Artificial Intelligence

Prof. P. Dasgupta

Department of Computer Science & Engineering

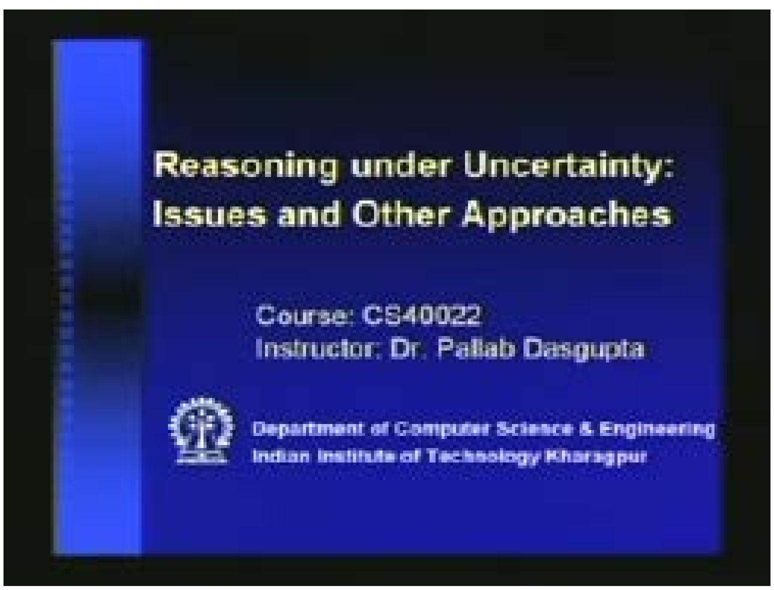
Indian Institute of Technology, Kharagpur

Lecture No - 25

Reasoning Under Uncertainty: Issues and other Approaches

This will be a last lecture on Reasoning Under Uncertainty. In this lecture, we will wrap by studying few extensions of the Bayesian kind of probabilistic analysis and also studying a few other different kinds of approaches towards reasoning under uncertainty.

(Refer Slide Time: 01:14)



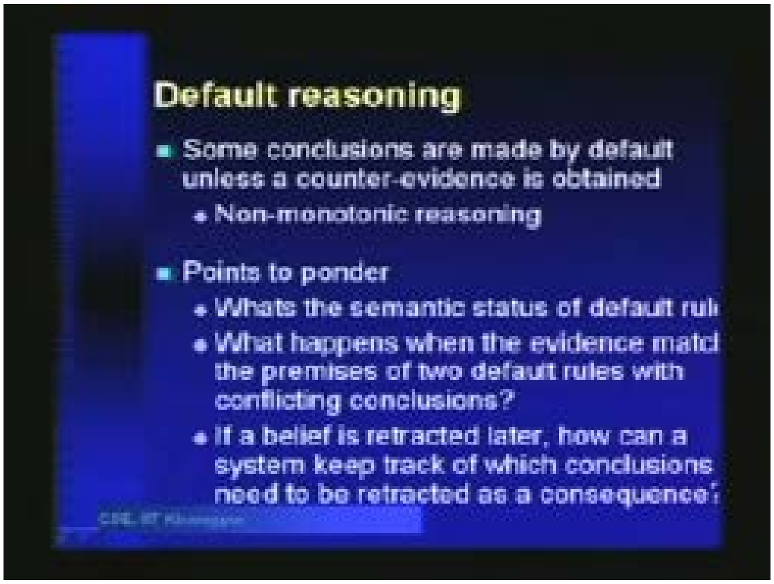
There are, of course, in this subject, a lot of other literature which we are not going to touch up. For example, we are not going to touch upon reasoning which extends over time, like temporal reasoning, and things like mark of models or mark of change, which are used for reasoning with uncertainty over times or over many successive time steps. What we are going to do here is going to study a few things which are closely related to

Bayes kinds of inferencing. The first thing that we are going to look at is default reasoning. A default reasoning is a paradigm where we make some default assumptions about some facts.

For example, if you see a car and you can see 2 wheels from this side, then, by default, you can infer that there are 2 other wheels on the other side. So, we make such default assumptions and add them into the knowledge base but of course, we do know that the default reasoning that we are applying here may later on get retracted, because we might find that 1 of the tires is stolen. In such cases, we find the default reasoning is a kind of non-monotonic reasoning.

