A pragmatical view on the certainty factor model

Linda van der Gaag

RUU-CS-90-42
December 1990

Utrecht University
Department of Computer Science
Padualaan 14, P.O. Box 80.089,
3508 TB Utrecht, The Netherlands,
Tel. : ... + 31 - 30 - 531454

ISSN:0924-3275

A Pragmatical View on the Certainty Factor Model

Linda van der Gaag
Utrecht University, Department of Computer Science
P.O. Box 80.089, 3508 TB Utrecht, The Netherlands

Abstract

Since its introduction in the 1970s, the certainty factor model has enjoyed widespread use in rule-based expert systems. Many present-day, commercially available expert system shells offer the model for modelling and reasoning with uncertain information. Among researchers in the field of plausible reasoning it is widely known that the model is mathematically incorrect: many have investigated the probabilistic foundation of the model. The detailed results of these theoretical analyses of the model, however, are relatively unknown to engineers interested in applying or developing an expert system. In this paper, we take a pragmatical look at the theoretical results and use them to formulate some guidelines for the representation of expert knowledge in production rules which, if adhered to, allow the certainty factor model to behave satisfactorily from a probabilistic point of view.

1 Introduction

When building expert systems it becomes evident that in many real-life domains expert knowledge is not precisely defined, but instead is of an imprecise nature. Yet, human experts typically are able to form judgements and take decisions from uncertain, incomplete and sometimes even contradictory information. In order to be useful in an environment in which only such deficient information is available, an expert system has to capture and exploit not only the highly-specialized expert knowledge, but the uncertainties that go with the represented pieces of information as well. Researchers in artificial intelligence therefore have sought methods for representing uncertainty and have developed reasoning procedures for manipulating uncertain information.

The early diagnostic expert systems of the 1970s mostly used production rules as a formalism for representing expert knowledge in a modular form and employed a heuristic reasoning method for applying the rules, yielding an 'intelligent' problem-solving behaviour by concentrating only on those hypotheses that were actually suggested by the evidence. In these systems, the production rules typically were used in selectively gathering evidence and heuristically pruning the search space of possible diagnoses. We have mentioned that, to be useful for real-life applications, these originally deterministic rule-based systems had to be extended with some notion of uncertainty. As probability theory is one of the oldest mathematical theories concerning uncertainty, it is no wonder that this formal theory was chosen as the first point of departure in the pioneering work on automated reasoning with uncertainty. However, it soon became evident that probability theory could not be applied in a rule-based setting in a straightforward manner: in a rule-based system, an expert typically is asked to associate probabilities only with the rules he has provided, that is, only a partial specification of a probability distribution is given. For computing probabilities from such a partial specification for all (intermediate) results derived from applying the production rules, however, probability theory does not provide explicit computation rules. To overcome this problem, in the 1970s several modifications of probability theory for efficient application in a rule-based environment were developed, the most well-known of which is the *certainty factor model* that was originally designed for dealing with uncertainty in the MYCIN system, [Shortliffe84]. This model is not well-founded from a mathematical point of view: it offers computation rules which do not always accord with the axioms of probability theory but which render the model to some extent insensitive to partial specification and inconsistency of a probability distribution. Even though it is widely known that the certainty factor model is mathematically flawed, it has since its introduction enjoyed wide-spread use in rule-based systems built after MYCIN. The relative success can be accounted for by its computational simplicity; furthermore, the certainty factors used in the model are intuitively appealing and easy to handle, [Shultz90].

Now, the last few years research on plausible reasoning has progressed considerably: several (mathematically correct) probabilistic models have been proposed based on so-called *belief networks*, see for example [Pearl88]. Informally speaking, a belief network is a graphical representation of a problem domain consisting of the statistical variables discerned in the domain and their probabilistic interrelationships; these relationships are quantified by means of 'local' probabilities. So, these models employ a knowledge representation scheme other than the production rule formalism; furthermore, they require a fully and consistently specified probability distribution. From a knowledge acquisition point of view, these belief network models are much more demanding that the 'rule-based' certainty factor model; in addition, they are not supported as yet by engineering methodologies. The certainty factor model therefore is still employed frequently in present-day rule-based expert systems. Its frequent use motivated a pragmatical look at the certainty factor model and its foundation in probability theory. This enabled us to formulate some guidelines for the representation of expert knowledge in production rules allowing the model to behave satisfactorily. After briefly reviewing the certainty factor model in Section 2 and its probabilistic foundation in Section 3, we will state these guidelines in Section 4.

