Annales Univ. Sci. Budapest., Sect. Comp. 10 (1989) 221-241

A SHELL STRUCTURE OF PROBABILISTIC TYPE FOR EXPERT SYSTEMS

CASTILLO E. and ALVAREZ E.

Universidad de Cantabria. 39005 Santander. Spain

SUMMARY

The paper presents some probabilistic expert system shells, which allow for all or some dependences of events to be included and are fully coherent, include a linear programming tool able to advise the expert to keep coherence, when assigning new probabilities, and have reasoning explanation facilities. The learning capacity of the models is shown through the appropriate updating formulas or by a log-linear model simplyfying assumption. Finally, an application to medical diagnosis illustrates the models.

Key Words: Expert systems, explanation, learning, log-linear models, medical diagnosis, probability-based models.

1. INTRODUCTION

The practical feasibility or unfeasibility of probability, as opposed to rule based, expert system models has been discussed by convinced defendors (Adams (1976), Cheeseman (1985), Lindley (1987), etc.) and detractors (Shafer (1982), Zadeh (1983), etc.). As an example Lindley (1987) when referring to expert systems states: "the only satisfactory description of uncertainty is probability. By this is meant that every uncertainty statement must be in the form of a probability; and that the calculus of probabilities is adequate to handle all situations involving uncertainty. In particular, alternative descriptions of uncertainty are unnecessary".

We are much more on Lindley's side and we spouse probabilistic methods though simultaneously understanding and accepting the problems they actually have. However, when this debate arises, a variety of reasons can be argued in favor of both sides.

Probability models are criticized because either of the huge number of parameters involved or the difficulties in their estimation from data. In fact, there is a general agreement in that the most general probability model, including all possible dependences, is unfeasible in practice. Other source of criticisms comes from the weakness of the independence model that is too simple to give good results in a general context. However, between these two extremes, an extensive range of possibilities exist.

Nevertheless, one must be aware that many of the alternative methods need a probability space framework. Thus, they can be equivalently formulated in terms of probability. Alternative methods could be defendible because of the claim that more understandable and easily handable concepts than probabilities are used, but this is not the case. There is no problem in defining new concepts in terms of probabilities as certainty factors, belief functions, etc.). The problem consists in defining complicated concepts or forgetting that, once defined, they are constrained by probability rules.

We are in full agreement with Lindley (1987) in that, under no circumstances, probability axioms must be violated if a probability framework is, directly or indirectly, utilized. Thus, efforts should be done in order to help experts to clarify what and what is not coherent. The fact that a human expert has no coherent knoledge does not justify an incoherent model. The aim of the expert system should be the adquisition of the human expert knowledge but not of the human expert errors. In this sense, the development of tools supporting coherence and giving advise to experts at this stage must be encouraged.

To us, the problem arises when some probability axioms are violated, either directly or when ad hoc solutions, lacking axiomatic basis, for propagation of uncertainties are proposed. This problem arises in practically all the existing rule based expert systems. In fact, usual uncertainty propagation formulas are not better than the assumption of independence. Similarly, encouragement of future research on this problem is convenient.

One of the most desirable capabilities of any expert system is that of learning by experience. There are two types of learning: parameter learning and structural learning. When we talk about parameter learning we assume that we have a model with defined structure dependent on some parameter values. The learning process consists in improving the parameter estimates from new experiences. In the case of structural learning, we allow for model structure to be changed (for example, new dependences of rules can be included in it). The last possibility is much more powerful and implies testing of the structural hypotheses of the model from time to time, in the light of data, and modification of them when necessary. For this to be possible, the model has to be initially opened to the new structure.

One typical difficulty of probability models in expert systems is that of reasoning explanation, which becomes almost trivial for rule based models, because at every step the active rules are well known. On the contrary, simulation and related education and training possibilities become trivial for probability models while undergo some difficulties for rule based models.

All the above suggests that more than a model, a full shell to solve all mentioned problems is needed.

