The Certainty-Factor Model*

David Heckerman

Departments of Computer Science and Pathology

University of Southern California

HMR 204, 2025 Zonal Ave

Los Angeles, CA 90033

dheck@sumex-aim.stanford.edu

1 Introduction

The *certainty-factor* (CF) model is a method for managing uncertainty in rule-based systems. Shortliffe and Buchanan (1975) developed the CF model in the mid-1970s for MYCIN, an expert system for the diagnosis and treatment of meningitis and infections of the blood. Since then, the CF model has become the standard approach to uncertainty management in rule-based systems.

When the model was created, many artificial-intelligence (AI) researchers expressed concern about using Bayesian (or subjective) probability to represent uncertainty. Of these researchers, most were concerned about the practical limitations of using probability theory. In particular, information-science researchers were using the *idiot-Bayes* model to construct expert systems for medicine and other domains. This model included the assumptions that (1) faults or hypotheses were mutually exclusive and exhaustive, and (2) pieces of evidence were conditionally independent, given each fault or hypothesis. (See Bayesian Inference

*This work was supported by the National Cancer Institute under Grant RO1CA51729-01A1.

Methods (*qv*) for a definition of these terms.) The assumptions were useful, because their adoption made the construction of expert systems practical. Unfortunately, however, the assumptions were often inaccurate in practice.

The CF model was created to avoid the unreasonable assumptions in the idiot-Bayes model. In this article, however, we see that the CF model is no more useful than is the idiot-Bayes model. In fact, in certain circumstances, the CF model implicitly imposes assumptions of conditional independence that are stronger than those of the idiot-Bayes model. We trace the flaws in the CF model to the fact that the model imposes the same sort of *modularity* on uncertain rules that we ascribe to logical rules, although uncertain reasoning is inherently less modular than is logical reasoning. In addition, we examine the belief network, a graphical representation of beliefs in the probabilistic framework. We see that this representation overcomes the difficulties associated with the CF model.

2 The Mechanics of the Model

To understand how the CF model works, let us consider a simple example taken from Bayesian Inference Methods (*qv*):

> Mr. Holmes receives a telephone call from his neighbor Dr. Watson stating that he hears a burglar alarm sound from the direction of Mr. Holmes' house. Preparing to rush home, Mr. Holmes recalls that Dr. Watson is known to be a tasteless practical joker, and he decides to first call his other neighbor, Mrs. Gibbons, who, despite occasional drinking problems, is far more reliable.

A miniature rule-based system for Mr. Holmes' situation contains the following rules:

R_1: if WATSON'S CALL then ALARM, $CF_1 = 0.5$

R_2: if GIBBON'S CALL then ALARM, $CF_2 = 0.9$

R_3: if ALARM then BURGLARY, $CF_3 = 0.99$

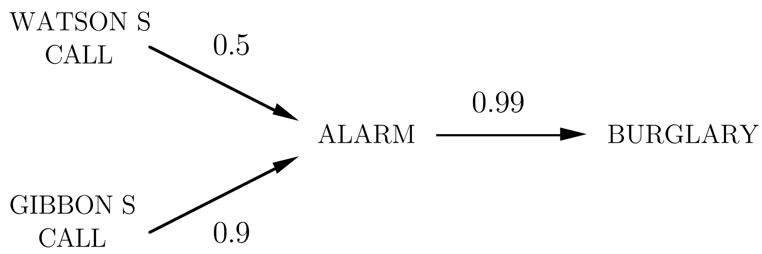


Figure 1: An inference network for Mr. Holmes' situation.

Each arc represents a rule. For example, the arc from ALARM to BURGLARY represents the rule R_3 ("if ALARM then BURGLARY"). The number above the arc is the CF for the rule.

In general, rule-based systems contain rules of the form "if e then h," where e denotes a piece of evidence for hypothesis h. Using the CF model, an expert represents her uncertainty in a rule by attaching a single CF to each rule. A CF represents a person's (usually, the expert's) *change* in belief in the hypothesis given the evidence. In particular, a CF between 0 and 1 means that the person's belief in h given e increases, whereas a CF between -1 and 0 means that the person's belief decreases. Unlike a probability, a CF does not represent a person's *absolute* degree in belief of h given e. In Section 4.3, we see exactly what a CF is in probabilistic terms.

Several implementations of the rule-based representation display a rule base in graphical form as an *inference network*. Figure 1 illustrates the inference network for Mr. Holmes' situation. Each arc in an inference network represents a rule; the number above the arc is the CF for the rule.

