Lattices for Studying Monotonicity of Bayesian Networks

Linda C. van der Gaag, Silja Renooij, and Petra L. Geenen
Department of Information and Computing Sciences, Utrecht University,
P.O. Box 80.089, 3508 TB Utrecht, the Netherlands.
e-mail: {linda,silja,petrag}@cs.uu.nl

Abstract

In many real problem domains, the main variable of interest behaves monotonically in terms of the observable variables, in the sense that higher values for the variable of interest become more likely with higher-ordered observations. Unfortunately, establishing whether or not a Bayesian network exhibits these monotonicity properties is highly intractable in general. In this paper, we present a method that, by building upon the concept of assignment lattice, provides for identifying any violations of the properties of (partial) monotonicity of the output and for constructing minimal offending contexts. We illustrate the application of our method with a real Bayesian network in veterinary science.

1 Introduction

In many problem domains, the variables of importance have different roles. Often, a number of observable input variables and a single output variable are distinguished. In a biomedical diagnostic application, for example, the input variables capture the findings from different diagnostic tests and the output variable models the possible diseases. Multiple input variables and a single output variable in fact are typically found in any type of diagnostic problem.

For many problems, the relation between the output variable and the observable input variables is monotone in the sense that higher-ordered values for the input variables give rise to a higher-ordered output for the main variable of interest. In a biomedical diagnostic application, for example, observing symptoms and signs that are more severe will result in a more severe disease being the most likely value of the diagnostic variable. The concept of monotonicity in distribution has been introduced to capture this type of knowledge for Bayesian networks (Van der Gaag *et al.*, 2004). More specifically, a network is said to be isotone if the conditional probability distribution computed for the output variable given specific observations is stochastically dominated by any such distribution given higher-ordered observations.

Unfortunately, the problem of verifying whether or not a Bayesian network exhibits the properties of monotonicity from its domain of application is highly intractable in general and remains to be so even for polytrees. Although an approximate anytime algorithm is available for deciding if a given network is monotone, we found that its runtime requirements tend to forestall use in a practical setting.

In this paper, we present a new, more practicable method for studying monotonicity of Bayesian networks. The method builds upon a lattice of all joint value assignments to the observable variables under study; this lattice moreover, is enhanced with information about the effects of the various assignments on the probability distribution over the main variable of interest. The assignment lattice is used for identifying any violations of the properties of monotonicity in the network. Subsequently, minimal offending contexts are constructed that characterise the identified violations and provide for further investigation.

The runtime complexity of our method is exponential in the number of observable variables under study. For larger networks including a large number of observable variables therefore, our method provides for verifying partial monotonicity for limited subsets of variables only. The unfavourable computational complexity of the problem, however, is likely to forestall the design of essentially more efficient methods.

We applied our verification method for studying monotonicity to a Bayesian network in veterinary science. In recent years, we developed a network for the detection of classical swine fever in pigs. Both the network's structure and its associated probabilities were elicited from two experts. During the elicitation interviews, the experts had produced several statements that suggested properties of monotonicity. We studied these properties in the network using our verification method. We found a small number of violations of the suggested properties of monotonicity, which proved to be indicative of modelling inadequacies.

The paper is organised as follows. In Section 2, we review the concept of monotonicity. In Section 3, we present our method for studying monotonicity. We report on the application of our method in Section 4. The paper ends with our concluding observations in Section 5.

2 The Concept of Monotonicity

Upon reviewing the concept of monotonicity, we assume that a Bayesian network under study includes a single output variable C and n observable variables E_i, $i = 1, \ldots, n$, $n \geq 1$; in addition, the network may include an arbitrary number of intermediate variables which are not observed in practice. Each variable X_i in the network adopts one of a finite set $\Omega(X_i) = \{x_i^1, \ldots, x_i^m\}$, $m \geq 1$, of values. We assume that there exists a total ordering $\leq$ on this set of values; without loss of generality, we assume that $x_i^j \leq x_i^k$ whenever $j \leq k$. The orderings per variable are taken to induce a partial ordering $\preceq$ on the set of joint value assignments to any subset of the network's variables.

