A New Uncertainty Calculus
For Rule-Based Expert Systems

Keith Wright, (E-mail: wrightm@uhd.edu), University of Houston, Downtown
Richard C. Hicks, (Email: rhicks@tamiu.edu), Texas A&M International University

ABSTRACT

The solution of non-deterministic expert systems consists of two components –the solution reached and a calculated measure of belief in each solution. This measure of belief is often the most critical factor in analyzing the solution. Unfortunately, as this paper reviews, the issue of how best to implement uncertainty calculi in expert systems has never been settled. Some popular rule-based approaches have in fact been shown to produce results no better than random guessing. To improve the accuracy of rule-based systems, we propose a new calculus we call gamma factors. This calculus combines ideas from two popular certainty factor calculi the product method, and the probability sum method.   It includes a tuning mechanism which the expert can use in a rule pre-processing step to compensate for dependent parallel evidence combination.

INTRODUCTION

$\mathcal{T}$he output of a rule-based expert system consists of two components – the solutions set and the belief in each solution, which is usually a single number called a probability, certainty, or confidence factor. The most common *normative* theories of uncertainty propagation in expert systems are *Bayesian Decision Theory BDT* [1] [3][5][6][10][19][20][25][33][34][35] the *Dempster Schaffer Theory* [7][9]28][29], and *Fuzzy logic* [8][24][30][36]. These theories are all based on rationality principles such as coherence, transitivity of choices, or the maximization of subjective utility.

Although there is a paucity of empirical research comparing the accuracy of implementations of normative theory-based systems, some basic conclusions are generally accepted [22].   For example Tong and Shapiro [31] showed the following for rule based systems:

- A rule-base can be molded to suit a particular calculus.
- Consultation accuracy decreases as the length of the inference chain increases.

Furthermore Wise & Henrion [32] and Henrion, et. al. [16] showed the following:

- Bayesian Network [23] calculi performed the best of the calculi tested, in the sense that the its worst performance was better than the worst performance of each of the other calculi.
- Fuzzy logic and *certainty factors* [29] calculate intermediate confidences as too extreme.
- *Certainty factor* calculi interpreted data as less diagnostic than it really was and responded in the wrong direction 25% of the time.
- The worst performances were in cases where items of evidence were strongly dependent for the hypothesis.

Our paper is organized as follows. First, we review the complexity of normative theory based non-deterministic expert system architecture, using Bayesian theory. This review reveals that normative applications today remain largely cost prohibitive. Next, we discuss an alternative descriptive approach we call the *gamma factor calculus*, which is an extension of the Certainty Factor approach pioneered in MYCIN. We will then show how the Gamma Factor can be used to compensate for dependent rule clusters [17] and to tune individual components of the rule base for more useful measures of belief.

NORMATIVE UNCERTAINTY MODELS FOR EXPERT SYSTEMS

Uncertain inputs in expert systems arise from three basic sources. First, the user may be uncertain of his factual knowledge. An example would be when a clinician thinks there is a 0.002 chance that any arbitrary patient has rheumatic heart disease. Second, the expert may be uncertain of the accuracy of a particular heuristic.   An example of this would be when a physician believes an adult patient has a 0.6 chance of having rheumatic heart disease, given a positive test for that disease. Third, the user or expert may be uncertain as to the trustworthiness of a particular piece of evidence. An example would be a physician's belief in the accuracy of particular medical test.

The network shown in Figure 1 represents a hypothetical medical diagnostic belief system. The upper nodes represent possible heart diseases. Coronary Artery Disease (*CAD*), is a condition that affects the supply of blood to the heart. Rheumatic heart disease (*RHD*) involves damage to the heart and heart vessels caused by rheumatic fever. *ChildRFev* represents whether or not rheumatic fever was present in childhood. *Test1* represents a specific diagnostic test indicating rheumatic fever was present in childhood. *Test2* represents another such test.
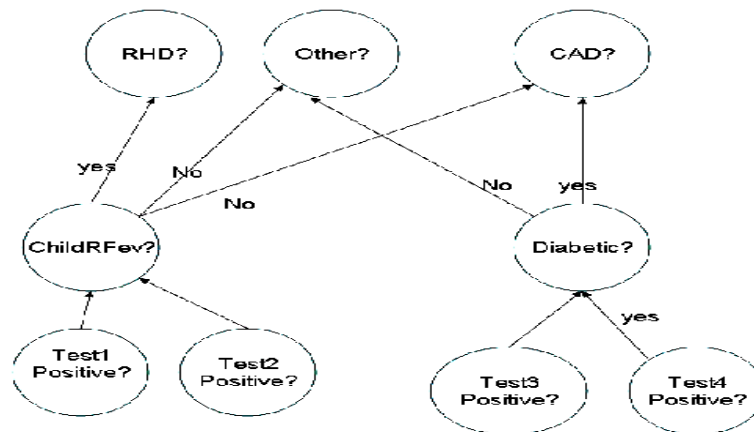