2 The Certainty Factor Model Revisited

Although we assume that the reader is acquainted with production rules and top-down inference, we start with a brief description of these notions in order to introduce some terminology. For a more elaborate introduction, the reader is referred to for example [Jackson90] or [Lucas91]. In a rule-based, top-down reasoning expert system applying the certainty factor model, three major components are discerned:

- *production rules* and associated *certainty factors*. Basically, an expert in the domain in which the expert system is to be used models his knowledge of the field in a set of production rules of the form $e \rightarrow h$. The left-hand side e of such a production rule is a Boolean combination of conditions; e as well as its constituting parts will be called *(pieces of) evidence*. In general, the right-hand side h of a production rule is a conjunction of conclusions; however, for ease of exposition we will restrict ourselves to single-conclusion production rules. From now on, a conclusion will be called a *hypothesis*. A production rule has the following meaning: if evidence e has been observed, then the hypothesis h is true.

An expert associates with the hypothesis h of each production rule $e \rightarrow h$ a (real) number $CF(h, e, e \rightarrow h)$, quantifying the degree to which the actual observation of evidence e confirms the hypothesis h. The values $CF(x, y, z)$ of the (partial) function CF are called *certainty factors*; $CF(x, y, z)$ should be read as 'the certainty factor of x, given y and the derivation z of x from y'. (Note that in [Shortliffe84] the developers of the model, E.H. Shortliffe and B.G. Buchanan, use for certainty factors the two-argument notation $CF(h, e)$. For syntactical as well as semantical reasons, we considered it necessary to introduce the notion of a derivation in the notational convention; for our motivation for doing so, the reader is referred to [Gaag89]). Certainty factors range from -1 to $+1$. A certainty factor greater than zero is associated with a hypothesis h given some evidence e if the hypothesis h is confirmed to some degree by the observation of e; a negative certainty factor is suggested if the observation of e disconfirms the hypothesis h. A certainty factor equal to zero is suggested by the expert if the observation of evidence e does not influence the confidence in h.

- *user-supplied data* and associated certainty factors. During a consultation of the expert system, the user is asked to supply actual case data. The user attaches a certainty factor $CF(e, u, u \rightarrow e)$ to every piece of evidence he supplies the system with (u is taken to represent the user's de facto knowledge).
- a (*top-down*) *inference engine* and a (*bottom-up*) *scheme for propagating uncertainty*. Top-down inference is a well-known goal-directed reasoning technique in which the production rules are applied exhaustively to prove one or more goal hypotheses. Due to the application of production rules, several intermediate hypotheses will be confirmed or disconfirmed to some degree during the inference process. The certainty factor to be associated with such an intermediate hypothesis h is calculated from the certainty factors associated with the production rules used in deriving h. For the purpose of thus propagating uncertainty, several functions for combining certainty factors are defined.

From now on we will focus on the scheme for propagating uncertainty. For those readers who are already familiar with the certainty factor model it is noted that in the sequel we abstract from several pragmatical issues added to the model such as for example the discontinuity of the evaluation of the left-hand side of a production rule (that is, the 0.2 threshold).

As has been mentioned before, an expert associates a function value $CF(h, e, e \rightarrow h)$ with the conclusion h of a production rule $e \rightarrow h$. Recall that this value expresses the degree to which the actual occurrence of evidence e influences the confidence in the hypothesis h. When using the production rules, however, the evidence e used may be an intermediate hypothesis that has been confirmed to some degree $CF(e, u, D^{u,e})$ not necessarily equalling $+1$; $D^{u,e}$ is some derivation of e from the user's knowledge u with respect to a given set of production rules. That is, it may be the case that the truth of the evidence e is not known with certainty. After application of the rule $e \rightarrow h$ described above, the actual certainty factor for h is computed by means of

$$CF(h, u, D^{u,e} \circ (e \rightarrow h)) = CF(h, e, e \rightarrow h) \cdot max\{0, CF(e, u, D^{u,e})\}$$

where $D^{u,e} \circ (e \rightarrow h)$ denotes the sequential composition of the derivations $D^{u,e}$ of e from u and $e \rightarrow h$ of h from e. This computation rule is called the *combination function for uncertain evidence.*

