Int. J. Man–Machine Studies (1990) **32,** 293–301

Soliciting weights or probabilities from experts for rule-based expert systems

DANIEL E. O'LEARY
Graduate School of Business, University of Southern California, Los Angeles, California 90089-1421, USA

(Received 9 February 1988 and in revised form 12 February 1989)

Rule-based expert systems attach a weight to each rule in order to represent uncertainty or strength of association. There are a number of schemes that are used to represent uncertainty in expert systems. Some of these methods allow the system designer to solicit either the probabilities, used to compute the weights, or to solicit the weights directly, or both.

This paper presents results that indicate that if the weights are gathered directly, rather than using probabilities, then the weights may not meet the underlying conditions of the mathematical model of uncertainty on which the weights are based or the weights may imply highly unusual behavior for the underlying probabilities and implicit utility function.

In one system it is found that there were violations of the mathematical properties of the model in over forty percent of the weights on the rules of the system. If the weights do not meet the constraints of the underlying mathematical models then such violations may yield inappropriate parameterization of other weights in order to make the model work. Further, such violations can lead to an inappropriate estimation of the probabilities of events by the system and yield inappropriate inferred weights.

In another case it was found that a system was dominated by weights that suggest highly unusual behavior for the underlying probabilities.

From an operational perspective these inconsistencies indicate the importance of the method used in gathering the weights, e.g. indirectly through the probabilities or directly through the weights. It also indicates the importance of validating and verifying the weights to ensure that the weights meet the needs of the underlying theory and do not force unusual relationships onto the underlying probabilities.

1. Introduction

Rule-based expert systems are expert systems where the knowledge base is represented by a set of "If (condition) then (consequence)" rules. Oftentimes weights are associated with the rules in an effort to measure, e.g. the strength of belief or extent of uncertainty of the rule. Although these weights generally are based on probability theory, the weights are often not probabilities. Instead, these weights typically incorporate probabilities and go under such names as "certainty factors."

One of the primary ways of representing the weights is the approach implemented in the expert system shell AL/X (Spiegelhalter, 1986). The AL/X approach is developed in Duda, Hart, Nilsson & Sutherland (1976) and Duda, Gaschnig & Hart (1979). Implementation of the AL/X approach is discussed in more detail in Paterson (1981) and Kidd & Cooper (1985).

AL/X has used either weights based on likelihood ratios or weights based on logarithms of likelihood ratios. In order to generate the weights on the rules, the

0020-7373/90/030293 + 9$03.00/0

expert system developer can solicit either the probabilities underlying the weights or the weights directly. Unfortunately, if the weights are developed directly, then the weights may not conform to the underlying mathematical structure on which the weights are based or the weights may imply a restrictive functional relationship between the probabilities in the likelihood ratios. Four case studies from the literature document these findings.

Violations of the mathematical structure of the particular model, by inappropriate choice of weights on some rules, can impair the quality of overall model and the inferencing process in the model. Violations of the underlying structure on which these models are based or implied functional relationship between probabilities in the likelihood ratios can yield inappropriate parameterizations on the other rules, as the system is "fine-tuned." Further, as will be seen later in the paper, the violations discussed here can lead to substantial mis-estimation of inferred probabilities of events.

Accordingly, the resulting systems may yield mediocre or incorrect decision results, or the system may mis-estimate the uncertainty in events because the systems do not use the weights as they were intended to be used. Thus, from an operational perspective, this indicates the importance of the method of developing the weights (probabilities directly or weights directly) and the importance of validating and verifying the weights (e.g. O'Leary, 1987; 1988) to ensure that they meet the underlying mathematical structure of the model.

In addition, these violations may cast doubt on the ability of directly solicited weights in AL/X to *simulate* human assessment of the strength of the relationship of a rule. Since the model is based on Bayes' Theorem, this does *not* say that the approach should not be used to develop normative decision making models or that probability-based assessments are inappropriate.

