# QUALITATIVE AND QUANTITATIVE REFINEMENT OF PARTIALLY SPECIFIED BELIEF NETWORKS BY MEANS OF STATISTICAL DATA FUSION

**BY**

**EUGENE GROIS**

**B.S., University of Illinois at Urbana-Champaign, 1995**

**THESIS**

**Submitted in partial fulfillment of the requirements for the degree of Master of Science in Computer Science in the Graduate College of the University of Illinois at Urbana-Champaign, 1998**

**Urbana, Illinois**

# ACKNOWLEDGEMENTS

I would, first and foremost, like to thank my research advisor, Prof. David Wilkins, for providing me with the computing facilities, technical support, and funds to carry out this research. Without his sponsorship, this work would not have been possible.

Next, I would like to thank the members of the Knowledge-based Systems Group at the University of Illinois for their help on many aspects of this project ranging from computer system administration support, to technical contributions. In particular, I would like to thank Bill Hsu and Ole Mengshoel for introducing me to the wonderful and exciting world of belief networks, and for contributing a great deal to my understanding of the field. Also, appreciation goes to Mikhail Voloshin, who implemented a portion of the experimental code, thereby allowing me to spend more time on theoretical aspects of the project.

Finally, I would like to thank the graduate students of the Beckman Institute AI Group, whose contributions through enlightening and stimulating discussions have been invaluable to formulating my view of Artificial Intelligence, both on a technical and philosophical levels. Most especially, I would like to express by deep appreciation to Ricardo Vilalta, who volunteered many hours from his busy schedule to help me understand the concepts and theory of Machine Learning.

# TABLE OF CONTENTS

# 1. INTRODUCTION

In the 1970s systems were introduced that could make reasonably intelligent decisions, and produce satisfactory diagnoses in uncertain domains. These systems have become known as *expert systems*. Expert systems are domain-specific knowledge-based systems that can offer solutions and advice at the level comparable to that of a human expert in the same field. Example domains in which modern expert systems have been successful include medical dagnosis [Heckerman et al., 1990], system fault diagnosis [Palmer, 1998], battlefield reasoning [Mengshoel and Wilkins, 1997], and damage control [Bulitko and Wilkins, 1998]. These and other real-world domains typically require considerable human expertise. Furthermore, in dealing with real-world domains, human experts must form judgments and take decisions from uncertain, incomplete, and sometimes even contradictory information. In order to be effective in an environment that is characterized by such imprecision, an expert system must be capable of capturing and exploiting not only the highly specialized expert knowledge, but the uncertainty inherent in represented knowledge, as well. This need has given rise to a field of research known as *uncertainty reasoning* (see for example [Shafer and Pearl, 1990]).

Pioneering efforts aimed at developing intelligent problem-solving systems for real-world domains (primarily medical diagnosis) were characterized by the use of probability theory as the principal vehicle for decision making under uncertainty [Horvitz, 1988]. These early systems were devised for clear-cut problem domains with a small number of hypotheses and restricted evidence. As a result, most of the necessary probabilities were obtained from statistical analysis of empirical data. Despite the simplifying assumption underlying these systems, they performed considerably well. However, it quickly became evident that this approach would not scale well. For larger and more complex domains pure statistical methods become prohibitive either computationally, or from a probability assessment point of view. As a result, interest in probabilistic approaches to uncertainty reasoning declined.

The next generation of expert systems were mostly rule-based. They relied on production rules as a formalism for the representation of expert knowledge, the application of which was governed by heuristic reasoning methods. To be effective in real-world domains, the originally deterministic rule-based systems had to be extended with some notion of uncertainty. Two systems developed in the 1970s were especially influential in this regard. These were the MYCIN system for assisting physicians in the diagnosis and treatment of bacterial infections [Buchanan and Shorliffe, 1984], and the PROSPECTOR system for aiding non-expert geologists in the identification of mineral deposits [Duda et al., 1979]. The former relied on the *certainty factor model* [Shortliffe and Buchanan, 1984], while the latter used the *subjective Bayesian method* [Duda et al., 1976]. The certainty factor model, in particular turned out to be quite successful, and has enjoyed widespread use in expert systems built since MYCIN.

The strength of this approach lies in its computational simplicity, and in the straightforwardness of expert probability assessments required for its implementation. However, to obtain these improvements in performance, certain modification to the axioms governing the combination of probabilities were required. These led to considerably improved efficiency, but less accurate and often *ad hoc* results.

Nonetheless, on the whole the approach seemed to perform ‘satisfactorily’ in the problem domains to which it was applied [Shortliffe and Buchanan, 1984]. Despite reasonably good results in practice, some researchers heavily criticized such *quasi-probabilistic* approaches due to their lack of mathematical correctness, and the often unreasonable independence assumptions that were required to obtain the promised efficiency.

In the past decade, *belief networks* have gained popularity as a paradigm for the representation of uncertain knowledge in knowledge-based systems. The introduction of this graphical approach to the representation and evaluation of probabilistic knowledge has revived wide-spread interest in probability-based uncertainty reasoning. Belief networks are firmly founded in the same well-understood theory as the early probabilistic expert systems. However, due to the localization of causal interactions explicated by the graphical structure, only a fraction of the probabilities needed by the early probabilistic approaches is required here [Pearl, 1988].

We next introduce the notion of a belief network more systematically.

## 1.1. Belief Network Basics

Consider the following situation. By looking out the window we can see that our driveway is wet. However, we are not sure if this was caused by rain, or by a member of the family having turned the sprinkler on. If the sky is cloudy we can deduce that there is a good chance that it did indeed rain, and simultaneously, a fairly small chance that the sprinkler would have been turned on. Also, we can look over at the neighbor’s driveway. If it too is wet, chances are pretty good that it rained. The belief network in Figure 1.1 models this situation.

Belief networks are directed acyclic graphs (DAGs), such that the nodes represent random variables, and the arcs encode causal relationships among them. Each random variable can take on one of a finite set of values. In the network under discussion variables only take on values from the set {TRUE, FALSE}, but in general any finite number of states can be defined for a random variable. In fact, variable values do not even need to be discrete. See [] for a discussion of continuous variables in belief networks. They are not, however, addressed in this work.

The network in Figure 1.1 contains five nodes representing the random variables: Cloudy-Sky, Sprinkler-On, Rain, Driveway-Wet, and Neighbor's-Driveway-Wet*.

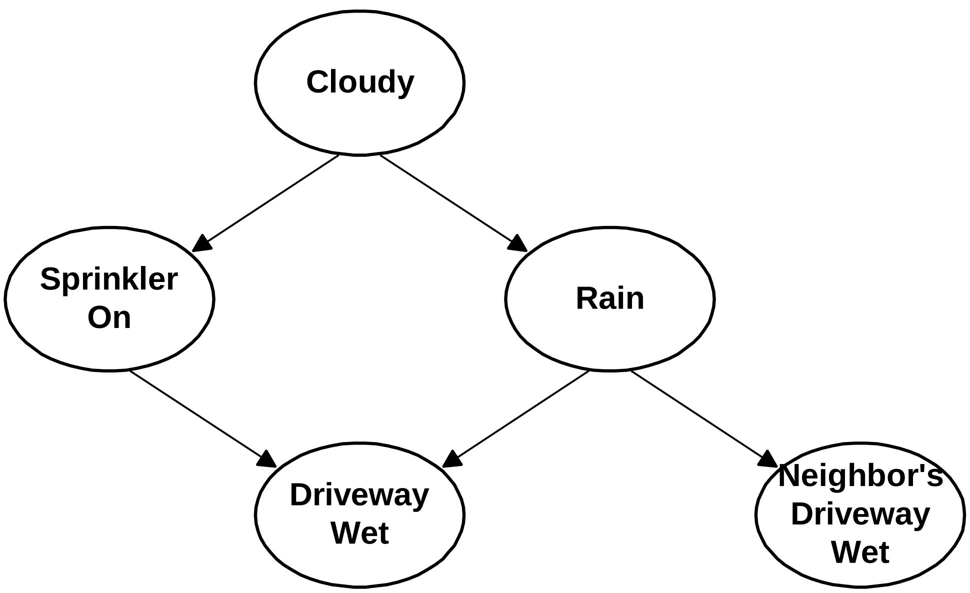



**Figure 1.1** Belief network representing the rain-sprinkler domain.

As stated, the arcs encode causal relationships among these random variables. Thus, for example, a cloudy sky affects the likelihood of rain, and of the sprinkler being turned on. Similarly, both the sprinkler and rain can cause our driveway to be wet. Notice that the random variable Neighbor's-Driveway-Wet has an arc to it only from Rain. This is because no causal relationship exists between our sprinkler being on or off, and whether or not the neighbor's driveway is wet. In other words, the random variables Sprinkler-On and Neighbor's-Driveway-Wet are *causally independent*. Thus, we see that not only does a belief network encode causal dependence, but it encodes causal independence as well. It will shortly become evident how this comes about, and why it is important.