Remember how we are defining non-monotonic reasoning. Non-monotonic reasoning is a kind of reasoning where you derive facts at some point of time which can later on be retracted. The other, as compared to monotonic reasoning, where whenever we derive something, then, that is true throughout the history of time. But in non-monotonic reasoning, what is true now or what we believe to be true now, can later on become untrue.

(Refer Slide Time: 04:22)



Default reasoning is where some conclusions are made by default, unless a counter evidence is obtained. It is a kind of non-monotonic reasoning. Now, whenever we go for non-monotonic reasoning, the points to ponder are that what are the semantic status of default rules? In the sense that default rules are not always applicable, because if you have evidence to the contrary, then, you cannot apply the default rule. If you know that 1 tire is stolen, then, you cannot apply the default rule to infer that there are 2 other tires on that side.

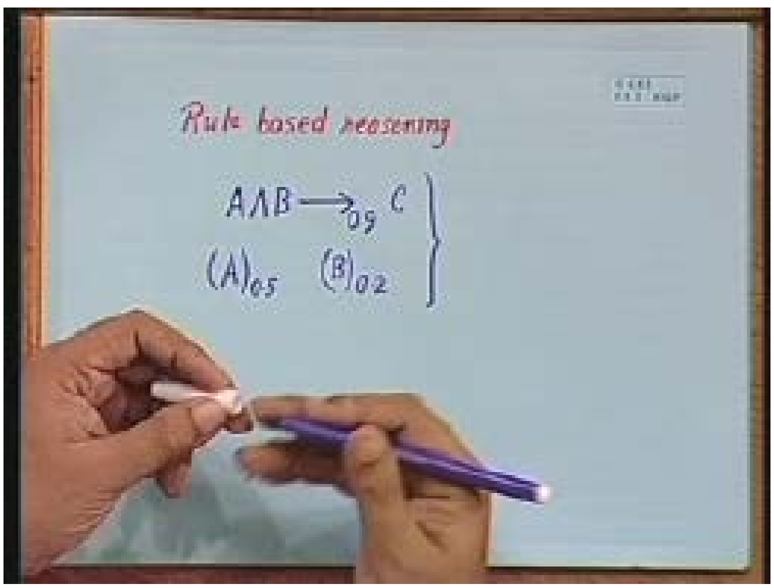
Then, what happens when the evidences matches the premises of 2 default rules with conflicting conclusions? I can have 2 default rules with conflicting conclusions and I have evidence which matches both of the rules. We have to decide what to do on those kinds of cases and most importantly, if a belief is retracted later, then, how can a system keep track of which conclusions need to be retracted? As a consequence, because when we add a fact into the knowledge base, we can use that fact to deduce other facts. So, when something gets retracted, then, all those other side effects will have to be also retracted.

There are things called truth maintenance systems insides in such diagnostic systems which keep track of what is derived from what, and if something gets retracted, then, it follows those links to retract out the other things. 1 of the most classical means of reasoning is called rule based reasoning. Modus ponens, for example, is 1 kind of rule based reasoning, where we deduce, we have certain rules, we have certain facts in the knowledge base, we apply those rules and deduce new facts. There was a lot of research done to find out whether probabilistic analysis could be extended to rule based reasoning.

The rule based reasoning that we have studied in the context of proportional logic and first order logic- in those, the rules were certainties in the sense that if the pre-condition of the rule matched, then, the post condition of the rule is always true. There is no probability associated with that. But when we move into the probabilistic reasoning domain, then, associated with the rule also, you can have a probability. You can say that if A and B, then, with certain probability, we have C.

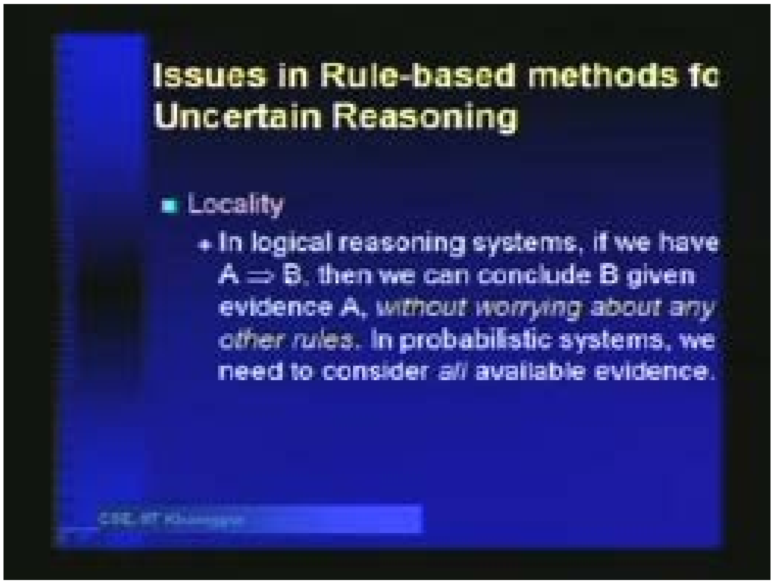
This was the extension that people try to think and then, people wanted to see whether the kind of reasoning that we were doing in first order logic could also be extended to the probabilistic domain. The basic idea is like this: we have that if A and B, then, let us say with 0.9 certainty, we have C. Now, we can have A in our knowledge base, but remember that because we are dealing with a probabilistic situation, we could have A with 0.5 probability and we may have B with, say, 0.2 probability and then, we want to deduce that with what probability can we have C. 1 way to infer is that I find out what is the probability of A and B, that is 0.5 into 0.2. That is a probability with which the pre-condition is satisfied times 0.9, give me the probability of C.