Although this paper is concerned with weights developed using the AL/X approach, it can be *generalized* to other such systems where there are weights and an underlying mathematical theory on which the weights are based. In addition, it is likely to be of interest to those who use likelihood ratios to measure uncertainty in decision making.

This paper proceeds as follows. Section 2 presents background on the types of weights on the rules used in AL/X. Section 3 presents three theorems relating to the correct behavior of the two types of weights, based on the underlying mathematical structure of the model. Then the weights generated from two expert systems in the literature are investigated. It is found that when weights are gathered directly, rather than as probabilities, they may not meet the requirements of the underlying mathematical model on which the weights are based. Section 4 presents a theorem for a special relationship between sets of weights that implies a highly restricted relationship between the probabilities, and thus, a highly limited set of feasible probabilities. Section 5 discusses the implications of the discrepancy of developing weights directly or using the underlying probabilities, and provides a summary of the paper.

2. Weights in AL/X-based expert systems

In AL/X, knowledge is specified as a set of rules and weights on the rules of the form:

"If E then H (to degree S,N)"

S and N are numeric values that represent the strength of association between E and H. S is called the sufficiency factor because a large S means that a high probability for E is sufficient to produce a high probability of H. N is called a necessity factor because a small value of N means that a high probability for E is necessary to produce a high probability of H.

In AL/X, S and N can be specified directly. Alternatively, S and N can be specified indirectly by establishing the probabilities in the following likelihood ratios:

$$S = P(E \mid H)/P(E \mid H')$$

and

$$N = P(E' \mid H)/P(E' \mid H'),$$

where, E' and H' correspond to "not E" and "not H," and $P(E \mid H)$ is the probability of E given H. Early systems using AL/X used S and N directly. However, in later versions of AL/X, S and N are used as logarithms (to the base ten) of the likelihood ratios, as respectively, negative weights (NW's) and positive weights (PW's).

The PW's and NW's provide the means by which to express quantitatively the corroborative or discorroborative, respectively, value of the evidence. PW's and NW's can take on negative or positive numbers from -100 to 100. However, some applications limit the scale from -30 to 30 (e.g. Dungan & Chandler, 1985).

The use of likelihood ratios is a generally accepted approach to decision making (Lindley, 1985). Thus, although this approach is AL/X specific, it remains a general approach to evaluating evidence. In addition, the existence of two ways (weights directly or probabilities directly) to assess the weights is typical of other expert system weighting schemes, e.g. MYCIN (Buchanan & Shortliffe, 1985).

2.1. SUBJECTIVE BAYESIAN UPDATING

Once the weights have been associated with individual rules, the AL/X inference approach uses those weights to establish weights on chains of inference (Duda *et al.*, 1976; Duda *et al.*, 1979; Dungan, 1983). As above, assume the rule is "if E then H." Although AL/X employs a scheme that allows the probability of E to be less than 1, without loss of generality it is assumed that $P(E) = 1$. Define the prior odds of H, $O(H)$, to be $P(H)/P(H')$ and similarly, the posterior odds $O(H \mid E)$ as $P(H \mid E)/P(H' \mid E)$. Odds and probabilities are related through the relationship $O(H) = P(H)/1 - P(H)$. Updating the prior can then be accomplished using Bayes' rule as $O(H \mid E) = S * O(H)$ and $O(H \mid E') = N * O(H)$.

In the initial version of AL/X the prior odds of the hypotheses were updated using the multiplicative version. A later version of the system that employs the PW's and NW's replaces the multiplicative approach, with an equivalent additive approach. Using that approach, the odds (called "degrees of belief") $DB = 10 * \mathrm{Log}_{10}(O(H)))$ are added appropriately to the weights.