Each variable taking on a value from its value set defines an *instantiation* of the model. The set of all possible instantiations defines a *joint probability distribution* over the set of variables. Thus, there are two ways in which to interpret the semantics of a belief network. The first is to view it as an encoding of conditional independence statements. The second is to see it as a representation of the joint probability distribution. These two interpretations are fundamentally equivalent, but the first serves to facilitate inference and belief revision procedures, while the second is useful in the network construction process.

## 1.2. Research Problem Description

As we saw in the previous section, exact computations in belief networks require that the joint probability distribution over the network variables be fully specified, or at least, that probability distributions for each of the variables conditional on their direct predecessors in the directed graph be available. The information needed for computing these distributions may be gleaned from literature, elicited from experts, or learned directly from data. In fact, in many domains some combination of these sources may be required. This can occur in a situation where only part of the required information is available from

* We shall use node and variable names interchangeably

prior knowledge. For example, an expert may be able to provide reliable information concerning only a subset of the variables. If statistical data about the remaining variables are available, it would be highly advantageous to have a method for combining it with the expert's assessment to obtain a fully quantified network.

Another way the need to fuse varying types of information can arise is when not all the information in numerical in nature. An expert may be certain of the fact that some values of a statistical variable *A* make some values of variable *B* more likely, and perhaps have an idea of the lower and upper bounds on the numerical strength of the influence, yet may not be able to give exact numbers. Also, available probabilities may not match the probabilities to be assessed. Medical literature, for example, often reports probabilities of symptoms given diseases and not the probabilities of symptoms given no diseases, and not necessarily the specific probabilities required for the intermediate disease states modeled in the network. Moreover, experts may feel more confident providing estimates of conditional probabilities in the diagnostic direction than in the causal direction of probabilistic influence. Thus, we would like to have a method for quantifying belief networks through the fusion of qualitative and quantitative prior knowledge, as well as any available statistical data.

In the current work we present a method that meets these needs. It is based on an existing approach for fusing qualitative and quantitative prior knowledge only. We adapt this approach, and expand it to accommodate both iterative and batch fusion of statistical data. Additionally, whereas the original method only addresses the quantification of the belief network, we propose an approach for utilizing the newly devised framework we present to refine the structure of the belief network as well.

## 1.3. Overview of Thesis

Now that we have covered the relevant introductory material and described the problem at hand, we shall proceed to formulate a solution. The following section presents an overview of prior work related to the issue of belief network construction and information fusion. It points out significant advances, and notes important shortfalls, which characterize the mentioned approaches. Finally, it provides the context for the present work, and explicates its contributions. Section 3 outlines our approach to flexible prior knowledge-data fusion, and gives examples of its application. Section 4 presents a description of the implementation of the method, and the results of an experimental evaluation. Section 5 concludes the work.

# 2. RELATED WORK

Ever since the initial introduction of belief networks to the AI community [Pearl, 1988], a great deal of research effort has been expended on devising methods for the efficient synthesis of these structures. For the most part this work has concentrated either on designing effective methods of knowledge elicitation from experts, or on developing methods for learning belief networks from data. In this section we review some of the key results emanating from these efforts, and relate them to the research problem at hand.

## 2.1. Knowledge Elicitation for Belief Network Synthesis

One approach is to rely on a human expert to provide both the causal structure of the network, as well as the quantification (see for example [Beinlich, *et al.*, 1989; Buntine, 1995]). In simple, well-understood domains this approach works quite well. However, in large, complex, and uncertain domains experts often find it difficult to provide complete and accurate information. In the case of the network structure, this may occur in domains where the causal relationship among variables is poorly understood. Also, in large and complex domains it may be difficult for an expert to keep track of all variables and the causal relationships among them. As a result certain important influences may be overlooked, leading to inaccuracies in the final model.

To help deal with these difficulties, a formalism called *similarity networks* has been proposed [Heckerman, 1991; Geiger and Heckerman, 1996]. This formalism permits the model builder a greater degree of flexibility than standard belief networks in specifying relationships among domain variables. Specifically, similarity networks are a natural means for representing *asymmetric independence* assertions. Such assertions state that variables are independent for some, but not all of their values. The representation employs multiple cross-correlated networks. This can be illustrated by an example adapted from [Geiger and Heckerman, 1996]:

> A guard of a secured building expects four types of persons to approach the building's entrance: workers in the building, company executives, approved visitors, and spies. As a person approaches the building, the guard can note their gender (*g*), whether or not they are wearing a badge (*b*), and whether or not they arrive in a limousine (*l*). Spies are mostly men. Spies always wear badges in an attempt to fool the guard. Visitors don't wear badges because they don't have one. Female workers tend to wear badges more often than do male workers. Only executives arrive in limousines, and male and female executives wear badges just as do regular workers. The task of the guard is to identify the person approaching the building.

Figure 2.1 (a) shows a standard belief network that represents this story. Variable *h* in the network represents the correct identification. It has four values: *worker, visitor*, *executive*, and *spy*. Variables *g*, *b*, and *l* represent, respectively, the person's gender, whether or not they are wearing a badge, and whether or not they arrive in a limousine. While this network correctly represents all dependencies, its topology hides certain asymmetric independence relationships. In particular, this network fails to represent the fact that gender and badge are conditionally independent given that the person is a spy or a visitor.

Figure 2.1 (b) shows a similarity network representation for the same story. Here the topology of the network is broken up into three *subsets*: {*spy*, *visitor*}, {*visitor*, *worker*}, {*worker*, *executive*}. The structure comprising each subset discriminates between two values of the hypothesis variables. Importantly, each *local network* contains only those variables which are relevant to the hypothesis values of the corresponding subset. Thus, we see that the main advantage of similarity networks, from the perspective of knowledge acquisition, is that a domain expert who provides the parameters of the network is not required to quantify the dependence between variables that are not related to the hypothesis under consideration.

In order not to loose information needed for correct diagnosis, the local networks (and corresponding subsets) must comprise a *collected cover* of hypotheses (see [Geiger and Heckerman, 1996] for definitions and proofs). Thus, a second important advantage of similarity networks representation is that it helps prevent the model builder from omitting relevant information. Every variable that is useful for discriminating between

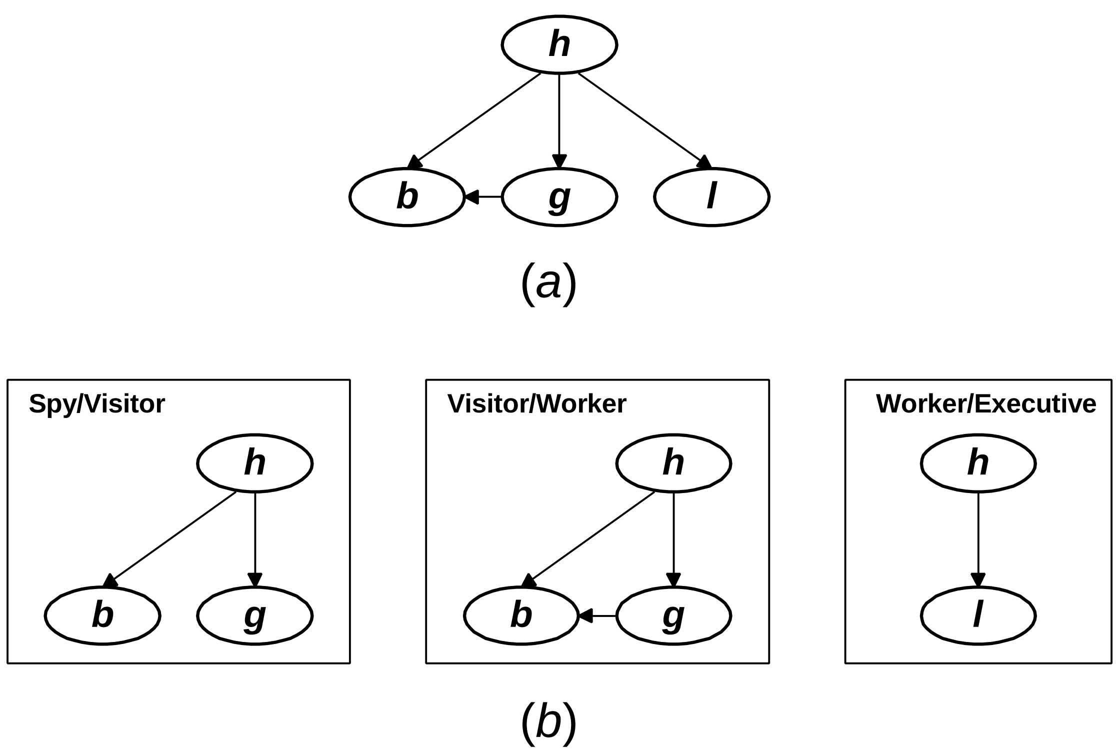



**Figure 2.1** A similarity network representation of the secured-building story

some pair of hypotheses will appear in at least one local network.

