Chap-8

57

Statistical reasoning

Date ___/___/_____

→ Till now, we have described several repreⁿ techniques that can be used to model beliefs in which, a particular fact is believed to be true or false or not anything.

→ For some kinds of prob.s solving, it is useful to describe beliefs that are not certain but for which there is some supporting evidence. Let us consider two classes of such prob.s.

→ The first class contains prob.s in which there is randomness in the world. For ex., "Playing bridge" is a good example of this class.

→ In this chap., we explore several techniques that can be used to augment knowledge repreⁿ with statistical measure that describe levels of evidence & belief.

8.1 Probability & Baye's theorem Q-14

→ An imp. goal for many prob. solving systems is to collect evidence as the system goes along & modify its behavior on the basis of evidence.

→ To model this behavior, we need a statistical theory of evidence. Bayesian statistics is such a theory.

→ The fundamental notation of Bayesian statistics is "conditional probability:

$$\boxed{P(H \mid E)}$$

Date ___ /___ /______

→ It can be read as:- "Probability of hypothesis 'H' given that we have observed evidence 'E'.

→ To compute this, we need to take into account the prior probability of H and the extent to which 'E' provides evidence of 'H'.

→ Baye's theorem states that,

$$\boxed{P(H_i|E) = \frac{P(E|H_i) \cdot P(H_i)}{\sum_{i=1}^{k} P(E|H_n) \cdot P(H_n)}}$$

$P(H_i|E)$ = The probability that hypo. H_i is true given evidence E.

$P(E|H_i)$ = The prob. that we will observe evidence E given that hypo. 'i' is true.

$P(H_i)$ = The prob. that hypo. i is true is absence of any specific evidence.

k = The no. of possible hypotheses.

→ Second example we can take is of data mining to find class label of given sample X. (age, student any example).

From Data-mining

Date ___ / ___ / ______

→ The key to using Baye's theorem for uncertain reasoning is to recognize exactly what it says. Specially, when we say $P(A|B)$, we describe conditional probability of A given the evidence that is 'B'.

→ Consider foll. assertions:

S : Patient has spots.
M : Patient has measles.
F : Patient has high fever.

→ Without any additional evidence, the presence of "spots" as evidence in "favor" of "measles".

→ It also serves as evidence of ~~feb~~ fever, since measles would cause fever.

→ But, suppose we already know that patient has "measles", then additional evidence "~~sp~~ having spots" doesn't tell anything about likelihood of fever.
(એટલે કે આપણે "measle" માટે sure હોઈએ અને તેને "fever" હોય, તો "having spots" કોઈ additional proof આપતું નથી, b'coz "measles" હોય તો "~~spots~~ fever" હોય, પણ આમાં "spots" નો role ક્યાંય ના આવ્યો)

→ In short, either "spots" or "fever" alone would contain evidence in favor of measles.

Scanned by CamScanner

Date ___/___/___

→ If both are present, we need to take both into account in determining the total weight of evidence.

→ But, since "spots" & "fever" are dependent events, we can't just sum their effects.

→ Instead, we need to represent explicitly the conditional probability that arises from their conjuction.

→ In general, given a prior body of evidence 'e' & some new observation 'E', we need to compute,

$$\boxed{P(H|E, e) = P(H|E) \cdot \frac{P(e|E, H)}{P(e|E)}}$$

8.2 ☆ **Certainty factors & rule-based systems** Q-15

→ Here we describe one practical way of compromising a pure Bayesian system. The approach we discuss here is founded ~~by~~ in "MYCIN" system, which attempts to recommend appropriate therapies for patients with bacterial infections.