The aim of this paper is to present expert systems probability shells solving some or all of those problems. In particular, some of the shells to be presented:

a) have a model with a reasonable number of parameters to make it practically feasible,

b) have full coherence,

c) advise to experts in order to maintain coherence of the model,

d) have an exact uncertainty propagation mechanism,

e) have parameter learning capacity,

f) allow for structural learning capacity to be included,

g) allow for explanation capabilities, and

h) allow for easy simulation, education and training facilities

2. LEARNING PROBABILITY MODELS

This section is devoted to probability models. The models are called "learning probability models" because special attention is giving to updating formulas for their parameters once given data is known.The models range from the most general dependence model to the independence one. In order to make them more clearly understandable, we shall refer to symptoms and diseases when referring to event related to data or goals, respectively, because of their remembrance of medical diagnosis. Nevertheless, the validity of these models goes beyond this particular example.

Let us assume that we are interested in the problem of diagnosis of the set of diseases $\{E_i; i = 1,2\ldots,n\}$, not necessarily disjoint, in a given population Ω, and that data is collected in the form of symptoms $\{S_j; j = 1,2,\ldots,m\}$ (see figure 1). One symptom, S_j, is a partition, $S_j = \{B_{j1}, B_{j2}, \ldots, B_{jr_j}\}$ of Ω.

The goal of the inference engine is to obtain the probabilities:

$$\begin{aligned} &P(E_i/A_1 \cap A_2 \cap \ldots \cap A_m) = \\ &= P(E_i \cap A_1 \cap A_2 \cap \ldots \cap A_m)/P(A_1 \cap A_2 \cap \ldots \cap A_m) \\ &i = 1,2,\ldots,n \end{aligned} \tag{1}$$

where $A_j \in \mathbf{P}(S_j) - \{\phi\}$ and $\mathbf{P}(S_j)$ is the set of parts of S_j.

Note that $A_j = \Omega$ means lack of information about the symptom S_j.

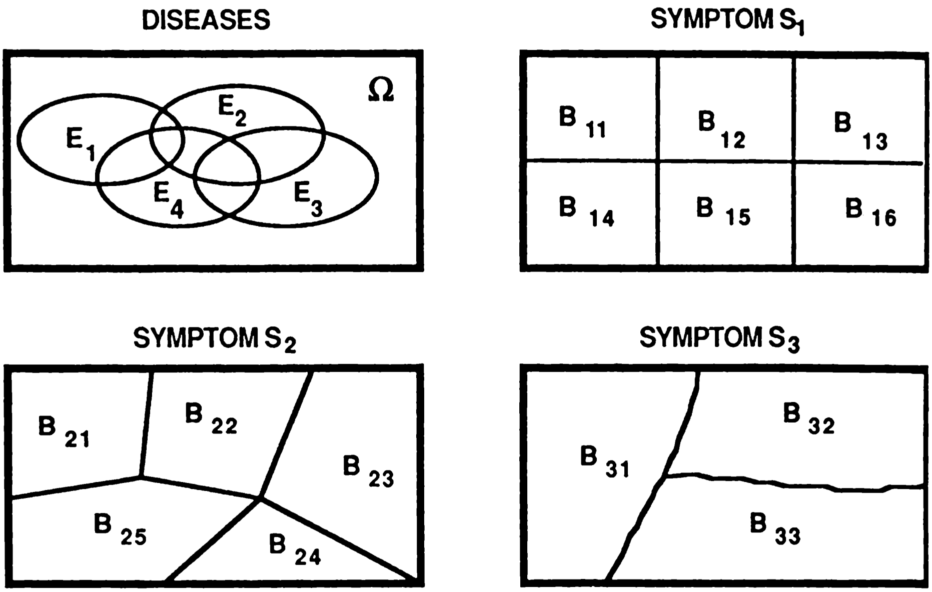

FIGURE Graphycal representation of diseases and symptoms

The role of the probability $P(A_1 \cap A_2 \cap \ldots \cap A_m)$ is just to act as a normalizing constant, and a decision based on the largest value of $P(E_i \cap A_1 \cap A_2 \cap \ldots \cap A_m)$ is not only equivalent but it simplifies the set of constraints to which parameters must satisfy.

