\mathbf{Q}$.

Heckerman and Breese (Heckerman, 1993) identified other forms (or rather properties) of the ICI models that are interesting from the practical point of view. We would like to note that those forms (decomposable, multiple decomposable, and temporal ICI) are related to properties of the function f, and can be applied to the PICI models in the same way as they are applied to the ICI models.

5 NOISY-AVERAGE MODEL

In this section, we propose a new local distribution model that is a PICI model. Our goal is to propose a model that (1) is convenient for knowledge elicitation from human experts by providing a clear parameterization, and (2) is able to express interactions that are impossible to capture by other widely used models (like the noisy-MAX model). With this model we are interested in modeling positive and negative influences on the effect variable that has a distinguished state in the middle of the scale.

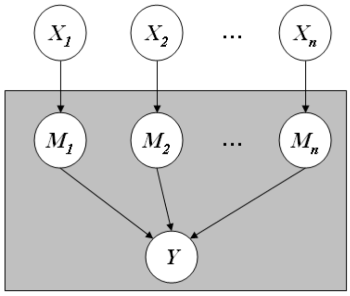

Figure 2: BN model for probabilistic independence of causal interactions, where $P(Y|\mathbf{M}) = f(\mathbf{Q}, \mathbf{M})$.

We assume that the parent nodes X_i are discrete (not necessarily binary, nor an ordering relation between states is required), and each of them has one distinguished state, that we denote as x_i^*. The distinguished state is not a property of a parent variable, but rather a part of a definition of a causal interaction model — a variable that is a parent in two causal independence models may have different distinguished states in each of these models. The effect variable Y also has its distinguished state, and by analogy we will denote it by y^*. The range of the mechanism variables M_i is the same as the range of Y. Unlike the noisy-MAX model, the distinguished state may be in the middle of the scale.

In terms of parameterization of the mechanisms, the only constraint on the distribution of M_i conditional on $X_i = x_i^*$ is:

$$\begin{aligned} P(M_i = m_i^* | X_i = x_i^*) &= 1 \\ P(M_i \neq m_i^* | X_i = x_i^*) &= 0 \,, \end{aligned} \qquad (2)$$

while the other parameters in the CPT of M_i can take arbitrary values.

The definition of the CPT for Y is a key element of our proposal. In the ICI models, the CPT for Y was by definition constrained to be a deterministic function, mapping states of M_is to the states of Y. In our proposal, we define the CPT of Y to be a function of probabilities of the M_is:

$$P(y|\mathbf{x}) = \begin{cases} \prod_{i=1}^{n} P(M_i = y^*|x_i) & \text{for } y = y^* \\ \frac{\alpha}{n} \sum_{i=1}^{n} P(M_i = y|x_i) & \text{for } y \neq y^* \end{cases}$$

where α is a normalizing constant discussed later. Let $q_i^{j^*} = q_i^*$. For simplicity of notation assume that $q_i^j = P(M_i = y^j|x_i)$, $q_i^* = P(M_i = y^*|x_i)$, and $D = \prod_{i=1}^{n} P(M_i = y^*|x_i)$. Then we can write:

$$\sum_{j=1}^{m_y} P(y_j|\mathbf{x}) = D + \sum_{j=1, j\neq j^*}^{m_y} \frac{\alpha}{n} \sum_{i=1}^{n} q_i^j =$$

$$= D + \frac{\alpha}{n} \sum_{j=1, j\neq j^*}^{m_y} \sum_{i=1}^{n} q_i^j = D + \frac{\alpha}{n} \sum_{i=1}^{n} (1 - q_i^*),$$

where m_y is the number of states of Y. Since the sum on the left must equal 1, as it defines the probability distribution $P(Y|\mathbf{x})$, we can calculate α as:

$$\alpha = \frac{n(1-D)}{\sum_{i=1}^{n} (1 - q_i^*)} \ .$$

The definition above does not define $P(Y|\mathbf{M})$ but rather $P(Y|\mathbf{X})$. It is possible to calculate $P(Y|\mathbf{M})$ from $P(Y|\mathbf{X})$, though it is not needed to use the model. Now we discuss how to obtain the probabilities $P(M_i|X_i)$. Using definition of $P(y|\mathbf{x})$ and the amechanistic property, this task amounts to obtaining the probabilities of Y given that X_i is in its non-distinguished state and all other causes are in their distinguished states (in a very similar way to how the noisy-OR parameters are obtained). $P(y|\mathbf{x})$ in this case takes the form:

$$P(Y = y|x_1^*, \ldots, x_{i-1}^*, x_i, \ldots, x_{i+1}^*, x_n^*) =$$
$$= P(M_i = y|x_i) \ , \qquad (3)$$

and, therefore, defines an easy and intuitive way for parameterizing the model by just asking for conditional probabilities, in a very similar way to the noisy-OR model. It is easy to notice that $P(Y = y^*|x_1^*, \ldots, x_i^*, \ldots, x_n^*) = 1$, which poses a constraint that may be unacceptable from a modeling point of view. We can address this limitation in a very similar way to the noisy-OR model, by assuming a dummy variable X_0 (often referred to as *leak*), that stands for all unmodeled causes and is assumed to be always in some state x_0. The leak probabilities are obtained using:

$$P(Y = y|x_1^*, \ldots, x_n^*) = P(M_0 = y) \ .$$

However, this slightly complicates the schema for obtaining parameters $P(M_i = y|x_i)$. In the case of the leaky model, the equality in Equation 3 does not hold, since X_0 acts as a regular parent variable that is in a non-distinguished state. Therefore, the parameters for other mechanism variables should be obtained using conditional probabilities $P(Y = y|x_1^*, \ldots, x_i, \ldots, x_n^*)$, $P(M_0 = y|x_0)$ and the combination function. This implies that the acquired probabilities should fulfil some non-trivial constraints. Because of space limitation, we decided to skip the discussion of these constraints. In a nutshell, these constraints are similar in nature to constraints for the leaky noisy-MAX model. These constraints should not be a problem in practice, when $P(M_0 = y^*)$ is large (which implies that the leak cause has marginal influence on non-distinguished states).

Now we introduce an example of the application of the new model. Imagine a simple diagnostic model for an engine cooling system. The pressure sensor reading (S) can be in three states high, normal, or low $\{hi, nr, lo\}$, that correspond to pressure in a hose. Two possible faults included in our model are: *pump failure* (P) and *crack* (C). The pump can malfunction in two distinct ways: work non-stop instead of adjusting its speed, or simply fail and not work at all. The states for *pump failure* are: $\{ns, fa, ok\}$. For simplicity we assume that the crack on the hose can be *present* or *absent* $\{pr, ab\}$. The BN for this problem is presented in Figure 3. The noisy-MAX model is not appropriate here, because the distinguished state of the effect variable (S) does not correspond to the lowest value in the ordering relation. In other words, the neutral value is not one of the extremes, but lies in the middle, which makes use of the MAX function over the states inappropriate. To apply the noisy-average model, first we should identify the distinguished states of the variables.

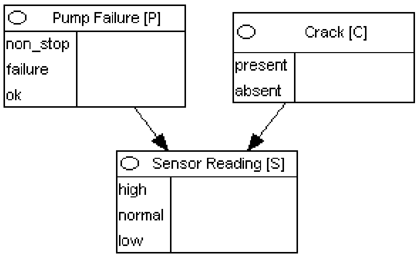

Figure 3: BN model for the pump example.

In our example, they will be: *normal* for *sensor reading*, *ok* for *pump failure* and *absent* for *crack*. The next step is to decide whether we should add an influence of non-modeled causes on the sensor (a leak probability). If such an influence is not included, this would imply that $P(S = nr * | P = ok*, C = ab*) = 1$, otherwise this probability distribution can take arbitrary values from the range $(0, 1]$, but in practice it should always be close to 1.

Assuming that the influence of non-modeled causes is not included, the acquisition of the mechanism parameters is performed directly by asking for conditional probabilities of form $P(Y|x_1^*, \ldots, x_i, \ldots, x_n^*)$. In that case, a typical question asked of an expert would be: *What is the probability of the sensor being in the low state, given that a crack was observed but the pump is in state ok?* However, if the unmodeled influences were significant, an adjustment for the leak probability is needed. Having obtained all the mechanism parameters, the noisy-average model specifies a conditional probability in a CPT by means of the combination function.