Now recall that the evidence e in a production rule $e \rightarrow h$ in general is a Boolean combination of atomic pieces of evidence. Before the combination function for uncertain evidence can be applied, a certainty factor for the entire combination of evidence e has to be known. This certainty factor is computed from the separate certainty factors for each of the atomic pieces of evidence e comprises, using

$$CF(e_1 \wedge e_2, u, D^{u,e_1} \,\&\, D^{u,e_2}) = min\{CF(e_1, u, D^{u,e_1}), CF(e_2, u, D^{u,e_2})\}$$

where $D^{u,e_1} \,\&\, D^{u,e_2}$ denotes the conjunction of the two derivations D^{u,e_1} and D^{u,e_2}, and

$$CF(e_1 \vee e_2, u, D^{u,e_1} \mid D^{u,e_2}) = max\{CF(e_1, u, D^{u,e_1}), CF(e_2, u, D^{u,e_2})\}$$

where $D^{u,e_1} \mid D^{u,e_2}$ denotes the disjunction of the two derivations D^{u,e_1} and D^{u,e_2}. These functions are called the *combination functions for composite hypotheses.*

When different successful production rules $e_i \rightarrow h$, $i > 1$, (that is, rules with different left-hand sides e_i), conclude on the same hypothesis h, a certainty factor $CF(h, u, D^{u,e_i} \circ (e_i \rightarrow h))$ is derived from the application of each of these rules. The net certainty factor for h is computed using

$$CF(h, u, D_1^{u,h} \parallel D_2^{u,h}) = \begin{cases} CF(h,u,D_1^{u,h}) + CF(h,u,D_2^{u,h})(1 - CF(h,u,D_1^{u,h})), \\ \text{if } CF(h,u,D_i^{u,h}) > 0,\ i = 1,2 \\ \\ \frac{CF(h,u,D_1^{u,h}) + CF(h,u,D_2^{u,h})}{1 - min\{|CF(h,u,D_1^{u,h})|, |CF(h,u,D_2^{u,h})|\}}, \\ \text{if } -1 < CF(h,u,D_1^{u,h}) \cdot CF(h,u,D_2^{u,h}) \le 0 \\ \\ CF(h,u,D_1^{u,h}) + CF(h,u,D_2^{u,h})(1 + CF(h,u,D_1^{u,h})), \\ \text{if } CF(h,u,D_i^{u,h}) < 0,\ i = 1,2 \end{cases}$$

where $D_1^{u,h} \parallel D_2^{u,h}$ denotes the parallel composition of the separate derivations $D_1^{u,h}$ and $D_2^{u,h}$. This computation rule is called the *combination function for co-concluding production rules.*

3 The Probabilistic Foundation of the Model

In [Shortliffe84], E.H. Shortliffe and B.G. Buchanan have suggested a mathematical foundation for their model in probability theory. The certainty factor function we have introduced in the preceding section is not the basic notion of uncertainty employed in the certainty factor model: this function is defined in terms of two basic measures of uncertainty, the measures of belief and disbelief, which in turn are defined in terms of probability theory. The *measure of belief* MB is the three-argument function defined by

$$MB(h, e, D^{e,h}) = \begin{cases} 1 & \text{if } Pr(h) = 1 \\ max\{0, \frac{Pr(h \mid e \wedge D^{e,h}) - Pr(h)}{1 - Pr(h)}\} & \text{otherwise} \end{cases}$$

expressing the degree to which the observation of evidence e increases the belief in the hypothesis h. It is noted that we assume an appropriate probabilistic interpretation of derivations; for details of the interpretation chosen, the reader is referred to [Gaag90]. The *measure of disbelief* MD is the function defined by

$$MD(h,e,D^{e,h}) = \begin{cases} 1 & \text{if } Pr(h) = 0 \\ max\{0, \frac{Pr(h) - Pr(h \mid e \wedge D^{e,h})}{Pr(h)}\} & \text{otherwise} \end{cases}$$

expressing the degree to which the observation of evidence e increases the disbelief in the hypothesis h. The certainty factor function CF then is defined in terms of these measures of belief and disbelief

$$CF(h,e,D^{e,h}) = \frac{MB(h,e,D^{e,h}) - MD(h,e,D^{e,h})}{1 - min\{MB(h,e,D^{e,h}), MD(h,e,D^{e,h})\}}$$

as mentioned.

This probabilistic foundation of the certainty factor model motivated many researchers to analyse the relation between the model and probability theory, see for example [Adams84], [Wise86] and [Gaag90]; in some cases such an analysis led to the formulation of counter-proposals for some parts of the model, as in [Heckerman86]. These analyses taken together show that the model is mathematically flawed and, what's more, that it is not possible to devise a mathematically sound probabilistic model based on the principles underlying the certainty factor model and similar models developed for rule-based expert systems in the 1970s.

