Computational Statistics Quarterly 4,269-282(1990)

Bayesian Updating in Causal Probabilistic Networks by Local Computations

F. V. Jensen, S. L. Lauritzen and K. G. Olesen

Department of Mathematics and Computer Science, Institute for Electronic Systems, Aalborg University, Frederik Bajers Vej 7, DK-9220 Aalborg Ø, Denmark

Summary

An object-oriented version of the computational scheme in Lauritzen and Spiegelhalter (1988) is presented and proof of its correctness is given. The approach was motivated by the need for flexible programming and specification in the development of the HUGIN shell and the MUNIN expert system. Experience from this work is reported.

Key words: Belief network, causal probabilistic network, evidence, expert systems, influence diagram, junction tree, knowledge based systems, probability propagation, reasoning under uncertainty, recursive graphical model.

1 Introduction

In recent years the focus of expert system development has diverged from rule-based systems to systems based on a representation of so-called deep knowledge of the domain in question.

For domains with inherent uncertainty statisticians have for decades been working with graphs where variables are represented as nodes and statistical dependencies are represented as edges (Wright 1921; Wold 1954; Darroch *et al.* 1980). When in particular the dependencies are causal relations the graph becomes directed, and the causal relations are represented as conditional probabilities. This knowledge representation scheme has been termed differently in the literature: Recursive graphical model (Wermuth and Lauritzen 1983), Bayes belief network (Pearl 1986), causal probabilistic network (Andreassen *et al.* 1987), causal network (Lauritzen and Spiegelhalter 1988), probabilistic causal network (Cooper 1984), influence diagram (Howard and Matheson 1981). We have in this paper chosen the term causal probabilistic network, or CPN for short.

CPNs have many virtues in connection with expert systems, mainly due to the transparency of the knowledge embedded in a CPN and their ability to unify almost all domain knowledge relevant for an expert system (Andersen *et al.* 1986; Jensen *et al.* 1987; Horvitz *et al.* 1988).

0723-712X/90/4/269-282 $2.50© Physica-Verlag, Heidelberg

However, the calculation of specific probabilities was for a long period intractable and therefore an obstacle for pursuing the virtues. In the early 80's a first breakthrough was established (Kim and Pearl 1983) giving efficient calculation methods when the CPN is a tree, and later methods were constructed taking care of arbitrary CPNs without directed cycles (Lauritzen and Spiegelhalter 1988; Andersen *et al.* 1987; Jensen 1988; Jensen *et al.* 1989).

Incidentally a similar development had taken place in genetics some years before (Cannings *et al.* 1978) apparently unnoticed by the AI-community until recently (Thomas 1988), see also Thompson (1986) and Spiegelhalter (1989).

In the MUNIN project (Andreassen *et al.* 1987, 1989; Olesen *et al.* 1989) CPNs are used for knowledge representation. In the first small prototype (Jensen *et al.* 1987) the methods of Kim and Pearl were used. However, the domain could not be represented as a causal probabilistic tree, and it was therefore decided to use the methods developed by Lauritzen and Spiegelhalter.

When the methods were to be implemented they were met by several requests. The main request was that an object-oriented style of specification (and programming) should be used as was done in the first prototype of MUNIN. This request together with mere performance considerations led to a series of simplifications and changes of the methods. These modifications led to a rather different conceptual framework for the entire approach, which is now used in MUNIN (Jensen *et al.* 1989, Olesen *et al.* 1989) and built into the expert system shell HUGIN (Andersen *et al.* 1989). A development with many similarities to our approach has been made by Shafer and Shenoy (1988, 1989), see also Perez and Jiroušek (1985) and Dempster and Almond (1988).

In this paper we give a self-contained presentation of this approach together with proofs of the correctness of the methods. Section 2 describes the framework in terms of notation and basic definitions. Through Sections 3 and 4 the methods are described in detail and proofs of correctness are given. Section 5 outlines extensions to general state spaces, and in Section 6 we briefly describe the expert system shell HUGIN and the application MUNIN.

2 Definitions and Notation

A causal probabilistic network (CPN) is constructed over a *universe* U, consisting of a set of *nodes* each node having a finite set of *states*. The universe is organized as a *directed acyclic graph*, i.e. the graph has no directed cycles. The set of *parents* of A is denoted by $\mathrm{pa}(A)$ and $\mathrm{fm}(A)$ denotes the *family* $\mathrm{pa}(A) \cup \{A\}$. To each node $A \in U$ is attached a conditional probability table for $P(A \mid \mathrm{pa}(A))$. Note that if $\mathrm{pa}(A) = \emptyset$ the table reduces to unconditional probabilities.