Figure 1

So we see that the diagnostic hypotheses that our example system strives to resolve are as follows:

H1: patient has Rheumatic Heart Disease; or
H2: patient has Coronary Artery Disease; or
H3: patient has neither disease.

Each node in the network represents a proposition that has an associated certainty. For example, the node childRRev represents the expert's subjective judgment that a person had rheumatic fever as a child. (The base rate of childhood rheumatic fever in the population would be an objective measure for this.) An intermediate node like *Diabetic* can play two roles: it is diagnostic evidence for the connected disease nodes above, and a cause of the connected symptom nodes below.   Each arc represents a belief dependence between the propositions. In rule-based expert systems these dependencies are represented as IF-THEN rules.

An expert system designer must choose a calculus that will cause the various certainty measures in the belief system to change in a reasonable way as new evidence is obtained. Such calculations must handle sequential evidence combination (i.e. effect of nodes *Test1* and *ChildRRev* on *RHD*), as well as parallel combination (i.e. the effect of *Test1* and *Test2* on *ChildRFev*). The calculus chosen must also handle updates to the rules and propositions.

Our paper uses a Bayesian framework to explain the issues of rule-based uncertainty calculi, because of compelling mathematical analysis supporting it as the most parsimonious theory of uncertainty propagation[5][6][21][25]. Using Bayesian terminology, each arc in Figure 1 represents a set of rules, each with an associated conditional probability. In our example one rule is, "If a rheumatic fever was present in childhood, then doctor is somewhat certain that patient has rheumatic heart disease." During the expert system consultation, rules fire, and certainty measurements must ultimately be assigned to one of the three hypotheses.

The complexity of normative uncertainty propagation depends on the amount of dependent rules in the rule base. An understanding of why this is so begins with an understanding of the Bayesian concept of conditional probability. It is given by given by equation (1) below,

$$p(H \mid E) = \frac{p(E \wedge H)}{p(E)} \qquad (1)$$

where $p(H \mid E)$ is the *posterior probability* of H, given evidence E. $p(E)$ is the total probability of E, an example of which would be the physician's belief that a certain medical test of any arbitrary person would be positive. The $p(E \wedge H)$ is the joint probability that both E and H are true. The definition of joint probability is given in equation (2)

$$p(E \wedge H) = p(E \mid H) * p(H) \qquad (2)$$

where $p(H)$ is the *prior probability* of H prior to evidence presented. An example would be a physician's belief about the base rate of a certain disease incidence in the general population. $p(E \mid H)$ is the conditional probability of E given H. An example would be a physician's belief about the probability that a medical test would be positive for a patient known to have the disease. Substituting (2) in to (1) we get the simple form of Bayes' theorem.

$$p(H \mid E) = p(H) * \frac{p(E \mid H)}{p(E)} \qquad (3)$$

The ratio $p(E \mid H) \ / \ p(E)$ is called the *likelihood ratio*. Bayesian theory prescribes that a person when presented with new evidence should update his prior belief such that his posterior belief equals his prior belief times the likelihood ratio.

Let's take a minute to see why it seems reasonable for a physician use Bayes' theorem to update his opinions as he receives evidence. First of all it seems intuitive that a physician has a *prior* opinion about the person's health, given no other specific medical information about that person. For example, let's suppose my physician thinks any arbitrary person has a .002 chance of having CAD. That's his belief about me *prior* to when I give him any other information. Furthermore, let's suppose that the physician believes that a 0.0025 fraction of the general population would test positive for CAD, regardless of whether or not the person actually has CAD. That is a measure of the test's accuracy, and a rational prior belief about my test result, before the test. Finally let's suppose that my physician believes that a 0.75 fraction of people who have CAD would test positive for it. That's another measure of the test's accuracy, and would be his conditional belief that a person who has CAD would test positive for it.