Intuitively, the noisy-average combines the various influences by averaging probabilities. In case where all active influences (the parents in non-distinguished states) imply high probability of one value, this value will have a high posterior probability, and the synergetic effect will take place similarly to the noisy-OR/MAX models. If the active parents will 'vote' for different effect's states, the combined effect will be an average of the individual influences. Moreover, the noisy-average model is a decomposable model — the CPT of Y can be decomposed in pairwise relations (Figure 4) and such a decomposition can be exploited for sake of improved inference speed in the same way as for decomposable ICI models.

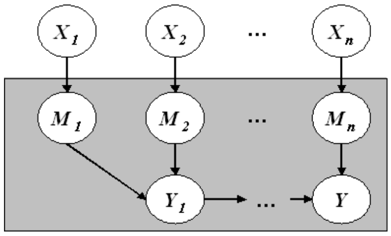

Figure 4: Decomposition of a combination function.

It is important to note that the noisy-average model does not take into account the ordering of states (only the distinguished state is treated in a special manner). If a two causes have high probability of *high* and *low* pressure, one should not expect that combined effect will have probability of normal state (the value in-between).

6 AVERAGE MODEL

Another example of a PICI model we want to present is the model that averages influences of mechanisms. This model highlights another property of the PICI models that is important in practice. If we look at the representation of a PICI model, we will see that the size of the CPT of node Y is exponential in the number of mechanisms (or causes). Hence, in general case it does not guarantee a low number of distributions. One solution is to define a combination function that can be expressed **explicitly** in the form of a BN but in such a way that it has significantly fewer parameters. In the case of ICI models, the decomposability property (Heckerman and Breese, 1996) served this purpose, and can do too for in PICI models. This property allows for significant speed-ups in inference.

In the average model, the probability distribution over Y given the mechanisms is basically a ratio of the number of mechanisms that are in given state divided by the total number of mechanisms (by definition Y and $\mathbf{M}$ have the same range):

$$P(Y = y|M_1, \ldots, M_n) = \frac{1}{n}\sum_{i=1}^{n} I(M_i = y) \ . \quad (4)$$

where I is the identity function. Basically, this combination function says that the probability of the effect being in state y is the ratio of mechanisms that result in state y to all mechanisms. Please note that the definition of how a cause X_i results in the effect is defined in the probability distribution $P(M_i|X_i)$. The pairwise decomposition can be done as follows:

$$P(Y_i = y|Y_{i-1} = a, M_n = b)$$

$$= \frac{i}{i+1} I(y = a) + \frac{1}{i+1} I(y = b) \ ,$$

for $Y_2, \ldots, Y_n$. Y_1 is defined as:

$$P(Y_1 = y|M_1 = a, M_2 = b) =$$

$$= \frac{1}{2} I(y = a) + \frac{1}{2} I(y = b) \ .$$

The fact that the combination function is decomposable may be easily exploited by inference algorithms. Additionally, we showed that this model presents benefits for learning from small data sets (Zagorecki et al., 2006).

Theoretically, this model is amechanistic, because it is possible to obtain parameters of this model (probability distributions over mechanism variables) by asking an expert only for probabilities in the form of $P(Y|\mathbf{X})$. For example, assuming variables in the model are binary, we have $2n$ parameters in the model. It would be enough to select $2n$ arbitrary probabilities $P(Y|\mathbf{X})$ out of 2^n and create a set of $2n$ linear equations applying Equation 4. Though in practice, one needs to ensure that the set of equations has exactly one solution what in general case is not guaranteed.

As an example, let us assume that we want to model classification of a threat at a military checkpoint. There is an expected terrorist threat at that location and there are particular elements of behavior that can help spot a terrorist. We can expect that a terrorist can approach the checkpoint in a large vehicle, being the only person in the vehicle, try to carry the attack at rush hours or time when the security is less strict, etc. Each of these behaviors is not necessarily a strong indicator of terrorist activity, but several of them occurring at the same time may indicate possible threat.

The average model can be used to model this situation as follows: separately for each of suspicious activities (causes) a probability distribution of terrorist presence given this activity can be obtained which basically means specification of probability distribution of mechanisms. Then combination function for the average model acts as "popular voting" to determine $P(Y|\mathbf{X})$.

The average model draws ideas from the linear models, but unlike the linear models, the linear sum is done over probabilities (as it is PICI), and it has explicit hidden mechanism variables that define influence of single cause on the effect.