In order to measure how likely is disease E_i in the light of symptoms $A_1 \cap A_2 \cap \ldots \cap A_m$ we shall use likelihood ratios similar to the following

$$V_i = \frac{P(E_i \cap A_1 \cap A_2 \cap \ldots \cap A_m)}{\max_j P(E_j \cap A_1 \cap A_2 \cap \ldots \cap A_m)}; i = 1, 2, \ldots, n \quad (2)$$

Note that they range from zero and one and that the disease with maximum $P(E_i \cap A_1 \cap A_2 \cap \ldots \cap A_m)$ will have a value of one. Note also that if two or more diseases are present the associated values could be close to one.

In the following we shll call **C** the class of sets of the form $(A_1 \cap A_2 \cap \ldots \cap A_m)$. This class includes sets with any level of information about symptoms ranging from no information at all ($A_j = \mathbf{\Omega}$, for $j = 1, 2, \ldots, m$) to complete information about the presence of all symptoms ($A_j = B_{jk}$ for all j).

If all A_j are distinct, there are $\prod_{j=1}^{m} (2^{r_j} - 1)$ different sets in the class Cand all of them can be written as unions of sets of the subclass **D** of **C** sets such that $A_j = B_{jk}$ for all j. Class **D** only includes sets with information about all symptoms. Note that **D** is a partition of $\mathbf{\Omega}$ and its cardinal is $\prod_{j=1}^{m} r_j$. We assume that in the diagnosis problem only sets of **C** are involved.

In order to present the following models we systematically analyze:

a) how many independent parameters exist in each model.

b) what are they (one of the possible choices is selected).

c) what are the feasible regions for those parameters.

d) how the updating of those parameters can be done when new information is available.

In this section we include models which allow the consideration of dependences between symptoms or sets of symptoms. We start with the most general case and then, by assuming independence for those symptoms that are not relevant, we derive some simplified but extremely powerful models.

2.1. General dependence model(GD)

In the most general dependence model we allow for any kind of dependence among symptoms $S_1, S_2, \ldots, S_m$. This implies that we have, for any disease, the following $\prod_{j=1}^{m} r_j$ degrees of freedom or parameters $P(E_i \cap A_1 \cap A_2 \cap \ldots \cap A_m)$where$(A_1 \cap A_2 \cap \ldots \cap A_m) \in \mathbf{D}$.

In order to have updating possibilities when a new patient is known we need the extra parameter N (size of the population). Thus, the total number of parameters become

$$P_{GD} = n \prod_{j=1}^{m} r_j + 1 \tag{3}$$

Expression (3) shows the unfeasibility of the general dependence model for most practical cases. Note, as an example, that for all $r_j = 2, n = 200$ and $m = 300, P_{GD} = 4 \cdot 10^{92}$ and no existing computer could be able to store such amount of information.

The above parameters because of their probability character, are not completely free and they must satisfy the following constraints

$$\left.\begin{array}{l} \sum\limits_{D \in \mathbf{D}} P(E_i \cap D) \leq 1 \\ P(E_i \cap D) \geq 0 \text{ for all } D \in D \end{array}\right\} i = 1, 2, \ldots, n \tag{4}$$

$$N > 0$$

Note that restrictions for different i values are uncoupled. These conditions define the feasible set for parameters in this model. They will play a fundamental role if it is to remain internally consistent. The updating of parameters, when a new patient with symptoms $B = A_2^+ \cap A_2^+ \cap \ldots \cap A_m^+ \in C$ and given diseases (D) is known can be done by

$$P^*(E_i \cap A) =$$

$$\begin{cases} P(E_i \cap A)[N + 1/P(E_i \cap B)]/(N+1) \text{if} i \in D; A \in D^+ \\ [P(E_i \cap A)N]/(N+1) \quad \text{otherwise} \end{cases}$$

$$N^* = N + 1 \tag{5}$$

where $B = A_1^+ \cap A_2^+ \cap \ldots \cap A_m$, $D = \{j/\text{the patient has disease } j\}$, $\mathbf{D}^+ = \{A_1 \cap A_2 \cap \ldots \cap A_m \in \mathbf{D}/A_j \subset A_j^+ \text{ for all } j\}$ and the asterisks in P and N refer to updated values.

