26:198:722 Expert Systems

- Representing Uncertainty
- Certainty factors - MYCIN
- PROSPECTOR

Representing Uncertainty

- Sources of Uncertainty
  - Unreliable sources of data and information
  - Abundance of irrelevant data
  - Imprecision of language and perception
  - Lack of understanding
  - Faulty equipment
  - Conflicting sources of data
  - Hidden or unknown variables
  - Unknown or poorly specified rules or procedures
  - Data difficult or expensive to obtain

Representing Uncertainty

- Inexact methods
  - Exact methods are not known
  - Exact methods are impracticable
- Epistemological adequacy
  - Interaction of probabilities with quantifiers
  - Probabilities require information that is not available
- Also
  - Imprecise or vague terms not handled
  - Too many numbers
  - Expensive, intractable

Representing Uncertainty

- Conditional probability

$$\mathrm{P}(d \mid s) = \frac{\mathrm{P}(d \,\&\, s)}{\mathrm{P}(s)}$$

- Bayes' Rule

$$\mathrm{P}(d \mid s) = \frac{\mathrm{P}(s \mid d) \cdot \mathrm{P}(d)}{\mathrm{P}(s)}$$

Representing Uncertainty

- Criticisms of relevance and applicability of *objective* probabilities (based on long-run frequencies)
- Consideration of *subjective* probabilities
  - Bayesian updating important here
  - Subjective probabilities must exhibit
    - Coherence
    - Total Evidence
    - Conditionalization
  - *In practice bounded rationality makes this difficult*

Representing Uncertainty

- The more general form of Bayes' rule

$$\mathrm{P}\left(d \middle| s_1 \,\&\ldots\&\, s_k\right) = \frac{\mathrm{P}\left(s_1 \,\&\ldots\&\, s_k \middle| d\right)\cdot \mathrm{P}\left(d\right)}{\mathrm{P}\left(s_1 \,\&\ldots\&\, s_k\right)}$$

  requires computation of $(mn)^k + m + n^k$ probabilities (for m diseases and n symptoms)

- Tractability requires independence assumptions

Representing Uncertainty

- Probability theory thus leaves us with a trade-off
  - assume data are independent
    - fewer numbers
    - simpler calculations
    - sacrifice accuracy
  - track dependencies
    - pay computational price

Representing Uncertainty

- Kahneman & Tversky etc.
  - Humans are poor Bayesian reasoners
  - Discount prior odds
  - Recency effects
  - Over-confident in judgments
  - Poor understanding of sampling theory

- *N.B. Constructive probabilities*

Representing Uncertainty

- Vagueness and possibility
  - Fuzzy set theory
    - crisp sets
    - fuzzy sets
    - degrees of membership
    - relates to many-valued logic

Representing Uncertainty

- Vagueness and possibility
  - Fuzzy logic
    - min for conjunction
    - max for disjunction
    - commutative
    - associative
    - mutually distributive
    - compositional
  - Possibility theory
    - precise questions - imprecise knowledge

Certainty Factors

- Designed originally for use in MYCIN
- CF: {propositions} --> [-1, +1]
  - CF(X) = 1   X is certainly true
  - CF(X) = -1   X is certainly false
  - CF(X) = 0   X is entirely unknown
- Generally:
  CF(action)= CF(rule) x CF (Premise)

Certainty Factors

- As applied in MYCIN
  - IF patient has symptoms s_1 & . . . & s_k
    and background conditions t_1 & . . . & t_m
    THEN conclude patient has disease d_i
    with certainty τ
- Background knowledge constrains application of the rules
- Buchanan & Shortcliffe argue that rigorous application of Bayes' rule would not be more accurate because conditional probabilities are subjective
- They intend CFs and their associated manipulations as *approximations* of probabilistic reasoning

Certainty Factors

- Computation of certainty factors is *modular* (Pearl)
  - i.e., we don't need to consider information not contained in the rule
  - conditional probabilities are not modular in this sense
  - thus, when A is true, we cannot conclude $\mathrm{P}(B)=\tau$ from $\mathrm{P}(B|A)=\tau$ unless A is all that we know
  - otherwise, if we acquire additional knowledge E, we may need to consider $\mathrm{P}(B|A,E)$

Certainty Factors

- In order to combine support provided by two different rules, Shortcliffe & Buchanan looked for a method that was
  - commutative
    - independent of order of firing
  - asymptotic
    - certainty arises only from an absolute proof
- Note also the argument in S & B (1975) that imperfect evidence in favor of a hypothesis is not to be construed as evidence against it