7 CONCLUSIONS

In this paper, we formally introduced a new class of models for local probability distributions that is called PICI. The new class is an extension of the widely accepted concept of independence of causal influences. The basic idea is to relax the assumption that the combination function should be deterministic. We claim that such an assumption is not necessary either for clarity of the models and their parameters, nor for other aspects such as convenient decompositions of the combination function that can be exploited by inference algorithms.

To support our claim, we presented two conceptually distinct models for local probability distributions that address different limitations of existing models based on the ICI. These models have clear parameterizations that facilitate their use by human experts. The proposed models can be directly exploited by inference algorithms due to fact that they can be explicitly represented by means of a BN, and their combination function can be decomposed into a chain of binary relationships. This property has been recognized to provide significant inference speed-ups for the ICI models. Finally, because they can be represented in form of hidden variables, their parameters can be learned using the EM algorithm.

We believe that the concept of PICI may lead to new models not described here. One remark we shall make here: it is important that new models should be explicitly expressible in terms of a BN. If a model does not allow for compact representation and needs to be specified as a CPT for inference purposes, it undermines a significant benefit of models for local probability distributions – a way to avoid using large conditional probability tables.

Acknowledgements

This research was supported by the Air Force Office of Scientific Research grants F49620–03–1–0187 and FA9550–06–1–0243 and by Intel Research. While we take full responsibility for any errors in the paper, we would like to thank anonymous reviewers for several insightful comments that led to improvements in the presentation of the paper, Tsai-Ching Lu for inspiration, invaluable discussions and critical comments, and Ken McNaught for useful comments on the draft.

References

C. Boutilier, R. Dearden, and M. Goldszmidt. 1995. Exploiting structure in policy construction. In Chris Mellish, editor, *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pages 1104–1111, San Francisco. Morgan Kaufmann.

C. Boutilier, N. Friedman, M. Goldszmidt, and D. Koller. 1996. Context-specific independence in Bayesian networks. In *Proceedings of the 12th Annual Conference on Uncertainty in Artificial Intelligence (UAI-96)*, pages 115–123, San Francisco, CA. Morgan Kaufmann Publishers.

K.C. Chang, P.E. Lehner, A.H. Levis, Abbas K. Zaidi, and X. Zhao. 1994. On causal influence logic. *Technical Report for Subcontract no. 26-940079-80.*

F. J. Díez and Marek J. Druzdzel. 2002. Canonical probabilistic models for knowledge engineering. Forthcoming.

I. Good. 1961. A causal calculus (I). *British Journal of Philosophy of Science*, 11:305–318.

D. Heckerman and J. Breese. 1994. A new look at causal independence. In *Proceedings of the Tenth Annual Conference on Uncertainty in Artificial Intelligence (UAI–94)*, pages 286–292, San Francisco, CA. Morgan Kaufmann Publishers.

D. Heckerman and J. Breese. 1996. Causal independence for probability assessment and inference using Bayesian networks. In *IEEE, Systems, Man, and Cybernetics*, pages 26:826–831.

D. Heckerman. 1993. Causal independence for knowledge acquisition and inference. In *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence (UAI–93)*, pages 122–127, Washington, DC. Morgan Kaufmann Publishers.

D. Heckerman. 1999. A tutorial on learning with Bayesian networks. In *Learning in Graphical Models*, pages 301–354, Cambridge, MA, USA. MIT Press.

M. Henrion. 1989. Some practical issues in constructing belief networks. In L.N. Kanal, T.S. Levitt, and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence 3*, pages 161–173. Elsevier Science Publishing Company, Inc., New York, N. Y.

J. F. Lemmer and D. E. Gossink. 2004. Recursive noisy-OR: A rule for estimating complex probabilistic causal interactions. *IEEE Transactions on Systems, Man and Cybernetics*, (34(6)):2252 – 2261.

J. Pearl. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers, Inc., San Mateo, CA.

J. A. Rosen and W. L. Smith. 1996. Influence net modeling and causal strengths: An evolutionary approach. In *Command and Control Research and Technology Symposium.*

A. Zagorecki, M. Voortman, and M. Druzdzel. 2006. Decomposing local probability distributions in Bayesian networks for improved inference and parameter learning. In *FLAIRS Conference.*