(Refer Slide Time: 08:08)



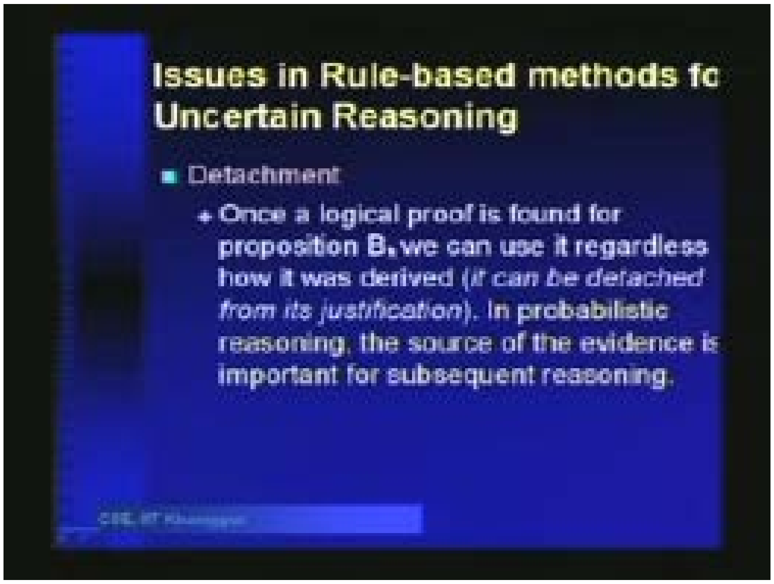
This seems to be very attractive, but it ran into quite a few different kinds of problems, so, we will just briefly see that what are the kinds of problems that will crop up when we try to use this kind of reasoning over the probabilistic domain. Issues in rule based methods for uncertain reasoning: now, these are the problems that I am highlighting out. First problem is the problem of locality in logical reasoning systems. If we A implies B, then, we get conclude B, given evidence A, without worrying about any other rules.

(Refer Slide Time: 09:02)



But we will see that in probabilistic systems, we need to consider all available evidence. It is not just the evidence A that is going to determine the probability of B. There can be other evidence also, which can influence it, that the reason is that see, there is no certainty with respect to A and with respect to the implication. It is not 100 percent certain, then, I will give examples for this, then, we have the problem of detachment, so, once a logical proof is found for proposition B, I am referring to that A implies B.

(Refer Slide Time: 09:21)