When my doctor sees my test result, Bayes' theorem prescribes that he re-evaluate his confidence that I have CAD, and in what direction, and by how much. Would it be rational for his posterior belief to elevate all the way to 0.75? No, because he believes the test to be somewhat in-accurate. The likelihood ratio (LR) prescribes what he should do. The LR thus summarizes the expert's belief in the accuracy of the test. The LR here is 0.75 / 0.0025 = 300, indicating moderate accuracy. So, since we have $p(H) = 0.002$, Bayes' theorem tells the physician to update his belief to 0.002 * 300 = 0.6, tempering his confidence by his low *prior* belief which was only 0.002. To drive his prior all the way to 1.0 would require a likelihood ratio of 500, the maximum theoretically possible in this example.

In rule based systems the LR can be applied recursively as rules fire.   However this can be exponentially complex in the number of hypotheses and evidence elements. To see why, lets explore the form of Bayes' theorem given in (4) below

$$p(H \mid E, e) = \frac{p(E \mid H, e) * p(H \mid e)}{p(E \mid e)} \tag{4}$$

where E denotes a new evidential proposition, and e the conjunctive *set* of old evidential propositions. For example suppose we have a rule with two conditions, each evaluating to true or false. Then e would be a set with four elements,

$$\{ (e_1 \wedge e_2), (e_1 \wedge \sim e_2), (\sim e_1 \wedge e_2), (\sim e_1 \wedge \sim e_2) \}.$$

Because of the numerator in equation (4), certainty propagation would require four elicitations from the expert, each corresponding to one element in the possible set of conjunctive conditions. In general, the number of certainty elicitations per rule in such a Bayesian system would be r^n, where r is the number of values for a condition and n is the number of conditions.   A rule with five ternary conditions effecting one conclusion could require $3^5 = 243$ elicitations from an expert. This complexity can be avoided only if most of the items of evidence are conditionally independent.

To see why, let us examine the corresponding formula for updating the negation of the hypothesis.

$$p(\sim H \mid E, e) = \frac{p(E \mid \sim E, e) * p(\sim H \mid e)}{p(E \mid e)} \tag{5}$$

Dividing (4) by (5) we get

$$\frac{p(H \mid E, e)}{p(\sim H \mid E, e)} = \frac{p(E \mid H, e)}{p(E \mid \sim E, e)} * \frac{p(H \mid e)}{p(\sim H \mid e)} \tag{6}$$

Viewing (6) as three ratios, we see that the ratio on the left is the posterior odds on H, and the one on the right is the prior odds on H. The ratio in the middle is the likelihood ratio.
Converting to odds, we can rewrite (6) as

$$O(H \mid E, e) = \frac{P(E \mid H, e)}{P(E \mid \sim H, e)} * O(H \mid e) \tag{7}$$

Now let's see what happens to the likelihood ratio when we can assume independent evidence. Suppose that, if you already knew which of H or ˜H was true, the old evidence (e) would not influence your belief that future evidence E will be present. That is:

$$P(E \mid H, e) = p(E \mid H) \tag{8}$$

$$P(E \mid \sim H, e) = p(E \mid \sim H) \tag{9}$$

Substituting the right hand sides of (8) and (9) into (7), the e drops out of the likelihood ratio and we get:

$$O(H \mid E, e) = \frac{p(E \mid H)}{P(E \mid \sim H)} * O(H \mid e) \qquad (10)$$