Although the similarity network formalism simplifies development of the network structure, and even aids in the quantification process, it does not address a crucial aspect of the latter. The difficulty of assessing probabilities is well know as a result of the human decision-making process under uncertainty (see for example [Meyer and Booker, 1991; Cooke, 1991; Kahneman *et al*., 1982]). Although it is, in principle, possible for a domain expert to assess a probability for any proposition even if he knows little about it, in practice experts are often uncertain and uncomfortable about the probabilities they are providing. Thus, we would like to conduct knowledge elicitation in such a way as to accommodate whatever degree of detail the expert is willing to specify. However, such flexibility poses a new set of difficulties. In the traditional approach to knowledge elicitation, where the expert is forced to specify point probabilities for all variables, the resulting belief network is fully and consistently quantified. In other words, the initially assessed conditional probabilities define a unique joint probability distribution on the statistical variables concerned. If we allow the expert more flexibility in the quantification process, we may not end up with a unique joint probability distribution. This ambiguity can have several causes. For example, some of the information provided may not be numerical in nature. An expert may be certain of the fact that some values of a statistical variable *A* make some values of variable *B* more likely, and perhaps have an idea of the lower and upper bounds on the numerical strength of the influence, yet may not be able to give exact numbers. Also, available probabilities may not match the probabilities to be assessed. Medical literature, for example, often reports probabilities of symptoms given diseases and not the probabilities of symptoms given no diseases, and not necessarily the specific probabilities required for the intermediate disease states modeled in the network. Moreover, experts may feel more confident providing estimates of conditional probabilities in the diagnostic direction than in the causal direction of probabilistic influence.

Thus, we see that probabilistic information is often available in many different forms. It ranges from numerical point and interval probabilities, through order of magnitude estimates and signs of influence synergies, to purely qualitative statements concerning independence variables. [van der Gaag, 1990; van der Gaag, 1991] proposed a method of dealing with probabilities specified in the form of intervals and/or comparisons, and [Druzdzel and van der Gaag, 1995] extended this approach to accommodate information that is even more qualitative in nature.

Their work is based on a particular constraint-based formulation of probability theory originally introduced by Boole [Boole, 1958; Hailperin, 1986]. The formulation conceptualizes a *distribution hyperspace* of all possible joint probability distributions over the set of domain variables. The true, yet unknown distribution is a point in this hyperspace. If no information is available about the true distribution, then it can be any point in the distribution hyperspace. Information about the true distributions, whether qualitative or quantitative, expresses a constraint on the hyperspace since certain distributions become incompatible with this information. Probability elicitation can now be looked upon as constraining the distribution hyperspace as much as possible. To this end, we express all the information about the unknown distribution as constraints. Assuming that all joint probability distributions that are compatible with this information are equally likely, second order distributions over the probabilities can then be assessed.

These second order distributions can be used directly or superimposed onto an existing belief network structure. In either case they can serve as a starting point for further refinement.

We next examine the paradigm in more detail since it serves as a basis for the key results of the present work. We introduce the approach by means of an example adapted from [Druzdzel and van der Gaag, 1995].

Consider building a highly simplified belief network modeling causes of HIV virus infection. The network would include four variables: *HIV infection* (*H*), *needle sharing* (*N*), *sexual intercourse* (*I*), and *use of a condom* (*C*). It is assumed for the sake of simplicity that these variables are binary; for example *H* has two outcomes, denoted h and $\overline{\text{h}}$, representing "HIV infection present" and HIV infection absent", respectively. The causal relationships and independences among these variables can now be used to synthesize the structure of the network. For example, let us assume that sharing needles

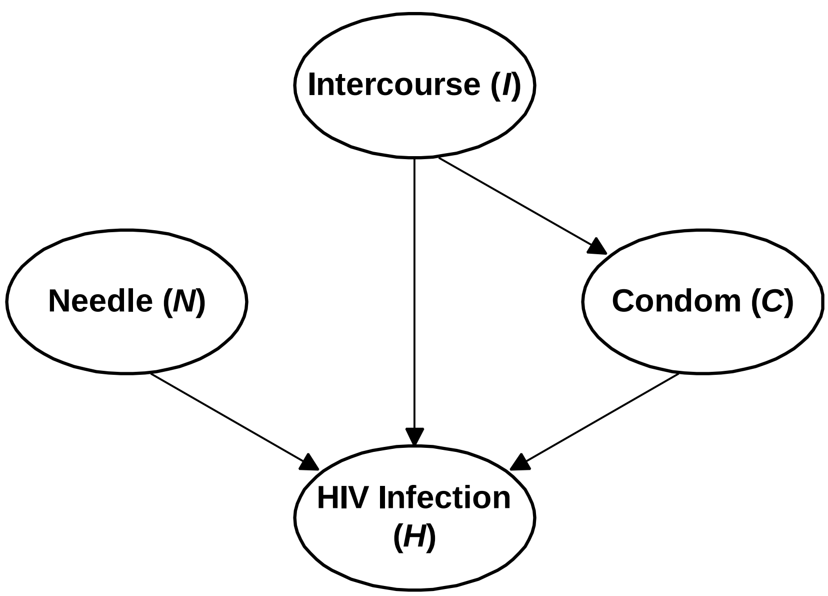



**Figure 2.2** An example belief network for HIV infection.

and condom usage are independent. Similarly, let us consider whether or not a person shares needles to be independent of whether this person engages in sexual intercourse. Figure 2.2 shows one possible causal graph reflecting these independence assertions.

Now that the structure of the network has been defined, it can be quantified. This necessitates assessing the probabilities of each variable conditioned upon its direct predecessors in the graph. For the network shown in Figure 2.2, numbers are required for $\Pr(N)$, $\Pr(I)$, $\Pr(C|I)$, and $\Pr(H|NIC)$. As mentioned earlier, determining these numbers is not a trivial matter, due to difficultiesin obtaining statistical data and in eliciting probabilities from domain experts. Several sources of information that can be used for the example at hand. Morbidity tables can provide $\Pr(h)$, a point estimate of the prevalence of HIV in the population of interest. Approximate estimates on frequencies of sexual intercourse and condom usage in intercourse, that is, $\Pr(i)$ and $\Pr(c|i)$. Also, since condoms are used primarily during intercourse, $\Pr(\bar{i}|c)$. In addition, various populations of intravenous drug users have been studied with respect to their needle sharing habits. Findings from these studies may help in assessing $\Pr(n)$. Also, statistics may be obtained on the different ways of contracting HIV from among the infected population, yielding estimates for $\Pr(n|h)$ and $\Pr(i|h)$, and perhaps even for $\Pr(ic|h)$ and Pr($i$ ` $c$ |$h$). There is also semi-numerical information available. For example, the probability of contracting HIV be needle sharing is higher than the probability of contracting it in sexual intercourse, that is, $\Pr(h|n) > \Pr(h|i)$. Also, the relatively small number of intravenous drug users compared to the size of the sexually active population suggests that $\Pr(i) > \Pr(n)$.

In order to accommodate such a variety of uncertainty information, probabilities are represented in a canonical form. This canonical form builds on the property that any joint probability distribution on a set of variables is uniquely defined by the probabilities of all possible combinations of values for all variables from the set. If these probabilities are known, any other (secondary) probability from the distribution can be computed from them by applying the basic rules of marginalization and conditioning from probability theory. The individual combinations of values for all variables are called *constituent assignments*. The corresponding probabilities for constituent assignments in a joint probability distribution are called *constituent probabilities*. Given these definitions it becomes clear that the set of all possible joint probability distributions over the domain variables spans a hyperspace whose dimensionality corresponds to the number of constituent probabilities.