3. Relationship between positive weights and negative weights

If the expert system developer gathers information on the probabilities, $P(E \mid H)$ and $P(E \mid H')$, then as long as the probabilities take on values so that S and N are

defined, there will be no violation of the mathematical structure of the model. However, if the developer gathers the *PW*'s and *NW*'s independent of the probabilities then there may be violations with the model. This would result because directly gathered weights have implications for the probabilities because of the "interlinking" nature of the weights.

There are some important relationships between the *PW*'s and the *NW*'s that allow us to validate those weights to ensure that the weights (and thus the probabilities) have been assigned in concert with the assumptions of the AL/X model. These relationships are summarized in Theorems 1–3.

Theorem 1. $PW > 0$ if and only if $NW < 0$.

Proof. If PW is >0 then $S > 1$. Thus, $P(E \mid H)/P(E \mid H') > 1$. This implies that $P(E \mid H) > P(E \mid H')$.

As a result, $1 - P(E \mid H) < 1 - P(E \mid H')$. Thus, $N < 1$ and $NW = \mathrm{Log}(N) < 0$.

If NW is <0 then $N < 1$. Thus, $(1 - P(E \mid H))/(1 - P(E \mid H')) < 1$. This implies that $(1 - P(E \mid H)) < (1 - P(E \mid H'))$.

As a result, $P(E \mid H) > P(E \mid H')$. Thus, $S > 1$ and $PW = \mathrm{Log}(S) > 0$.

Theorem 2. $PW < 0$ if and only if $NW > 0$.

Proof. Similar to Theorem 1.

Theorem 3. $PW = 0$ if and only if $NW = 0$.

Proof. If $PW = 0$ then $\mathrm{Log}\, S = 0$. Thus, $S = 1$. Accordingly, $P(E \mid H) = P(E \mid H')$ and thus, $(1 - P(E \mid H)) = (1 - P(E \mid H'))$, assuming $1 > P(i \mid j) > 0$ for all i and j. Thus $N = 1$ and $NW = 0$.

If $NW = 0$ then $\mathrm{Log}\, N = 0$. Thus, $N = 1$. Accordingly, $(1 - P(E \mid H)) = (1 - P(E \mid H'))$ and thus, $P(E \mid H) = P(E \mid H')$ assuming $1 > P(i \mid j) > 0$ for all i and j. Thus $S = 1$ and $PW = 0$.

These theorems summarize some relationships between the signs of the weights, that result from the likelihood ratios. In particular, if the expert has assigned a positive weight to the *PW* then the *NW* for that rule should be negative, and conversely. Otherwise there is a violation in the assumption of the weights of the model. Similarly, if the expert has assigned a negative weight to the *PW* then the *NW* is positive, and conversely. Otherwise there is a violation of the assumptions of the model. Finally, if the expert assigns a zero weight to either the *PW* or the *NW* then the other also is zero.

Generally the weights are nonzero so that Theorems 1 and 2 define the relationship between the *PW*'s and the *NW*'s. However, theorem 3 is indicative of an uncertain situation. $PW = NW = 0$ indicates that $P(E \mid H) = P(E \mid H')$. This says that the probability of event E given event H is the same as the probability of event E given H'. This indicates that the expert has a highly uncertain view of the world, with respect to these events.

These theorems can be extended to the earlier version of AL/X where the values of S and N were used directly. Theorem 1 would become $S > 1$ if and only if $N < 1$. Theorem 2 would become $S < 1$ if and only if $N > 1$. Finally, Theorem 3 would

become $S = 1$ if and only if $N = 1$. For a system based on the original set of weights see Duda *et al.* (1979).

These theorems and extensions provide a basis with which to validate and verify the weights developed for use in AL/X-based systems. If the theorems are violated then there is a discrepancy between the way the expert views the world and the way the model is designed to view the world.

For example, if an expert had both a positive *PW* and a zero *NW*, this would be a violation of both Theorems 1 and 3. It would imply first, that $P(E' \mid H) > P(E' \mid H')$ and second, that $P(E' \mid H) = P(E' \mid H')$. Clearly this is a contradiction and an impossible situation. Unfortunately, there are systems that have been developed using AL/X with weights that do not satisfy the above theorems and extensions.