If (10) holds in a domain, the effect of the old evidence on the likelihood ratio can be ignored, the number of elicitations needed per rule reduces to two, and likelihood ratio uncertainty propagation is not computationally explosive. But we see that (10) is a very restrictive assumption, and is not robust across domains. For (10) to hold, a new item of evidence must be *conditionally independent* of the old items of evidence. In lay terms, conditional independence means two things. First it means that new evidence (E) is independent of the old evidence (e), if certain knowledge of the applicable hypothesis would render the new information (E) useless. This is a strong assumption, not usually holding in medical test domains. For example, suppose two medical tests would increase physicians belief about the likelihood of disease 1, but that neither is conclusive. If the physician knew the patient had disease1, then the two tests are dependent simply because they are performed on the same person. Only in the case where the physician knew the patient did not have disease 1, would the two tests be independent. Second, it means that two hypotheses are independent given certain knowledge of their causes. Suppose the physician believes that disease 1 or disease 2 arise from independent processes in the body. Thus the two diseases seem a priori independent. But furthermore let's assume the he believes that either disease 1 or disease 2 can cause positive results on test 1. Then the results of test 1 can explain away the results on test 2. And thus the physician's subjective likelihood of disease 2 becomes dependent on his belief about the likelihood of disease 2.

In summary, we see that applying normative theories of normative uncertainty propagation is very difficult, because of the large number of elicitations the expert has to give, and the inherent computational complexity. The most successful normative approaches have been Bayesian networks [28]. Leveraging graph theory, these algorithms address the computationally complexity by parallel computing. Furthermore the algorithms reduce the number expert elicitations needed, by assuming coherent values for those the expert is unable to give. However, there have been demonstrated cases where Bayesian Networks performed no better than random guessing [13][16] [31].

In response to the above challenges with strict adherence to Bayesian reasoning, other normative models were proposed. For example many systems today are based loosely on *fuzzy logic*. However fuzzy logic has been well explained using Bayesian theory [3], so has not shed much light on normative theory. Furthermore early proponents of MYCIN's *certainty factor* system claimed their calculus avoided the above difficulties with dependent evidence [29]. However since then the *certainty factor* algebra has been shown to be a special case of Bayesian theory [14][15].

In brief, to our knowledge today, effective normative implementations of numeric uncertainty calculi in expert systems remain cost prohibitive. As a possible cost-effective alternative, we propose a new calculus which is instead based on how experts untrained in statistics actually seem to make decisions. We call this calculus the *gamma factor* calculus.

THE GAMMA FACTOR CALCULUS

Our proposal is loosely based on the pioneering work of MYCIN [29]. Instead of a probability, Shortliffe believed that uncertainties are better understood by most system builders as Confidence Factors (CNF). The CNF expresses belief in the accuracy of a rule or fact, on a scale from one to zero; where one means complete accuracy, and zero means total inaccuracy. Using our approach, system builders would use CNFs to measure the accuracy of rules. Users would use CNFs to measure the accuracy of variable values they input at run time. Our *gamma factor* calculus uses CNFs more like likelihood ratios than probabilities.

We base the *gamma factor* calculus in part on that of *confidence factor* calculi, first discussed in [18] and later embodied in many expert systems shells, including GURU [26]. Unlike the MYCIN system, which was designed for medical diagnosis only, GURU and many others like it are general purpose systems, and thus are highly configurable. These types of systems offer the builder many different uncertainty propagation calculi. However they offer little guidance on how to select among them. The four book User's Guide for CLIPS [12] makes no mention of CNF strategies.

Many of these general purpose systems elicit expert beliefs in the form of *confidence factors [CNFs]*. The scale of these ranges from -1 to +1. The calculi shown below in table 1 are typical [26]. Table 1's calculi are sequenced from the most pessimistic to the most optimistic, for the case where the CNF of the conditions is 50 and the CNF of the rule is 90.

Table 1

Calculus Name	Calculus	Result
Product	(a*b)/100	45
Balance	((a*b)/100)*(2-MAX(a,b)/100)	49.5
Minimum	min(a,b)	50
Average	(a+b)/2	70
Maximum	max(a,b)	90
Probability Sum	a+b – (a*b)	95

CNF algorithms use the above calculi in various combinations, to calculate:

- the confidence in the conditions entered by the user;
- the confidence in a fired rule, considering the confidence in the conditions and the confidence in the rule assigned by the rule builder;
- the combined confidence of rules that reach the same conclusion, and
- the overall confidence in the chain of reasoning which supports the conclusion.