Let $V \subseteq U$. The *space* of V is the Cartesian product of the state sets of the nodes of V and is denoted by $\mathrm{Sp}(V)$. For later notational convenience we think of the probability tables as functions, and denote them by the greek letters ϕ and ψ. If $A \in U$ then $\phi_A = P(A \mid \mathrm{pa}(A))$ maps $\mathrm{Sp}(\mathrm{fm}(A))$ into the unit interval $[0, 1]$. Later it becomes convenient to consider functions that are not normalized and therefore take on arbitrary non-negative values. So, in the sequel, ϕ and ψ denote such functions. Exploiting these convention we define a number of basic operations.

2.1 Basic operations

2.1.1 Extension

Let $V \subseteq W \subseteq U$ and $\phi : \mathrm{Sp}(V) \rightarrow \mathbf{R}_0$. The function ϕ is extended to W in the following way: Let $x \in \mathrm{Sp}(W)$. If y is the projection of x on $\mathrm{Sp}(V)$ then we let $\phi(x) = \phi(y)$.

When it is clear from the context we will not distinguish between a function and its various extensions.

2.1.2 Restriction

Let $W \subseteq V \subseteq U$ and $\phi : \mathrm{Sp}(V) \to \mathbf{R}_0$ and let $z \in \mathrm{Sp}(W)$. The restriction ϕ^z of ϕ to $\mathrm{Sp}(V \setminus W)$ is defined as

$$\phi^z(x) = \phi(z.x)$$

where $z.x$ is the element in $\mathrm{Sp}(V)$ with projections z to $\mathrm{Sp}(W)$ and x to $\mathrm{Sp}(V \setminus W)$.

2.1.3 Multiplication

Multiplication is defined the obvious way, i.e. for ϕ and ψ defined on $\mathrm{Sp}(V)$ and $\mathrm{Sp}(W)$ respectively we define $\phi * \psi$ on $\mathrm{Sp}(V \cup W)$ as

$$(\phi * \psi)(x) = \phi(x)\psi(x)$$

where ϕ and ψ on the right-hand side first have been extended to the relevant space, cf. the remark at the end of 2.1.1.

2.1.4 Addition

Addition is defined analogously.

2.1.5 Marginalization

Let $W \subseteq V \subseteq U$ and $\phi : \mathrm{Sp}(V) \to \mathbf{R}_0$. The expression $\sum_{V \setminus W} \phi$ denotes the marginalization of ϕ to W and is defined as

$$\sum_{V \setminus W} \phi = \sum_{z \in \mathrm{Sp}(V \setminus W)} \phi^z.$$

2.1.6 Division

This is likewise defined in the obvious way, just that special care has to be taken when dividing by zero

$$(\phi/\psi)(x) = \begin{cases} 0 & \text{if } \phi(x) = 0 \\ (\phi(x)/\psi(x)) & \text{if } \psi(x) \neq 0 \\ \text{undefined} & \text{otherwise} \end{cases}$$

With these definitions we are now able to give a precise meaning to a causal probabilistic network.

2.2 The joint probability function

Let U be the universe of nodes for a CPN. We define the (a priori) joint probability function ϕ_U as the product

$$\phi_U = \prod_{A \in U} \phi_A.$$

This definition makes ϕ_U satisfy the causal (or directed) Markov property (Kiiveri *et al.* 1984; Lauritzen *et al.* 1989) thus conforming to the 'causal' interpretation of the network, see also Pearl (1988).

2.3 Evidence

CPNs are in expert systems used dynamically. Initially the CPN holds a set of causal relations and prior probabilities, but when information on the universe is achieved, it is fed into the CPN yielding new (posterior) probabilities.

Let V be a set of nodes. By *evidence* on V we mean a function

$$E_V : \mathrm{Sp}(V) \rightarrow \mathbf{R}_0.$$

Thus evidence is represented as a likelihood function, giving relative weights to elements in the space of V.

In the particular case, where E_V is a function into $\{0, 1\}$ it represents a statement that some elements of $\mathrm{Sp}(V)$ are impossible. In that case we call E_V a *finding*. Typically a finding is a statement that a certain node is in one particular state.

If the prior joint probability function for the CPN is ϕ then the posterior joint probability function is defined to be $\mu(\phi * E_V)$ where μ is a normalizing constant.

Note: If E_V is a finding then the normalizing constant μ is the reciprocal of the prior probability of E_V and the posterior joint probability function is the conditional joint probability, given the findings.

2.4 The calculation problem

Given a CPN with universe U, a set of (pieces of) evidence, and let $A \in U$. What is the probability distribution for A given the evidence?

In principle it is possible to calculate ϕ_U, multiply it with the evidence functions, and then to marginalize the resulting function to A. However, this calculation is linear in the cardinality of $\mathrm{Sp}(U)$ and in practice intractable even for small universes. We therefore have to exploit the local structure of the network, which is the theme of the next sections.

3 Trees of Belief Universes

The aim of the implementation of efficient methods for solving the calculation problem is to have a set of objects which can send messages to each other and can perform actions as results of received messages.