Certainty Factors

- This is expressed rather more formally:
  $\mathrm{C}[h,e] \neq 1 - \mathrm{C}[\neg h,e]$
  confirmation is not 1 - disconfirmation
- This is an idea we will re-visit e.g. when we consider Dempster-Shafer Belief Functions and their potential application in auditing

Certainty Factors

- Measure of Belief
  - the measure of increased belief in the hypothesis h, based on the evidence e, is x

    $$\mathrm{MB}[h,e] = x$$

- Measure of Disbelief
  - the measure of increased disbelief in the hypothesis h, based on the evidence e, is y

    $$\mathrm{MD}[h,e] = y$$

Certainty Factors

- Formal definitions in terms of probability

$$\mathrm{MB}[h,e]=\begin{cases}1 & \text{if } \mathrm{P}(h)=1\\ \dfrac{\max\left[\mathrm{P}(h),\mathrm{P}(h|e)\right]-\mathrm{P}(h)}{\max[1,0]-\mathrm{P}(h)} & \text{otherwise}\end{cases}$$

$$\mathrm{MD}[h,e]=\begin{cases}1 & \text{if } \mathrm{P}(h)=0\\ \dfrac{\min\left[\mathrm{P}(h),\mathrm{P}(h|e)\right]-\mathrm{P}(h)}{\min[1,0]-\mathrm{P}(h)} & \text{otherwise}\end{cases}$$

$$\mathrm{CF}[h,e]=\mathrm{MB}[h,e]-\mathrm{MD}[h,e]$$

Certainty Factors

- Characteristics

$$0 \le \mathrm{MB}[h,e] \le 1, \quad 0 \le \mathrm{MD}[h,e] \le 1, \quad -1 \le \mathrm{CF}[h,e] \le 1$$

$$\text{If } \mathrm{P}\left[h \middle| e\right] = 1$$

$$\mathrm{MB}[h,e] = 1, \quad \mathrm{MD}[h,e] = 0, \quad \mathrm{CF}[h,e] = 1$$

$$\text{If } \mathrm{P}\left[\neg h \middle| e\right] = 1$$

$$\mathrm{MB}[h,e] = 0, \quad \mathrm{MD}[h,e] = 1, \quad \mathrm{CF}[h,e] = -1$$

$\mathrm{MB}[h,e] = 0$ if h is not confirmed by e
$\mathrm{MD}[h,e] = 0$ if h is not disconfirmed by e
$\mathrm{CF}[h,e] = 0$ if h is neither confirmed nor disconfirmed by e

Certainty Factors

- CF as defined here has the desired property
  - confirmation is not 1 - disconfirmation
- In fact
  - confirmation + disconfirmation = 0
- CF judgments must be elicited carefully from experts to ensure that they respect the constraints implied by these formal definitions

Certainty Factors

- Defining criteria
  - Limits

$$\mathrm{MB}[h,e+]\to 1,\quad \mathrm{MD}[h,e-]\to 1,$$

$$\mathrm{CF}[h,e-]\le \mathrm{CF}[h,e-\&e+]\le \mathrm{CF}[h,e+]$$

  - Absolutes

$$\mathrm{MB}[h,e+]=1\Rightarrow \mathrm{MD}[h,e-]=0$$

$$\mathrm{MD}[h,e-]=1\Rightarrow \mathrm{MB}[h,e+]=0$$

$$\mathrm{MB}[h,e-]=\mathrm{MD}[h,e-]\,\mathrm{is undefined}$$

Certainty Factors

- Defining criteria
  - Commutativity

$$\mathrm{MB}[h, s_1 \,\&\, s_2] = \mathrm{MB}[h, s_2 \,\&\, s_1]$$

$$\mathrm{MD}[h, s_1 \,\&\, s_2] = \mathrm{MD}[h, s_2 \,\&\, s_1]$$

$$\mathrm{CF}[h, s_1 \,\&\, s_2] = \mathrm{CF}[h, s_2 \,\&\, s_1]$$

  - Missing information

$$\mathrm{MB}[h, s_1 \,\&\, s_?] = \mathrm{MB}[h, s_1]$$

$$\mathrm{MD}[h, s_1 \,\&\, s_?] = \mathrm{MD}[h, s_1]$$

$$\mathrm{CF}[h, s_1 \,\&\, s_?] = \mathrm{CF}[h, s_1]$$

Certainty Factors

- Combining functions
  - Incrementally acquired evidence

$$\mathrm{MB}[h, s_1 \,\&\, s_2] = \begin{cases} 0 & \text{if } \mathrm{MD}[h, s_1 \,\&\, s_2] = 1 \\ \mathrm{MB}[h, s_1] + \mathrm{MB}[h, s_2] \cdot (1 - \mathrm{MB}[h, s_1]) & \text{otherwise} \end{cases}$$