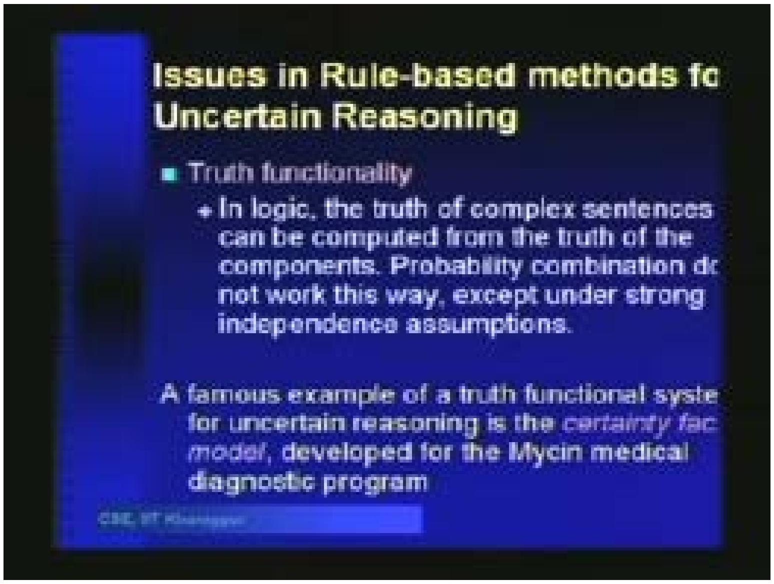


Once a logical proof is found for proposition B, we can use it regardless of how it was derived. If we have A implies B and then, somehow we have been able to deduce B, then, we do not care how we deduce B; we can just use the fact B in other rules for further deduction. But in probabilistic reasoning, the source of the evidence is important for subsequent reasoning. Again, I will explain that why this is so. We will see cases where we have to remember that what was the source of the evidence which lead to the deduction of a particular fact.

The third thing is truth functionality in logic; the truth of complex sentences can be computed from the truth of the components. If you have A and B or C, then, given A, B and C, you can compute the truth of this Boolean formula. It says that the truth of a complex sentence, which means first order or Boolean formula, can be computed given the truth of the components. Probability combination does not work this way, except under strong independence assumptions. If you have them to be independent, then, you can have the probability of A and B as probability of A times probability of B, otherwise not.

When we said here, that that the probability of A is 0.5 and probability of B is 0.2, so, we can just take 0.5 into 0.2 into 0.9. That was under the assumption that these 2 are independent, but if they are not independent, then, the probability of A and B cannot be computed so easily. Let us start by looking at some of these issues. The first issue that we are going to look at is the last 1: the truth functionality.

(Refer Slide Time: 28:28)

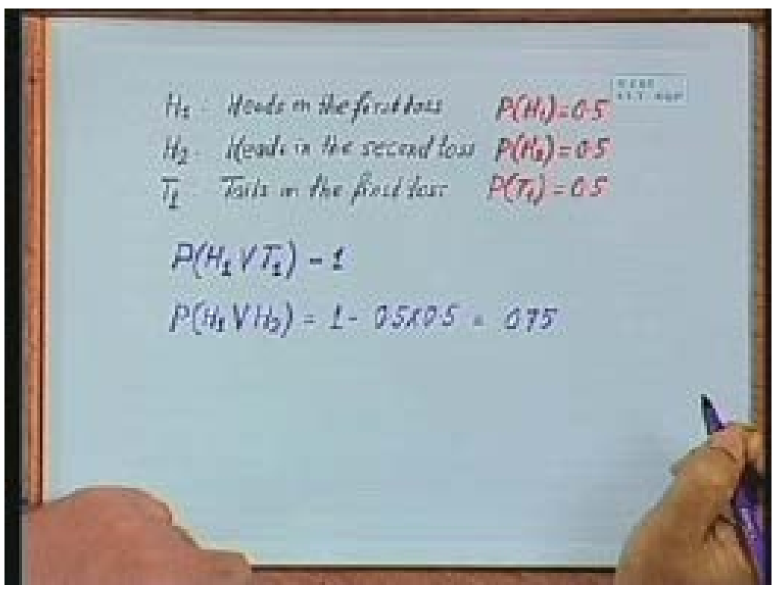


Suppose we have 2 events: H1 and H2, and okay, so, let us say that I toss a coin 2 times, and H1 is the event that we have head in the first toss, H2 is heads in the second toss, and T1 is tails in the first toss. Can we write down that assuming that the coin is unbiased, probability of H1 is 0.5, probability of H2 is 0.5, probability of T1 is 0.5? Now, let us look at the probabilities of a few compound events, that if we look at the probability of heads, see, these 3 events- these 3 probabilities- are all 0.5.

Ideally, we would like to do a reasoning where any the disjunction of any of these gives me the same probability. But that will hold only if these are independent events. Otherwise, for example, if you look at probability of head 1 or tail one, this is one. Probability of head 1 or head 2 is 1 minus 0.5 into 0.5. That gives me- so, this is problem number 1. Problem number 1 is that we cannot simply take each of the facts and then, deduce the probability of any conjunction or disjunction of those events by a simple formula. We have to know that which of these events are the complements of the other.

For example, here, the information which was lacking is that H1 and T1 are actually complements of each other and which in a deductive system you have to explicitly specify, because unless you specify that, how is a reasoning system going to know that if you toss a coin, you can have only 2 outcomes, either head or tail, and that the coin cannot stand like this?