Note that

$$P(E_i \cap A_1^+ \cap A_2^+ \cap \ldots \cap A_m^+) = \sum_{A \in D^+} P(E_i \cap A)$$

It is worthwhile mentioning that updated parameters in (5) satisfy constraints (4) if initial parameters do.

2.2. Dependence on relevant symptoms model (DR)

In order to reduce the huge number of parameters of the general dependence model, we allow for full dependence only to the so called relevant symptoms for a given disease, i.e. we select for each disease E_i a subset of symptoms for a given disease, i.e. we select for each disease E_i a subset of symptoms $\{S_j; j \in Q_i\} = \{k_1^i, k_2^i, \ldots, k_{l_i}^i\}$ (those which are relevant for the identification of that disease) for which all dependences are allowed and we assume that other symptoms are independent of this set for this disease. In other words, we assume

$$P(E_i \cap A_1 \cap A_2 \cap \ldots \cap A_m) = P(E_i \cap (\bigcap_{j \in Q_i} A_j)) \prod_{j \notin Q_i} r_{ij} \tag{7}$$

where

$$r_{ij} = P(A_j/E_i) = P(B_{jk}/E_i) = p_{ijk} \quad \text{if} \quad A_j = B_{jk} \tag{8}$$

and $p_{ijk} \in [0,1]$.

In the following we shall call $\mathbf{D}_j$ to the subclass of $\mathbf{C}$, such that $A_j \in S_j$ if $j \in Q_i$ and $A_j = \Omega$ if $j \notin Q_i$. Note that this class is a partition of Ω with cardinal $\prod_{j=1}^{l_i} r_{k_j^i}$ and that it includes only sets with complete information about all relevant symptoms to E_i.

Accordingly, the parameters of the model are those of the form $P(E_i \cap D)$ with $D \in \mathbf{D}_i, p_{ijk} (i = 1,2,\ldots,n; k = 1,2,\ldots,r_j; j \notin Q_i)$ and N.

They make a total of

$$P_{DR} = \sum_{i=1}^{n} \left[\prod_{j=1}^{l_i} r_{k_j^i} + \sum_{j \notin Q_i} r_j \right] + 1 \tag{9}$$

The constraints defining the feasible set for parameters are

$$\left.\begin{array}{l} \sum_{D \in \mathbf{D}_i} P(E_i \cap D) \leq 1 \\ P(E_i \cap D) \geq 0; \text{ for all } D \in \mathbf{D}_i \\ \left.\begin{array}{l} \sum_{k=1}^{r_j} p_{ijk} = 1 \\ p_{ijk} \geq 0 \end{array}\right\} k = 1,2,\ldots,r_j \; j \notin Q_i \end{array}\right\} i = 1,2,\ldots,n$$

$$N > 0$$

Consistent updating of parameters can be done by means of

$$P^*(E_i \cap (\underset{j \in Q_i}{\cap A_j})) =$$

$$\begin{cases} [P(E_i \cap (\underset{j \in Q_i}{\cap} A_j))[N + 1/P(E_i \cap (\underset{j \in Q_i}{\cap} A_j^+))]/(N+1) \\ \qquad\qquad \text{if } i \in D \ \text{ for all } j \in Q_i A_j \subset A_j^+) \\ [P(E_i \cap (\underset{j \in Q_i}{\cap} A_j))N]/(N+1) \quad \text{otherwise} \end{cases} \tag{11}$$

$$P^*_{ijk} =$$

$$\begin{cases} [p_{ijk}(NP(E_i) + P(E_i)/P(E_i \cap A_j^+))]/(NP(E_i)+1) \\ \qquad\qquad \text{if } i \in D \text{ and } A_j \subset A_j^+ \\ p_{ijk} NP(E_i)/[NP(E_i)+1] \text{ if } i \in D \text{ and } A_j \subset A_j^+ \\ p_{ijk} \text{ otherwise} \end{cases}$$

$N^* = N + 1$

where $P(E_i) = \sum_{D \in \mathbf{D}_i} P(E_i \cap D)$

Note that $P(E_i \cap A_j^+) = \sum_{A \subset A_j^+} P(E_i \cap A)$.