$$\mathrm{MD}[h, s_1 \,\&\, s_2] = \begin{cases} 0 & \text{if } \mathrm{MB}[h, s_1 \,\&\, s_2] = 1 \\ \mathrm{MD}[h, s_1] + \mathrm{MD}[h, s_2] \cdot (1 - \mathrm{MD}[h, s_1]) & \text{otherwise} \end{cases}$$

Certainty Factors

- Combining functions
  - Conjunctions of hypotheses

$$\mathrm{MB}[h_1 \,\&\, h_2, e] = \min\left(\mathrm{MB}[h_1, e], \mathrm{MB}[h_2, e]\right)$$

$$\mathrm{MD}[h_1 \,\&\, h_2, e] = \max\left(\mathrm{MD}[h_1, e], \mathrm{MD}[h_2, e]\right)$$

  - Disjunctions of hypotheses

$$\mathrm{MB}[h_1 \vee h_2, e] = \max\left(\mathrm{MB}[h_1, e], \mathrm{MB}[h_2, e]\right)$$

$$\mathrm{MD}[h_1 \vee h_2, e] = \min\left(\mathrm{MD}[h_1, e], \mathrm{MD}[h_2, e]\right)$$

Certainty Factors

- Strength of evidence
  - Suppose evidence s_1 is not known with certainty, but a CF based on prior evidence e is known. If MB' and MD' are the degrees of belief and disbelief when s_1 is known with certainty, then the actual degrees of belief and disbelief are given by

$$\mathrm{MB}[h, s_1] = \mathrm{MB}'[h, s_1] \cdot \max(0, \mathrm{CF}[h, s_1])$$

$$\mathrm{MD}[h, s_1] = \mathrm{MD}'[h, s_1] \cdot \max(0, \mathrm{CF}[h, s_1])$$

Certainty Factors

- Note that in S & B (1975) MYCIN computes and maintains MBs and MDs separately, only computing CFs at the end, although CFs are then used to generate recommendations
- This differs from “simplifed” explanation e.g. in Durkin Chapter 12

Certainty Factors

- In accordance with the limiting properties, multiple items of confirming evidence will result in MB --> 1 (say, 0.99)
- Suppose, however, we have a single item of disconfirming evidence with MD = 0.8
- Then CF = MB - MD = 0.19, i.e., many sources of confirmation have been almost completely offset by a single disconfirming item

Certainty Factors

- To de-sensitize this effect, the definition of CF was subsequently modified to

$$\mathrm{CF}[h,e]=\frac{\mathrm{MB}[h,e]-\mathrm{MD}[h,e]}{1-\min\left[\mathrm{MB}[h,e],\mathrm{MD}[h,e]\right]}$$

- Using this definition, the CF for our example becomes $\frac{0.99-0.8}{1-\min[0.99,0.8]}=\frac{.19}{.20}=0.95$

Certainty Factors

- If we are only interested in updating CFs without retaining MBs and MDs, we can perform incremental updating using

$$\mathrm{CF}_{COMBINE} = \begin{cases} \mathrm{CF}_1 + \mathrm{CF}_2 \cdot \left(1 - \mathrm{CF}_1\right) & \text{if both} > 0 \\ \mathrm{CF}_1 + \mathrm{CF}_2 \cdot \left(1 + \mathrm{CF}_1\right) & \text{if both} < 0 \\ \dfrac{\mathrm{CF}_1 + \mathrm{CF}_2}{1 - \min\left(\left|\mathrm{CF}_1\right|, \left|\mathrm{CF}_2\right|\right)} & \text{otherwise} \end{cases}$$

Certainty Factors

- CFs may be used
  - to direct a best-first search
  - to control search explicitly
  - to prune the search
    - e.g., to drop goals with when their CFs fall within the range [-0.2, +0.2]
  - to rank order hypotheses