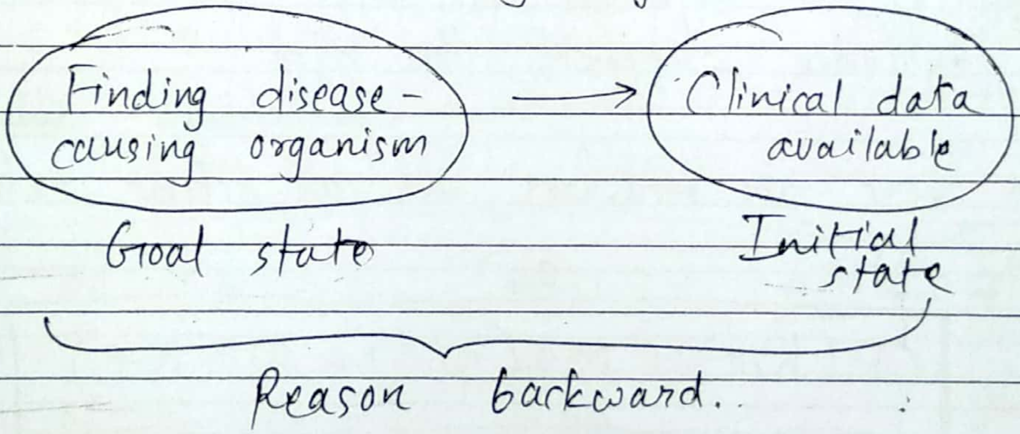
→ It interacts with the physician to acquire clinical data it needs.

→ "MYCIN" is an example of expert system, since it performs a task normally done by a human expert.

→ Here, we focus on the use of "probabilistic reasoning."

→ "MYCIN" represents most of its diagnostic knowledge as a set of rules. Each rule is associated with it a "certainty factor"

→ "MYCIN" uses these rules to reason backward to the clinical data available from its goal finding significant disease-causing organism.



→ Once it finds identities of such organism, it then attempts to select a "therapy" for disease treatment.

→ To understand, "how MYCIN" exploits uncertain info",
we need answers to two questions:-

1. What do certainty factor mean?
2. How does "MYCIN" combines estimates of certainty to produce final estimate of certainty of its conclusion?

→ Let's start first with a simple answer to first question i.e about "certainty factor".
A certainty factor (CF[h,e]) is defined in terms of two components.

- **MB[h,e]** :- A measure betⁿ 0 and 1 of **belief** in hypothesis 'h' given the evidence 'e'. MB measures the extent to which evidence supports the hypothesis. It's zero if evidence fails to support the hypo.

- **MD(h,e)** :- A measure betⁿ 0 & 1 of **disbelief** in 'h' hypo. given the evidence 'e'. MD measures the extent to which the evidence supports the negation of hypo. It is zero if evidence supports the hypo.

→ From these two measures, we can define "certainty factor" as,

$$\boxed{CF(h,e) = \overset{0.8}{MB(h,e)} - \overset{0.2}{MD(h,e)}}$$

→ The measures of belief and disbelief of a hypo. given two observations s_1 and s_2 are computed from

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

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

Date __ / __ / ____

→ We can interprete these formula in this manner :- The measure of belief in 'h' is '0' if 'h' is disbelieved with certainty. Otherwise, the measure of belief in 'h' given two observations S1 & S2 is the measure of belief for given one observation plus some increment for the second observation. The increment is computed by first taking the diff. betn '1' (certainty) and the belief given only the first observation. This diff. is the most that can be added by second observation.

→ A simple example shows how these func^n's operate. Suppose we make an first ~~initial~~ observation that confirms our belief in h with MB = 0.3. Then, $MD[h, S_1] = 0$ and $CF[h, S_1] = 0.3$. Now, we make a second observation, which also confirms 'h', with $MB[h, S_2] = 0.2$. Now:

$$\boxed{\begin{aligned} MB[h, S_1 \wedge S_2] &= 0.3 + (0.2 * 0.7) \\ &= 0.44 \\ MD[h, S_1 \wedge S_2] &= 0.0 \\ CF[h, S_1 \wedge S_2] &= 0.44 \end{aligned}}$$

→ All the above calculation is for several rules all provide evidence that relates to a single hypo. For ex., fig. (8.1) illustrates several rules & all provide evidence that relates to a single hypo.

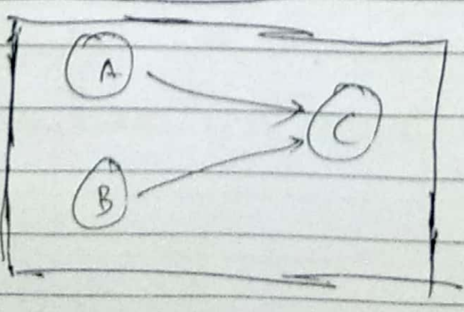


Fig. 8.1

Date ___ / ___ / _____

→ In last section, we studied that pure Bayesian statistics usually leads to intractable systems. But "MYCIN" works. Why?