Using the CF model, we can compute the change in belief in any hypothesis in the network, given the observed evidence. We do so by applying simple *combination functions* to the CFs that lie between the evidence and the hypothesis in question. For example, in Mr. Holmes' situation, we are interested in computing the change in belief of BURGLARY, given WATSON'S CALL and GIBBON'S CALL. We combine the CFs in two steps. First, we combine CF_1 and CF_2, the CFs for R_1 and R_2, to give the CF for the rule R_4:

R_4: if WATSON'S CALL and GIBBON'S CALL then ALARM, CF_4

We combine CF_1 and CF_2 using the function

$$CF_4 = \begin{cases} CF_1 + CF_2 - CF_1 CF_2 & CF_1, CF_2 \geq 0 \\ CF_1 + CF_2 + CF_1 CF_2 & CF_1, CF_2 < 0 \\ \frac{CF_1 + CF_2}{1 - \min(|CF_1|, |CF_2|)} & \text{otherwise} \end{cases} \tag{1}$$

In Mr. Holmes' case, we have

$$CF_4 = 0.5 + 0.9 - (0.5)(0.9) = 0.95$$

Equation 1 is called the *parallel-combination function*. In general, we use this function to combine two rules that share the same hypothesis.

Second, we combine CF_3 and CF_4, to give the CF for the rule R_5:

R_5: if WATSON'S CALL and GIBBON'S CALL then BURGLARY, CF_5

The combination function is

$$CF_5 = \begin{cases} CF_3 CF_4 & CF_3 > 0 \\ 0 & CF_3 \leq 0 \end{cases} \tag{2}$$

In Mr. Holmes' case, we have

$$CF_5 = (0.99)(0.95) = 0.94$$

Equation 2 is called the *serial-combination function*. We use this function to combine two rules where the hypothesis in the first rule is the evidence in the second rule.

If all evidence and hypotheses in a rule base are simple propositions, we need to use only the serial and parallel combination rules to combine CFs. The CF model, however, also can accommodate rules that contain conjunctions and disjunctions of evidence. For example, suppose we have the following rule in an expert system for diagnosing chest pain:

R_6: if CHEST PAIN *and*
SHORTNESS OF BREATH
then HEART ATTACK, $CF_6 = 0.9$

Further, suppose that we have rules that reflect indirect evidence for chest pain and shortness of breath:

R_7: if PATIENT GRIMACES then CHEST PAIN, $CF_7 = 0.7$

R_8: if PATIENT CLUTCHES THROAT then SHORTNESS OF BREATH, $CF_8 = 0.9$

We can combine CF_6, CF_7, and CF_8 to yield the CF for the rule R_9:

R_9: if PATIENT GRIMACES *and*
PATIENT CLUTCHES THROAT
then HEART ATTACK, CF_9

The combination function is

$$CF_9 = CF_6 \min(CF_7, CF_8) = (0.9)(0.7) = 0.63 \tag{3}$$

That is, we compute the serial combination of CF_6 and the minimum of CF_7 and CF_8. We use the minimum of CF_7 and CF_8, because R_6 contains the conjunction of CHEST PAIN and SHORTNESS OF BREATH. In general, the CF model prescribes that we use the minimum of CFs for evidence in a conjunction, and the maximum of CFs for evidence in a disjunction.

There are many variations among the implementations of the CF model. For example, the original CF model used in MYCIN treats CFs less than 0.2 as though they were 0 in serial combination, to increase the efficiency of computation. For the sake of brevity, we will not describe other variations.

3 An Improvement over Idiot Bayes?

In the simple case of Mr. Holmes, the CF model is an improvement over the idiot-Bayes model. In particular, WATSON'S CALL and GIBBON'S CALL are not conditionally independent, given BURGLARY, because even if Mr. Holmes knows that a burglary has occurred, receiving Watson's call increases Mr. Holmes belief that Mrs. Gibbons will report the alarm sound. The lack of conditional independence is due to the fact that the calls are triggered by the alarm sound, and not by the burglary. The CF model represents accurately this lack of independence through the presence of ALARM in the inference network.

Unfortunately, the CF model cannot represent most real-world problems in a way that is both accurate and efficient. In the next section, we shall see that the assumptions of conditional independence associated with the parallel-combination function are stronger (i.e., are less likely to be accurate) than are those associated with the idiot-Bayes model.

4 Theoretical Problems with the CF Model

Rules that represent logical relationships satisfy the *principle of modularity.* That is, given the logical rule "if e then h," and given that e is true, we can assert that h is true (1) no matter how we established that e is true, and (2) no matter what else we know to be true. We call (1) and (2) the *principle of detachment* and the *principle of locality,* respectively. For example, given the rule

R_{10}: if L_1 and L_2 are parallel lines then L_1 and L_2 do not intersect

we can assert that L_1 and L_2 do not intersect once we know that L_1 and L_2 are parallel lines. This assertion depends on neither how we came to know that L_1 and L_2 are parallel (the principle of detachment), nor what else we know (the principle of locality).