The above model can be simplified by assuming the following log-linear models for $P(E_i \cap D)(D \in \mathbf{D}_i, i = 1, 2, \ldots, n)$, i.e.

$$\log(P(E_i \cap (\underset{j \in Q_i}{\cap} A_j)) = u_i + \sum_{k \in \mathbf{K}_i} v_{ik} \tag{12}$$

where u_i and $v_{ik}(k \in \mathbf{K}_i; i = 1, 2, \ldots, n)$ are constants and the cardinals, m_i, of $\mathbf{K}_i(i = 1, 2, \ldots, n)$ depends on the selected log-linear models.

Now, expressions (7) and (9) can be written

$$logP(E_i \cap A_1 \cap A_2 \cap \ldots \cap A_m) =$$

$$= log[P(E_i \cap (\bigcap_{j \in Q_i} A_j))] + \sum_{j \notin Q_i} log r_{ij} = u_i + \sum_{k \in \mathbf{K}_i} v_{ik} + \sum_{j \notin Q_i} log r_{ij} \tag{13}$$

$$P_{DR} = \sum_{i=1}^{n} [m_i + \sum_{j \notin Q_i} r_j] + 1 \tag{14}$$

and the first two restrictions in (10) reduce to

$$U_i \leq -log\{ \sum_{D \in \mathbf{D}_i} exp[\sum_{k \in \mathbf{K}_i} v_{ik}]\} \tag{15}$$

Note that the log-linear model is excellent for parametric and structural learning. In fact, if enough data is available, the usual model selection and estimation procedures allows for realistic models. This is an important contribution to existing expert systems.

2.3. Independence model

If independence between any set of symptoms for a given disease is assumed, we have

$$P(E_i \cap A_1 \cap A_2 \cap \ldots \cap A_m) = P(E_i)P(A_1 \cap A_2 \cap \ldots \cap A_m / E_i) =$$

$$= P(E_i)P(A_1/E_i)P(A_2/E_i)\ldots P(A_m/E_i) \tag{16}$$

which gives an illustration of the contribution of each symptom to the value of $P(E_i \cap A_1 \cap A_2 \cap \ldots \cap A_m)$; when no symptom is known it takes the value $P(E_i)$ and any new symptom, A_j, modifies this value by the factor $P(A_j/E_i)$.

In this model, we have the following parameters: $P(E_i)(i = 1,2,\ldots,n), P(B_{jk}/E_i)(i = 1,2,\ldots,n; k = 1,2,\ldots,r_j,$ $j = 1,2,\ldots,m)$ and N.

The number of parameters of this model is

$$P_G = n\sum_{i=1}^{m} r_j + n + 1 \tag{17}$$

The constraints defining the feasible set of parameters are

$$\left.\begin{array}{c} 0 \leq P(E_i) \leq 1 \\ 0 \leq P(B_{jk}/E_i) \quad k = 1,2,\ldots,r_j, j = 1,2,\ldots,m \\ \sum_{k=1}^{r_j} P(B_{jk}/E_i) = 1 \end{array}\right\} i = 1,2,\ldots,n \tag{18}$$

$$N > 0$$

and consistent updating can be done as follows

$$P^*(E_i) = \begin{cases} [P(E_i)N + 1]/(N+1) & \text{if } i \in D \\ [P(E_i)N]/(N+1) & \text{if } i \notin D \end{cases}$$

$$P^*(B_{jk}/E_i) =$$

$$\begin{cases} P(B_{jk}/E_i)[P(E_i)N + 1/P(A_j^+/E_i)]/(NP(E_i)+1) \\ \qquad\qquad \text{if } i \in D \text{ and } B_{jk} \subset A_j^+ \\ P(B_{jk}/E_i)P(E_i)N/(NP(E_i)+1) \quad \text{if } i \in D \text{ and } B_{jk} \not\subset A_j^+ \\ P(B_{jk}/E_i) \text{ otherwise} \end{cases} \tag{19}$$

$$N^* = N + 1$$

3. A LINEAR PROGRAMMING MODEL VALIDATION TOOL

As mentioned above, coherence of models is a necessary requirement. Once a probabilistic framework has been adopted, the freedom of possible actions is limited by probability axioms.