Ans → Each CF in a MYCIN rule represents the contribution of an individual rule to MYCIN's belief in a hypo. (CF In MYCIN = contribution of individual rules) Sometimes it also represents a "conditional probability", P(H|E).

→ But, in a pure Bayesian system, P(H|E) describes the conditional probability of H ~~only relevant~~ given that only the relevant evidence is E. If there is other evidence, then "join probabilities" need to be considered. This is where MYCIN diverges from pure Bayesian system.

8.3 ☆ Bayesian networks Q-16

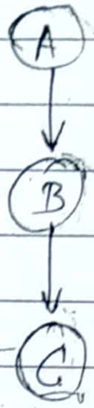
→ In last section, we described CF's as a mechanism for reducing the complexity of a Bayesian reasoning. In this section, we describe an alternative approach. That is Bayesian N/W.

→ The main idea is that, in real world, it is not necessary to use a joint probability table in which we list the probabilities of combⁿs of events. Most events are conditionally independent of most others, so their interactions need not be considered.

Saathi

Date ___ / ___ / ______

Example → Let's consider what happens in scenario like foll. figure :-



A concrete example for this one is :

પાણીનો ફુવારો

S: Sprinkler was on last night.

W: Grass is wet.

R: It rained last night.

→ It is worth pointing out here that this example illustrates one specific rule structure that almost always cause trouble (Problem) and should be avoided.

→ Notice that our first rule describes a "causal relaship" (Sprinkler causes wet grass). The second rule describes an inverse caulality relationship ("wet grass is caused by rain and thus Rain is evidence for its cause). So, both these rules are totally contradictory.

→ To avoid this prob., many rule-based systems should either limit their rules to one structure or clearly partition the two kinds so that they can't interfere with each other. (આનો મતલબ કે કાં તો આપણે S and W વાળી અને બંને or statements limited રાખવાનાં, કાં તો બંને વચ્ચે એક clear partition રાખવાનું, એટલે કે (S, W) અને (W, R) બંને સાથે ના રાખી શકાય). Therefore, we discuss "Bayesian N/w" in this section to describe systematic solⁿ to this problem.

Scanned by CamScanner

Date ___ / ___ / ______

→ Consider again the example of "sprinkler", "rain" and "grass". Fig. (8.2) shows the flow of constraints we described in MYCIN style.

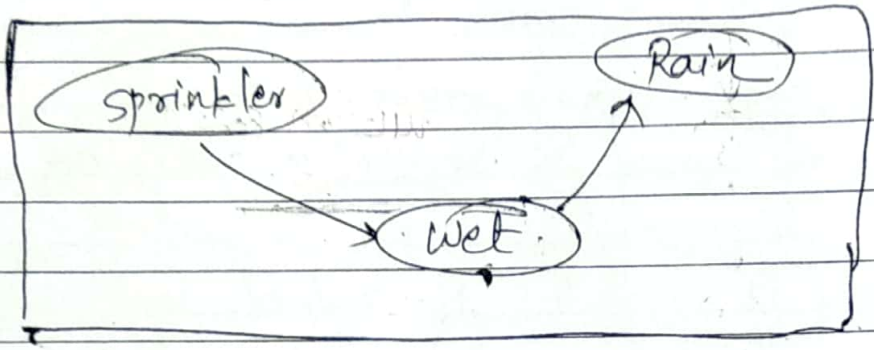


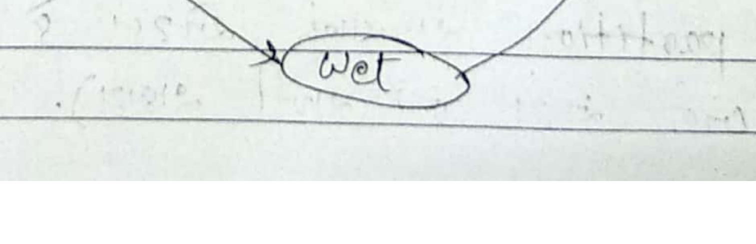
Fig. 8.2

→ The problem that we encountered with that example was the constraint flowed incorrectly from "sprinkler on" to "rained last night".