We have mentioned before that the certainty factor model has been incorporated as a special feature in many present-day, commercially available rule-based expert system shells and therefore is likely to be applied to any type of domain. This observation motivated a closer look at the analyses of the probabilistic foundation of the model, this time with the aim of identifying conditions under which the model will behave satisfactorily from a probabilistic point of view. In this section, we will report the results of our analysis of the model; for full details the reader is once more referred to [Gaag90]. In Section 4 we will translate these results into guidelines for knowledge representation in production rules.

We begin by looking at the combination function for co-concluding production rules. Given two derivations $D_1^{u,h}$ and $D_2^{u,h}$ with respect to a given set of production rules of a hypothesis h, we can discern three possibilities for their relationship with the belief in h:

- both $D_1^{u,h}$ and $D_2^{u,h}$ do not increase the disbelief in h, that is, $CF(h,u,D_1^{u,h}) \geq 0$ as well as $CF(h,u,D_2^{u,h}) \geq 0$,
- both $D_1^{u,h}$ and $D_2^{u,h}$ do not increase the belief in h, that is, $CF(h,u,D_1^{u,h}) \leq 0$ as well as $CF(h,u,D_2^{u,h}) \leq 0$,
- one of $D_1^{u,h}$ and $D_2^{u,h}$ increases the disbelief in h while the other one increases the belief in h, that is, either $CF(h,u,D_1^{u,h}) > 0$ and $CF(h,u,D_2^{u,h}) < 0$, or $CF(h,u,D_1^{u,h}) < 0$ and $CF(h,u,D_2^{u,h}) > 0$.

Now suppose that the two certainty factors $CF(h,u,D_1^{u,h})$ and $CF(h,u,D_2^{u,h})$ yielded by the derivations $D_1^{u,h}$ and $D_2^{u,h}$, respectively, are combined into a net certainty factor $CF(h,u,D_1^{u,h} \parallel D_2^{u,h})$ by means of the combination function for co-concluding production rules. In the first case mentioned above, that is, in case both $D_1^{u,h}$ and $D_2^{u,h}$ do not increase the disbelief in h, the result yielded is consistent with the probabilistic foundation of the model if the two derivations are mutually independent and conditionally independent given the negation of the hypothesis. In case both $D_1^{u,h}$ and $D_2^{u,h}$ do not increase the belief in h, the result yielded by the combination function is correct if the two derivations are mutually independent and conditionally independent given the hypothesis. In the case of 'conflicting' derivations, however, the combination function for co-concluding production rules cannot be shown to respect the probabilistic foundation of the model.

Several authors have analysed the combination function for combining the certainty factors yielded by co-concluding production rules. The other combination functions have received far less attention in the literature. We feel, however, that these combination functions may also have a considerable impact on the computed certainty factors. Practical experience in using the certainty factor model has learned for example that the combination functions for composite hypotheses are applied at least as often as the function for co-concluding production rules. An analysis of the functions for composite hypotheses unfortunately has not enabled us to formulate 'natural' conditions under which these functions can be shown to be correct with respect to the probabilistic foundation of the model. Our analysis, however, suggested that for two hypotheses e_1 and e_2 and their respective derivations D^{u,e_1} and D^{u,e_2} relative to a given set of production rules, the certainty factor $CF(e_1 \wedge e_2, u, D^{u,e_1} \,\&\, D^{u,e_2})$ yielded by the combination function for a conjunction of hyptheses is a satisfactory approximation of the actual certainty factor of $e_1 \wedge e_2$ if e_1 and e_2 are strongly correlated; a similar result holds for the combination function for a disjunction of hypotheses.

To conclude this section on the probabilistic foundation of the certainty factor model, we consider a production rule $e \rightarrow h$ and a derivation $D^{u,e}$ of the evidence e with respect to a given set of production rules, and investigate the combination function for uncertain evidence. Our analysis of this function has shown that the certainty factor $CF(h, u, D^{u,e} \circ (e \rightarrow h))$ yielded by this combination function respects the probabilistic definition of the certainty factor function if, among other less relevant conditions, we have $h \wedge e = h$.

4 Guidelines for Using the Model

In the previous section we have reported some results of an in-depth study of the certainty factor model and its probabilistic foundation. Here, we exploit these results and derive from them some guidelines for knowledge representation that will allow the certainty factor model to perform satisfactorily.