In this section we concentrate on GD and DR models and we describe a tool able to help experts in giving initial coherent parameters or rules. This tool is an important component of the expert system shell.

Assuming one of the dependence models in secion 2 has been selected, the aim of the knowledge engineer would be to get from experts the maximum of information about its parameters. The most direct way of doing this work is by asking the expert about their values. However, in many occasions, they are difficult of being answered and they prefer to answer alternative simpler questions instead. We shall assume in the following that the expert will be asked about exact values or intervals containing exact values of $P(E_i \cap A)$ or $P(E_i \cap A)/P(E_i \cap B)$ where A and B belong to C.

Note that these values are linear combinations of the parameter values. Note also that conditional probabilities are of the second form. Thus, they are included as possible questions.

Therefore, the expert knowledge will be recorded as a set of restrictions of the type

$$b \leq P(E_i \cap A) \leq c \tag{20}$$

or

$$b \leq P(E_i \cap A)/P(E_i \cap B) \leq c \tag{21}$$

where a, b and c are real constants in the interval [0,1], and we assume $P(E_i \cap B) > 0$.

Note that for the GD model inequalitis (20) and (21) can be written as inequalities of linear combinations of the parameters of the model, i.e.

$$\mathbf{b} \leq \sum_{D \in \mathbf{D}} \alpha(D) P(E_j \cap D) \leq c \tag{22}$$

where

$$\alpha(D) = \begin{cases} 1 & \text{if } A \cap D \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

and

$$\begin{aligned} &\sum_{D \in \mathbf{D}} \gamma(D) P(E_i \cap D) \leq 0 \\ &\sum_{D \in \mathbf{D}} \delta(D) P(E_i \cap D) \leq 0 \end{aligned} \tag{23}$$

where

$$\begin{aligned} \gamma(D) &= \begin{cases} 1 & \text{if } A \cap D \neq 0 \text{ and } B \cap D = 0 \\ 1 - c & \text{if } A \cap D \neq 0 \text{ and } B \cap D \neq 0 \\ -c & \text{if } A \cap D = 0 \text{ and } B \cap D \neq 0 \\ 0 & \text{otherwise} \end{cases} \\ \delta(D) &= \begin{cases} -1 & \text{if } A \cap D \neq 0 \text{ and } B \cap D = 0 \\ b - 1 & \text{if } A \cap D \neq 0 \text{ and } B \cap D \neq 0 \\ b & \text{if } A \cap D = 0 \text{ and } B \cap D \neq 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{24}$$

The set of constraints above, together with those in section 2 (we shall call these initial constraints), define the set of feasible values for parameters once the knowledge adquisition from the expert has taken place.

If the expert knowledge is not coherent, the above constraints will lead to no feasible solution and he/she must be forced to redefine constraints. In order to avoid this problem, before giving each piece of information (constraint), the expert can be informed about the maximum and minimum values leading to feasible solutions. If the probability $P(E_i \cap A)$ is required from the expert at this stage, this can be done by solving the two linear programming problems

$$\underset{P(E_i \cap D)}{\text{Maximize}} \quad \sum_{D \in \mathbf{D}} \alpha(D) P(E_i \cap D)$$

and

$$\underset{P(E_i \cap D)}{\text{Minimize}} \quad \sum_{D \in \mathbf{D}} \alpha(D) P(E_i \cap D)$$

subject to the initial plus the additonal constraints (those defining the expert previously given information).