→ Specially, we construct a directed acyclic graph (DAG) that represents causality relaⁿ among variables. The idea of graph (N/w) has proved to be very useful in several systems specially medical diagnosis systems.

→ In fig. (8.2 (b)), we show a causality graph for the wet grass example. In addition to the three nodes, the graph contains a new node tells us whether it is currently rainy season.

8.2 (b)



Scanned by CamScanner

Date __/__/____

8.4 ☆ Dempster - Shafer theory

→ Till now we have described individual propositions and assign to each of them a point estimate (a single no.) of the degree of belief as per given the evidence.

(Probability)

→ In this section, we consider alternative technique called "Dampster - Shafer theory."

→ This new approach considers sets of propositions and assigns to each of them an interval

Plausibility એટલે, જે ખરેખર સાચું છે તે.

Belief એટલે ધારણા, અથવા evidence પરથી આપણે કહીએ તે.

[Belief, Plausibility]

(Plausibility = Perceptibility
પ્રત્યક્ષ રીતે લાગતું હોય (obvious))

This is an interval, in which degree of belief must lie. Belief (Bel) measures the strength of evidence in favor of a set of propositions. Belief ranges from 0 (indicating no evidence) to 1 (denoting certainty).

→ Plausibility (Pl) is to be

$$\boxed{Pl(S) = 1 - Bel(\neg S)}$$

"Plausibility એટલે 'S'ના opposeમાં જેટલા લોકો હોય, એટલે કે 'S'ના supportમાં ના હોય, તેને 1 માંથી બાદ કરી દઈએ, તો બાકીના લોકો 'S'ના supportમાં જ હોય, એટલે કે 'S'ના favourમાં જ હોય, તો તેને "Plausibility" કહેવાય.

→ This also ranges from 0 to 1 and measures the extent to which evidence in favor of $\neg S$ leaves room for belief in 'S'. For ex., we have "certain" evidence in favor of $\neg S$, then $Bel(\neg S)$ will be 1 and $Pl(S)$ will be 0.

$Bel(S) = 0$

Saathi

Date ___/___/_____

→ Suppose that we are currently considering three competing hypo: A, B & C. If we have ~~info~~ no info., we represent that by saying, that the true probability is in range of [0, 1]. This interval can also shrink (get short & limited).

→ This is totally opposite to pure Baye's approach, in which we begin by distributing prior probability equally among the hypo's and assert for each $P(h) = 0.33$.

→ The interval approach (belief - plausibility) does not have any info. when we start.

→ So far we have talked about "Bel" as a measure of belief in some hypo. given some evidence. Let's now define it more precisely. For ex., we need to start with Baye's theorem with an exhaustive universe of mutually exclusive hypo's. We call this the "frame of discernment" & noted as Θ.

કોઈ રોગનું નિદાન કરવું. Capital theta

→ For ex., in simple diagnosis prob., Θ might consist of set {All, flu, Cold, Pneu}:

> All : Allergy
> Flu : Flu
> Cold : Cold
> Pneu : pneumonia

Date __/__/____ Capital theta

Saathi

→ Our goal is to attach some measure of belief to elements of Θ. However, not all evidence is directly supportive of individual elements ~~to~~ in Θ. For ex., in our diagnosis prob., "fever" might support {Flu, Cold, Pneu}.

→ In a pure Bayesian system, we can list all comb^n's by conditional probabilities. "Dampster-Shafer theory" lets us to handle interactions by manipulating hypos directly.

→ The key func^n we use is probability ~~dist~~ density func^n, which we denote m. The func^n "m" is defined for Θ & its all subsets.

→ If Θ has "n" elements, then there are 2^n subsets of Θ.