The CF model employs the same principles of detachment and locality to belief updating. For example, given the rule

R_3: if ALARM then BURGLARY, $CF_3 = 0.99$

and given that we know ALARM, the CF model allows us to update Mr. Holmes' belief in BURGLARY by the amount corresponding to a CF of 0.99, no matter how Mr. Holmes established his belief in ALARM, and no matter what other facts he knows.

Unfortunately, uncertain reasoning often violates the principles of detachment and locality. Use of the CF model, therefore, often leads to errors in reasoning.[1] In the remainder

[1]Heckerman and Horvitz (1987 and 1988) first noted the nonmodularity of uncertain reasoning, and the relationship of such nonmodularity to the limitations of the CF model. Pearl (1988, Chapter 1) first

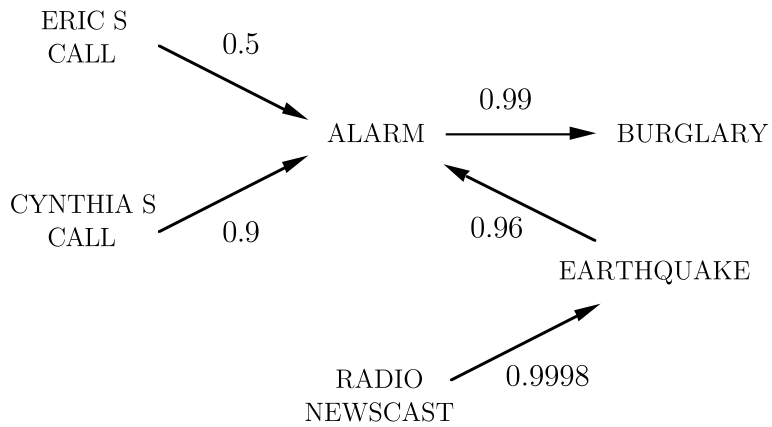


Figure 2: Another inference network for Mr. Holmes' situation. In addition to the interactions in Figure 1, RADIO NEWSCAST increases the chance of EARTHQUAKE, and EARTHQUAKE increases the chance of ALARM.

of this section, we examine two classes of such errors.

4.1 Multiple Causes of the Same Effect

Let us consider the simple embellishment to Mr. Holmes' problem given in Bayesian Inference Methods (*qv*):

> Mr. Holmes remembers having read in the instruction manual of his alarm system that the device is sensitive to earthquakes and can be triggered by one accidentally. He realizes that if an earthquake had occurred, it would surely be on the news. So, he turns on his radio and waits around for a newscast.

Figure 2 illustrates a possible inference network for his situation. To the original inference network of Figure 1, we have added the rules

R_{11}: if RADIO NEWSCAST then EARTHQUAKE, $CF_{11} = 0.9998$

R_{12}: if EARTHQUAKE then ALARM, $CF_{12} = 0.95$

---

decomposed the principle of modularity into the principles of detachment and locality.

The inference network does not capture an important interaction among the propositions. In particular, the modular rule R_3 ("if ALARM then BURGLARY") gives us permission to increase Mr. Holmes' belief in BURGLARY, when his belief in ALARM increases, no matter how Mr. Holmes increases his belief for ALARM. This modular license to update belief, however, is not consistent with common sense. If Mr. Holmes hears the radio newscast, he increases his belief that an earthquake has occurred. Therefore, he decreases his belief that there has been a burglary, because the occurrence of an earthquake would account for the alarm sound. Overall, Mr. Holmes' belief in ALARM increases, but his belief in BURGLARY decreases.

When the evidence for ALARM came from WATSON'S CALL and GIBBON'S CALL, we had no problem propagating this increase in belief through R_3 to BURGLARY. In contrast, when the evidence for ALARM came from EARTHQUAKE, we could not propagate this increase in belief through R_3. This difference illustrates a violation of the detachment principle in uncertain reasoning: the *source* of a belief update, in part, determines whether or not that update should be passed along to other propositions.

Pearl (1988, Chapter 1) describes this phenomenon in detail. He divides uncertain inferences into type types: diagnostic and predictive.[2] In an *diagnostic inference*, we change the belief in a cause, given an effect. All the rules in the inference network of Figure 2, except R_{12}, are of this form. In a *predictive inference*, we change the belief in an effect, given a cause. R_{12} is an example of such an inference. Pearl describes the interactions between the two types of inferences. He notes that, if the belief in a proposition is increased by a diagnostic inference, then that increase can be passed through to another diagnostic inference—just what we expect for the chain of inferences from WATSON'S CALL and GIBBON'S CALL to BURGLARY. On the other hand, if the belief in a proposition is increased by a predictive inference, then that belief should not be passed through a diagnostic inference. Moreover, when the belief in one cause of an observed effect increases, the beliefs in other causes should decrease—just what we expect for the two causes of ALARM.