If, on the contrary $P(E_i \cap A)/P(E_i \cap B)$ is required, we need to solve the two non-linear programming problems

$$\underset{P(E_i \cap D)}{\text{Maximize}} \quad \sum_{D \in \mathbf{D}} \alpha(D) P(E_i \cap D) / \sum_{D \in \mathbf{D}} \beta(D) P(E_i \cap D)$$

and

$$\underset{P(E_i \cap D)}{\text{Minimize}} \quad \sum_{D \in \mathbf{D}} \alpha(D) P(E_i \cap D) / \sum_{D \in \mathbf{D}} \beta(D) P(E_i \cap D)$$

where

$$\beta(D) = \begin{cases} 1 & \text{if } B \cap D \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

subject to the same constraints.

Even though the last two problems are non-linear, they can be reduced to a series of linear programming problems (parametric linear programming).

All the above refers to the GD model. For the case of the DR model, expressions (22) and (23) are not linear any more in all parameters. However, they remain linear in the relevant symptoms $P(E_i \cap D)$. So, if p_{ijk} are assumed fixed and the expert is not allowed to change them, all the above is valid for the DR model too, i.e. we have a linear programming problem.

4. A LINEAR PROGRAMMING INFERENCE ENGINE

The tool just described is a true inference engine. In fact, questions formulated to the expert system can be stated in the form of simple or conditional probabilities. We remind the reader that diagnosis is based $P(E_i \cap A_1 \cap A_2 \cap ... \cap A_m)$, which can be written as a linear function of parameters. Thus, its maximization and minimization under costraints will lead to sharp upper and lower bounds. If the users feel unconfortable with intervals and they pfefer point estimates, degenerated (zero-whidth) intervals can be used.

The size of the associated linear programming problem is closely related to the number of parameters of the model. Thus, if the latter is reasonably low, so will be the first too.

Note that upper and lower bounds (25) and (26) (or (27) and (28)) are extremely useful because they give a valuable information not only about the uncertainty of $P(E_i \cap A)$ or $P(E_i \cap A)/P(E_i \cap B)$ but on the associated ignorance. Uncertainty reveals as upper and lower bounds far from unity and ingnorance by means of large differences between both bounds.

Correct diagnosis requires certanity (bounds close to one) and lack of ignorance (bounds close together). The appearance of wide intervals during the inference process is an indication of a poor definition of the probability measure. However, this fact can be unavoidable in many practical cases and the decision maker must be used to deal with this situation. The appearance of probability values far from unity reveals lack of information (not enough symptoms are known for a diagnosis), or an incorrect discrimination power of the probability structure (selected symptoms do not allow a correct discrimination among diseases or goals).

5. LEARNING ABOUT STRUCTURE

Apart from the simplified log-linear model case, testing for structures can be easily performed with the models described in section 2 by means of the likelihood ratio test.

In order to test the quality of a given model as opposed to a more general model, it is enough to estimate, by the maximum likelihood method, the parameters of both models and to calculate the likelihood ratio, i.e. assuming M_1 and M_2 are the two models with r_1 and r_2 parameters, respectively, and such that model M_2 generalizes model M_1, we calculatae the ratio

$$I = \frac{\max\limits_{M_2} \sum_{j=1}^{n} P(E_j \cap A_{1j} \cap A_{2j} \cap \ldots \cap A_{mj})}{\max\limits_{M_1} \sum_{j=1}^{n} P(E_j \cap A_{1j} \cap A_{2j} \cap \ldots \cap A_{mj})} \tag{25}$$

where n is the size of the sample (number of patients with symptoms and diseases known) and the maximizations are with respect to the parameters of models M_1 and M_2, respectively, and subject to their repective constraints. To this end, data must be conveniently recorded.

Significance levels can then be calculated by the $\mathcal{X}^2$ with the corresponding degrees of freedom. ($-2log$I converges in probability to a $\mathcal{X}^2(r_1 - r_2)$).