In order to avoid a global control structure it is convenient to have the objects organized in a tree such that messages only can be passed between neighbours in the tree. Then a global operation can be started in any object and successive message passing to neighbouring objects will spread the operation to the entire tree and stop by itself when this is done. Another effect would be that the objects can perform their tasks in parallel and thus exploit computer architectures with several processors.

3.1 Basic notions

A *tree of belief universes* consists of a collection C of sets of nodes organized in a tree. The union of all the sets in the collection is called the *total universe* and denoted U. We only consider trees with finite total universe U.

The intersections of neighbours in the tree are called *separators*. The collection of separators is called $\mathcal{S}$. Both the universes and the separators have *belief potentials* ϕ_W attached to them, where ϕ_W maps $\mathrm{Sp}(W)$ to $\mathbf{R}_0$. The *joint system belief* ϕ_U is defined as a function on $\mathrm{Sp}(U)$ given by

$$\phi_U = \frac{\prod_{V \in \mathcal{C}} \phi_V}{\prod_{S \in \mathcal{S}} \phi_S}.$$

A belief potential ϕ_W is said to be *normalized* if $\sum_W \phi_W = 1$. A *normalized tree* of belief universes is one where all belief potentials are normalized.

Note: If the tree is normalized, then so is the joint system belief ϕ_U.

3.2 Supportive trees

Let $\phi_V : \mathrm{Sp}(V) \to \mathbf{R}_0$. The *support* of ϕ_V is

$$\mathrm{Spt}(\phi_V) = \{x \in \mathrm{Sp}(V) \mid \phi_V(x) \neq 0\}.$$

When no confusion is possible we write $\mathrm{Spt}(V)$ instead of $\mathrm{Spt}(\phi_V)$. If V and W are sets of nodes with belief potentials ϕ_V and ϕ_W then

$$\mathrm{Spt}(V) \subseteq \mathrm{Spt}(W)$$

denotes the statement that if ϕ_V and ϕ_W are extended to $V \cup W$ then the support of ϕ_V is included in the support of ϕ_W.

A tree of belief universes is said to be *supportive* if for any V and for any neighbouring separator S to V we have $\mathrm{Spt}(V) \subseteq \mathrm{Spt}(S)$.

3.3 Construction

With these definitions it is easy to construct tree structures representing the same joint probability function as CPNs. Let N be a CPN with universe U and probability functions

$$\phi_A : \mathrm{Sp}(\mathrm{fm}(A)) \to \mathbf{R}_0, \quad A \in U.$$

Let further T be a tree of belief universes with collection $\mathcal{C}$ and separators $\mathcal{S}$ constructed such that

i) to each $A \in U$ we assign a $V \in \mathcal{C}$ with $\mathrm{fm}(A) \subseteq V$;

ii) for $V \in \mathcal{C}$ let $A, \ldots, B$ be the nodes to which V is assigned. Let

$$\phi_V = \phi_A * \cdots * \phi_B;$$

iii) for all $S \in \mathcal{S}$ let ϕ_S be any constant positive function (this makes T supportive).

Note: To each $A \in U$ there might be several $V \in \mathcal{C}$ such that $\mathrm{fm}(A) \subseteq V$ but we only assign one of them to A. The construction is illustrated in Figure 1.

From the definitions it is easy to see that the joint system belief for T is proportional to the joint probability function for N and that the quotient between them is the product of the values for the belief potentials of the separators.

Now the basic structure is established. The belief universes are the objects and the separators are the communication channels. Next we define the basic operations for the objects.

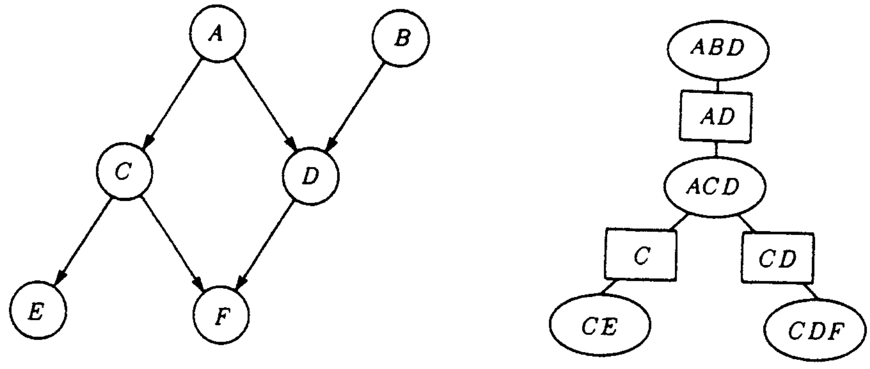

Figure 1: Construction of the tree of belief universes. The universe ABD is assigned to B and D, ACD is assigned to A and C, CE to E and CDF to F. Other assignments were possible.