→ let's see how "m" works for our diagnosis prob. Assume that we have no info. about how to choose among hypo.s when we start the diagnosis task. Then we define m as:

$\boxed{\{\Theta\} \quad (1.0)}$

All evidences are present then P = 1.0

→ Now suppose we have an evidence that suggest at level 0.6 that the correct diagnosis in the set {Flu, Cold, Pneu}. "Fever" might be such a piece of evidence. So, we update "m" as follow.

{Flue, Cold, Pneu} (0.6)

Θ (0.4)

Page No.

Suppose we are given two belief functions m_1 and m_2. Let X be the set of subsets of Θ to which m_1 assigns a nonzero value and let Y be the corresponding set for m_2.

We define the combination ~~of~~ m_3 of m_1 and m_2.

~~This gives us a new belief function that we can apply to any subset Z of Θ we can describe what this formula is doing by looking first at the simple case~~

~~fora~~

This gives us a new belief function that we can apply to any subset Z of Θ

for ex: suppose m_1 corresponds to our belief after observing fever:

{Flue, Cold, Pneu} (0.6)

Θ (0.4)

suppose m_2 corresponds to our belief after observing a runny nose/throat infection:

{All, Flu, Cold} (0.8)

& Θ (0.2).

Then we can compute their combination m_3 using the following table:

	{A, F, C} (0.8)	Θ (0.2)
{F, C, P} (0.6)	{F, C} (0.48)	{F, C, P} (0.12)
Θ (0.4)	{A, F, C} (0.32)	Θ (0.8)

As a result of applying m_1 and m_2, we produced m_3:

{Flu, Cold} (0.48)

{All, Flu, Cold} (0.32)

{Flu, Cold, Pneu} (0.12)

Θ (0.08)

Scanned by CamScanner

Fuzzy logic

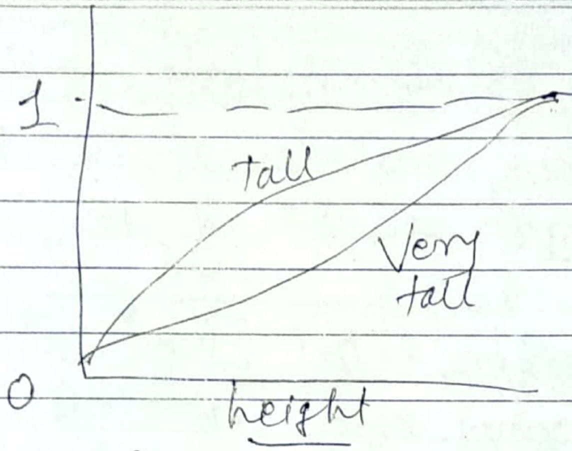
→ Fuzzy logic is a very Convenient and Precise way of expressing imprecise Knowledge.

↳ for Example, the classical logic is not sufficient to represent the facts like "The Coffee is very hot" or "Sameer is tall".

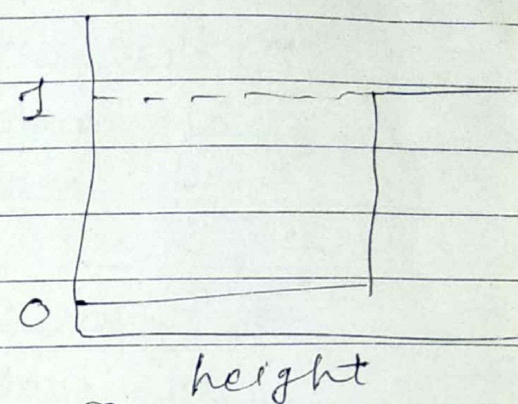
→ But using fuzzy logic we can very well ~~re~~ represent ambiguities or uncertainties in effective manner.

↳ A ~~number~~ Numeric Value between 0 and 1, denoting how true the fact is.

↳ Let's Consider the Sentence "Sameer is tall" and understand how Common Sense can be mapped to formal logic.