The concept of *monotonicity in distribution* now builds upon the posterior probability distributions over the output variable given the various joint value assignments to the network's observable variables (Van der Gaag *et al.*, 2004). The concept, more specifically, is defined in terms of stochastic dominance among these distributions. For a probability distribution $\Pr(C)$ over the output variable, the cumulative distribution function $F_{\Pr}$ is defined as $F_{\Pr}(c^i) = \Pr(C \leq c^i)$ for all values c^i of C. For two distributions $\Pr(C)$ and $\Pr'(C)$ over C, associated with $F_{\Pr}(C)$ and $F_{\Pr'}(C)$ respectively, we say that $\Pr'(C)$ is *stochastically dominant* over $\Pr(C)$, denoted $\Pr(C) \leq \Pr'(C)$, if $F_{\Pr'}(c^i) \leq F_{\Pr}(c^i)$ for all $c^i \in \Omega(C)$ (Berger, 1980). We now say that the network is *isotone* in its set of observable variables E if

$$e \preceq e' \rightarrow \Pr(C \mid e) \leq \Pr(C \mid e')$$

for all joint value assignments e, e' to E. Informally speaking, we have that a network is isotone in distribution if entering a higher-ordered value assignment to the observable variables cannot make higher-ordered values of the output variable less likely. The concept of antitonicity has the reverse interpretation: the network is said to be *antitone* in E if

$$e \preceq e' \rightarrow \Pr(C \mid e) \geq \Pr(C \mid e')$$

for all value assignments e, e' to E. Note that if a network is isotone in its observable variables given the orderings $\leq$ on their sets of values, then the network is antitone given the reversed orderings. Although antitonicity thus is (reversely) equivalent to isotonicity, we explicitly distinguish between the two types of monotonicity since a domain of application may exhibit an intricate combination of isotonicity and antitonicity for interrelated observable variables.

Building upon the above definitions, we have that deciding whether or not a Bayesian network is isotone amounts to verifying that entering any higher-ordered value assignment to its observable variables results in a stochastically dominant probability distribution over the output variable. Establishing antitonicity amounts to verifying that entering any such assignment results in a dominated distribution.

The problem of deciding monotonicity for a Bayesian network is known to be coNP^{PP}-complete in general, and in fact remains coNP-complete for polytrees (Van der Gaag *et al.*, 2004). In view of these complexity considerations, Van der Gaag *et al.* designed an approximate anytime algorithm for verifying whether or not a given network is monotone. This algorithm studies the relation between the output variable and each observable variable separately, in terms of the sign of the qualitative influence between them (Wellman, 1990); upon inconclusive results, iteratively tightened numerical bounds are established on the relevant probabilities using an anytime method available from Liu and Wellman (1998). Unfortunately, the overall runtime requirements of the algorithm tend to forestall use in a practical setting.

3 Studying Monotonicity

Our method for studying monotonicity in Bayesian networks now builds upon the construct of an assignment lattice. The lattice includes all joint value assignments to the set of observable variables and is enhanced with probabilistic information computed from the network under study. From the lattice, all violations of the properties of monotonicity are identified, from which minimal offending contexts are constructed for further investigation.

3.1 The assignment lattice

The *assignment lattice* for the set E of observable variables of a Bayesian network captures all joint value assignments to E, along with the partial ordering between them. For each value assignment e to E, an element $L(e)$ is included in the lattice. The bottom of the lattice encodes the assignment e for which we have that $e \preceq e'$ for all $e' \in \Omega(E)$; the bottom thus encodes the lowest-ordered value assignment to E. The top of the lattice encodes the assignment e'' for which we have that $e' \preceq e''$ for all $e' \in \Omega(E)$; the top thus encodes the highest-ordered value assignment to E. In the lattice, we further have that an element $L(e)$ precedes an element $L(e'')$ if $e \preceq e''$; we moreover say that $L(e)$ precedes $L(e'')$ directly if there is no assignment e' with $e \prec e' \prec e''$. The partial ordering defined by the lattice thus coincides with the partial ordering $\preceq$ on the joint value assignments to E. Figure 1 depicts, as an example, the assignment lattice

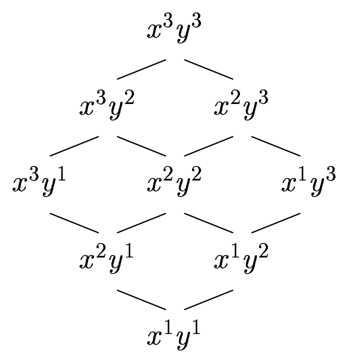

Figure 1: An example assignment lattice for two ternary observable variables.

for the two ternary variables X and Y. For further information about lattices in general, we refer to (Grätzer, 1971).