[2]Henrion (1987) also makes this distinction.

We might be tempted to repair the inference network in Figure 2, by adding the rule

$$R_{13}\text{: if EARTHQUAKE then BURGLARY, } CF_{13} = -0.7$$

Unfortunately, this addition leads to another problem. In particular, suppose that Mr. Holmes had never received the telephone calls. Then, the radio newscast should not affect his belief in a burglary. The modular rule R_{13}, however, gives us a license to decrease Mr. Holmes' belief in BURGLARY, whether or not he receives the phone calls. This problem illustrates that uncertain reasoning also can violate the principle of locality: The validity of an inference may depend on the truth of other propositions.

To represent accurately the case of Mr. Holmes, we must include a rule for every possible combination of observations:

if WATSON'S CALL *and*
GIBBON'S CALL *and*
RADIO NEWSCAST
then BURGLARY

if NOT WATSON'S CALL *and*
GIBBON'S CALL *and*
RADIO NEWSCAST
then BURGLARY

⋮

This representation is inefficient and is difficult to modify, and needlessly clusters propositions that are only remotely related. Ideally, we would like a representation that encodes only direct relationships among propositions, and that infers indirect relationships. In Section 6, we examine the belief network, a representation with such a capability.

We find the same difficulties in representing Mr. Holmes' situation with the CF model, whenever we have multiple causes of a common effect. For example, if a friend tells us that our car will not start, we initially may suspect that either our battery is dead or the gas tank is empty. Once we find that our radio is dead, however, we decrease our belief that the

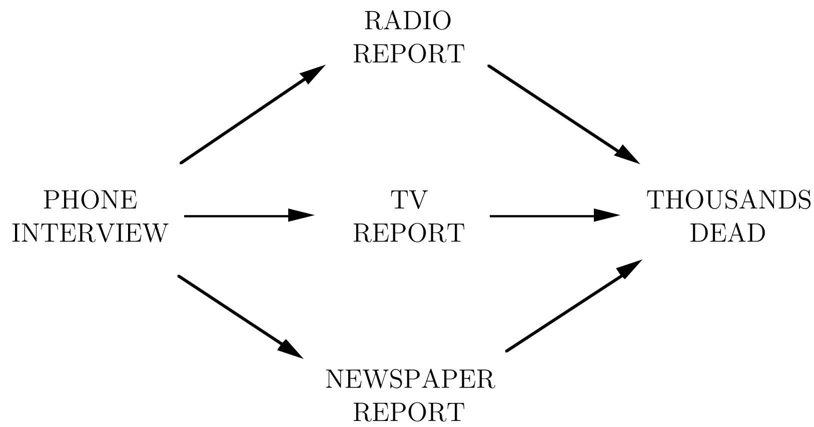


Figure 3: An inference network for the Chernobyl disaster (from Henrion, 1986). When we combine CFs as modular belief updates, we overcount the chance of THOUSANDS DEAD.

tank is empty, because now it is more likely that our battery is dead. Here, the relationship between CAR WILL NOT START and TANK EMPTY is influenced by RADIO DEAD, just as the relationship between ALARM and BURGLARY is influenced by RADIO NEWSCAST. In general, when one effect shares more than one cause, we should expect violations of the principles of detachment and locality.

4.2 Correlated Evidence

Figure 3 depicts an inference network for news reports about the Chernobyl disaster. On hearing radio, television, and newspaper reports that thousands of people have died of radioactive fallout, we increase substantially our belief that many people have died. When we learn that each of these reports originated from the same source, however, we decrease our belief. The CF model, however, treats both situations identically.

In this example, we see another violation of the principle of detachment in uncertain reasoning: The sources of a set of belief updates can strongly influence how we combine those updates. Because the CF model imposes the principle of detachment on the combination of belief updates, it overcounts evidence when the sources of that evidence are positively correlated, and it undercounts evidence when the sources of evidence are negatively correlated.