These general purpose systems are very difficult for a builder to understand and work with [26]. In most shells, one strategy is used for every rule in the rule base. Furthermore they do not take explicitly into account dependencies among rules. Our gamma factor calculus offers a simple straight forward, backward chaining alternative that takes into account rule dependencies. Here is a typical gamma factor consultation example.

Let's assume the expert has input the following rules:

```
Rule 1.  IF      childRFev      = true
         THEN   CAD            = true         CNF  = 0.85


Rule 2.  IF      Diabetic       = true
         THEN   CAD            = true         CNF 0 .9


Rule 3.  IF        test1        = ChildRFev
         AND      test2        = ChildRFev
         THEN  ChildRFev       = true         CNF = 0.75


Rule 4.  IF        test3        = Diabetic
         AND      test4        = Diabetic
         THEN  Diabetic        = true         CNF = 0.75
```

Furthermore let's assume that the proposition base contains the following:

test1 = ChildRFev   CNF = 0.8
test2 = ChildRFev   CNF = 0.7
test3 = Diabetic   CNF = 0.8
test4 = Diabetic   CNF = 0.7

Suppose we ask our system the following, "What is the accuracy of a diagnosis of CAD?" First our system tests Rule 1. Then our engine would chain to rule 3, and determine the minimum CNF from it's condition list -- in this case, 0.7. Next we use the *product method* to calculate rule 3's accuracy as .52.5. = 0 .7 * .75. This seems reasonable because rule 2's conditions are uncertain. Now we calculate rule 1's CNF using the product method, to yield a cumulative CNF in CAD for the inference chain using rules 1 and 3 of **0.446**. = 0.52.5 * 0.85.

Now rule 2 is tested. It will chain to rule 4, and determine the minimum CNF from it's condition list, again 0.7. Again the product method calculates rule 4's CNF to 0.525= 0.7 * 0.75. Then we use the product method to calculate rule 2's accuracy to **0 .47.** = 0.9 * 0.525.

So now our system has traversed two inference chains, each yielding a different accuracy for a diagnosis of CAD, **0.446** and **0.47**. How should we combine accuracy measurements of two parallel evidence paths? Choose the minimum? Maximum? If we use the product method again, the result would not seem reasonable because 0.47 * 0.446 = **0.21,** clearly too low. Most existing implementations either pick the maximum, or use the *probability sum* method [4][11] using the logic that the conclusion is reinforced because it is reached from two independent sources. The probability sum would yield **0.6** = (0.446 + .47) – (.71*0.446), more reasonable perhaps. But what if the expert believes that diabetes and childhood rheumatic fever are highly correlated? This dependency means that a high CNF for diagnosis one would render the other virtually useless. In that case, a CNF of **0.6** might seem too high.

The above example shows that the *probability sum* method tends to produce optimistic diagnostic accuracies when pertinent items of evidence are dependent: it is easy to construct cases that when a sufficient number of rules fire, the probability sum method would produce a final consultation accuracy of 1.0 regardless of the CNFs in the knowledge base. On the other hand, the *product method* tends to produce pessimistic CNFs: With that method, the consultation CNF can never rise above the least of one condition of any fired rule, even if every other CNF in the knowledge base is higher.

To address this issue of dependent parallel inference in rule based expert systems, we propose an extension to typical certainty factor calculi we call the *gamma factor (GF)*. The gamma factor results in a calculation, which is a weighted average of the *product method* and the *probability sum method*. In the research prototype, a Gamma Factor of 0 uses pure Product Method, a Gamma Factor of 1 uses pure Probability Sum method, and any value between 0 and 1 creates a weighting between the two.

The Gamma Factor may be implemented at many different levels of granularity for the four different CNF calculation situations; conditions, rule, rules which reach the same conclusion, and overall confidence. The research prototype partitions the rule base into Rule Clusters. Each individual rule cluster contains all of the rules that reach the same conclusion. The sum of the Rule Clusters is the Rule Base. The Gamma Factors are set at the Rule Cluster level, meaning that four different Gamma Factors are assigned to each cluster. It is also, of course, possible to assign Gamma Factors to individual rules.