To describe the effects of the various value assignments e on the probability distribution over the output variable, the assignment lattice is enhanced with probabilistic information. With each element $L(e)$ of the lattice, the conditional probability distribution $\Pr(C \mid e)$ over the output variable C is associated. We note that these distributions are readily computed from the Bayesian network under study. We will return to the complexity of the computations involved in Section 3.4.

3.2 Exploiting the assignment lattice

By building upon the assignment lattice, verifying monotonicity in distribution amounts to checking whether particular dominance properties hold among the probability distributions associated with the lattice's elements.

We recall that a Bayesian network is isotone in its set of observable variables E if entering a higher-ordered value assignment to E results in a stochastically dominant probability distribution over the output variable C. We further recall that the partial ordering $\preceq$ on the joint value assignments to E is encoded directly in the assignment lattice for E. We now consider all pairs of conditional probability distributions $\Pr(C \mid e)$ and $\Pr(C \mid e')$ associated with elements $L(e)$ and $L(e')$ in the lattice such that $L(e)$ precedes $L(e')$. If for each such pair we have that $\Pr(C \mid e) \leq \Pr(C \mid e')$, then the network is isotone in E. The network is antitone in E if for each such pair of distributions we have that $\Pr(C \mid e') \leq \Pr(C \mid e)$. We note that, since the property of stochastic dominance is transitive, we have to study the dominance properties of the distributions of directly linked pairs of elements in the lattice only to decide upon isotonicity or antitonicity.

As mentioned in Section 2, a domain of application may exhibit an intricate combination of properties of isotonicity and antitonicity for interrelated observable variables. As an example, we consider the two variables X and Y such that the output variable of interest behaves isotonically in X and antitonically in Y. We focus on the value assignments

$$\begin{array}{rcl} e_{ij} & \equiv & x^i y^{j+1} \\ e'_{ij} & \equiv & x^{i+1} y^j \\ e''_{ij} & \equiv & x^{i+1} y^{j+1} \end{array}$$

to these variables. Note that for the three assignments we have that $e_{ij} \preceq e''_{ij}$ and $e'_{ij} \preceq e''_{ij}$; the assignments e_{ij} and e'_{ij} have no ordering. For studying the monotonicity properties involved, we now have to compare the probability distributions over the output variable C that are computed from the Bayesian network under study, given the various assignments to X and Y. If for the pair of distributions $\Pr(C \mid e_{ij})$ and $\Pr(C \mid e''_{ij})$, we have that $\Pr(C \mid e''_{ij})$ is stochastically dominant over $\Pr(C \mid e_{ij})$, then the network exhibits the associated property of isotonicity in X. If for the distributions given e'_{ij} and e''_{ij} we have that $\Pr(C \mid e'_{ij})$ is stochastically dominant over $\Pr(C \mid e''_{ij})$, then the network shows the associated property of antitonicity in Y. The network thus reveals both properties if

$$\Pr(C \mid e_{ij}) \leq \Pr(C \mid e''_{ij}) \leq \Pr(C \mid e'_{ij})$$

Verifying whether or not the network is isotone in X and antitone in Y now amounts to studying the above inequalities for all values x^i and y^j of the two variables.

3.3 Deriving offending contexts

Upon using the assignment lattice as described above, some properties of monotonicity may not show in the Bayesian network under study. We then say that these properties are violated. More formally, we say that a pair of joint value assignments e, e' with $e \preceq e'$ *violates* the property of isotonicity if we have $\Pr(C \mid e) > \Pr(C \mid e')$ for their associated probability distributions; violation of the property of antitonicity is defined similarly. We observe that for any violating pair e, e' whose elements are directly linked in the assignment lattice, there is a *unique* variable E_i in the set of observable variables E to which e and e' assign a different value. Let e_i^j be the value of this variable within the assignment e and let e_i^k, $k \neq j$, be its value in e'. The violation now is denoted by $(e, e' \mid E_i^{j \to k})$. We say that the violation has a change of the value of E_i from e_i^j to e_i^k for its *origin*, written $E_i^{j \to k}$. The value assignment e^- to $E \setminus \{E_i\}$ that is shared by e and e', is called the *context* of the violation.