(Refer Slide Time: 14:38)

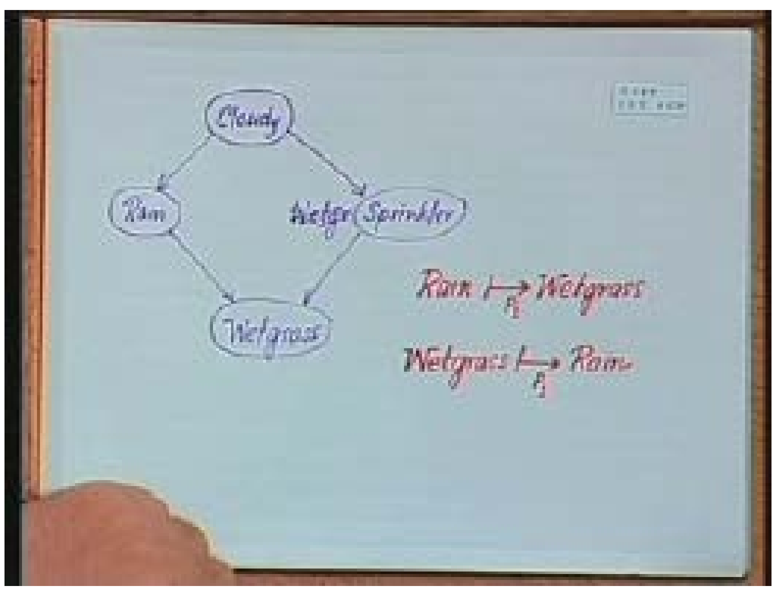


Unless you specify that explicitly, the reasoning system will not know, so, the problem that we are arriving at here is that for every events that we have, if a set of events together form the universe, in the sense that there is no other possible outcome, then, we have to able to specify that, otherwise we will end up in peculiar kinds of problem. Let us look at another case; you recall that we had an example where we had cloudy as an event? Then, we had rain, we had wet grass, sorry, which was sprinkler.

Always, in a reasoning system, we can put in rules like rain. So, if rain, then, wet grass and we can associate some probability with this; some probability, say P1, with this. We can also have wet grass. This says that if we find red wet grass, then, most probably, it had rained. Now see, if you have both of these rules in the knowledge base, then, we have a problem. The problem is that when you have rain, that is going to strengthen, the probability of wet grass that is going to, because that is increased evidence that we will have wet grass.

And then, the reasoning system can use that increased evidence on wet grass to again increase the evidence of rain, because if you have more evidence that we have wet grass, then, you have more evidence that there is rain and we are going to go around in circles strengthening the evidence of having rain and wet grass. That is very strength, because that should not happen, so therefore, the reasoning system should be able to deduce that, because I deduced wet grass from rain.

(Refer Slide Time: 17:53)



I should not strengthen rain based on the evidence of wet grass or, in other words, it should detect this cycle and this is typically the case, because we have put 2 rules: 1 is on the causal inferencing and the other is a diagnostic. 1 is from the cause to the effect and the other is from the effect to the cause. We can have typically have rules of both types: 1 is diagnostic and the other is causal. But we have to be careful that this cycle does not takes place and even if there is such a cycle, then, the reasoning system should not go around that cycle and that is why we were saying that it is also important to remember that how we deduce the certain thing. There can be more peculiar things.

Suppose we can have sprinkler with a certain probability, P3, to wet grass. If we turn on the sprinkler, then, with certain probability, we will have wet grass. The probability can come for facts like, there is no water, so, if there is no water, then, you still not have wet grass. So, the probability that there is water is what determines the probability of this rule, and then, we can also have wet grass with some P4 or P2. We had this rule already, rain. Now, look at the funny thing that is going to happen. If we have evidence that the sprinkler was turned on, then, that is going to increase the evidence on wet grass and if you have evidence that you have wet grass, that is going to increase the evidence that you have rain.

(Refer Slide Time: 22:09)

By turning on the sprinkler, we are actually increasing the probability of rain. That is ridiculous. By turning on the sprinkler, you will end up finding more evidence that there was rain. Actually, it should be the other way around, because you know that the sprinkler was turned on, so, the fact that you have wet grass is explained by the sprinkler being turned on. The evidence of rain should actually go down; you should have less evidence that there was rain, because the fact that the grass is wet is explained by the fact that we had the sprinkler on.