4.3 A Probabilistic Interpretation for Certainty Factors

Heckerman (1986) delineated precisely the limitations of the CF model. He proved that we can give a probabilistic interpretation to any scheme—including the CF model—that combines belief updates in a modular and consistent fashion. In particular, he showed that we can interpret a belief update for hypothesis h, given evidence e, as a function of the likelihood ratio

$$\lambda = \frac{p(e|h,\xi)}{p(e|\text{NOT } h,\xi)} \tag{4}$$

In Equation 4, $p(e|h,\xi)$ denotes the probability (i.e., degree of belief) that e is true, given that h is true, and ξ denotes the background knowledge of the person to whom the belief belongs. Using Bayes' theorem (Bayesian Inference Methods, *qv*), we can write λ as the ratio of the *posterior odds* to *prior odds* of the hypothesis:

$$\lambda = \frac{O(h|e,\xi)}{O(h|\xi)} = \frac{\frac{p(h|e,\xi)}{1-p(h|e,\xi)}}{\frac{p(h|\xi)}{1-p(h|\xi)}} \tag{5}$$

Equation 5 shows more clearly that λ represents a change in belief in a hypothesis, given evidence. Bayesian Inference Methods (*qv*) contains a detailed description the likelihood ratio.

For the CF model, Heckerman showed that, if we make the identification

$$CF = \begin{cases} \frac{\lambda-1}{\lambda} & \lambda \geq 1 \\ \lambda - 1 & \lambda < 1 \end{cases} \tag{6}$$

then the parallel-combination function (Equation 1) follows exactly from Bayes' theorem. In addition, with the identification in Equation 6, the serial-combination function (Equation 2) and the combination functions for disjunction and conjunction are close approximations to the axioms of probability.

In developing this probabilistic interpretation for CFs, Heckerman showed that each combination function imposes assumptions of conditional independence on the propositions involved in the combinations. For example, when we use the parallel-combination function to combine CFs for the rules "if e_1 then h" and "if e_2 then h," we assume implicitly that e_1 and e_2 are conditionally independent, given h and NOT h. Similarly, when we use the serial-combination function to combine CFs for the rules "if a then b" and "if b then c," we assume implicitly that a and c are conditionally independent, given b and NOT b.

With this understanding of the CF model, we can identify precisely the problem with the representation of Mr. Holmes' situation. There, we use serial combination to combine CFs for the sequence of propositions EARTHQUAKE, ALARM, and BURGLARY. In doing so, we make the inaccurate assumption (among others) that EARTHQUAKE and BURGLARY are conditionally independent, given ALARM. No matter how we manipulate the arcs in the inference network of Figure 2, we generate inaccurate assumptions of conditional independence.

The assumptions of independence imposed by the CF model are not satisfied by most real-world domains. Moreover, the assumptions of the parallel-combination function are stronger than are those of the idiot-Bayes model. That is, when we use the idiot-Bayes model, we assume that evidence is conditionally independent given each hypothesis. When we use the parallel-combination function, however, we assume that evidence is conditionally independent given each hypothesis and the *negation* of each hypothesis. Unless the space of hypotheses consists of a single proposition and the negation of that proposition, the parallel-combination assumptions are essentially impossible to satisfy, even when the idiot-Bayes assumptions are satisfied (Johnson, 1986).

For example, let us consider the task of identifying an unknown aircraft. Let us suppose that the aircraft could be any type of commercial or military airplane. Further, let us suppose that we have clues to the identity of the aircraft such as the airspeed, the fuselage size, and the distribution of the plane's heat plume. It may be reasonable to assume that the clues are conditionally independent, given each possible aircraft type. Under this idiot-Bayes assumption, however, the clues cannot be conditionally independent, given each aircraft type and the negation of each aircraft type.

4.4 A Fundamental Difference

We can understand the problems with the CF model at a more intuitive level. Logical relationships represent what we can observe directly. In contrast, uncertain relationships encode invisible influences: exceptions to that which is visible. For example, a burglary will not always trigger an alarm, because there are hidden mechanisms that may inhibit the sounding of the alarm. We summarize these hidden mechanisms in a probability for ALARM given BURGLARY. In the process of summarization, we lose information. Therefore, when we try to combine uncertain information, unexpected (nonmodular) interactions may occur. We should not expect that the CF model—or any modular belief updating scheme—will be able to handle such subtle interactions. Pearl (1988, Chapter 1) provides a detailed discussion of this point.

5 A Practical Problem with the CF Model

In addition to the theoretical difficulties of updating beliefs within the CF model, the model contains a serious practical problem. Specifically, the CF model requires that we encode rules in the direction in which they are used. That is, an inference network must trace a trail of rules from observable evidence to hypotheses.

Unfortunately, we often do not use rules in the same direction in which experts can most accurately and most comfortably assess them. Kahneman and Tversky have shown that people are usually most comfortable when they assess predictive rules—that is, rules of the form

if CAUSE then EFFECT

For example, expert physicians prefer to assess the likelihood of a symptom, given a disease, rather than the likelihood (or belief update) of a disease, given a symptom (Tversky and Kahneman, 1982). Henrion attributes this phenomenon to the nature of causality. In particular, he notes that a predictive probability (the likelihood of a symptom, given a disease) reflects a stable property of that disease. In contrast, a diagnostic probability (the likelihood of a disease, given a symptom) depends on the incidence rates of that disease and of other diseases that may cause the manifestation. Thus, predictive probabilities are a more useful and parsimonious way to represent uncertain relationships—at least in medical domains (see Horvitz et al., 1988, pages 252–3).