3.1. EXAMPLE SYSTEMS THAT VIOLATE THE THEOREMS

There are very few sets of rules and corresponding weights available to determine if users of AL/X-based systems develop weights that violate Theorems 1–3 and the extensions. However, there has been at least one system, AUDITOR (Dungan, 1983; Dungan & Chandler, 1983; Dungan & Chandler, 1985) for which the entire set of rules and weights has been published. For this expert system, the weights were gathered directly. The probabilities were not developed independently. A summary of the violations of the weights in AUDITOR are given in Table 1. Three of the rules with weight in violation are given in Table 2.

In that system there were sixteen rules (out of thirty-eight rules) with contradictions to Theorems 1 through 3. These contradictions took two forms $PW > 0$ and $NW = 0$ (fourteen occurrences), $NW < 0$ and $PW = 0$ (one occurrence) and $PW < 0$ and $NW = 0$ (one occurrence).

AUDITOR is not the only system with such deviations from the above theorems. Kidd & Cooper (1985) list the weights for an AL/X system for fault finding in radio equipment. It appears that in the analysis of "output timing tail light on" two sets of weights (10, 0) and (10, 0) also are in violation of the above theorems. Thus, if the weights are developed directly, there is a real danger that the resulting weights will not be developed in concert with the underlying mathematical model in AL/X.

3.2. IMPLICATIONS OF ERRORS IN THE WEIGHTS

The impact of such errors in the weights can be displayed, in part by use of an example, based on Spiegelhalter (1986). The example is a multiplicative version but

TABLE 1

Violations to theorems of the weights in AUDITOR

Type of violation	Number of rules violating†
No violations	22
Total number of rules violating theorems	16
Violate Theorem 1	14
Violate Theorem 2	2
Violate Theorem 3	16

† All rules that violated a theorem, violated Theorem 3.

TABLE 2
Sample rules from AUDITOR that violated theorems

"While small payments are being received on this account, the outstanding delinquent balance is growing." *PW* 1 and *NW* 0.
"The debtor has issued notes for the unpaid portions of his account." *PW* 0 and *NW* − 0·5
"Collection effort being applied by your client is less rigorous than is desirable." *PW* 1 and *NW* 0

TABLE 3
Example†

	Value of NW				
$Q(H)$	−1·0	−0·5	−0·1	−0·01	0·0
0·25	0·0244	0·073	0·166	0·196	0·200
0·50	0·0476	0·136	0·284	0·328	0·333
0·75	0·0697	0·191	0·373	0·423	0·429
0·90	0·0825	0·222	0·417	0·468	0·473
0·95	0·0867	0·231	0·430	0·481	0·487
10·0	0·5000	0·760	0·888	0·907	0·909

† Example assumes that $PW = 0{\cdot}954$. Minor computational differences are due to rounding.

will be extended to the logarithmic additive case. Because the occurrence of 0 in one of the weights was the major difficulty, the example explores the impact of using a 0 instead of a relatively small, but proper value for one of the weights.

Assume that the prior probability of H is 0·2. Assume that the weights are $PW = 0{\cdot}954$ and $NW = -1$. Thus, $O(H) = 0{\cdot}2/(1 - 0{\cdot}2) = 0{\cdot}25$. The log of 0·25 is $-0{\cdot}602$. This leads to $\log O(H \mid E) = 0{\cdot}954 - 0{\cdot}602 = 0{\cdot}352$, thus, $O(H \mid E) = 2{\cdot}25$ and $P(H \mid E) = 2{\cdot}25/(2{\cdot}25 + 1) = 0{\cdot}6923$. Similarly, $O(H \mid E')$ can be computed. $\text{Log}\, O(H \mid E') = -1{\cdot}000 - 0{\cdot}602 = -1{\cdot}602$, thus, $O(H \mid E') = 0{\cdot}025$. This implies that $P(H \mid E') = 0{\cdot}025/1 + 0{\cdot}025 = 0{\cdot}0244$.