Certainty Factors

- Durkin recommends
  - Obtain CFs from expert's use of qualified terms
  - Don't elicit CFs directly
  - Avoid deep inference chains (because approximate departs increasingly from probabilistic values)
  - Avoid many rules with the same hypothesis
  - Avoid rules with many premises - split into multiple rules

Certainty Factors

- Adam (1976) criticized certainty factors
  - CF associated with a hypothesis by MYCIN does not correspond to a simple probability model based on Bayes' rule
    - did S & B (1975) claim that it did?
  - Degrees of belief from different evidence cannot always be chosen independently
    - e.g., absolute diagnostic indicators
  - min and max are not always appropriate for conjunctions
    - e.g., mutually exclusive alternatives

Certainty Factors

- CF ranking may reverse probability ranking
  - Suppose $P(h_1)=0.8 \qquad P(h_2)=0.2$

    $P(h_1|e)=0.9 \quad P(h_2|e)=0.8$
  - Note $P(h_1|e)=0.9 > P(h_2|e)=0.8$
  - But $CF(h_1,e)=\dfrac{P(h_1|e)-P(h_1)}{1-P(h_1)}=\dfrac{0.9-0.8}{0.2}=0.5$

    $CF(h_2,e)=\dfrac{P(h_2|e)-P(h_2)}{1-P(h_2)}=\dfrac{0.8-0.2}{0.8}=0.75$
  - Hence $CF(h_1,e) < CF(h_2,e)$

Certainty Factors

- Transitivity across chains of reasoning is not generally valid
- CFs are defined from MBs and MDs in terms of *increases* or *decreases* in belief, but elicited for MYCIN as *absolute values*

Certainty Factors

- Heckerman (1986)
  - Provides an example to show that the S & B (1975) definition of CFs, in conjunction with the rules for combining (incremental updating), lead to non-commutativity
  - His conclusion from this is that we should take desirable properties of CFs as axiomatic, retain the combination rules, and seek an alternative formulation of CFs in probabilistic terms

Certainty Factors

- Heckerman (1986)
  - Axiomatizes the “desiderata” for certainty factors using a somewhat modified (simplified) notation, but formally conditioning on prior evidence,
  - Exhibits an example of non-commutativity
  - States a formal requirement for a probabilistic interpretation of CFs
  - Gives the odds-likelihood form of Bayes’ Theorem

$$\mathrm{O}\left(h \mid e, e_p\right) = \frac{\mathrm{P}\left(e \mid h, e_p\right)}{\mathrm{P}\left(e \mid \neg h, e_p\right)} \cdot \mathrm{O}\left(h \mid e_p\right) = \lambda\left(h, e, e_p\right) \cdot \mathrm{O}\left(h \mid e_p\right)$$

Certainty Factors

- Heckerman (1986)
  - Defines conditional independence of e and e_p given H and $\neg H$
  - Shows that λ is a candidate for a probabilistic interpretationof CFs except that it ranges from 0 to ∞
  - Shows that any monotonic increasing transformation of the likelihood ratio satisfying
    $\mathrm{F}\left(1/x\right) = -\mathrm{F}(x) \quad \text{and} \quad \mathrm{F}(\infty) = 1$
    is a probabilistic interpretation for CFs (and conversely)

Certainty Factors

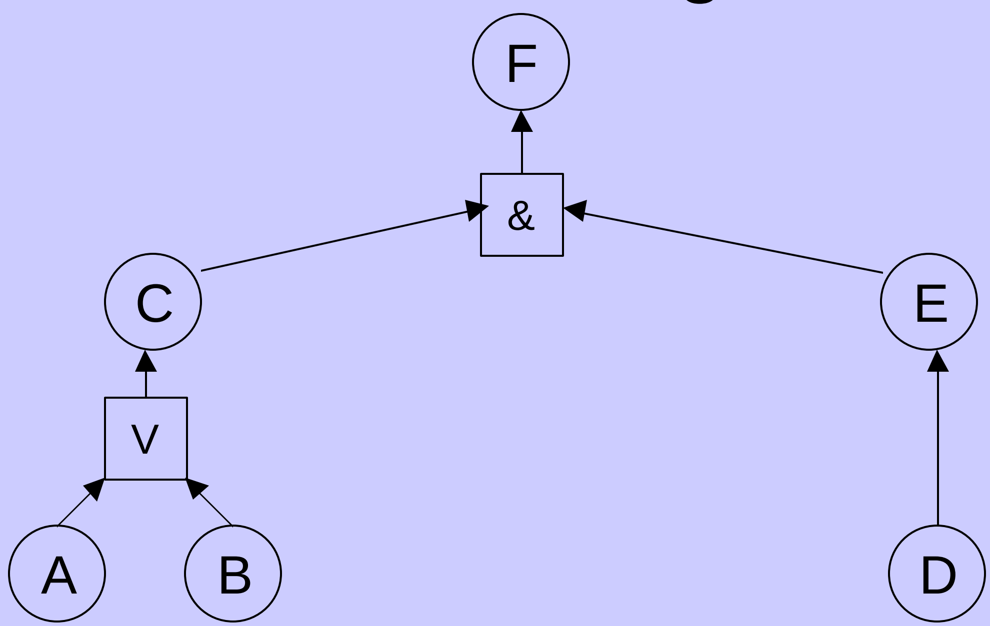
- Heckerman (1986)
  - Gives specific examples of such transformations
  - Observes that evidence combined using the S & B combination functions is required to be conditionally independent given both the hypothesis *and its negation*
  - Argues by example that the latter condition often fails in practice
  - Introduces axioms for sequential combination (corresponding to strength of evidence in S & B)