In general, using the assignment lattice for a Bayesian network may result in a set $\mathcal{V}$ of violations of the properties of monotonicity. Some of the identified violations may show considerable regularity, in the sense that if a violation originates from a change of value in a particular context, then this same change causes a violation in all higher-ordered contexts as well. Such violations will be termed *structural.* Other violations from the set $\mathcal{V}$ do not have such a regular structure and may be considered *incidental.* We distinguish between the two types of violation to allow a compact representation of the entire set of identified violations.

We consider the subset $\mathcal{V}(E_i^{j \to k}) \subseteq \mathcal{V}$ of all violations of the properties of monotonicity that originate from a change of the value of the variable E_i from e_i^j to e_i^k. The context e^- of such a violation now is called a *structural context of offence* if we have that there is a violation $(e, e' \mid E_i^{j \to k}) \in \mathcal{V}(E_i^{j \to k})$ for all value assignments e to E that include this context e^- or a higher-ordered one. If there is no lower-ordered structural context of offence for the same change of value $E_i^{j \to k}$, we say that the context of offence is *structurally minimal.* A structurally minimal context of offence thus characterises a set of related violations. A structurally minimal context e^- for the violations originating from $E_i^{j \to k}$ more specifically defines a sublattice of the assignment lattice in which each element includes e_i^j and in which for each element e there exists a violation $(e, e' \mid E_i^{j \to k})$ in the set $\mathcal{V}(E_i^{j \to k})$; the context e^- is the bottom of this sublattice.

In addition to the violations that are covered by structurally minimal contexts of offence, the set of all identified violations may include elements that are not structurally related. A context e^- is said to be an *incidental context of offence* if it is not a structural context of offence. We note that any set of violations identified from a Bayesian network is now characterised by a set of structurally minimal contexts of offence and a set of incidental contexts.

As an example, we consider the lattice from Figure 1 for the two ternary observable variables X and Y. Suppose that upon verifying isotonicity of the Bayesian network under study, the three violations $(x^1y^2, x^2y^2 \mid X^{1 \to 2}), (x^1y^3, x^2y^3 \mid X^{1 \to 2})$ and $(x^2y^2, x^3y^2 \mid X^{2 \to 3})$ are identified. The first two of these violations have the same origin: the two violations arise if the value of the variable X is changed from x^1 to x^2. These violations are characterised by the structurally minimal context of offence y^2: in both the context y^2 and the higher-ordered context y^3, changing the value of X from x^1 to x^2 gives rise to a violation. The third violation mentioned above is not yet covered by the identified minimal context since it has a different origin. This violation originates from a change of the value of the variable X from x^2 to x^3 and again has y^2 for its offending context. The context y^2 now is merely incidental since no violation has been identified for the pair of assignments x^2y^3 and x^3y^3 with include the higher-ordered context y^3.

To conclude, we would like to note that for a given set of violations, the structurally minimal contexts of offence are readily established by traversing the assignment lattice under study. Starting at the bottom of the lattice, the procedure finds an element $L(e)$ such that e is the lowest-ordered joint value assignment occurring in a violation $(e, e' \mid E_i^{j \to k}) \in \mathcal{V}$. The procedure subsequently isolates from the assignment lattice the sublattice with the element $L(e)$, with $e \equiv e_i^j e^-$, for its bottom; the sublattice further includes all elements $L(e_i^j e^+)$ with $e_i^j e^- \preceq e_i^j e^+$. Now, if each element of the sublattice occurs in a violation with $E_i^{j \to k}$ for its origin, then the context e^- is a structurally minimal context of offence; otherwise it is incidental. The procedure is iteratively repeated for all yet uncovered violations.

3.4 Complexity considerations

The runtime complexity of our method for studying monotonicity of a Bayesian network is determined by the size of the assignment lattice used. We observe that this lattice encodes an exponential number of value assignments to the set of observable variables. Constructing the lattice and computing the probability distributions to be associated with its elements, therefore, takes exponential time. Moreover, the dominance properties of an exponential number of pairs of probability distributions have to be compared. For n binary observable variables, for example, already

$$\sum_{i=0}^{n} \binom{n}{i} \cdot (n - i) > 2^n$$

comparisons are required. From these considerations we have that our method has a very high runtime complexity. Although its requirements can be reduced to at least some extent by exploiting the independences modelled in a Bayesian network, for larger networks including a large number of observable variables studying monotonicity inevitably becomes infeasible. The unfavourable computational complexity of the problem, however, is likely to forestall the design of essentially more efficient methods.