Unfortunately for the CF model, effects are usually the observable pieces of evidence, and causes are the sought-after hypotheses. Thus, we usually force experts to construct rules of the form

if EFFECT then CAUSE

Consequently, in using the CF model, we force experts to provide judgments of uncertainty in a direction that makes them uncomfortable. We thereby promote errors in assessment. In the next section, we examine the belief network, a representation that allows an expert to represent knowledge in whatever direction she prefers.

6 Belief Networks: A Language of Dependencies

The examples in this paper illustrate that we need a language that helps us to keep track of the sources of our belief, and that makes it easy for us to represent or infer the propositions on which each of our beliefs are dependent. The belief network is such a language.[3] Several researchers independently developed the representation—for example, Wright (1921), Good (1961), and Rousseau (1968). Howard and Matheson (1981) developed the influence diagram, a generalization of the belief network in which we can represent decisions and the preferences of a decision maker. Probabilistic Networks (*qv*) and Influence Diagrams (*qv*), respectively, contain detailed descriptions of these representations.

Figure 4 shows a belief network for Mr. Holmes' situation. The belief network is a directed acyclic graph. The nodes in the graph correspond to uncertain variables relevant to the problem. For Mr. Holmes, each uncertain variable represents a proposition and

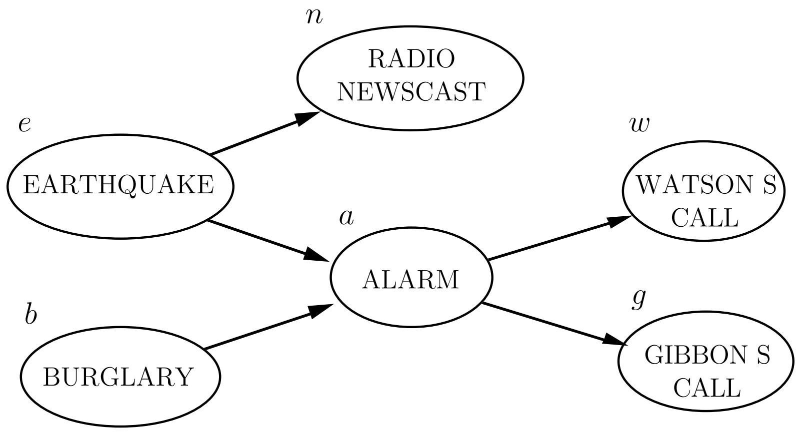
[3]Other names for belief networks include probabilistic networks, causal networks, and Bayesian networks.

that proposition's negation. For example, node b in Figure 4 represents the propositions BURGLARY and NOT BURGLARY (denoted b_+ and b_-, respectively). In general, an uncertain variable can represent an arbitrary set of mutually exclusive and exhaustive propositions; we call each proposition an *instance* of the variable. In the remainder of the discussion, we make no distinction between the variable x and the node x that represents that variable.

Each variable in a belief network is associated with a set of probability distributions. (A probability distribution is an assignment of a probability to each instance of a variable.) In the Bayesian tradition, these distributions encode the knowledge provider's beliefs about the relationships among the variables. Mr. Holmes' probabilities appear below the belief network in Figure 4.

The arcs in the directed acyclic graph represent direct probabilistic dependencies among the uncertain variables. In particular, an arc from node x to node y reflects an assertion by the builder of that network that the probability distribution for y may depend on the instance of the variable x; we say that x *conditions* y. Thus, a node has a probability distribution for every instance of its conditioning nodes. (An instance of a set of nodes is an assignment of an instance to each node in that set.) For example, in Figure 4, ALARM is conditioned by both EARTHQUAKE and BURGLARY. Consequently, there are four probability distributions for ALARM, corresponding to the instances where both EARTHQUAKE and BURGLARY occur, BURGLARY occurs alone, EARTHQUAKE occurs alone, and neither EARTHQUAKE nor BURGLARY occurs. In contrast, RADIO NEWSCAST, WATSON'S CALL, and GIBBON'S CALL are each conditioned by only one node. Thus, there are two probability distributions for each of these nodes. Finally, EARTHQUAKE and BURGLARY do not have any conditioning nodes, and hence each node has only one probability distribution.