3.4 Absorption

Let T be a tree of belief universes with collection $\mathcal{C}$ and separators $\mathcal{S}$. Let $V \in \mathcal{C}$ and let $W_1, \ldots, W_m$ be neighbours of V with separators $S_1, \ldots, S_m$ respectively. Suppose that $\text{Spt}(V) \subseteq \text{Spt}(S_i)$, $i = 1, \ldots, m$. The universe V is said to ***absorb*** from $W_1, \ldots, W_m$ if the belief potentials ϕ_{S_i} and ϕ_V are changed to ϕ'_{S_i} and ϕ'_V, where

$$\begin{aligned} \phi'_{S_i} &= \sum_{W_i \setminus V} \phi_{W_i}, \quad i = 1, \ldots, m \\ \phi'_V &= \phi_V * (\phi'_{S_1}/\phi_{S_1}) * \cdots * (\phi'_{S_m}/\phi_{S_m}). \end{aligned}$$

Note the following important points:

a) After an absorption the belief potential for S_i is the marginal of W_i with respect to S_i.

b) We have that

$$\text{Spt}(\phi_{W_i}) \subseteq \text{Spt}\left(\sum_{W_i \setminus V} \phi_{W_i}\right) = \text{Spt}\left(\phi'_{S_i}\right).$$

Hence the supportiveness of a tree T is invariant under absorption.

c) We also have that

$$\phi_V/(\phi_{S_1} * \cdots * \phi_{S_m}) = \phi'_V/(\phi'_{S_1} * \cdots * \phi'_{S_m})$$

and hence the joint system belief is invariant under absorption.

d) If $m = 1$ and $\phi_S \propto \sum_{V \setminus W} \phi_V$ - where $\propto$ means 'proportional to' - then

$$\sum_{V \setminus W} \phi'_V \propto \phi'_S = \sum_{W \setminus V} \phi_W.$$

In that case we say that V has *calibrated* to W. The reason is, that V has been forced to take over the information that W holds on their common nodes.

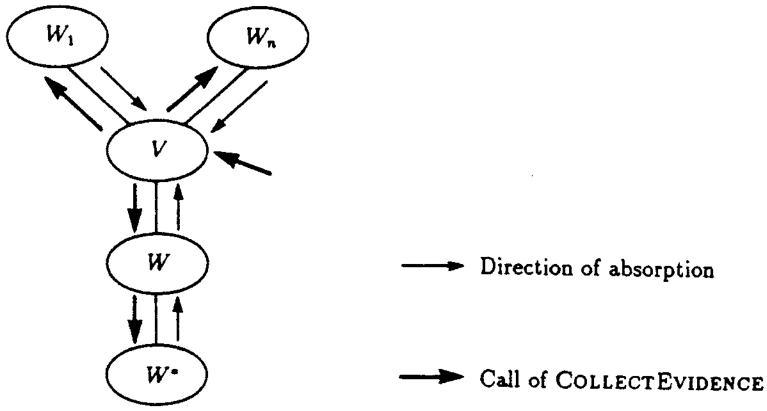

Figure 2: The calls and message passing in COLLECTEVIDENCE.

e) In the actual implementation, the separators play a slightly more active role: first they receive the marginal from W_i, then they send the fraction ϕ'_S/ϕ_S to V and finally they update their own belief potential.

3.5 Entering evidence

Let T be a tree of belief universes and $V \subseteq U$. Let E_V be an evidence function. This can be entered to T if there is a $W \in C$ such that $V \subseteq W$. This is done simply by multiplying ϕ_W by E_V.

Note: If T is constructed from a causal probabilistic network N as in Section 3.3 then the posterior joint system belief for T is proportional to the posterior joint probability function for N.

If there is no W with $V \subseteq W$, then the entering of evidence is more subtle. We will not deal with that problem in this paper.

3.6 Collecting evidence

Based on the local operation of absorption we can now construct the propagation operations.

Each $V \in C$ is given the action COLLECTEVIDENCE: When COLLECTEVIDENCE in V is called from a neighbour W then V calls COLLECTEVIDENCE in all its other neighbours and when they have finished their COLLECTEVIDENCE, V absorbs from them (see Figure 2).

Note the following points:

a) Since COLLECTEVIDENCE is composed of absorptions only, supportiveness and the joint system belief is invariant under COLLECTEVIDENCE (see notes b) and c) in Section 3.4).

b) Suppose COLLECTEVIDENCE is envoked in V from the outside. Let W and W^* be neighbours with separator S such that W is closer in the tree to V than W^* is and such that they are not on the branch from which COLLECTEVIDENCE is envoked, see Figure 2. Then the COLLECTEVIDENCE from V will cause W to absorb from W^*. From note a) in Section 3.4 we have that after COLLECTEVIDENCE from V, the belief potential for S is the marginal of W^* with respect to S.