Any information about the true, yet unknown joint probability distribution can now be expressed as a system of (in)equalities involving this distribution's constituent probabilities as unknowns. Any solution to this system of (in)equalities is a joint probability distributions that is compatible with the available information. If the system has a unique solutions, then the information provided suffices for uniquely defining the

| | | | |
|---|---|---|---|
| $c_1 = hnic$ | $c_5 = hni\bar{c}$ | $c_9 = h\bar{n}\bar{i}c$ | $c_{13} = \bar{h}\bar{n}i\bar{c}$ |
| $c_2 = \bar{h}nic$ | $c_6 = \bar{h}\bar{n}ic$ | $c_{10} = h\bar{n}i\bar{c}$ | $c_{14} = \bar{h}n\bar{i}\bar{c}$ |
| $c_3 = h\bar{n}ic$ | $c_7 = \bar{h}n\bar{i}c$ | $c_{11} = hn\bar{i}\bar{c}$ | $c_{15} = h\bar{n}\bar{i}\bar{c}$ |
| $c_4 = hn\bar{i}c$ | $c_8 = \bar{h}ni\bar{c}$ | $c_{12} = \bar{h}\bar{n}\bar{i}c$ | $c_{16} = \bar{h}\bar{n}\bar{i}\bar{c}$ |

**Table 2.1** Constituent probabilities for the HIV example

joint. Conversely, if the system does not have any solution at all, then the information provided is inconsistent.

We now make these concepts concrete through an illustrative example. Referring again to the HIV infection domain described above, we see that there are sixteen constituent assignments for the variables involved. An ordered list of these appears in Table 2.1. Consider the assignment expressing a person's having sexual intercourse without using a condom, that is, $i\bar{c}$. This assignment can be written as

$$i\bar{c} = hni\bar{c} \vee \bar{h}ni\bar{c} \vee h\bar{n}i\bar{c} \vee \bar{h}\bar{n}i\bar{c}$$

$$= c_5 \vee c_8 \vee c_{10} \vee c_{13}$$

The probability $\Pr(i\bar{c})$ can now be expressed as

$$\Pr(i\bar{c}) = \Pr(c_5) + \Pr(c_8) + \Pr(c_{10}) + \Pr(c_{13})$$

$$= x_5 + x_6 + x_{10} + x_{13}$$

Besides the numerical and semi-numerical information, a substantial amount of purely qualitative knowledge may be available. For example, in the HIV domain it is known that both sharing a needle and sexual intercourse with an HIV carrier make infection more likely, while the use of a condom during intercourse has the opposite effect. These pieces of information express *qualitative influences* between pairs of variables [Wellman, 1990; Druzdzel and Henrion, 1993]. Such information, while very general, can still go a long way toward constraining the distribution hyperspace. [Druzdzel and van der Gaag, 1995] extend the representation proposed in [van der Gaag, 1990; van der Gaag, 1991] to facilitate the inclusion of qualitative influences. The resulting constraint equations, however, are neither linear nor even fractionally linear. Thus, their inclusion forces the interval calculation problem to become one of non-linear optimization, and therefore, given a potentially large number of constraint equations, extremely expensive. For this reason we choose not to address qualitative influences in the present work.

## 2.2. Learning Belief Networks From Data

A number of researchers have examined methods for learning belief networks from data. These can be divided into two main categories: those that learn conditional probabilities only, and those that learn both conditional probabilities and structure. Conditional probabilities can be learned directly by application of Bayes Rule to evidence nodes, and subsequent propagation of evidence throughout the network. This approach is also know as *belief revision* or *belief refinement*, and is addressed in [Spiegelhalter and Lauritzen, 1988; Spiegelhalter and Lauritzen, 1990; Huang and Darwiche, 1994], among others.

Methods for learning complete belief networks, i.e., conditional probabilities along with the underlying causal structure are discussed in [Cooper and Herskovits, 1992; Heckerman and Geiger, 1995]. These methods all have the same basic components: a scoring metric and a search procedure. The metric computes a score that is proportional to the posterior probability of the network structure, given the data. The search procedure generates networks for evaluation by the scoring metric. These methods use the two components to identify a network or a set of networks with high posterior probabilities, and these networks are then used to predict future events.

The general approach described above, of learning belief networks solely from data suffers from two key shortfalls. Firstly, the enormous search space of possible network structures for belief networks in realistic domains makes efficient search all but impossible, and limits the quality of the final structure [Heckerman, *et al.*, 1997]. The second problem involves the statistical nature of the learning process, and pertains to the large number of data that is required to guarantee the reliability of the posterior probability calculations.

These shortfalls can in part be alleviated by the application of human prior knowledge. Several approaches exist that take advantage of prior knowledge. [Cooper, 1995; Heckerman et al., 1995] discuss methods for using human knowledge to supply prior and initial conditional probabilities for variables, as well as at least part of the causal structure. The application of this information amounts to selection bias in the search process. The combined prior knowledge/data approach has been shown to result in more accurate final belief networks and rapid convergence, with a substantial reduction in the amount of data required.

A major barrier to the effective application of combined prior knowledge/data learning algorithms to real-world domains has for some time now been the difficulty inherent in accurate and complete knowledge elicitation [Heckemran, 1990]. [Druzdzel and van der Gaag, 1995] propose a representation techniques that can accommodate both qualitative and quantitative probabilistic information about a yet incompletely specified joint probability distribution . The basic approach is to represent the joint as a set of constituent probabilities, and encode all known probabilistic information as a series of constraints over this set of constituent probabilities. Constraint propagation techniques can then be used to prune away all those networks that are invalidated by the prior information. This serves to reduce the search space, and can result in improved search times. The two chief shortcoming of this approach is that it assumes a previously existing belief network structure, and also provides no facility for integrating data.

The present work uses the above representation as a basis for the elicitation of prior knowledge, extends it to facilitate the learning of network structure, and combines it with belief refinement techniques to permit the integration of data. The result is an efficient mechanism for the constructionof belief networks through the combination of prior knowledge and data.

# 3. The Multi-source Fusion Algorithm

Having covered the background material, we now present the Multi-source Algorithm for Belief Network Synthesis. The aim of this algorithm is to allow for the flexible integration of quantitative, semi-quantitative and purely qualitative information, as well as statistical data for the purpose of belief network construction. At the heart of this algorithm lies the canonical formulation of probability presented in [Druzdzel and van der Gaag, 1995], and described in some detail in section 2.2. This approach as presented, however, only supports the fusion of probabilistic information elicited from experts. It does not suggest a means for the integration of probabilistic information available from data, or address structural information in any form. The algorithm we present modifies and extends the approach based on constituent probabilities to facilitate both iterative and batch introduction of data. We further suggest how this algorithm can be extended to facilitate the structural refinement of a belief network.

We now develop the algorithm.

## 3.1. Fusing in Data For Quantification Refinement

There are several ways in which statistical data can help in belief network synthesis. First of all, in many complex domains experts are not able to provide reliable information about some of the variables in the domain. One reason for this may be that the variables are hidden and not normally observable. Another reason may simply be that the expert's knowledge only extends to a subset of all the variables, and he is thus not able to provide reliable quantification for the variables outside his area of expertise. In the first case, since the variable is hidden, data will not be available. However, in the second case, if data is available for unquantified variables, depending on its reliability, it can prove very useful in replacing information that the expert was unable to provide. Another way data can prove useful is in refining previously specified probabilities. It is well known that information provided by domain experts is not infallible. In the case when an expert provides a point probability, this value can be inaccurate. Data can be combined with the incorrect probability to increase its accuracy. Additionally, if a variable's value is defined by an interval, data can help narrow this interval by propelling the variable's bounds towards its true value. Since we are using the representation described in section 2.2, a point probability is in fact an interval where both the upper bound and the lower bound have the same value. Thus, only the second case is relevant.

The key question, however, is how to efficiently combine statistical data with interval-based quantification in a belief network. If we were dealing with numbers and not intervals, the solution would be quite straightforward. After the initial quantification was in place, we could directly apply an existing evidence propagation technique such as message passing [Pearl, 1989] or the Lauritzen-Spiegelhalter algorithm for local probability computations on graphical structures [Spiegelhalter and Lauritzen, 1990]. This is, in fact, the approach taken by [Heckerman, *et al.*, 1995]. However, because the values of variables are represented as intervals the entire issue of evidence propagation becomes extremely complicated.

As discussed in section 2.2, the probability intervals can be viewed as a set of constraints over the hyperspace of joint probability distributions. In effect, a belief network partially quantified in this manner represents the set of possible joint probability distributions over the domain variables. In the absence of other information, each of these distributions is equally likely. All evidence propagation methods in effect iteratively refine a joint probability distribution through the application of Bayes rule to numerical probabilities. Now, however, we do not have access to individual numbers. In fact, it is not possible to efficiently use Bayes rule in a continuous interval-based domain. A detailed argument in terms of convex sets is given in [van der Gaag, 1990].

We propose an all together different approach. It turns out that for the purpose of interval refinement through evidence propagation, it is not really necessary to operate in a continuous interval-based domain. In fact, we can easily deal with the interval bounds only! This puts us back in the domain of real numbers. This observation is a large step forward, but we are not out of the woods yet.