The lack of arcs in a belief network reflects assertions of conditional independence. For example, there is no arc from BURGLARY to WATSON'S CALL in Figure 4. The lack of this arc encodes Mr. Holmes' belief that the probability of receiving Watson's telephone call from his neighbor does not depend on whether or not there was a burglary, provided Mr. Holmes knows whether or not the alarm sounded.



$$p(b{+} \mid \xi) = 0.0001$$

$$p(e{+} \mid \xi) = 0.0003$$

$$p(w{+} \mid a{-}, \xi) = 0.4$$
$$p(w{+} \mid a{+}, \xi) = 0.8$$

$$p(g{+} \mid a{-}, \xi) = 0.04$$
$$p(g{+} \mid a{+}, \xi) = 0.4$$

$$p(n{+} \mid e{-}, \xi) = 0.0002$$
$$p(n{+} \mid e{+}, \xi) = 0.9$$

$$p(a{+} \mid b{-}, e{-}, \xi) = 0.01$$
$$p(a{+} \mid b{+}, e{-}, \xi) = 0.95$$
$$p(a{+} \mid b{-}, e{+}, \xi) = 0.2$$
$$p(a{+} \mid b{+}, e{+}, \xi) = 0.96$$

Figure 4: A belief network for Mr. Holmes' situation.

The nodes in the belief network represent the uncertain variables relevant to Mr. Holmes' situation. The arcs represent direct probabilistic dependencies among the variables, whereas the lack of arcs between nodes represents assertions of conditional independence. Each node in the belief network is associated with a set of probability distributions. These distributions appear below the graph. The variables in the probabilistic expressions correspond to the nodes that they label in the belief network. For example, $p\,(b_{+}|\xi)$ denotes the probability that a burglary has occurred, given Mr. Holmes' background information, ξ. The figure does not display the probabilities that the events failed to occur. We can compute these probabilities by subtracting from 1.0 the probabilities shown.

In Probabilistic Networks (*qv*), Geiger describes the exact semantics of missing arcs. Here, it is important to recognize, that, given any belief network, we can construct the joint probability distribution for the variables in any belief network from (1) the probability distributions associated with each node in the network, and (2) the assertions of conditional independence reflected by the lack of some arcs in the network. The *joint probability distribution* for a set of variables is the collection of probabilities for each instance of that set. The distribution for Mr. Holmes' situation is

$$p(e,b,a,n,w,g|\xi) = p(e|\xi)\;\; p(b|\xi)\;\; p(a|e,b,\xi)\;\; p(n|e,\xi)\;\; p(w|a,\xi)\;\; p(g|a,\xi) \tag{7}$$

The probability distributions on the right-hand side of Equation 7 are exactly those distributions associated with the nodes in the belief network.

6.1 Getting Answers from Belief Networks

Given a joint probability distribution over a set of variables, we can compute any conditional probability that involves those variables. In particular, we can compute the probability of any set of hypotheses, given any set of observations. For example, Mr. Holmes undoubtedly wants to determine the probability of BURGLARY (b_+) given RADIO NEWSCAST (n_+) and WATSON'S CALL (w_+) and GIBBON'S CALL (g_+). Applying the axioms of probability (Bayesian Inference Methods, *qv*) to the joint probability distribution for Mr. Holmes' situation, we obtain

$$\begin{aligned} p(b_+|n_+,w_+,g_+,\xi) &= \frac{p(b_+,n_+,w_+,g_+|\xi)}{p(n_+,w_+,g_+|\xi)} \\ &= \frac{\sum_{e_i,a_k}\; p(e_i,b_+,a_k,n_+,w_+,g_+|\xi)}{\sum_{e_i,b_j,a_k}\; p(e_i,b_j,a_k,n_+,w_+,g_+|\xi)} \end{aligned}$$

where e_i, b_j, and a_k denote arbitrary instances of the variables e, b, and a, respectively.

In general, given a belief network, we can compute any set of probabilities from the joint distribution implied by that network. We also can compute probabilities of interest directly within a belief network. In doing so, we can take advantage of the assertions of conditional independence reflected by the lack of arcs in the network: Fewer arcs lead to less computation. Probabilistic Networks (*qv*) contains a description of several algorithms that use conditional independence to speed up inference.

6.2 Belief Networks for Knowledge Acquisition

A belief network simplifies knowledge acquisition by exploiting a fundamental observation about the ability of people to assess probabilities. Namely, a belief network takes advantage of the fact that people can make assertions of conditional independence much more easily than they can assess numerical probabilities (Howard and Matheson, 1981; Pearl, 1986). In using a belief network, a person first builds the graph that reflects his assertions of conditional independence; only then does he assess the probabilities underlying the graph. Thus, a belief network helps a person to *decompose* the construction of a joint probability distribution into the construction of a set of smaller probability distributions.