3.7 Distributing evidence

Each $V \in C$ is given the action DISTRIBUTEEVIDENCE: When DISTRIBUTEEVIDENCE is called in V from a neighbour W then V absorbs from W and calls DISTRIBUTEEVIDENCE in all its other neighbours.

Note: The joint system belief and supportiveness is invariant under DISTRIBUTEEVIDENCE .

3.8 Local consistency

A tree of belief universes is said to be *locally consistent* if whenever V and W are neighbours with separator S then

$$\sum_{V \setminus S} \phi_V \propto \phi_S \propto \sum_{W \setminus S} \phi_W.$$

An important prerequisite for the methods is the following

Theorem 1 *Let V be any belief universe in a supportive tree of belief universes. If first* COLLECTEVIDENCE *is evoked from V and then* DISTRIBUTEEVIDENCE *is evoked from V, the resulting tree of belief universes will be locally consistent .*

Proof From note b) of Section 3.6 we have that after COLLECTEVIDENCE from V each of the belief potentials on the separators is the marginalization of the belief potential in the neighbour most distant from V. When DISTRIBUTEEVIDENCE is evoked from V, a series of calibrations will follow, see note d) of Section 3.4, and local consistency results. □

4 Junction Trees

4.1 Junction trees and consistency

What we aim for is a tree of belief universes such that the probability distributions can be directly inferred from the belief potentials without having to calculate the joint system belief. That is: If $W \subseteq V$ then $\sum_{V \setminus W} \phi_V$ is proportional to the probability distribution for W. In order to ensure this, local consistency is not sufficient. For example, if $W \subseteq V_1$ and $W \subseteq V_2$, then local consistency does not automatically ensure that

$$\sum_{V_1 \setminus W} \phi_{V_1} \propto \sum_{V_2 \setminus W} \phi_{V_2}$$

unless V_1 and V_2 are neighbours in the tree. We therefore define a tree of belief universes to be (globally) *consistent* if for each $V, W \in C$

$$\sum_{V \setminus W} \phi_V \propto \sum_{W \setminus V} \phi_W$$

that is, ϕ_V and ϕ_W coincide on $V \cap W$. Clearly, a consistent tree will always be locally consistent but the converse is false in general.

Call a tree of belief universes a *junction tree* if for any $V, W \in C$ and for any separator S on the path between V and W we have $V \cap W \subseteq S$. The important fact is that the junction tree property ensures the converse to hold. More precisely we have

Proposition 1 *A locally consistent junction tree is consistent.*

Proof Let V and W be universes in a locally consistent junction tree. Using the local consistency stepwise on the path from V to W we get the result. □

This adds a constraint on the transformation of a CPN to a tree of belief universes: The tree must be a junction tree. The main concept for establishing an appropriate junction tree is triangulation (Jensen 1988), a topic on which a wide range of literature exists (Berge 1973; Rose *et al.* 1976; Golumbic 1980; Yannakakis 1981; Lauritzen *et al.* 1984). A brief description of this process is given in Section 6.

The key result of the method now follows:

Theorem 2 *Let T be a consistent junction tree of belief universes with collection C. Let ϕ_U be the joint system belief for T and let $V \in C$. Then*

$$\sum_{U \setminus V} \phi_U \propto \phi_V. \tag{1}$$

Proof By induction on n, the number of sets $|C|$ in C. For $n = 1$ the statement is clearly true as $U = V$. Assume then that for all consistent junction trees of belief universes with $|C| \leq n$, we have that (1) holds for all $V \in C$.

Consider a consistent junction tree of belief universes where $|C| = n+1$. Let $V \in C$ be arbitrary and let $W \neq V$ be a leaf in T with separator S. Since T is a junction tree, only W contains nodes in $W \setminus S$. Hence

$$\sum_{W \setminus S} \phi_U = \sum_{W \setminus S} \frac{\prod_{V' \in C} \phi_{V'}}{\prod_{R \in S} \phi_R} = \frac{\prod_{V' \in C \setminus \{W\}} \phi_{V'}}{\prod_{R \in S \setminus \{S\}} \phi_R} * \frac{\sum_{W \setminus S} \phi_W}{\phi_S}.$$

By local consistency the latter factor is constant so we further obtain

$$\sum_{W \setminus S} \phi_U \propto \frac{\prod_{V' \in C \setminus \{W\}} \phi_{V'}}{\prod_{R \in S \setminus \{S\}} \phi_R} = \phi_{U'} \tag{2}$$

where $\phi_{U'}$ is the joint system belief for the consistent junction tree T' on the total universe $U' = U \setminus (W \setminus S)$ with collection $C \setminus \{W\}$. Hence, by the induction hypothesis we can further marginalize $\phi_{U'}$ over $U' \setminus V$ to obtain ϕ_V and the result follows. If T is normalized then $\sum_{W \setminus S} \phi_W = \phi_S$ and equality holds in (2). □

At this point we have overcome the calculation problem stated in Section 2.4. The theorem shows that it is not necessary to calculate and marginalize ϕ_U in order to find the belief in a particular node. When we have a consistent junction tree the joint probability function is for each universe V proportional to the belief potential for V. We can now find the belief in any node $A \in V$ by marginalizing ϕ_V to A and then normalizing the result.