If it is desired to assign Gamma Factors at the lowest level of granularity, all rules reaching the same conclusion be grouped in a cluster and presented to the expert so that he can assign a Gamma Factor to each component rule, taking into account the dependency relationships, while remaining cognizant of the rules firing order. In the example of figure 1, there would be two rule clusters for the diagnosis of CAD. We believe that improved inference can be achieved when the expert carefully tailors the Gamma Factor for each rule in the cluster.

For example, a GF of 1.0 would be used when the rules are completely independent, and result in a calculation equal to that of the *probability sum* method which would produce higher CNFs. A gamma factor of 1 in our example above would yield **0.6** = 1*.6 + 0*0.21 -- probably too high if the rules are very dependent. At the other extreme, a *gamma factor* of 0 would result in a calculation equal to that of the *product* method, and would thus yield **0.21** = (0*.6) + (1*0.21), too low if rules are independent.   On the other hand a GF of 0.5 will average the two methods. In the example above, that would be (0.21*.5) + (0.6*.5) = 0.41, perhaps the most reasonable in this example.

The major determinant of the Gamma Factor is the relative independence of the components. If two components are dependent, a Gamma Factor of 0.5, which averages the Product and Product Sum calculations become more appropriate. When two components are totally independent, a Gamma Factor of 0, which uses the Product Method, becomes more appropriate. In the case where the components are reinforcing, such as when two independent rules reach to same conclusion, a Gamma Factor of 1, which uses the Product Sum method, is favored.

We feel that the Gamma Factor approach has several advantages over current implementations. First, it is more granular, making it possible to more finely tune confidence calculations to achieve more accurate conclusions. Secondly, Gamma Factor recommendations are made possible through an analysis of independence. Thirdly, the Gamma Factor calculations are computationally inexpensive, unlike the Bayesian systems.

CONCLUSION

In conclusion, at the heart of intelligence is heuristics, and at the heart of heuristics is belief. Unfortunately, as we have shown, early results of the trustworthiness of numerical techniques for uncertainty propagation in expert systems were unimpressive. Since then there has been a paucity of research in the use of such numerical techniques. However we believe that the careful experimentation with numerical techniques continues to be an important area of research.   With this paper we hope to revive the debate about how best to implement numerical techniques in expert systems, and caution that implementations should take special care to make the expert aware of and given some control over the assumptions and possible biases that my result from a system's use. Our proposed Gamma Factor calculus offers a simple straightforward approach that simplifies and extends typical certainty factor calculi by adding the option to weight the diagnostic accuracy calculations for each rule in a cluster to minimize or remove the biases inherent in two popular certainty factor calculi for parallel evidence combination, probability sum, and product.

(A previous version of this paper was presented at he Applied Business Research Conference, Honolulu Hawaii, January 2007. It was entitled "Tuning the Numeric Uncertainty Calculi in Expert Systems.)

BIBLIOGRAPHY

1.   Alpert, A. Raiffa, H. A progress report on the training of probability assessors. Judgment Under Uncertainty, *Heuristics and Biases*, 295- 305, 1982.
2.   Bayes, T.   Essay Towards Solving a Problem in the Doctrine of Chances Philosophical, *Transactions of the Royal Society of London*, 1764.
3.   Cheeseman, P. Probabilistic versus Fuzzy Reasoning. In Uncertainty in *Artificial Intelligence*, North-Holland, pp. 85-102, 1986.
4.   *Clips Reference Manual Volume I Basic Programming Guide* Version 6.24 June 15th 2006. http://www.ghg.net/clips/download/documentation/bpg.pdf
5.   Cox, R., Probability, Frequency and Reasonable Expectation, *American Journal of Physics,* 14 (1), January-February (1946), pp. 1-13.
6.   De Finetti, B. *Theory of Probability* 1, 2. Wiley, New York, 1974.
7.   Dempster, A.P., Kong, A. Probabilistic Expert Systems in Medicine: Practical Issues in Handling Uncertainty, *Statistical Science*. 1987, Vol. 2, No. 1, pp 33.
8.   Dubois,D. & Prade, H. A class of fuzzy measures based on triangular norms, *International Journal of General Systems*, 1982.