In light of the observation we just made, is very tempting to try to represent a belief network with interval-based quantification as two separate, but structurally identical belief networks, one quantified by the lower bounds from the original network, and the other quantified by the corresponding upper bounds. Such a view would permit us to straightforwardly perform two parallel, but independent, standard propagation operations. In essence we would refine the set of lower bounds in complete isolation from the set of upper bounds. Best of all, this would give us a space-time complexity almost identical to that of propagating evidence in a belief network with a fully specified quantification (i.e., all point probabilities.)

Unfortunately, such an approach will not work. While the two interval bounds for a given value are mutually independent, this is not necessarily true across variables. It depends on the type of qualitative influence that exists between two variables [Druzdzel and Henrion, 1993]. For example, if some value of a statistical variable *A* makes some value of a variable *B more* likely, then the upper bound of *B*'s value depends on the upper bound of *A*'s value, and so the upper and lower bounds are independent across these particular values of these two variables. However, if some value of variable *A* makes some value of variable *B less* likely, then the *lower* bound of *B*'s value depends on the upper bound of *A*'s value, and the dissimilar bounds are no longer independent across the values.

All is not lost, however. We can still make use of the important fact that in our framework, for any pair of intervals, exactly one cross-interval bound relationship holds. That is, either there is a direct relationship between similar bounds, or there is a direct relationship between dissimilar bounds (and thus an inverse relationship between similar bounds). This is due to two basic properties of intervals. The first is that the refinement

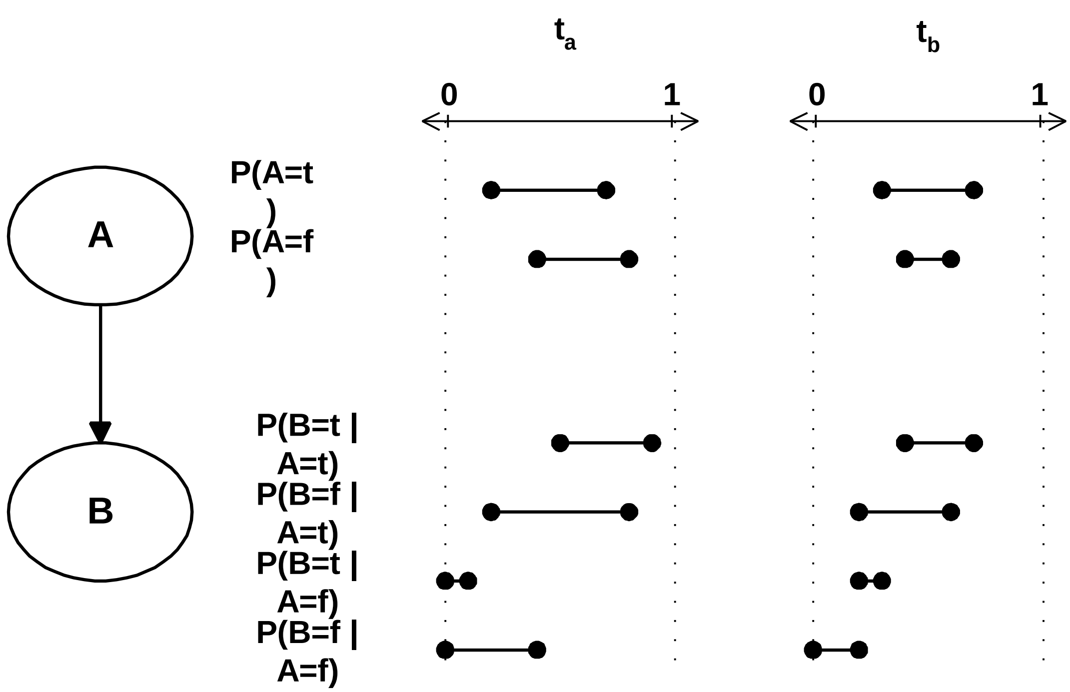



**Figure 3.1** Illustration of cross-interval influence relationships.

process can only shrink intervals, otherwise it is not sound. The second is that there are only two kinds of influences between values of variables, positive or negative. Now, because the same data is being applied to all of the intervals, only influence-preserving changes can occur. For example, Figure 3.1 shows two causally related variables, *A* and *B*. The probabilities of the values each variable can take on are represented by intervals, and two arbitrary snapshots in the midst of the refinement process are shown. Notice that for any two intervals for which a causal influence exists (e.g., P(*B*=f | *A*=t) and P(*A*=t) ), all bound movement occurs in accordance with the single bound relationship holding between those two intervals. For positive influences all four bounds move in the same direction, while for negative influences dissimilar intervals move in the same direction. Note that no movement at all of one bound can still be regarded as movement "with the influence" because as long as the other bound moves, the probability mass is shifting.

We now stop and formalize what we have just stated.

**Definition (Positive Influence)** *A positive influence exists between two values of distinct variables if a change of one results in a change of the other in the same direction.*

**Definition (Negative Influence)** *A negative influence exists between two values of distinct variables if a change of one results in a change of the other in the opposite direction.*

**Refinement Property 1** *An interval never increases in size.*

Note that this property only holds if the data is noise-free. Otherwise, intervals may indeed temporarily increase in size. However, as long as the noise is relatively small, the net change in interval size will be negative or zero.

**Refinement Property 2** *During refinement, the prevailing causal influence relationship between two values of distinct variables is preserved.*

**Lemma 1 (Soundness of Refinement)** *A refinement process in a set of causally related probability intervals is sound if and only if both Refinement Property 1 and Refinement Property 2 hold.*

We can now use the Soundness of Refinement Lemma to devise an approach for evidence propagation in semi-quantified belief networks. Let us take a look at an example. Figure 3.2 shows a simple inter-causal structure. The variables *A*, *B*, and *C* can each take on the values t and f. Let us assume that *C* has a positive influence relationship

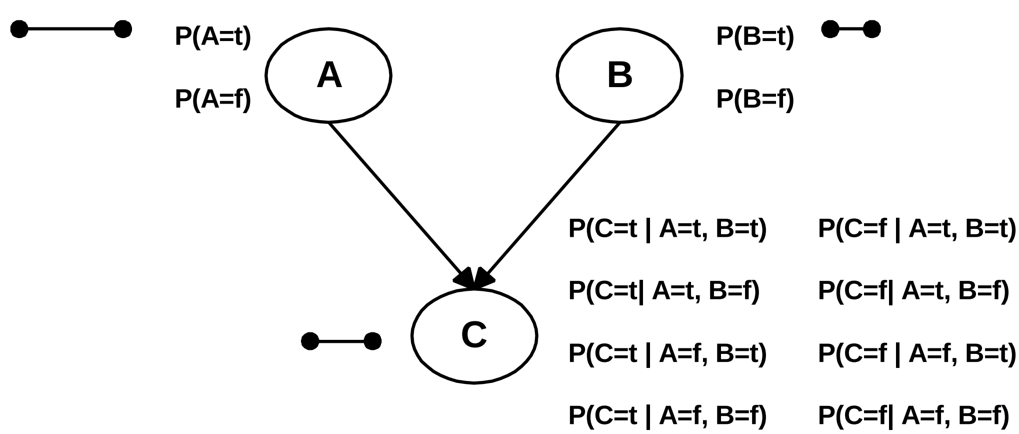



**Figure 3.2** A simple inter-causal structure.

with *A* and *B* (i.e., increasing the likelihood of either *A* or *B* will increase the likelihood of *C*). As was discussed in section 1.1, introducing evidence at *A* that increases its likelihood will resulting in the lowering of *B*'s likelihood. Thus, let's assume that we are interested in the probability intervals for P(A), P(C|A,B), and P(B|C). This is why in Figure 3.2 we show intervals for these probabilities only.