When evidence arrives to the CPN, it is entered to the junction tree as described in Section 3.5. Then the junction tree is made consistent by means of the operations CollectEvidence and DistributeEvidence and posterior probabilities can be found by the above procedure. If the original tree was normalized the prior probability of the evidence entered can be obtained by taking any belief universe and take the sum of values of its belief potential after propagation (see the note in Section 3.5).

4.2 Uniform junction trees

If we in the previous sections everywhere substitute '=' for '∝', most results still hold true. In particular, suppose V and W are neighbours with $S = V \cap W$, and $\phi_S = \sum_{V \setminus W} \phi_V$. If V calibrates to W then

$$\sum_{V \setminus W} \phi'_V = \phi'_S = \sum_{W \setminus V} \phi_W. \tag{3}$$

Junction trees with the property (3) holding for all pairs of neighbours are of particular interest. They are called *uniform* junction trees. By arguments completely analogous to those in the proof of Theorem 2, we get

Proposition 2 *If T is a uniform junction tree over the universe U, then for all belief potentials ϕ_X we have*

$$\phi_X = \sum_{U \setminus X} \phi_U.$$

In particular, $\sum_X \phi_X = \sum_U \phi_U$.

As the next proposition shows, uniform junction trees are easy to construct:

Proposition 3 *Whenever a junction tree is made consistent through by envoking* COLLECTEVIDENCE *followed by* DISTRIBUTEEVIDENCE *in the same belief universe, then the resulting tree is uniform.*

Proof As in Theorem 1. □

Note: If COLLECTEVIDENCE is called in a universe V, then the succeeding DISTRIBUTEEVIDENCE does not change ϕ_V. Therefore, after the call of COLLECTEVIDENCE, we already have that $\phi_V = \sum_{U \setminus V} \phi_U$.

An important consequence is the following result, also noted by Lauritzen and Spiegelhalter (1988) in their reply to the discussion.

Theorem 3 *Let T be a normalized junction tree and let $F_1, \ldots, F_n$ be a set of findings with prior probability $P(F_1, \ldots, F_n)$. Let $F_1, \ldots, F_n$ be entered to T and call* COLLECTEVIDENCE *in any belief universe V. Let ϕ^*_V be the resulting belief table for V. Then*

$$\sum_V \phi^*_V = P(F_1, \ldots, F_n).$$

Proof Let ϕ_U be the prior joint system belief. Then the posterior joint system belief is given by $\phi_U * F_1 * \cdots * F_n$. Since T is normalized, ϕ_U is the joint probability function and therefore

$$P(F_1, \ldots, F_n) = \sum_U (\phi_U * F_1 * \cdots * F_n).$$

The result now follows from Proposition 3. □

Theorem 3 can be used in a variety of ways. Most directly it gives an easy way of calculating the prior joint probabilities for *sets* of findings. But if findings are hypothetical, it also gives a way of calculating probabilities for complex hypotheses.

5 Extension to General State Spaces

For practical computations it is of interest to let some nodes in a universe correspond to real-valued variables. Most of the developments carry through with only small changes when we let each node A in the universe U carry a more general state space $\mathcal{X}_A$, say a complete, metrisable and separable space (Polish space, see e.g. Billingsley (1968)) or even more general than this.

Additionally we need to have a fixed σ-finite measure μ_A on $\mathcal{X}_A$ for each node $A \in U$ and we denote its products as $\mu_V = \otimes_{A \in V} \mu_A$ for $V \subseteq U$. The conditional probability tables ϕ_A shall then for fixed parent configuration be (equivalence classes of) probability densities with respect to μ_A.

Marginalization of functions is given as:

$$\sum_{V \setminus W} \phi = \int \phi^z \mu_{V \setminus W}(dz)$$

and requires integrability to be well-defined. The other operations are completely analogous to the case of finite state spaces.

A supportive tree of belief universes is one, where the measures $\phi_V \mu_V$ are absolutely continuous with respect to $\phi_S \mu_V$ for all separators $S \subseteq V$.

The joint system belief has to be assumed integrable which automatically will be the case after the initial construction as in Section 3.3. Only evidence functions preserving integrability are allowed to be entered. Note that all finding functions will be permissible since they are indicators of Borel sets and therefore bounded. Fubini's theorem ensures the marginalization integrals to be well defined under absorption and thus allows the general scheme to proceed in the usual fashion.