Certainty Factors

- Heckerman (1986)
  - Shows that these new axioms do not further constrain probabilistic interpretations of CFs
  - Demonstrates that although CFs have been applied to non-tree inference networks, updating is valid only in tree structures (rarely applicable in complex practical situations)

Certainty Factors

- Rules may conveniently be organized as an inference net, e.g.,



Certainty Factors

- Rules:
  - R1: A v B --> C CF = 0.8
  - R2: D --> E CF = 0.7
  - R3: C & E --> F CF = 0.9
- Facts
  - A CF = 0.4
  - B CF = 0.6
  - D CF = 0.9
  - C,E CF = 0
  - F CF = 0.2

Certainty Factors

- CF(A v B) = max(0.4, 0.6) = 0.6
- CF(R1’) = 0.8 x 0.6 = 0.48
- CF(C|A v B) = 0 + 0.48 x (1 - 0) = 0.48
- CF(R2’) = 0.7 x 0.9 = 0.63
- CF(E|D) = 0 + 0.63 x (1 - 0) = 0.63
- CF (C & E) = min (0.48, 0.63) = 0.48
- CF(R3’) = 0.9 x 0.48 = 0.432
- CF(F|C & E) = 0.2 + 0.432 x (1 - 0.2) = 0.5456

PROSPECTOR

- We have already seen

$$O(h|e) = \frac{P(e|h)}{P(e|\neg h)} \cdot O(h) = \lambda(h,e) \cdot O(h)$$

- Now, defining the Likelihood of Sufficiency by

$LS = \frac{P(e|h)}{P(e|\neg h)}$ we can write $O(h|e) = LS \cdot O(h)$

PROSPECTOR

- Similarly, if we define the Likelihood of Necessity by $\mathrm{LN} = \dfrac{\mathrm{P}(\neg e | h)}{\mathrm{P}(\neg e | \neg h)}$

  we can write $\mathrm{O}(h | \neg e) = \mathrm{LN} \cdot \mathrm{O}(h)$

- This enables us to develop rules of the form:

  IF e THEN h (LS, LN)

  with both factors provided by an expert

PROSPECTOR

- Mathematically, we have the constraints

$$\mathrm{LS} > 1 \Rightarrow \mathrm{LN} < 1$$

$$\mathrm{LS} < 1 \Rightarrow \mathrm{LN} > 1$$

$$\mathrm{LS} = 1 \Rightarrow \mathrm{LN} = 1$$

  but real-world problems may contradict this
- More generally, if we are uncertain of e itself, and it depends on observed evidence e', we can make adjustments

PROSPECTOR

- The probability of h given our belief e' is

  $$\mathrm{P}(h|e') = \mathrm{P}(h|e)\cdot\mathrm{P}(e|e') + \mathrm{P}(h|\neg e)\cdot\mathrm{P}(\neg e|e')$$

  from which the following derive

  $$\mathrm{P}(e|e') = \mathrm{P}(e) \Rightarrow \mathrm{P}(h|e') = \mathrm{P}(h)$$

  $$e\,\text{true} \Rightarrow \mathrm{P}(e|e') = 1 \text{ and } \mathrm{P}(h|e') = \mathrm{P}(h|e)$$

  $$e\,\text{false} \Rightarrow \mathrm{P}(\neg e|e') = 1 \text{ and } \mathrm{P}(h|e') = \mathrm{P}(h|\neg e)$$

  which in turn define a linear relationship between $\mathrm{P}(h|e')$ and $\mathrm{P}(e|e')$