6. MECHANISM FOR REASONING EXPLANATION

One of the most important and appreciated properties of one expert system is its reasoning explanation capability. While explanation in rule based expert systems is based on active rules, explanation in probabilistic expert systems relies upon whether or not relevant symptoms are present in the patient. For high likelihood diseases (diseases with a likelihood larger than a given threshold value a) all the symptoms favoring that disease can be listed. Note that not only relevant symptoms which are known to be present in the patient but all nonrelevant symptoms not present are included. For low likelihood diseases (diseases with a likelihood smaller than a) all symptoms favoring other diseases are listed.

7. ONE APPLICATON TO MEDICAL DIANOSIS

The above model for binary symptoms ($r_j = 2, j = 1, 2, \ldots, m$) has been implemented on an Apple-Maintosh Plus computer and utilized in order to aid the diagnosis of common and uncommon diseases by general practitioners, students and nurses. An initial example of 90 diseases,/140 binary symptoms and a maximum of 10 relevant symptoms/disease, making a total of 32022 parameters has been tested. Implementation of the model has been done in Lightspeed-Pascal and windows, alert and dialog boxes, pull down menus and selfexplanatory instructions, in english and spanish, have been included in order to get a friendly userinterface.

One of the pull down menus permits the access to the information of all diseases. Thus, at any time, the user can be informed of the relavant symptoms of any given disease. Similarly, given a symptom, a list of all diseases with that relevant symptom can be obtained.

One of the most important steps in order to have precise diagnosis is the initial definition of parameter values. Due to the great flexibility of the above model the number of parameters becomes very high. Thus, the initial definition becomes time consuming and requires an extremely careful work of the human expert and knowledge engineer. In order to facilitate this work the shell includes:

a) one tool controling the coherence of the parameter values given by the expert. Before any new parameter is given, the expert system calculates the minimum and maximum values that make the model coherent, inform the human expert and controls that the new information satisfies these constraints.

b) one option allowing the self-initialization of the independence model.

Input of the symptoms of a given patient can be easily done by using mouse operations only. By recursively pushing the mouse button the system understands yes, no or unknown to the selected symptom in the displayed list.

The inference engine has been optimized and is extremely efficient, giving an ordered list of the most likely diseases for given symptoms in a period ranging from 1 to 3 seconds. If the probability of the most likely disease is not close to one, the expert system ask for new information leading to a correct decision in only a few steps.

The model allows working either with probabilities or likelihood ratios (see expression (2)). If only one disease is expected in the patient probabilities are more convenient. If, on the contrary, more than one disease is possible the likelihoods give better advise.

Once the automatic diagnosis has been performed, the expert system can be required to give the adequate explanation for its decisions. The method described in figure 2 has been implemented, i.e. if the expert system has selected the disease, the list of symptoms favoring the decision is given. On the contrary, if the expert system excludes the disease, the symptoms contradicting that decision are included in the list.

The parameter learning formulas (11) have been included and allow the updating of parameters, when new patients with given symptoms and diseases are known, to be done.

Due to the fact that the shell includes a simulator, simulation, education and training becomes extremely easy. One of the pull down menus includes an option for simulating a patient with a random disease (either all or a random subset of the relevant symptoms for this disease are allowed). By consulting the simulated symptoms, the doctor, student or nurse can try a diagnosis of the disease and then make a comparison with the expert system decision and the actual disease, that can be known at any time.

Finally, new operational languages (french, german, italian, etc.) can be easily implemented by only modifying the associated resources and data (no modification of the program is needed).

8. REFERENCES

ADAMS, J.B. (1976). A Probability Model of Medical and the Mycin Model. Mathematical Biosciences, 32, 177-186.

CHEESEMAN, P. (1985). In Defense of Probability. IJCAI-85, 1002-1009.

LINDLEY, D.V. (1987). The Probability Approach to the Treatment of Uncertainty in Artificial Intelligence and Expert Systems. Statistical Science, Vol. 2,No. 1,17-24.

ZADEH, L.A. (1983). The Role of Fuzzy Logic in the Management of Uncertainty in Expert Systems. Fuzzy Sets and Systems 11,199-227.

Received 08.05.1988.