In view of the high runtime complexity involved, we propose to use our method for studying properties of *partial monotonicity* only. The concept of partial monotonicity applies to a subset X of the observable variables of a network. An assignment lattice is constructed for these variables as described above. The probability distributions associated with the elements of the lattice are conditioned on a *fixed* joint value assignment e^- to the observable variables $E^- = E \setminus X$ that are not included in the study. With each element $L(x)$ of the lattice thus is associated the conditional probability distribution $\Pr(C \mid xe^-)$. The lattice now provides for studying the monotonicity properties of the network for the set X given e^-. The assignment e^- then is termed the *background assignment* for studying the partial monotonicity.

We would like to note that partial monotonicity of a network given a particular background assignment e^- does not guarantee partial monotonicity given any other background assignment. Also, violations of the properties of monotonicity that are identified given a particular background assignment may not occur given another such assignment. The background assignment against which partial monotonicity is to be studied should therefore be chosen with care and be based upon considerations from the domain of application.

4 An Example Application

We applied our method for studying monotonicity to a real Bayesian network in veterinary science. We briefly introduce the network and describe the violations that we identified.

4.1 A network for classical swine fever

In close collaboration with two experts from the Central Institute of Animal Disease Control in the Netherlands, we are developing a Bayesian network for the detection of classical swine fever. Classical swine fever is an infectious disease of pigs, which has serious socio-economical consequences upon an outbreak. As the disease has a potential for rapid spread, it is imperative that its occurrence is detected in the early stages. The network under construction is aimed at supporting veterinary practitioners in the diagnosis of the disease when visiting pig farms with disease problems of unknown cause. Our network currently includes 42 variables for which over 2400 parameter probabilities have been assessed. The variables model the pathogenesis of the disease as well as the clinical signs observed in individual pigs. 24 of the total of 42 variables are observable. Figure 2 depicts the graphical structure of the current network.

4.2 Constructing the lattice

During the elicitation interviews, our veterinary experts produced various statements that suggested monotonicity. They indicated, for example, that the output variable *CSF Viraemia* should behave isotonically in terms of the five variables *Diarrhoea*, *Ataxia*, *Fever*, *Malaise*, and *Skin haemorrhages*. We focus on these observable variables to illustrate the application of our method for studying partial monotonicity.

From the five observable variables under study, we constructed an assignment lattice as described in the previous section. Since all variables involved were binary, adopting one of the values *true* and *false*, we used a slightly more concise encoding of their joint value assignments. For each assignment e to the set of observable variables E, we let $L(e)$ be the subset of E such that $E_i \in L(e)$ if and only if $E_i = true$ occurs in e. The elements of the resulting lattice thus are subsets of E; the bottom of the lattice is the empty set and the top equals the entire set E. The resulting lattice for the five variables under study is shown in Figure 3. It includes $2^5 = 32$ elements to capture all possible joint value assignments to the variables and further includes 80 direct set-inclusion statements.

Before the lattice could be enhanced with conditional probabilities of the presence of a viraemia of classical swine fever, we had to decide upon a background assignment for the other 19 observable variables against which the properties of partial monotonicity would be verified. We decided to take for this purpose the value assignment in which all other observable variables of the network had adopted the value *false*. We chose this particular assignment since the various clinical signs have a rather small probability of occurrence and it is highly unlikely to find a large number of these signs in a single live pig. Given this background assignment, we computed the various conditional probabilities to be associated with the elements of the assignment lattice.