6.3 Advantages of the Belief Network over the CF Model

The example of Mr. Holmes' illustrates the advantages of the belief network over the CF model. First, we can avoid the practical problem of the CF model that we discussed in Section 5; namely, using a belief network, a knowledge provider can choose the order in which he prefers to assess probability distributions. For example, in Figure 4, all arcs point from cause to effect, showing that Mr. Holmes prefers to assess the probability of observing an effect, given one or more causes. If, however, Mr. Holmes wanted to specify the probabilities of—say—EARTHQUAKE given RADIO NEWSCAST and of EARTHQUAKE given NOT RADIO NEWSCAST, he simply would reverse the arc from RADIO NEWSCAST to EARTHQUAKE in Figure 4. Regardless of the direction in which Mr. Holmes assesses the conditional distributions, we can use any of the available belief-network algorithms to compute the conditional probabilities of interest, if the need arises. (See Shachter and Heckerman, 1987, for a detailed discussion of this point.)

Second, using a belief network, the knowledge provider can *control* the assertions of conditional independence that are encoded in the representation. In contrast, the use of the combination functions in the CF model forces a person to adopt assertions of conditional independence that may be incorrect. For example, as we discussed in Section 4.3, the inference network in Figure 2 dictates the erroneous assertion that EARTHQUAKE and BURGLARY are conditionally independent, given ALARM.

Third, and most important, a knowledge provider does not have to assess indirect independencies, using a belief network. Such independencies reveal themselves in the course of probabilistic computations within the network.[4] Such computations can tell us—for example—that BURGLARY and RADIO NEWSCAST are normally independent, but become dependent, given WATSON'S CALL, GIBBON'S CALL, or both.

Thus, the belief network helps us to tame the inherently nonmodular properties of uncertain reasoning. Uncertain knowledge encoded in a belief network is not as modular as is knowledge about logical relationships. Nonetheless, representing uncertain knowledge in a belief network is a great improvement over encoding all relationships among a set of variables.

References

Geiger, D., Verma, T., and Pearl, J. (1990). Indentifying independence in Bayesian networks. *Networks*, 20.

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

Heckerman, D. (1986). Probabilistic interpretations for MYCIN's certainty factors. In Kanal, L. and Lemmer, J., editors, *Uncertainty in Artificial Intelligence*, pages 167–196. North-Holland, New York.

[4] In fact, we are not even required to perform numerical computations to derive such indirect independencies. An efficient algorithm exists that uses only the structure of the belief network to tell us about these dependencies (Geiger et al., 1990).

Heckerman, D. and Horvitz, E. (1987). On the expressiveness of rule-based systems for reasoning under uncertainty. In *Proceedings AAAI-87 Sixth National Conference on Artificial Intelligence,* Seattle, WA, pages 121–126. Morgan Kaufmann, San Mateo, CA.

Heckerman, D. and Horvitz, E. (1988). The myth of modularity in rule-based systems. In Lemmer, J. and Kanal, L., editors, *Uncertainty in Artificial Intelligence 2*, pages 23–34. North-Holland, New York.

Henrion, M. (1986). Should we use probability in uncertain inference systems? In *Proceedings of the Cognitive Science Society Meeting,* Amherst, PA. Carnegie–Mellon.

Henrion, M. (1987). Uncertainty in artificial intelligence: Is probability epistemologically and heuristically adequate? In Mumpower, J., editor, *Expert Judgment and Expert Systems*, pages 105–130. Springer-Verlag, Berlin, Heidelberg.

Horvitz, E., Breese, J., and Henrion, M. (1988). Decision theory in expert systems and artificial intelligence. *International Journal of Approximate Reasoning*, 2:247–302.

Howard, R. and Matheson, J. (1981). Influence diagrams. In Howard, R. and Matheson, J., editors, *Readings on the Principles and Applications of Decision Analysis*, volume II, pages 721–762. Strategic Decisions Group, Menlo Park, CA.

Johnson, R. (1986). Independence and Bayesian updating methods. In Kanal, L. and Lemmer, J., editors, *Uncertainty in Artificial Intelligence*, pages 197–202. North-Holland, New York.

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

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

Rousseau, W. (1968). A method for computing probabilities in complex situations. Technical Report 6252-2, Center for Systems Research, Stanford University, Stanford, CA.

Shachter, R. and Heckerman, D. (1987). Thinking backward for knowledge acquisition. *AI Magazine*, 8:55–63.

Shortliffe, E. and Buchanan, B. (1975). A model of inexact reasoning in medicine. *Mathematical Biosciences*, 23:351–379.

Tversky, A. and Kahneman, D. (1982). Causal schemata in judgments under uncertainty. In Kahneman, D., Slovic, P., and Tversky, A., editors, *Judgement Under Uncertainty: Heuristics and Biases*. Cambridge University Press, New York.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20:557–85.