ⓐ (Fuzzy membership)



ⓑ Conventional membership

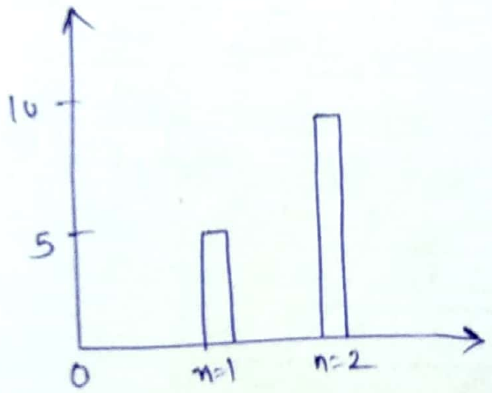
→ ⓐ Fuzzy allows us to represent set membership as a possibility dist. for set of tall & very tall people.

→ ⓑ Conventional denotes boolean defⁿ for tall people. In this, one is either tall or there must be a specific height that defines the boundary. The same is true for very tall.

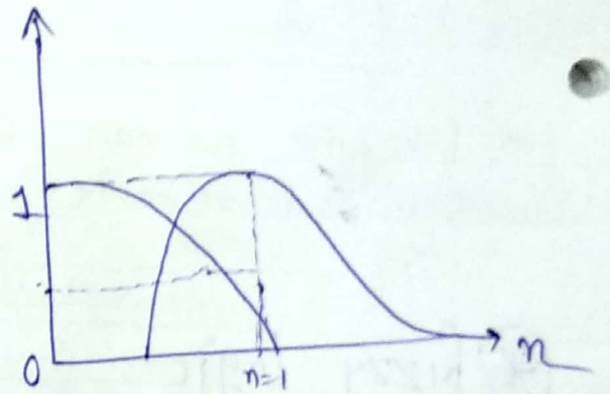
→ If we are referring to Indian standards for height of a person, this sentence will be Completely true if Sameer is 6 feet and above.

→ It will be totally false if Sameer is 5 feet and below.

→ But it will be partially true for the height range between 5 feet to 6 feet And in such cases we will need some measure to represent the degree of fullness.



classical logic



fuzzy logic

↳ figure is depicting the difference between how the same knowledge is represented in classical logic and in fuzzy logic.

Ex: Crisp set = {2, 4, 6, 8}

fuzzy set = $\left\{\frac{0.5}{2} + \frac{0.3}{4} + \frac{0.01}{6} + \frac{0.2}{8}\right\}$

{50% of 2 in set} {30% of 4 in set}

Q8 -

Crisp logic is like binary values.
That is either statement answer is 0 or 1.
It defines as either value is true or false.

and in fuzzy logic we could able to take the intermediate value.

Crisp set are the set in which an element is either a member of the set or not.

fuzzy set allows elements to be partially in a set. Each element is given a degree of membership in a set. This membership value can range from 0 to 1.

fuzzy logic :

6 ft = tall
5 - 6 ft = fairly tall
< 5 ft : short.

} fuzzy set

Sam is fairy tall (5 ft).

classic logic :

Speed = 0 slow
speed = 1 fast.

0 - 25 - slowest
25 - 50 - slow
50 - 75 - fast
75 - 100 - fastest.

} fuzzy set.

OR. Bayes' thm:

$$P(h/e) = \frac{P(h) \cdot P(e/h)}{P(e)}$$

$P(h/e)$: The prob. that hypothesis h is true for given evidence E.

$P(e/h)$: The prob. that we will observe evidence E given that hypothesis is true.

$P(h)$: The prob. that hypothesis is true in absence of any specific evidence.

Bayes' theorem :

R = Red
B = Black.

5 Balls = $\begin{bmatrix} 2R \\ 3B \end{bmatrix}$ A $\begin{bmatrix} 3R \\ 4B \end{bmatrix}$ B = 7 Balls.

1) $P(R/A) = 2/5$ (What is the prob. that Red Ball is chosen from Bag A)

2) → Red Ball drawn from bag A

$P(A \cap R) = P(A) P(R/A)$

3) What is Prob. of Red Ball.

$P(R) = P(A \cap R) + P(B \cap R)$

4) given that Red ball is drawn what is the prob. that the ball is from bag A. (Reverse Prob.)

$$P(A/R) = \frac{P(A \cap R)}{P(A \cap R) + P(B \cap R)}$$ (H/E) =

$$\boxed{P(A/R) = \frac{P(A) P(R/A)}{P(R)}}$$

Bay's theorem.