Once more we begin by looking at the combination function for co-concluding production rules. E.H. Shortliffe and B.G. Buchanan themselves have experimented with the model in the context of the MYCIN system using sampling data simulating several hundreds of patients, to compare the computed certainty factors with the correct probabilistic values, see [Shortliffe84]. In this experiment, they focussed on the function for co-concluding production rules. They observed that in most of the cases, the computed certainty factor did not differ radically from the theoretical probabilistic value. However, they have observed that the more the combination function is applied for a given hypothesis, the more the computed value tends to deviate from the theoretical one. Furthermore, their experiment showed that the most erroneous values arose from cases in which the different derivations of the hypothesis under consideration were strongly interrelated. We add to these observations that since the vast majority of the production rules of the MYCIN system contained positive certainty factors, the experiment of Shortliffe and Buchanan cannot have reflected the impact of conflicting derivations. Note that these observations are consistent with the ones we have found by a theoretical analysis of the model. We conclude that in representing domain-dependent problem-solving knowledge in production rules, it is advisory

- to specify the condition parts of production rules drawing opposite conclusions as 'mutually exclusive' as possible in order to minimize the occurrence of conflicting derivations for a single hypothesis, and
- in case several pieces of evidence pertain to a single hypothesis, to group these pieces of evidence in such a way that the Boolean combinations of evidence mentioned in separate production rules are as 'independent' as possible (in the sense described in the previous section) and the atomic pieces of such a Boolean combination of evidence within a production rule are as strongly correlated as possible.

Note that our observations concerning the combination functions for composite hypotheses are consistent with the latter guideline.

Our last observation concerns the combination function for propagating uncertain evidence. Informally speaking, the condition mentioned in the previous section shows that this function is correct in case the expert system is only able to narrow its focus and does not have the ability to turn to hypotheses slightly outside the scope of the derivation up till that moment. It is advisory therefore to specify production rules in such a way that a chain of rules that may arise when actually reasoning with the system, has this property of narrowing the focus of attention.

5 Summary and Conclusion

In this paper, we have considered the certainty factor model and its foundation in probability theory. Our aim has not been to present a detailed analysis showing the incorrectness of the model but merely to take a pragmatical look at it. Guided by an in-depth study of the probabilistic foundation of the model, we have formulated in Section 3 some conditions under which the model can be shown to be satisfactory from a probabilistic viewpoint. In Section 4 we have translated these conditions into guidelines for knowledge representation. If adhered to in practical applications, these guidelines will allow for the model to behave satisfactorily. We feel that as long as other, mathematically correct models for example based on belief networks, are too demanding either computationally or from an assessment point of view, the certainty factor model, if handled with proper care, is still a good alternative for dealing with uncertainty.

References

[Adams84] J.B. Adams (1984). Probabilistic reasoning and certainty factors, in: B.G. Buchanan, E.H. Shortliffe (eds). *Rule-Based Expert Systems. The MYCIN Experiments of the Stanford Heuristic Programming Project*, Addison-Wesley, Reading, Massachusetts, pp. 263 - 271.

[Gaag89] L.C. van der Gaag (1989). A conceptual model for inexact reasoning in rule-based systems, *International Journal of Approximate Reasoning*, vol. 3, no. 3, pp. 239 - 258.

[Gaag90] L.C. van der Gaag (1990). *Probability-Based Models for Plausible Reasoning*, Ph.D. thesis, University of Amsterdam.

[Heckerman86] D.E. Heckerman (1986). Probabilistic interpretations for MYCIN's certainty factors, in: L.N. Kanal, J.F. Lemmer (eds). *Uncertainty in Artificial Intelligence*, North-Holland, Amsterdam, pp. 167 - 196.

[Jackson90] P. Jackson (1990). *Introduction to Expert Systems*, 2nd edition, Addison-Wesley, Wokingham, England.

[Lucas91] P.J.F. Lucas, L.C. van der Gaag (1991). *Principles of Expert Systems*, Addison-Wesley, Wokingham, England, to appear.

[Pearl88] J. Pearl (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers, Palo Alto.

[Shortliffe84] E.H. Shortliffe, B.G. Buchanan (1984). A model for inexact reasoning in medicine, in: B.G. Buchanan, E.H. Shortliffe (eds). *Rule-Based Expert Systems. The MYCIN Experiments of the Stanford Heuristic Programming Project*, Addison-Wesley, Reading, Massachusetts, pp. 233 - 262.

[Shultz90] T.R. Shultz (1990). Managing uncertainty in rule based cognitive modeling, *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence*, pp. 21 - 26.

[Wise86] B.P. Wise, M. Henrion (1986). A framework for comparing uncertain inference systems to probability, in: L.N. Kanal, J.F. Lemmer (eds). *Uncertainty in Artificial Intelligence*, North-Holland, Amsterdam, pp. 69 - 83.