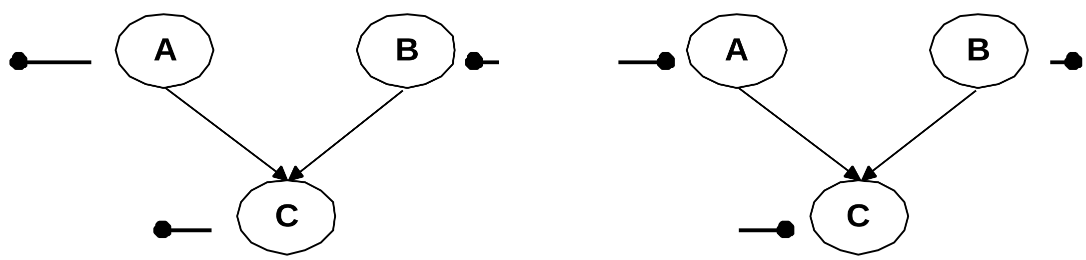



**Figure 3.3** A split-interval version of the inter-causal structure.

We now have three lower bounds and three upper bounds. Let us begin by mapping them onto two parallel belief networks, as was discussed at the beginning of the section. Figure 3.3 shows the resulting networks. If evidence concerning the true probability P(*A*) becomes available, we propagate it to node *C*, obtaining a new interval for P(*C*|*A*,*B*), and from there along to node *B*, calculating a new interval for P(*B*|*C*). If we do this with the split-interval structures shown in Figure 3.3, we will propagate from the new lower bound of node *A* to the lower bound of node *C*, and from there to the lower bound of node *B*. The same is done with the upper-bound structure. The results are shown in Table 3.1. As expected, the intervals for nodes *A* and *C* shrink because their respective bounds are being propelled towards the true probabilities. However, the interval for *B* becomes larger. This is a violation of the Soundness of Refinement Lemma. The reason it occurred is that we assumed there was a positive influence relationship between nodes *B* and *C*, where in reality it was a negative. Could we have known this in advance? The answer is yes; we could have determined this from the structure. The crucial observation to make here is that it does not matter whether the new evidence increases the probability of *A* or decreases it. The only thing we care about is how this *change* will affect the probability of *B*, and we know that P(*B*|*C*) will always change in the *opposite* direction of P(*A*).

| | | | |
|---|---|---|---|
| **lower bound A** | increase | **upper bound A** | decrease |
| **lower bound C** | increase | **upper bound C** | decrease |
| **lower bound B** | decrease | **upper bound B** | increase |

**Table 3.1** Results of first propagation.

At this point we have almost solved the problem. We can imagine an algorithm that first scans the structure, leaves all positively related intervals unchanged, but switches the bounds on the negatively related intervals. However, we cannot simply switch bounds directly, as a given interval may participate in more than one influence relationship, some of which are positive, while others are negative. What we need instead is a mechanism for indirectly referencing and changing bounds.

Figure 3.4 suggests how such a mechanism might work. We would have pointers connecting bounds, in essence indicating the direction of positive influence relationships. The negative influence relationships are the mirror image of the positive ones, so they

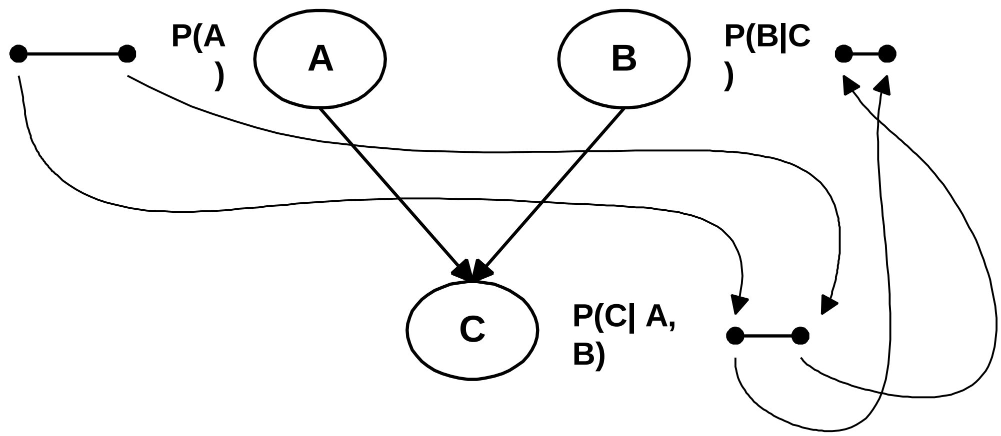



**Figure 3.4** A mechanism for evidence propagation in belief networks containing intervals.

need not be explicitly represented. Evidence affecting a particular interval would immediately propagate along the positive influence pointer paths from that interval's bounds. Thus, in effect, we are not propagating through the original belief network, but instead through a meta-network. The space-time complexity of the operation, however, does not change. We are still doing only about twice the work of a standard propagation algorithm because there are still exactly two pointers in the metanetwork for each link in the original belief network.

The algorithm is formally presented in the next section.

## 3.2. The Algorithm

We now provide a standard-form summary of the algorithm developed in the previous section.

INPUT: (1) Knowledge elicited from domain experts in varying formats. Acceptable formats are:

1. Point probability
2. Probability interval
3. Single probability bound

(2) Causal graph representing causal relationships among domain variables. (3) Statistical data for some, or all domain variables.

OUTPUT: Belief network with probability values for all variables, specified as probability intervals. Intervals are as wide as is consistent with all available information.

ALGORITH STEPS:

A. Preliminary

1. Obtain a causal graph qualitatively relating the variables.
2. Convert all domain expert information into constraint equations.
3. For each second-order prior or conditional probability value, compute and upper bound and a lower bound by solving a linear maximization problem and a linear minimization problem, respectively, over the constraint equations
4. Map the resulting intervals onto the casual graph to obtain a partially specified belief network.

B. Main Algorithm

1. Perform a breadth-first search of the belief network, examining every valid pair-wise relationship indicated in the graph.

   a. Where a *positive* influence is present, insert a pointer from the upper bound of the source value to the *upper* bound of the sink value.
   b. Where a *negative* influence is present, insert a pointer from the upper bound of the source value to the *lower* bound of the sink value.
2. Apply available statistical data to directly effected variable as evidence using the standard mechanism of Bayes rule.
3. Propagate effects of evidence to causally related variables along bound pointers using the Lauritzen-Spiegelhalter algorithm for local probability computations on graphical structures.

## 3.3. Refinement of Structure

In the previous section we presented a mechanism for using data to refine a semi-quantified belief network. We would like to say a few words about the possibility of using our framework to refine the structure as well. In section 2.2 we presented an overview of mechanisms for learning belief network structure. As was stated, learning a particular belief network's structure is essentially a process of search through the space of all possible structures. In fact, if the Bayesian approach to learning [Heckerman, *et al.*, 1995] is used, this is really just a search through the hyperspace of joint probability distributions, much like the hyperspace we use in section 3.1. This observation allows us to us constraints derived from out partial quantification to constrain and reduce the structure search space.

While this approach seems very promising, we have not investigated it in detail. One difficulty we foresee is the formulation of a common framework. The search space utilized by learning [Heckerman, *et al.*, 1995] is represented by numerical parameter sets, while we are working in an interval-based domain.

# 4. EXPERIMENTAL EVALUATION

In this section we describe the implementation of the data fusion algorithm, and report experimental results.

## 4.1. Implementation

The implementation of our proposed Multi-source Fusion Algorithm consisted of three major parts:

1. Constraint-based bound calculation.
2. Conversion between constraint-based representation and semi-quantified belief network.
3. Evidence propagation.

The first of these involved the implementation and adaptation of the Simplex algorithm for linear optimization. Since we dealt only with linear or fractionally linear equations, this seemed to be the logical choice.

The conversion mechanism was also fairly straightforward. In going from the constraint-based representation, each second-order probability value required for the belief network had to be extracted separately. This was done in two steps. First the upper bound for that probability value was calculated by solving a linear maximization problem involving all relevant constraints. Then, the lower bound for the probability value was calculated by solving a linear minimization problem in a similar manner. This was repeated for all second-order probability values of interest.

Finally, the evidence propagation mechanism was implemented according the description in section 3.1. Essentially, all of the interval bounds were placed into two arrays, and upper bound array and a lower bound array. Initially, all lower bound entries pointed to other lower bound entries, and similarly with the upper bound entries. A pre-scanner traversed the network in a breadth-first search manner and detected all pairs of nodes characterized by a negative influence relationship. Then these pairs of nodes had their mutual pointers switch to point to dissimilar bound entries. As noted before, nothing was changed in the original belief network. Only the meta-network pointers were switched around. Figure 4.1 shows how this implementation would work for the network in Figure 3.4.

## 4.2. Experimental Methodology

The goal of the experimental evaluation was to see how well the Multi-source Fusion Algorithm performs in practice. In a real-life situation, we would have some information about the domain in the form of elicited knowledge from experts. We would also have statistical data, which has been gathered from the domain environment. Since we did not have access to the required forms of information from a real-life domain, we had to do with a contrived one. We chose to adapt the ALARM domain [Beinlich, *et al.*, 1989] to our purpose. This involved using a version of the original ALARM network (shown in Figure 4.2) to represent the underlying true domain, and running simulations to generate information from it. Various filters were applied to the information thus obtained to simulate the required variety of knowledge sources. Two filters were in fact

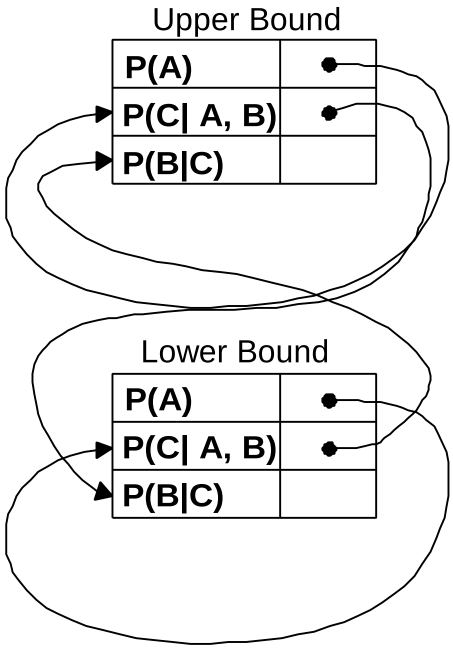



**Figure 4.1** Implementation of evidence propgation.

utilized, one to produce “human expert” information, and one to generate statistical data with varying amounts of uniform noise.