6 Experience from HUGIN and MUNIN

The methods described in this paper formed the basis for a shell for building expert systems, called HUGIN[1] (Andersen *et al.* 1989) which in turn is used for the MUNIN[2] expert system.

In HUGIN the user specifies a model of a domain as a causal probabilistic network. When the model is completed, a junction tree T is automatically constructed. This involves a moralization of the graph: for each node, links are added between all of its parents (if they are not connected already) and directions are removed.

The moral graph is then triangulated: Links are added until every cycle of length more than three has a chord. Based on this triangulation the junction tree is formed: The collection $\mathcal{C}$ is the set of cliques in the triangulated graph (a clique is a maximal set of nodes all of which are pairwise linked). Given the cliques the junction tree can be found through a maximal spanning tree algorithm (Jensen 1988).

The size of the cliques determines the runtime behaviour of the system, so the triangulation is the single most important step in the transformation. In order to optimize this, heuristic methods have been developed to obtain a small total clique size (Kjærulff 1989).

The moralization of the graph ensures that for each node A a set $V \in \mathcal{C}$ exists such that $\text{fm}(A) \subseteq V$. Hence the construction in Section 3.3 can be used and the joint system belief for T will be proportional to the joint probability function for the CPN.

[1] Handling Uncertainty in General Influence Networks
[2] MUscle and Nerve Inference Network

HUGIN will now maintain the joint belief for the CPN in the junction tree. Evidence can be entered and whenever requested the tree is made consistent.

In the Esprit project P599: "A Knowledge Based Assistant for Electromyography" (Andreassen *et al.* 1987, 1989; Olesen *et al.* 1989) a preliminary version of HUGIN has been used to construct large nets in order to diagnose muscle and nerve diseases. These nets constitute the MUNIN expert system.

In MUNIN we are dealing with nets of up to about one thousand nodes, having from 2 to 21 states. The largest net covers a total sample space of about 10^{600} states and updating in this net is done in a few minutes. Further, an approximation with an absolute error less than 0.1% has been made. This has reduced response times to about 10-20 seconds (Jensen and Andersen 1989). When handling networks on this scale new problems mainly concerning space arise. This has lead to work on optimal triangulation and other modifications of the methods as mentioned above. Current work indicates that even larger nets can be handled, a promising line of research being the extension of the methodology to include continuous variables in the network.

7 Conclusion

In the previous sections we have shown how an object-oriented specification of the evidence calculus allows a simple and self-contained description of the procedures involved.

This has enhanced efficient and flexible programming and conceptual clarification in the theoretical developments as well, thus preparing the road for future extensions of the methodology.

Experience with the shell HUGIN and the medical system MUNIN shows that large systems can be built in a short time, resulting in systems with reasonably small response times.

References

Andersen, S.K., Andreassen, S. and Woldbye, M. (1986). Knowledge representations for diagnosis and test planning in the domain of electromyography. In *Proceedings of the 7th European Conference on Artificial Intelligence*.

Andersen, S.K., Jensen, F.V. and Olesen, K.G. (1987). The HUGIN core - preliminary considerations on systems for fast manipulations of probabilities. In *Proceedings of Workshop on Inductive Reasoning: Managing Empirical Information in AI-Systems.* Risø, Denmark.

Andersen, S.K., Olesen, K.G., Jensen, F.V. and Jensen. F. (1989). HUGIN - a shell for building belief universes for expert systems. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*.

Andreassen, S., Jensen, F.V., Andersen, S.K., Falck, B., Kjærulff, U., Woldbye, M., Sørensen, A.R., Rosenfalck, A., Jensen, F. (1989). MUNIN - An EMG assistant. In *Computer-Aided Surface or Needle Electromyography and Expert Systems in Diagnosis.* J.E. Desmedt (editor), Elsevier Science Publishers, Amsterdam.

Andreassen, S., Woldbye, M., Falck, B. and Andersen S.K. (1987). MUNIN - A causal probabilistic network for interpretation of electromyographic findings. In *Proceedings of the 10th International Joint Conference on Artificial Intelligence*. 366-372.

Berge, C. (1973). *Graphs and Hypergraphs.* Transl. from French by E. Minieka. North-Holland, Amsterdam.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Cannings, C., Thompson, E.A. and Skolnick, M.H. (1978). Probability functions on complex pedigrees. *Adv. Appl. Probabil.* **10**, 26-61.

Cooper, G.F. (1984). NESTOR: A computer based medical diagnostic aid that integrates causal and probabilistic knowledge, Technical Report HHP-84-48, Medical Computer Science Group, Stanford University, Stanford, California.

Darroch, J.N., Lauritzen, S.L. and Speed, T.P. (1980). Markov-fields and log-linear models for contingency tables. *Ann. Statist.* **8**, 522-539.