Instead of $NW = -1{\cdot}0$ assume that $NW = 0$. This implies that $\log O(H \mid E') = 0{\cdot}0 - 0{\cdot}602 = -0{\cdot}602$, and, $O(H \mid E') = 0{\cdot}25$. Thus, $P(H \mid E') = 0{\cdot}25/(1 + 0{\cdot}25) = 0{\cdot}2$. In this case $P(H \mid E')$ increased 719·67%. In addition, the "distance" between $P(H \mid E)$ and $P(H \mid E')$ changed from $0{\cdot}6923–0{\cdot}0244 = 0{\cdot}6679$, to $0{\cdot}6923 - 0{\cdot}2 = 0{\cdot}4923$. This represents a substantial change in the uncertainty associated with those two different sets of events.

These and other values are summarized in Table 3. In addition to the impact of the difference of the weight from zero, it can be seen that the extent of change also is due, in part, to the value of the prior.

4. The impact of relationships between weights on probabilities

If the *PW*'s and *NW*'s are directly solicited then the weights on a rule may be related by the user with a functional relationship, e.g. $PW = -NW$, where *PW* is not equal to zero. (Selected examples of this situation are discussed below.) Because of the interlinking nature of the *PW*'s and *NW*'s, such a relationship between the *PW*'s and the *NW*'s has implications on the underlying probabilities, and resulting utility functions.

Theorem 4. If $PW = -NW$ then $P(E \mid H) - P(E \mid H)^2 = P(E \mid H') - P(E \mid H')^2$.

Proof. If $PW = -NW$ then

$$\log(P(E \mid H)/P(E \mid H')) + \log(P(E' \mid H)/P(E' \mid H')) = 0.$$

Thus,

$$\log(P(E \mid H)/P(E \mid H')) + \log((1 - P(E \mid H))/(1 - P(E \mid H'))) = 0.$$

As a result,

$$[P(E \mid H)/P(E \mid H')] * [(1 - P(E \mid H))/(1 - P(E \mid H'))] = 1.$$

Thus,

$$P(E \mid H) * (1 - P(E \mid H)) = P(E \mid H') * (1 - P(E \mid H')).$$

This is a highly restrictive requirement on $P(E \mid H)$ and $P(E \mid H')$ that forces the probabilities to be drawn from a set of lines. However, for a number of systems in the literature there is substantial use of weights of the type $PW = -NW$.

For example, in each of the two systems discussed above, some of the weights exhibited this behavior. In AUDITOR, there are seven such sets of weights. In Kidd & Cooper (1985) there are three such sets of weights. One additional system, EDP-XPERT (Hansen & Messier, 1982; Hansen & Messier, 1986; Hansen & Messier, 1987) also has some weights for which $PW = -NW$. However, no system found in the literature had all their weights meet this constraint.

The existence of weights under this relationship is not incorrect in and of itself. However, in a situation where the weights are directly developed there is a possibility that the expert will generate a functional relationship between the sets of weights, such as that relationship in Theorem 4. In that case there is substantial danger that the weights will imply unusual underlying utility functions.

5. Discussion and conclusion

This paper has summarized the mathematical relationships between the weights of an AL/X-based system. The results indicate that if either a *PW* or *NW* is greater than zero then the other is less than zero, and conversely. Further, the *NW* or *PW* can only take a zero value if the other has a zero value. An examination of example systems found a number of violations to the condition that both or neither of the rules were zero. Similar relationships can be developed for earlier versions of AL/X that employed a multiplicative inference model using values for *S* and *N*.

The violations impact the inference process. The extent of the difference in the inference process, resulting from those errors, is a function of the prior odds and the value of the weight that is appropriate. Investigation of an example found substantial differences as the value of the weight moved from -1 to 0.