The “human expert” filter took numbers coming from the “golden” ALARM network did one of three things to them:

1. Left them as point probabilities.
2. Changed them to a probability interval of width within ± 50%, chosen along a normal distribution. Thus, smaller intervals were more likely then larger ones.
3. Replaced by *one* probability interval bound by generating an offset value within ± 50%, chosen along a normal distribution. Thus, smaller offsets were more likely then larger ones.

The statistical data filter was used on the same setting for every value in a given trial. The available noise settings were 0%, 10%, and 30%.

Six types of experiments were run:

1. No “human expert” perturbation, no noise in data.
2. No “human expert” perturbation, 10% noise in data.
3. No “human expert” perturbation, 30% noise in data.
4. “Human expert” perturbation present, no noise in data.
5. “Human expert” perturbation present, 10% noise in data.
6. “Human expert” perturbation present, 30% noise in data.

The first experiment was for control purposes, to see how the algorithm performs with complete elicited knowledge, and error-free data. The remainder evaluated performance under varying degrees of realistic uncertainty.

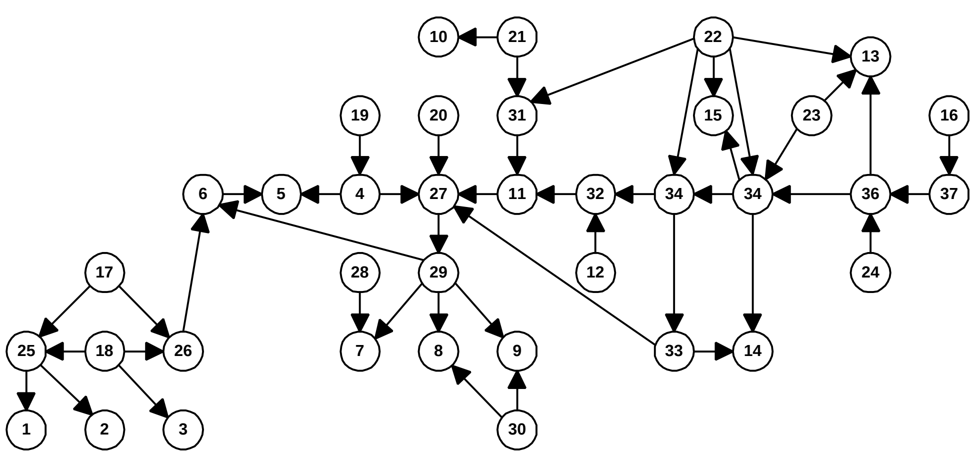



**Figure 4.2** ALARM network.

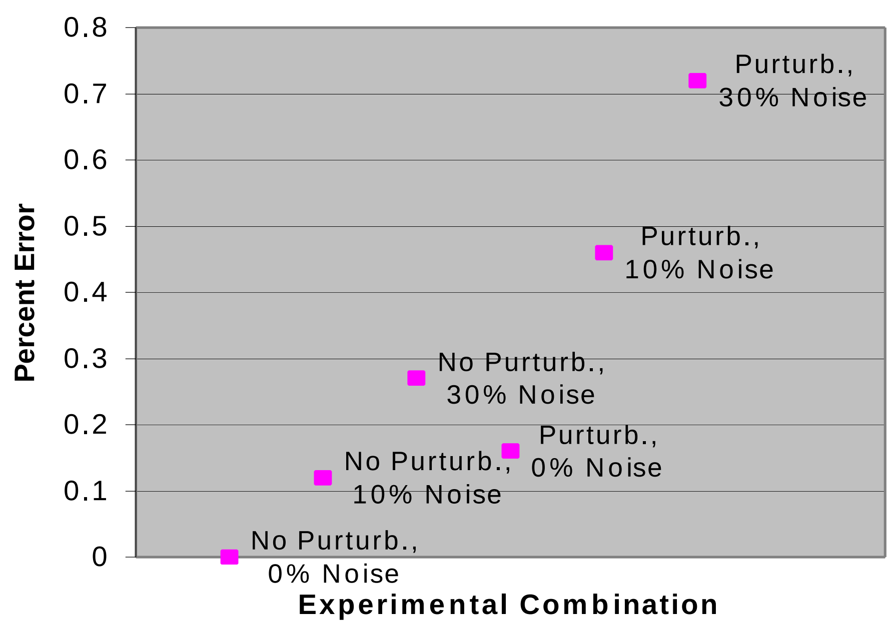
The metric for evaluation was the accuracy with which the synthesized network predicted the probabilities of three variable states. The three variables were chosen from widely dispersed sections of the original ALARM network:

1. P(*Node25*=T | *Node17*=T, *Node18*=F)
2. P(*Node11*=F | *Node31*=T, *Node32*=T)
3. P(*Node13*=T | *Node22*=T, *Node23*=F, *Node 36*=T)

A measure was obtained by averaging the three probabilities. The quality of the synthesized network was judged by the error in the measure with respect to that obtained for the same variables from the "golden" network.

For every experimental trial, 500 training examples were used, each having 37 T/F attributes, in accordance with the 37 random variables of the network in Figure 4.2.

**Plot 4.1** Experimental Evaluation of Multi-source Fusion Algorithm



The sample size of 500 comes from [Heckerman, *et al.*, 1995], in which this number worked well for refining point probabilities in the same causal graph structure. Though it was initially thought that this sample size would be sufficiently large for our experiments as well, it now seems that this is not the case.

Plot 4.1 shows the results for all six trials. As expected, the control case results in 0% error. This is because there is no uncertainty, and the algorithm simply calculates the requisite probabilities from exact information. As we increase the level of noise present in the data, the error rate goes up. This is also expected, but the particularly high error rate even at low levels of noise is somewhat worrying. This is most likely due to the fact that not enough statistical data was available. Apparently, the algorithm is not very tolerant to noise. This is an issue that should be investigated further if it is to be used in real-life domains. We notice further that when perturbation of "expert" information is added, but noise is eliminated, the error rate drops dramatically. Importantly, this is an indication that in low-noise environments the algorithm is quite effective at refining probabilistic intervals.

An additional worrying phenomenon is the fact that as both the perturbations of "expert" information and data noise increase, the error rate climbs beyond any degree of acceptability. It appears that in conjunction with the mentioned intolerance to noise, there is a second dynamic at work. It seems that the noisy data is not doing a very good job of refining the intervals. We believe that the problem is due to the violation of the Soundness of Refinement Lemma from section 3.1. As long as the noise level is very tiny, the local oscillations do not seems to have a significant effect on the quality of refinement, but as soon as the data noise level becomes even moderate, the oscillations become extremely noticeable. This points to an unboundedness condition. If this is the case, it is not clear at this point if there is a solution. It may be that drastically increasing the amount of statistical data will remove the observed effect. However, this is usually not the case with unbounded problems. In any case, this is anotherissue to be investigated further.

Thus, we see that the algorithm at least functions in an expected fashion. The error rate over-all is unacceptable, but we believe this problem can be addressed with minor changes to the framework. On the whole, we feelthat the results are encouraging in that the paradigm has been shown to be implementable and functional.

# 5. CONCLUSION

In this work we presented a methodology, and implemented an algorithm, for the fusion of statistical data with expert knowledge present in varying forms. This approach relies on a constraint-based representation of information. Each piece of expert knowledge is encoded as a constraint equation in terms of constituent probability assignments. These (in)equalities define a set of constraints over the space of all possible joint probability distributions. The distributions within this subspace are all consistent with the available knowledge. Given an existing network structure, the necessary second-order probability intervals can be extracted from this space of distributions by solving a series of linear optimization problems. Then, these second-order probability intervals can be mapped onto the waiting network structure. At this point statistical data can be introduced via the approach we outline and demonstrate.

We also suggested a means of refining network structure via this framework. We did not have the opportunity to develop this interesting and challenging problem further, but it seems that the approach is promising.

The experimental results were highly encouraging, but not definitive. The error rate was much to high where ever even moderate noise was present. However, given that the basic mechanism of the algorithm worked correctly, we feel that the noise sensitivity problem can be solved.