For each pair of directly linked elements from

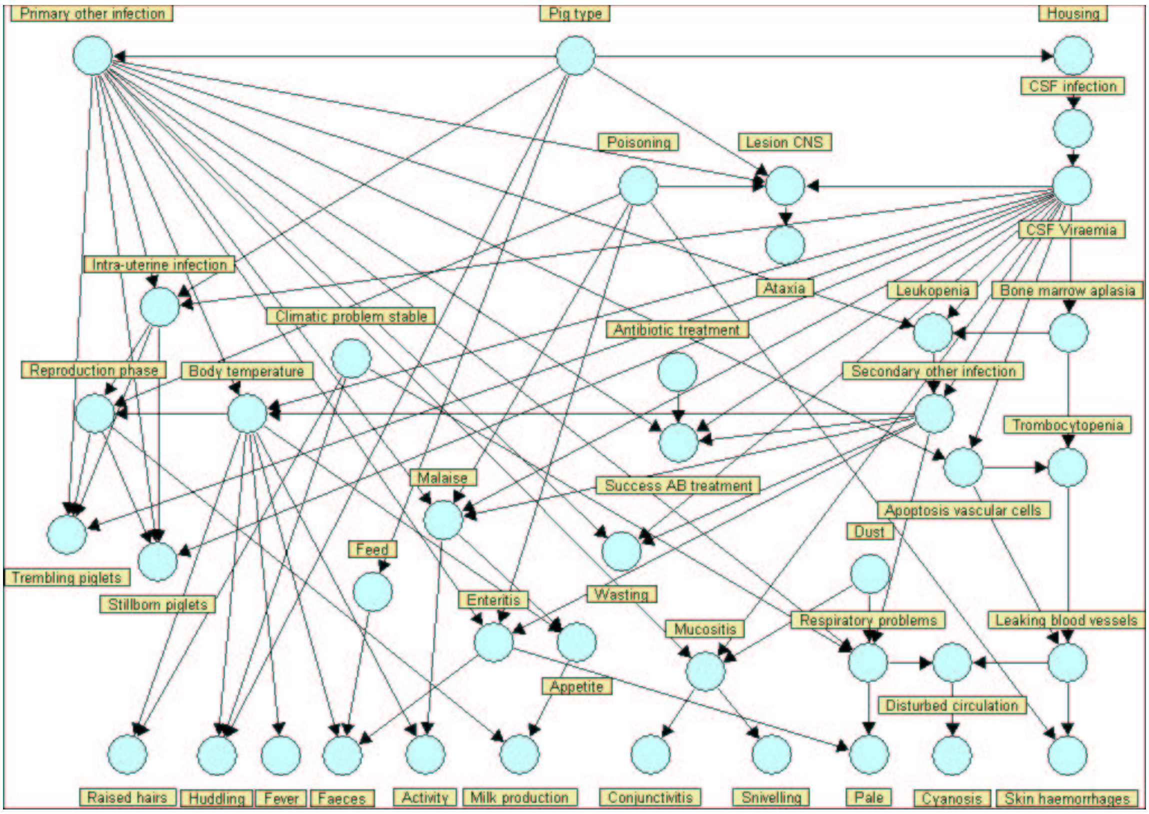

Figure 2: The graphical structure of our Bayesian network for classical swine fever in pigs.

the assignment lattice, we compared the conditional probabilities of a viraemia of classical swine fever. We found four violations of the properties of partial isotonicity given the selected background assignment; these violations are shown by dashed lines in Figure 3. The four violations all originated from adding the clinical sign of diarrhoea to the combination of findings of ataxia and malaise. The four violations thus were all covered by a single structurally minimal context of offence.

We presented the pairs of violating assignments to two veterinarians. Being confronted with the four violations, they independently and with conviction indicated that the probability of a viraemia of classical swine fever should increase upon finding the additional sign of diarrhoea. Both veterinarians mentioned that the combination of ataxia and diarrhoea especially pointed to classical swine fever; within the scope of the Bayesian network, they could not think of another disease that would be more likely to give rise to this combination of signs. They thus indicated that the network should indeed have been isotone in the five variables under study given the absence of any other signs. The four identified violations thus were indicative of modelling inadequacies in our current network.

5 Conclusions

In this paper, we have presented a method for studying monotonicity of Bayesian networks. In view of the unfavourable complexity of the problem in general, the method focuses on just a subset of the observable variables of a network and builds upon a lattice of all joint value assignments to these variables. The lattice is enhanced with information about the effects of these assignments on the probability distribution over the network's main output variable. The enhanced lattice then is used for identifying all violations of the properties of monotonicity and for constructing minimal offending contexts for further consideration.