This paper also has found that if the weights are directly related then this has implications for the relationship between the underlying probabilities. Functional relationships such as $PW = -NW$ indicate the probabilities in the underlying utility function are related in a very unusual manner. An examination of three systems found a number of weights functionally related in this manner.

Thus, it appears that it is likely that directly developed weights may not satisfy the underlying mathematical structure in AL/X or may imply very unusual underlying probabilities. This has a number of implications. First, the method of developing the weights is critical. Directly soliciting probabilities can ensure consistency with the underlying mathematical model. Second, if weights are directly solicited, rather than probabilities, then validating the weights to ensure that they meet the underlying structure in AL/X also is critical. In addition, such weights should be examined for user developed relationships between the weights, e.g. $PW = -NW$, for implications of the underlying structure of the probabilities. Third, if the directly solicited weights are inconsistent with the underlying probability theory in the model or they suggest highly unusual relationships between the probabilities, then there may be some question as to the ability of the *NW*'s and *PW*'s to *simulate* human decision making. Although this may be important to behavioralists or cognitive scientists studying human decision making, this does not impact the ability to use directly developed probabilities or the AL/X approach to build normative decision making models, because of the firm foundation of AL/X in decision making (e.g. Lindley, 1985).

The author would like to acknowledge the helpful comments of the anonymous referee.

References

BUCHANAN, B. & SHORTLIFFE, E. (1985). *Rule-Based Expert Systems,* Reading, Massachusetts: Addison-Wesley,

DUDA, R., GASCHNIG, J. & HART, P. (1979). Model design in the prospector consultant system for mineral exploration. In D. MITCHIE Ed., *Expert Systems for the Micro Electronic Age,* pp. 153–167. Edinburgh: Edinburgh University Press.

DUDA, R., HART, P., NILSSON, N. & SUTHERLAND, G. (1976). Subjective Bayesian methods for rule-based inference systems, *Proceedings of the National Computer Conference,* pp. 1075–1082.

DUNGAN, C. (1983). *A Model of Audit Judgement in the Form of an Expert System,* Unpublished Ph.D., University of Illinois.

DUNGAN, C. & CHANDLER, J. S. (1983). *Analysis of Audit Judgement Through an Expert System,* College of Commerce and Business Administration, University of Illinois, November.

DUNGAN, C. & CHANDLER, J. S. (1985). Auditor: a microcomputer-based expert system to support auditors in the field, *Expert Systems,* October.

HANSEN, J. & MESSIER, W. (1982). Expert systems for decision support in EDP auditing, *International Journal of Computer and Information Sciences,* October, 357–379.

HANSEN, J. & MESSIER, W. (1986). A knowledge-based expert system for auditing advanced computer systems, *European Journal of Operational Research,* September, 109–23.

HANSEN, J. & MESSIER, W. (1987). *A Case Study and Field Evaluation of EDP-XPERT,* Paper presented at the University of Southern California Symposium on Expert Systems in Auditing, February 1988.

KIDD, A. & COOPER, M. (1985). Man–machine interface issues in the construction and use of expert systems, *International Journal of Man–Machine Studies,* **22,** 91–102.

LINDLEY, D. (1985). *Making Decisions* (Second Edition), Wiley, London.

O'LEARY, D. (1987). Validation of expert systems, *Decision Sciences,* **18,** (4) 468–486.

O'LEARY, D. (1988). Methods of validating expert systems, *Interfaces,* November–December 1988, 72–79.

PATERSON, A. (1981). *AL/X User Manual,* Oxford, Intelligent Terminals Ltd.

SPIEGELHALTER, D. (1986). A statistical view of uncertainty in expert systems. In W. Gale, Ed., *Artificial Intelligence and Statistics,* pp. 17–55. Reading, Massachusetts: Addison-Wesley.