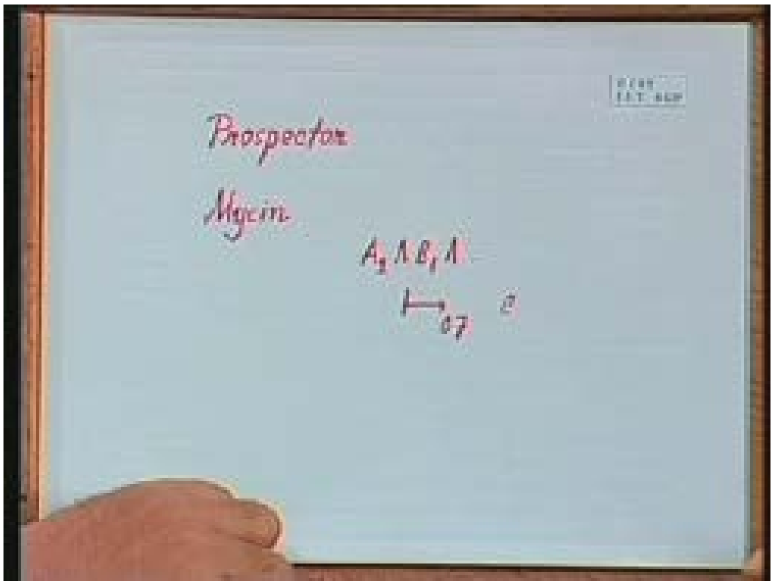
Therefore, this is another problem. So, it is not always that the probabilities should enhance the evidence of the conclusion of the rule. Yes. (Student speaking). Here, yes- (Student speaking). If you find it is cloudy, you are less likely to turn on the sprinkler because you will anticipate rain. (Student speaking). No, the sprinkler can be an automated sprinkler, which on sensing the sun at regular intervals of time, sets off the sprinkler. But if it is cloudy, then, that device which sets off the sprinkler may not go on. There is a probability- this probability of the sprinkler being turned on, depends on whether it is cloudy or not.

But what suddenly should not affect, is that, if the sprinkler gets turned on, then, there is absolutely no reason why it should rain. On the contrary, the evidence of rain should go down, the probability of rain should go down. (Student speaking). Right- so, if you use the Bayes network kind of analysis, you would see that if you have evidence of the sprinkler going off-going on, then, the probability of rain: it can go down. If you do a Bayes network analysis on the network that we had seen previously, then, you will have that, but in a rule based reasoning, we are not being able to mimic that, because of these kinds of problems.

So, these were some of the problems that people came up with rule based reasoning, but in spite of that, there was a point of time when rule based probabilistic deduction systems became very popular. And there are at least 2 very well-known tools which became quite popular; 1 of them is called the prospector- this was a tool which was used for basically zoological explorations, to find out whether there is some kind of mineral or this thing.

This was a rule based system which had rules and you could feed in the evidence and then, deduce that what is the probability with which you have a certain kind of mineral or an oil or whatever. Another thing, another tool, which became very famous, was called mycin. This is a diagnostic tool for pathological kind of inferencing.

(Refer Slide Time: 26:33)

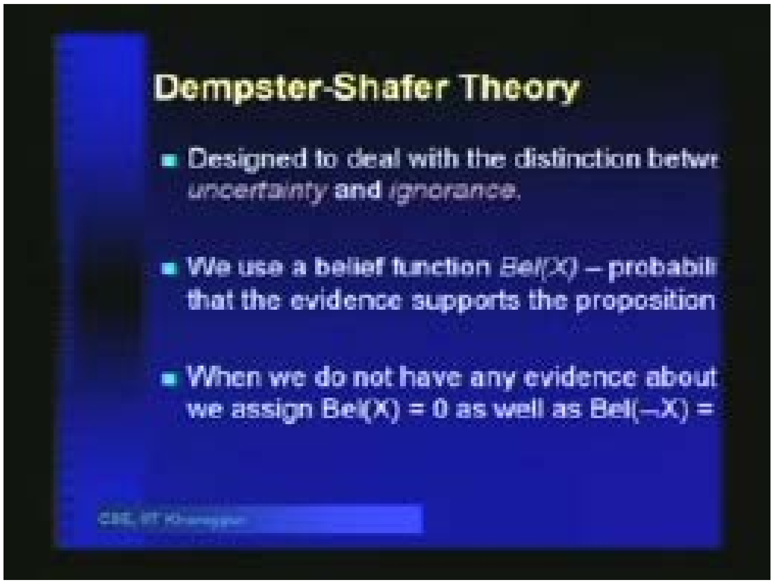


This mycin tool had rules which had associated certainty factor, so, if you have evidence of A1 and B1 and so on, and then, you can deduce with certain probability, say 0.7 or so, some C, some conclusion. The rules were of this form and in mycin specifically, the attempt was to keep all the rules of 1 type. If you keep all the rules to be either causal or diagnostic, then, the problem of those kinds of cycles do not come. Moreover, you try to keep the rules always, so that the conclusions are non-conflicting in nature, so that you do not have problems like sprinkler to wet grass and wet grass to rain.

