**Unit III CS6659 Artificial Intelligence**

**Other approaches to uncertain reasoning:**

Other sciences (e.g., physics, genetics, and economics) have long favored probability as a model for uncertainty.

A variety of alternatives to probability were tried for a variety of reasons:

1. One common view is that probability theory is essentially numerical, whereas human judgmental reasoning is more "qualitative."
2. One of the best studied is default reasoning, which treats conclusions not as "believed to a certain degree," but as "believed until a better reason is found to believe something else."
3. Rule-based approaches to uncertainty also have been tried.
4. Such approaches hope to build on the success of logical rule-based systems, but add a sort of "fudge factor" to each rule to accommodate uncertainty.
5. One area that we have not addressed so far is the question of ignorance, as opposed to uncertainty.
6. Consider the flipping of a coin. If we know that the coin is fair, then a probability of 0.5 for heads is reasonable. If we know that the coin is biased, but we do not know which way, then 0.5 is the only reasonable probability.
7. Obviously, the two cases are different, yet probability seems not to distinguish them. The Deinpster- Shafer theory uses interval-valued degrees of belief to represent an agent's knowledge of the probability of a proposition.
8. Probability makes the same ontological commitment as logic: those events are true or false in the world, even if the agent is uncertain as to which is the case. Researchers in fuzzy logic have proposed an ontology that allows vagueness: that an event can be "sort of" true.
9. **Rule-based methods for uncertain reasoning**

Rule-based systems emerged from early work on practical and intuitive systems for logical inference. Logical systems in general and logical rule-based systems in particular, have three desirable properties:

1. **Locality:** In logical systems, whenever we have a rule of the form A B, we can conclude B, given evidence A, *without worrying about any other rules.* In probabilistic systems, we need to consider all the evidence in the Markov blanket.
2. **Detachment:** Once a logical proof is found for a proposition B, the proposition can be used regardless of how it was derived. That is, it can be **detached** from its justification. In dealing with probabilities, on the other hand, the source of the evidence for a belief is important for subsequent reasoning.
3. **Truth-functionality:** In logic, the truth of complex sentences can be computed from the truth of the components. Probability combination does not work this way, except under strong global independence assumptions.

* The idea is to attach degrees of belief to propositions and rules and to devise purely local schemes for combining and propagating those degrees of belief. The schemes are also truth-functional; for example, the degree of belief in A V B is a function of the belief in A and the belief in B.
* The bad news for rule-based systems is that the properties of *locality, detachment, and truth-functionality are simply not appropriate for uncertain reasoning.*
* Let us look at truth functionality first. Let *HI* be the event that a fair coin flip comes up heads, let *TI* be the event that the coin comes up tails on that same flip, and let *H2* be the event that the coin comes up heads on a second flip.
* Clearly, all three events have the same probability, 0.5, and so a truth-functional system must assign the same belief to the disjunction of any two of them. But we can see that the probability of the disjunction depends on the events themselves and not just on their probabilities:



It gets worse when we chain evidence together. Truth-functional systems have rules of the form A B that allow us to compute the belief in B as a function of the belief in the rule and the belief in A. The belief in the rule is assumed to be constant and is usually specified by the knowledge engineer-for example, as A .*B*

Consider the wet-grass situation from Figure.

1. The basic idea of clustering is to join individual nodes of the network to form cluster nodes in such a way that the resulting network is a polytree.
2. For example, the multiply connected network shown in Figure (a) can be converted into a polytree by combining the *Sprinkler* and *Rain* node into a cluster node called *Sprinkler+Rain,* as shown in Figure (b).
3. The two Boolean nodes are replaced by a meganode that takes on four possible values: *TT, TF, FT,* and *FF.*
4. The meganode has only one parent, the Boolean variable Cloudy, so there are two conditioning cases.



Figure (a) A multiply connected network with conditional probability tables.

(b) A clustered equivalent of the multiply connected network.

1. If we wanted to be able to do both causal and diagnostic reasoning, we would need the two rules

Rain WetGrass and WetGrass Rain .

1. These two rules form a feedback loop: evidence for Rain increases the belief in WetGrass, which in turn increases the belief in Rain even more.
2. Clearly, uncertain reasoning systems need to keep track of the paths along which evidence is propagated.
3. Intercausal reasoning (or explaining away) is also tricky. Consider what happens when we have the two rules *Sprinkler WetGrass* and *WetGrass*  *Rain*
4. Suppose we see that the sprinkler is on. Chaining forward through our rules, this increases the belief that the grass will be wet, which in turn increases the belief that it is raining.
5. But this is ridiculous: the fact that the sprinkler is on explains away the wet grass and should reduce the belief in rain. A truth-functional system acts as if it also believes *Sprinkler*  *Rain.*
6. **Representing ignorance: Dempster-Shafer theory:**
7. The Dempster-Shafer theory is designed to deal with the distinction between uncertainty and ignorance.
8. Rather than computing the probability of a proposition, it computes the probability that the evidence supports the proposition. This measure of belief is called a belief function, written Be1 (X) .
9. Dempster-Shafer theory says that because you have no evidence either way, you have to say that the belief

*Bel (Heads)* == 0 and also that *Bel(⌐Heads)* = 0.