PROSPECTOR

- Real-world situations may result in experts providing values that contradict these assumptions, and some adjustment therefore needs to be made
- Duda et al. proposed an *ad hoc* assumption to relate $P(h|e')$ and $P(e|e')$ following a piecewise linear function
- This lead to PROSPECTOR

PROSPECTOR

- PROSPECTOR use two simple functions to avoid inconsistencies:

$$\mathrm{P}(h|e') = \mathrm{P}(h|\neg e) + \frac{\mathrm{P}(e|e')}{\mathrm{P}(e)} \cdot \left(\mathrm{P}(h) - \mathrm{P}(h|\neg e)\right) \quad \text{for } 0 \le \mathrm{P}(e|e') \le \mathrm{P}(e)$$

$$\mathrm{P}(h|e') = \frac{\mathrm{P}(h) - \mathrm{P}(h|e) \cdot \mathrm{P}(e)}{1 - \mathrm{P}(e)} + \mathrm{P}(e|e') \cdot \frac{\mathrm{P}(h|e) - \mathrm{P}(h)}{1 - \mathrm{P}(e)} \quad \text{for } \mathrm{P}(e) \le \mathrm{P}(e|e') \le 1$$

- PROSPECTOR is an expert system that assists geologists in mineral deposit exploration

PROSPECTOR

- A PROSPECTOR network is a set of nodes representing evidence or hypotheses and links connecting the nodes together with uncertain relationships represented by LS or LN values and prior probabilities for the nodes
- Probabilities are propagated upward to the topmost node

PROSPECTOR

- Where multiple nodes affect a single hypothesis, conditional independence is assumed, and rules combine conjunctively or disjunctively
  - Conjunctive rules
    - each e_i is based on the partial evidence e_i'
    - PROSPECTOR assumes $\mathrm{P}\left(e \middle| e'\right) = \min\left\{\mathrm{P}\left(e_i \middle| e'\right)\right\}$
    - the resulting value is combined using the linear function given above
  - Disjunctive rules
    - as above, but using max instead of min

PROSPECTOR

- Updating odds
  - Each time new evidence is provided, the odds are updated, assuming conditional independence

$$O\left(h \middle| e_1', e_2', \ldots, e_n'\right) = \prod_{i=1}^{i=n} \mathrm{LS}_i' \cdot O(h) \text{ where } \mathrm{LS}_i' = \frac{P\left(e_i \middle| h\right)}{P\left(e_i \middle| \neg h\right)}$$

$$O\left(h \middle| \neg e_1', \neg e_2', \ldots, \neg e_n'\right) = \prod_{i=1}^{i=n} \mathrm{LN}_i' \cdot O(h) \text{ where } \mathrm{LN}_i' = \frac{P\left(\neg e_i \middle| h\right)}{P\left(\neg e_i \middle| \neg h\right)}$$

PROSPECTOR

- Beliefs were elicited from users of PROSPECTOR using certainty measures, which were subsequently converted to conditional probabilities using the same piecewise linear approach outlined earlier

PROSPECTOR

- Using probabilities directly is a powerful but challenging technique
  - Probabilities must be known
  - Probabilities must be updated
  - Total probability must equal unity
  - Conditional independence is required

PROSPECTOR

- PROSPECTOR incorporates many simplifying assumptions, but it is still a demanding system
- A large number of probabilities are still typically required to be provided
  - difficult to obtain
  - computationally expensive
- Need to restart when new hypotheses are added: there is no incremental updating
- Such a system is called *intensional* or *global* - by contrast, MYCIN is *extensional* and has a *modular* structure

PROSPECTOR

- Other concerns about the updating methods
  - Rednault et al. (1981)
    - If A and B are intersections of the evidence $e_1 \ldots e_m$, then they are independent
  - Hussain (1972) sought to show
    - for exhaustive and mutually exclusive hypotheses $h_1 \ldots h_n$ and $e_1 \ldots e_m$ conditionally independent, no updating is possible
  - Gymour (1985)
    - gave a counter-example to disprove this
  - Johnson (1986)
    - showed that *multiple* updating of any hypothesis is impossible, i.e., there is at most one piece of evidence for which posteriors not the same as the prior