9.   Duda, R.O., Hart, P.E. & Nelson. N.J. Subjective Bayesian Methods for Rule-based Inference Systems, AFIPS Conference Proceedings, pp. 1075-1082. New York City, NY. June, 1976.
10.   Dyer, J.S. & Wright, M.K. Probabilistic Causal Reasoning in DSS. Proceedings of the NATO Advanced Study Institutes Program. Tuscany, Italy, June, 1991.
11.   EZ-Xpert Corporation. EZ-xPert Builder's Manual. 2007.
12.   Giarratano, Joseph C. *CLIPS 6.21 User's Guide*, International Thompson Publishing, 2003.
13.   Grunwald, P.D., Halpern, J.Y. When Ignorance is Bliss, Proceedings of the twentieth Conference on Uncertainty in Artificial Intelligence. AIAU Press. 2004.
14.   Heckerman, D. An axiomatic framework for belief updates UAI 1986: 11-22, 1986
15.   Heckerman, D. Horvitz. E. *The myth of modularity in rule-based systems for reasoning with uncertainty*, UAI 1986: 23-34.
16.   Henrion M,, Provan G., Favero, B. Sanders, G. An Experimental Comparison of Numerical and Qualitative Probabilistic Reasoning, Proceedings of the 10th Annual Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann,1994.
17.   Hicks, Richard C. Knowledge Base Management Systems - Tools For Creating Verified Intelligent Systems *Knowledge-Based Systems*, April 2003.
18.   Holsapple & Whinston. Certainty Factor Algebras, Chapter 10 in Irwin Ed., *Business Expert Systems*, pp. 211-241, 1987.
19.   Howard, R.A., and Matheson, J.E, Reading on the Principles and Applications of Decision Analysis, Strategic Decision Group, Menlo Park, CA, 1984.
20.   Jaynes, E.T. Prior probabilities, *IEEE Transaction on Systems Science and Cybernetics*, 4, 227-241, 1968.
21.   Lindley, D.V. Scoring rules and the inevitability of probability *International Statistical Review*. 50 1-26, 1982.
22.   McCune, B.P., Dean J.S., Tong, R.M. & Shapiro, D.G. RUBRIC: A system for rule-based information retrieval, *Advanced Information & Decision Systems, Final Technical Report* TR-1018-1, 1983.
23.   Pearl, J. A Constraint-Propagation Approach to Probabilistic Reasoning, In *Uncertainty in Artificial Intelligence*, North-Holland, pp. 357-370, 1986.
24.   Rescher, N. *Many Valued Logic*. New York: McGraw-Hill, 1969.
25.   Savage, L.J. *The foundations of statistics*. New York: Wiley, 1954.
26.   Savitor Corporation. Building Expert Systems Using GURU. http://www.savitar.com/manuals/guru/expert.pdf 2004.
27.   Shafer, G. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
28.   Shafer, G. Probability Judgment in Artificial Intelligence and Expert Systems, *Statistical Science*. V 2, No. 1, 3-44, 1987.
29.   Shortliffe. E.H. and Buchanan, B.G. A model of Inexact Reasoning in Medicine, *Mathematical Biosciences*, 23 pp. 351-379, 1975 Reprinted as Chapter 11 in Buchanan & Shortliffe, 1984.
30.   Smets, P. The degree of belief in a fuzzy set, *Information Sciences*, 25, 1-19, 1981.
31.   Tong, R.M. & Shapiro, D.G. Experimental investigations of uncertainty in a rule-based system for information retrieval, *International Journal of Man-Machine Studies*, pp. 265-282, March, 1985.
32.   Wise B. & Henrion, M. A Framework for Comparing Uncertain Inference Systems to Probability, In *Uncertainty in Artificial Intelligence*, North-Holland, pp. 69-84, 1986.
33.   Wright, M.K. Assessing Causal Probabilities for Decision Support Systems, Dissertation, University of Texas at Austin, 1992.
34.   Wright, M.K., Stokes, L. & Dyer (1994) J.S. Reliability and coherence of causal, diagnostic, and joint subjective probabilities. *Decision Sciences*, 25, 1994.
35.   Wright, M.K., Stokes, L. & Dyer (1995) Choosing probability inputs when building decision support systems, Proceedings of the Decision Sciences Institute Conference, Boston, November, 1995
36.   Zadeh, L.A. Fuzzy Sets, *Information and Control*, 8, 338-3 5 3 , 1965.

NOTES