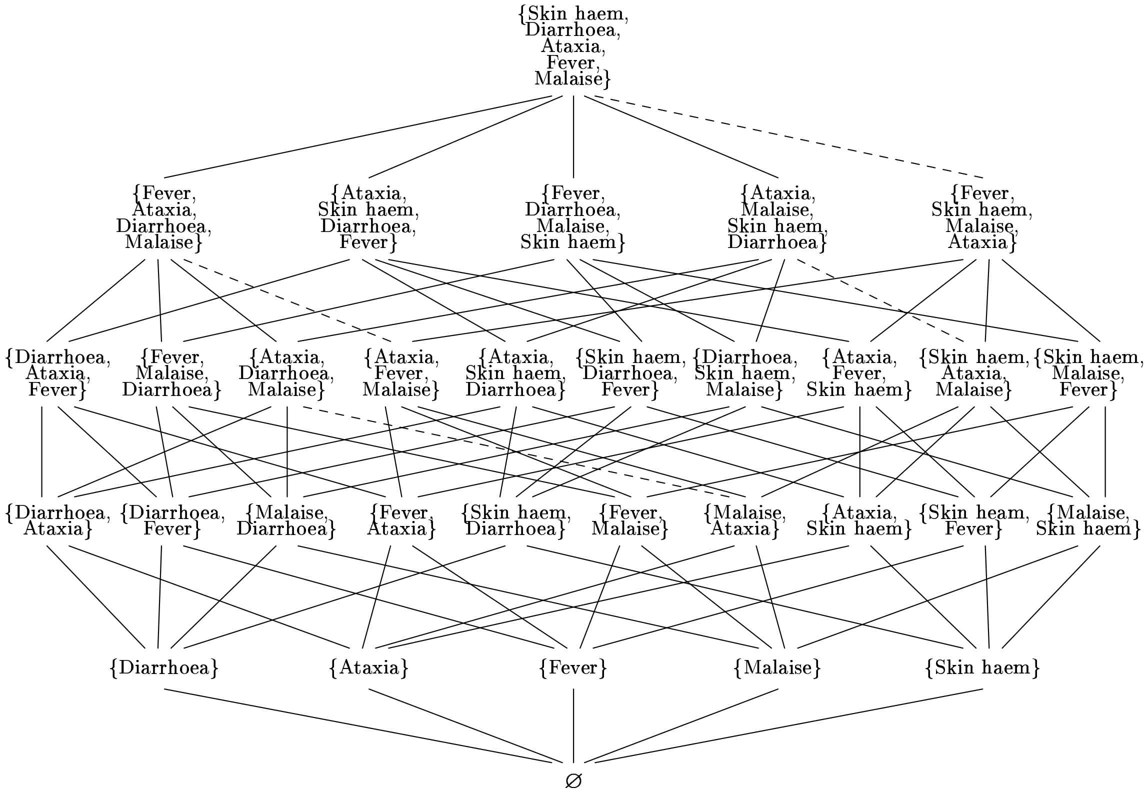

Figure 3: An assignment lattice for our Bayesian network for classical swine fever; the violations of the properties of partial monotonicity are indicated by dashed lines.

We would like to note that, as a supplement to our method for studying monotonicity, we designed a special-purpose elicitation technique that allows for discussing with domain experts whether or not the identified violations indeed can be construed as violations of commonly acknowledged patterns of monotonicity (Van der Gaag *et al.*, 2006). This technique has been designed specifically so as to ask little time and little cognitive effort from the experts in the verification of the identified violations.

The results that we obtained from applying our method for studying monotonicity to a real network in veterinary science, indicate that it presents a useful method for studying reasoning patterns in Bayesian networks. The next step now is to extend our method by techniques that exploit the constructed minimal offending contexts for identifying the modelling inadequacies in a network that cause the various violations of monotonicity.

References

J.O. Berger. 1980. *Statistical Decision Theory and Bayesian Analysis*, 2nd ed., Springer Verlag.

G. Grätzer. 1971. *Lattice Theory: First Concepts and Distributive Lattices*, W.H. Freeman.

C.-L. Liu and M.P. Wellman. 1998. Incremental tradeoff resolution in quantitative probabilistic networks. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, pages 338 – 345.

L.C. van der Gaag, H.L. Bodlaender and A. Feelders. 2004. Monotonicity in Bayesian networks. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, AUAI, pages 569 – 576.

L.C. van der Gaag, P.L. Geenen and H.J.M. Tabachneck-Schijf. 2006. Verifying monotonicity of Bayesian networks with domain experts. In *Proceedings of the 4th Bayesian Modelling Applications Workshop*, pages 9 – 15.

M.P. Wellman. 1990. Fundamental concepts of qualitative probabilistic networks. *Artificial Intelligence*, 44: 257 – 303.