Dempster, A.P. and Almond, R.G. (1988). In discussion of Lauritzen and Spiegelhalter. *J. Roy. Statist. Soc. B* **50**, 205.

Golumbic, M.C. (1980). *Algorithmic Graph Theory and Perfect Graphs.* Academic Press, London.

Horvitz, E.J., Breese, J.S. and Henrion, M. (1988). Decision theory in expert systems and artificial intelligence. *Journal of Approximate Reasoning.* Special issue: *Uncertain Reasoning.*

Howard, R.A. and Matheson, J.E. (1981). Influence diagrams. In *Readings in Decision Analysis.* Howard, R.A. and Matheson, J.E. (eds.), Strategic Decisions Group, Menlo Park, Ch. 38, 763 - 771.

Jensen, F. and Andersen, S.K. (1989). Approximations in Bayesian belief universes for knowledge based systems. Manuscript in preparation.

Jensen, F.V. (1988). Junction trees and decomposable hypergraphs. Judex Research Report, Aalborg, Denmark.

Jensen, F.V., Andersen, S.K., Kjærulff, U., and Andreassen, S. (1987). MUNIN: on the case for probabilities in medical expert systems - a practical exercise. In *Proceedings of 1st Conference of European Society for Artificial Intelligence in Medicine* (Fox, J., Fieschi, M., Engelbrecht, R. eds.). Springer-Verlag, Heidelberg, 149-160.

Jensen, F.V., Olesen, K.G. and Andersen, S.K. (1989). An algebra of Bayesian belief universes for knowledge based systems. To appear in *Networks.* Special issue: *Influence Diagrams.*

Kiiveri, H., Speed, T.P. and Carlin, J.B. (1984). Recursive causal models. *J. Austral. Math. Soc. A* **36**, 30-52.

Kim, J.H. and Pearl, J. (1983). A computational model for causal and diagnostic reasoning in inference systems. In *Proceedings of the 8th International Joint Conference on Artificial Intelligence,* 190-193.

Kjærulff, U. (1989). Triangulation of graphs - algorithms giving small total clique size. In preparation.

Lauritzen, S. L., Speed, T. P. and Vijayan, K. (1984). Decomposable graphs and hypergraphs. *J. Austral. Math. Soc. A* **36**, 12-29.

Lauritzen, S.L. and Spiegelhalter, D.J. (1988). Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *J. Roy. Statist. Soc. B* **50**, 157-224.

Lauritzen, S.L., Dawid, A.P., Larsen, B.N. and Leimer, H.-G. (1989). Conditional independence properties of directed Markov fields. To appear in *Networks*. Special issue: *Influence Diagrams*.

Olesen, K.G., Kjærulff, U., Jensen, F., Jensen, F.V., Falck, B., Andreassen, S. and Andersen, S.K. (1989). A MUNIN network for the median nerve - a case study on loops. *Applied Artificial Intelligence* **3**, 385-404. Special issue: *Towards Causal AI Models in Practice*.

Pearl, J. (1986). Fusion, propagation and structuring in belief networks. *Artificial Intelligence* **29**, 241-288.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Mateo.

Perez, A. and Jiroušek, R. (1985). Constructing an intensional expert system INES. In: *Medical Decision Making; Diagnostic Strategies and Expert Systems*. van Bemel, J.H., Grémy, F. and Zvárová, J., eds. North-Holland, Amsterdam, 307-315.

Rose, D.J., Tarjan, R.E. and Lueker, G.S. (1976). Algorithmic aspects of vertex elimination on graphs. *SIAM J. Comput.* **5**, 266-283.

Shafer, G. and Shenoy P.P. (1988). Local computation in hypertrees. School of Business Working Paper No. 201, University of Kansas.

Shafer, G. and Shenoy P.P. (1989). Probability propagation. To appear in *Annals of Mathematics and Artifical Intelligence*.

Spiegelhalter, D.J. (1989). Fast algorithms for probabilistic reasoning in influence diagrams with application in genetics and expert systems. In *Influence Diagrams, Belief Nets and Decision Analysis*. Oliver, R. and Smith, J. Q., eds. Wiley, Chichester.

Thomas, A. (1988). In discussion of Lauritzen and Spiegelhalter. *J. Roy. Statist. Soc. B* **50**, 215.

Thompson, E.A. (1986). *Pedigree Analysis in Human Genetics*. Johns Hopkins University Press, Baltimore.

Wermuth, N. and Lauritzen, S. L. (1983). Graphical and recursive models for contingency tables. *Biometrika* **70**, 537-552.

Wold, H. (1954). Causality and econometrics. *Econometrica* **22**, 162-177.

Wright, S. (1921). Correlation and causation. *J. Agricultural Research* **20**, 557-585.

Yannakakis, M. (1981). Computing the minimum fill-in is NP-complete. *SIAM J. Algebraic Discrete Methods* **2**, 77-79.