You are not able to deduce something which is conflicting with what you have, so, this careful structuring of the rules actually helped in this being successful. Later on also, there has been several other rule based systems, diagnostic and information systems, which use probabilistic reasoning, but they are all very carefully designed to avoid such kinds of problems. A different school of thought came up with an idea which was quite revolutionary at that time.

They say that besides the probability, there is also a thing called belief; what is important is the belief of a person when doing an inferencing. Instead of modeling the probability, they thought about modeling the belief. I will explain through an example that what is this change. We will briefly study a theory called Dempster-Shafer theory. This is designed to deal with the distinction between uncertainty and ignorance.

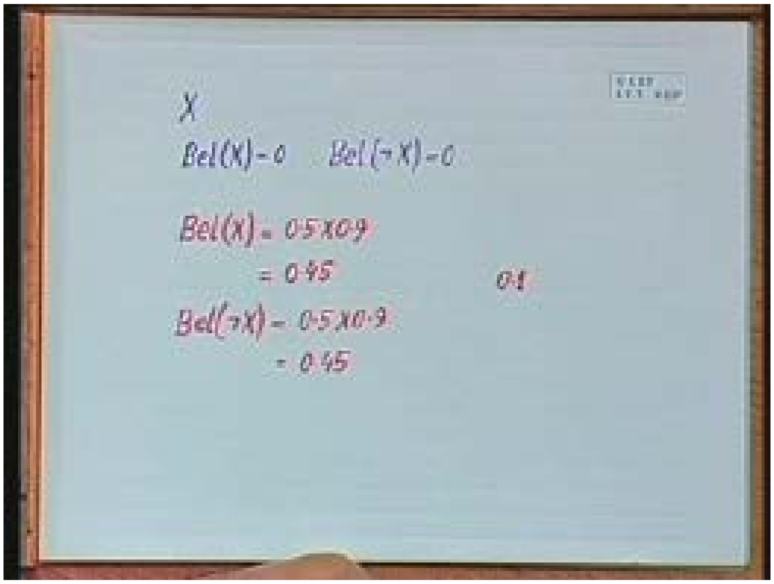
(Refer Slide Time: 29:38)



We use a belief function called Bel X, which represents the probability that the evidence supports the proposition, so, it is something like the kind of rules that we were saying that it is with. What probability does the evidence support the conclusion? When we do not have any evidence about X, we assign belief X to be 0, as well as belief of not X to be 0. Now, let me give you an example, that suppose- this is an example which are picked up from Norvig and Russell. The example goes like this: that suppose some person comes to you with a coin and says that if you toss the coin, then, you are more likely to get head.

Now, you do not believe that character; you do not believe that character, but that character says that if you toss the coin, you will get heads 90 percent of the time. Because you do not believe that person, so, if X is the event that you get heads, then, I will say that my belief that we will have X is 0, because I do not have any evidence to support the fact that X will be heads. Likewise, I do not have any evidence to support that it will not be heads either, so, belief of not X is also 0. Now, suppose a very reliable person comes and tells us that with 90 percent probability, the coin is fair, so, a reliable person comes to me and says that I am 90 percent certain that this coin is fair.

(Refer Slide Time: 35:10)