# REFERENCES

[Beinlich, *et al.*, 1989] Beinlich, I., Suermondt, H., Chavez, R., and Cooper, G. (1989). The ALARM monitoring system: a case study with two probabilistic inference techniques for belief networks. In *Proceedings of the Second European Conference on Artificial Intelligence in Medicine, London*. Springer-Verlag, Berlin.

[Boole, 1958] Boole, G. (1958). *An Investigation of the Laws of Thought on Which Are Founded the Mathematical Theories of Logic and Probabilities*. (Originally published in 1854 by Macmillan) Dover Publications, New York, NY.

[Buchanan and Shortliffe, 1984] Buchanan, B. and Shortliffe, E. (1984). *Rule-based Expert Systems. The MYCIN Experiments of the Stanford Heuristic Programming Project*, Addison-Wesley, Reading, Massachusetts.

[Bulitko and Wilkins, 1998] Bulitko, V. and Wilkins, D. (1998). Minerva: A Blackboard Expert System for Real-Time Problem-Solving and Critiquing, Technical Report, UIUC-BI-KBS-98-003, Beckman Institute, University of Illinois at Urbana-Champaign, Champaign, IL, 1998.

[Buntine, 1995] Buntine, W. (1995). A guide to literature on learning probabilistic networks from data. In *IEEE Transactions on Knowledge and Data Engineering*.

[Cestnik, B., 1990] Cestnik, B. (1990). Estimating probabilities: A crucial task in machine learning. In *Proceedings of the Ninth European Conference on Artificial Intelligence*, London.

[Charniak, 1991] Charniak, E. (Winter, 1991). Bayesian networks without tears. *AI Magazine*.

[Cooke, 1991] Cooke, R. (1 991). *Expert Uncertainty: Opinion and Subjective Probability in Science*. Oxford University Press.

[Cooper and Herskovits, 1992] Cooper, G. and Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9:309-347.

[Cooper, 1995] Cooper, G. (1995). Causal discovery from data in the presence of selection bias. In *Proceedings of the Fifth International Workshop on Artificial Intelligence and Statistics*, pages 140-150, Fort Lauderdale, FL.

[Druzdzel and Henrion, 1993] Druzdzel, M. and Henrion, M. (1993). Intercausal reasoning, with uninstantiated acestor nodes. In *Proceedings of the Ninth Annual Conference on Uncertainty in Artificial Intelligence*, Washington, D. C., pages 317-325. Morgan Kaufmann.

[Druzdzel and van der Gaag, 1995] Druzdzel, M. and van der Gaag, L. (1995). Elicitation of Probabilities for Belief Networks: Combining Qualitative and Quantitative Information. In *Proceedings of the Eleventh Annual Conference on*

*Uncertainty in Artificial Intelligence*, Montreal, QU, pages 141-148. Morgan Kaufmann.

[Duda *et al*., 1976] Duda, R., Hart, P., and Nilsson, N. (1976). Subjective Bayesian methods for rule-based inference systems. Technical Note 124, Artificial Intelligence Center, SRI International, Menlo Park.

[Duda *et al*., 1979] Duda, R., Gashing, J., and Hart, P. (1979). Model design in the PROSPECTOR consultant system for mineral exploration. In Michie, D. editor, *Expert Systems in the Micro Electronic Age*, Edinburgh University Press, Edinburgh, pp. 153-167.

[Geiger and Heckerman, 1996] Geiger, D. and Heckerman, D. (1996). Knowledge representation and inference in similarity networks and Bayesian multinets. *Artificial Intelligence*, 82:45-74.

[Hailperin, 1986] Hailperin, T. (1986). *Boole's Logic and Probability*, 2nd Edition, North-Holland, Amsterdam.

[Heckerman and Geiger, 1995] Heckerman, D. and Geiger, D. (February, 1995). Learning Bayesian networks. Technical Report MSR-95-02, Advanced Technology Division, Microsoft Research.

[Heckerman, *et al*., 1995] Heckerman, D., Geiger, D., Chickering, D. (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, 20:197-243.

[Heckerman, 1991] Heckerman, D. (1991). *Probabilistic Similarity Networks*. MIT Press, Cambridge, MA.

[Heckerman, *et al*., 1990] Heckerman, D., Horvitz, E., and Nathwani, B. (1990). Toward normative expert systems: The Pathfinder project. Technical Report KSL-90-08, Medical Computer Science Group, Section on Medical Informatics, Stanford University, Stanford, CA.

[Heckerman, *et al*., 1997] Heckerman, D., Meek, C., Cooper, G. (February, 1997). A Bayesian approach to causal discovery. Technical Report MSR-TR-97-05, Advanced Technology Division, Microsoft Research.

[Horvitz, 1988] Horvitz, E., Breese, J., Henrion, M. (1988). Decision theory in expert systems and artificial intellience, *International Journal of Approximate Reasoning*, vol. 2, pp. 247-302.

[Huang and Darwiche, 1994] Huang, C. and Darwiche, A (1994). Inference in belief networks: a procedural guide, *International Journal of Approximate Reasoning*, 11:1-158.

[Kahneman *et al*., 1982] Kahneman, D., Slovic, P, and Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge University Press, Cambridge, MA.

[Mengshoel and Wilkins, 1997] Mengshoel, O. and Wilkins, D. (1997). Visualizing Uncertainty in Battlefield Reasoning Using Belief Networks. In *1st Annual*

*Federated Laboratory Symposium, Advanced Displays and Interactive Displays*, Adelphi, MD, January 1997.

[Meyer and Booker, 1991] Meyer, M. and Booker, J. (1991). Eliciting and analyzing expert judgment: a practical guide. *Knowledge Based Systems Volume 5*. Academic Press.

[Mitchell, 1997] Mitchell, T. M. (1997). *Machine Learning*. WCB/McGraw-Hill, New York.

[Neapolitan, 1990] Neapolitan, R. E. (1990). *Probabilistic Reasoning in Expert Systems: Theory and Algorithms*. John Wiley & Sons, Inc., New York.

[Pearl, 1988] Pearl, J. (1988*). Probabilistic reasoning in intelligent systems: networks of plausible inference*, Morgan Kaufmann, San Mateo, CA.

[Palmer, 1998] Palmer, G. (1998). Combining Bayesian Belief Networks with Gas Path Analysis for Test Cell Diagnostics and Overhaul. Enter Software Inc. Available at http://www.enter.com/pub_dwnld/ohexpert.doc.

[Russel and Norvig, 1995] Russel, S. and Norvig, P. (1995). *Artificial Intelligence: A Modern Approach*. Prentice Hall, Upper Saddle River, New Jersey.

[Shafer and Pearl, 1990] Shafer, G. and Pearl, J. (1990). Readings in Uncertainty Reasoning. Morgan Kaufman, San Mateo, CA.

[Shortliffe and Buchanan, 1984] Shortliffe, E. and Buchanan, B. (1984). A model of inexact reasoning in medicine. In Buchanan, B. and Shortliffe, E. editors, *Rule-based Expert Systems. The MYCIN Eexperiments of the Stanford Heuristic Programming Project*, Addison-Wesley, Reading, Massachusetts, pp. 233-262.

[Spiegelhalter and Lauritzen, 1990] Spiegelhalter, D. and Lauritzen, S. (1990). Sequential updating of conditional probabilities on directed graphical structures. *Networks*, 20:579-605.

[Spirtes and Meek, 1995] Spirtes, P. and Meek, C. (1995). Learning Bayesian networks with discrete variables from data. In *Proceedings of First International Conference on Knowledge Discovery and Data Mining*, Montreal, QU. Morgan Kaufmann.

[Spirtes *et al*., 1993] Spirtes, P., Glymour, C., and Scheines, R. (1993). *Causation, Prediction , and Search*. Springer-Verlag, New York.

[Spirtes *et al*., 1995] Spirtes, P., Meek, C, and Richardson, T. (1995). Causal inference in the presence of latent variables and selection bias. In *Proceedings of Eleventh Conference of Uncertainty in Artificial Intelligence*, Montreal, QU, pages 499-506. Morgan Kaufmann.

[van der Gaag, 1990] van der Gaag, L. (1990). *Probability-Based Models for Plausible Reasoning*. Ph.D. Thesis, University of Amsterdam.

[van der Gaag, 1991] van der Gaag, L. (1991). Computing Probability Intervals Under Independency Constraints. In Bonissone, P., Henrion, M., Kanal, L, and Lemmer, J., editors, *Uncertainty in Artificial Intelligence 6*. Elsevier Science Publishers B.V. (North Holland), pages 457-466.

[Wellman, 1990] Wellman, M. P. (1990). Fundamental concepts of qualitative probabilistic networks, *Artificial Intelligence*, 44:257-303.