1. This makes Dempster-Shafer reasoning systems skeptical in a way that has some intuitive appeal.
2. Now suppose you have an expert at your disposal who testifies with 90% certainty that the coin is fair (i.e., he is 90% sure that P(Heads) = 0.5).
3. Then Dempster-Shafer theory gives Bel(Heads) = 0.9 x 0.5 = 0.45 and likewise Bel(⌐Heads) = 0.45.
4. There is still a 10 percentage point "gap" that is not accounted for by the evidence. "Dempster's rule" (Dempster, 1968) shows how to combine evidence to give new values for Bel, and Shafer's work extends this into a complete computational model.
5. As with default reasoning, there is a problem in connecting beliefs to actions. With probabilities, decision theory says that if P(Heads) = P(⌐Heads) = 0.5, then (assuming that winning $10 and losing $10 are considered equal magnitude opposites) the reasoned will be indifferent between the action of accepting and declining the bet.
6. A Dempster- Shafer reasoner has Bel(⌐Heads) = 0 and thus no reason to accept the bet, but then it also has Bel(Heads) = 0 and thus no reason to decline it. Thus, it seems that the Dempster- Shafer reasoner comes to the same conclusion about how to act in this case.
7. One interpretation of Dempster-Shafer theory is that it defines a probability interval: the interval for *Heads* is [0,1] before our expert testimony and [0.45,0.55] after.
8. **Representing vagueness: Fuzzy sets and fuzzy logic**
9. **Fuzzy set theory** is a means of specifying how well an object satisfie3 a vague description.
10. For example, consider the proposition "Nate is tall." Is this true, if Nate is 5' lo"? Most people would hesitate to answer "true" or "false," preferring to say, "sort of." Note that this is not a question of uncertainty about the external world-we are sure of Nate's height.
11. The issue is that the linguistic term "tall" does not refer to a sharp demarcation of objects into two classes-there are degrees of tallness. For this reason, *fuzzy set theory is not a method for uncertain reasoning at all*.
12. Rather, fuzzy set theory treats Tall as a fuzzy predicate and says that the truth value of Tall(Nate) is a number between 0 and 1, rather than being just true or false.
13. The name "fuzzy set" derives from the interpretation of the predicate as implicitly defining a set of its members-a set that does not have sharp boundaries.
14. Fuzzy logic is a method for reasoning with logical expressions describing membership in fuzzy sets.
15. For example, the complex sentence Tall(Nate) ˄ Heavy(Nate) has a fuzzy truth value that is a function of the truth values of it!; components. The standard rules for evaluating the fuzzy truth, T, of a complex sentence are

T(A ˄ B) = min(T(A), T(B))

T(A V B) = max(T(A), T(B))

T(⌐A) = 1 - T(A) .

1. Fuzzy logic is therefore a truth-functional system-a fact that causes serious difficulties. For example, suppose that T(Tall(Nate)) =0.6 and T(Heavy(Nate)) = 0.4. Then we have T(Tall(Nate) ˄ T(Heavy(Nate)) = 0.4, which seems reasonable, but we also get the result T( Tall (Nate) ˄ ⌐ Tall (Nate)) = 0.4, which does not.
2. Clearly, the problem arises from the inability of a truth-functional approach to take into account the correlations or anticorrelations among the component propositions.
3. **Fuzzy control** is a methodology for constructing control systems in which the mapping between real-valued input and output parameters is represented by fuzzy rules.
4. Fuzzy control has been very successful in commercial products such as automatic transmissions, video cameras, and electric shavers.
5. Critics argue that these applications are successful because they have small rule bases, no chaining of inferences, and tunable parameters that can be adjusted to improve the system's performance.
6. The fact that they are implemented with fuzzy operators might be incidental to their success; the key is simply to provide a concise and intuitive way to specify a smoothly interpolated, real-valued function.
7. There have been attempts to provide an explanation of fuzzy logic in terms of probability theory. One idea is to view assertions such as "Nate is Tall" as discrete observations made concerning a continuous hidden variable, Nate's actual Height.
8. A posterior distribution over Nate's height can then be calculated in the usual way, for example if the model is part of a hybrid Bayesian network. Such an approach is not truth-functional, of course. For example, the conditional distribution

P(0bserver says Nate is tall and heavy I Height, Weight)

allows for interactions between height and weight in the causing of the observation.

1. Thus, someone who is eight feet tall and weighs 190 pounds is very unlikely to be called "tall and heavy," even though "eight feet" counts as "tall" and "190 pounds" counts as "heavy."
2. Fuzzy predicates can also be given a probabilistic interpretation in terms of **random** sets-that is, random variables whose possible values are sets of objects.
3. For example, Tall is a random set whose possible values are sets of people. 'The probability

P(Tall= S1),where S1 is some particular set of people, is the probability that exactly that set would be identified as "tall" by an observer.

1. Then the probability that "Nate is tall" is the sum of the probabilities of all the sets of which Nate is a member.
2. Both the hybrid Bayesian network approach and the random sets approach appear to capture aspects of fuzziness without introducing degrees of truth.