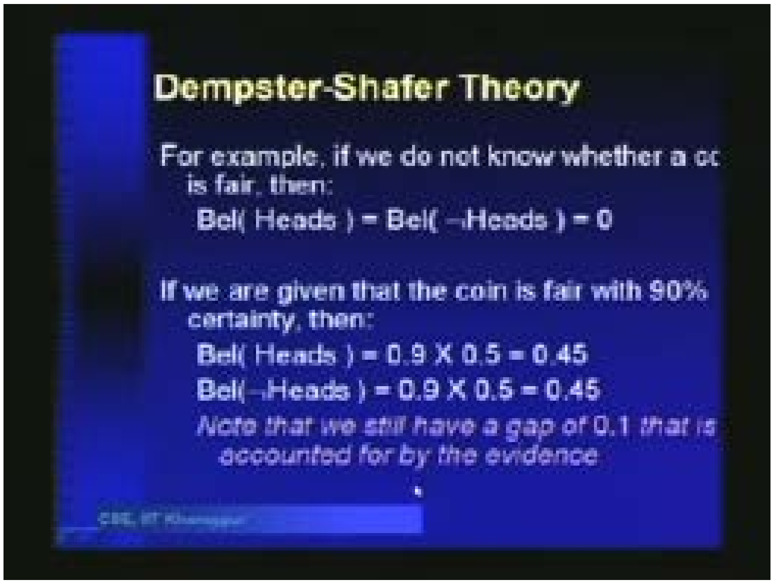
Then, I will say that belief- that it will be heads is 0.5 into 0.,9 90 percent certainty of the rule and 0.5 is the probability that I have heads, because he is saying it is fair, so, the probability of heads if it is fair, is 0.5. This gives me 0.45. If the coin is fake; I do not know whether it is fake or not, I do not know, so, I am just going by the evidence that I receive. The whole idea is that how can I use the evidence and the belief of that evidence to deduce my belief of the conclusion?

My confidence in that person is 90 percent, so, I have this 0.9 and that person says that the coin is fair, so, if the coin is fair, then, the probability of head is 0.5. Previously, it was 0 because I had no idea whether the coin was fair or not. I did not have any evidence that the coin was fair. My belief that it would be heads is 0 and the belief that it would be tail is also 0, but now that this person on whom I have 90 percent confidence comes and tells me that it is fair with 0.5 probability.

So, I conclude that it is 0.45- that is my belief that it will be heads, and what is my belief that it will be tails? 0.5 again for the tail and 0.9 because I, again, that confidence in that person, so, I have 0.45 again. See that the gap between the belief of X and not X is still noT1. And how much gap do I have between the 2? What is this 0.1 accounting for? This 0.1 is accounting for my lack of confidence on the person who is saying.

So, I am not 100 percent certain about the evidence; I am only 90 percent certain. I have 90 percent confidence in that person who is telling me that the coin is fair. If I had been 100 percent certain, then, the sum of these 2 beliefs would have given me one. So, this was the basic idea around which Dempster-Shafer theory was built. This is that example. If you are given that the coin is fair with 90 percent certainty, then, we have 0.45 and 0.45 and we still have a gap of 0.1.

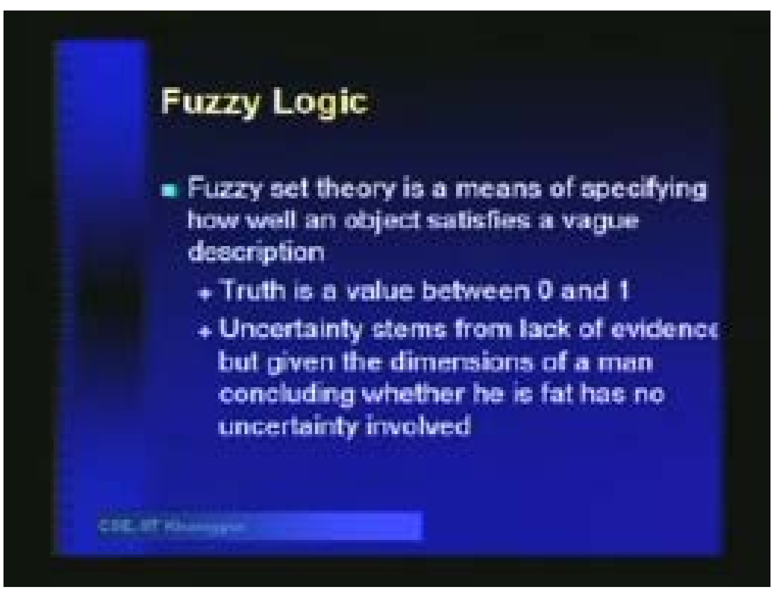
(Refer Slide Time: 35:33)



This was a paradigm shift from probability to belief, but this theory did not really pick up very much, mainly because it was not very clear how to include this kind of belief inside a rule based system, and also, the actual meaning of this belief and how we should interpret that in deducing conclusions was not well understood. So, it is more because of the lack of people's understanding and the belief on the theory that lead almost to the obsolescence of the Dempster-Shafer theory. But we do have 1 kind of theory, which has come up in a big way. But again, I have personally quite a bit of reservations about this theory also. Nevertheless, let us just get a flavor of what this theory is about.

Now, fuzzy logic makes again a paradigm shift from probability to possibility, so again, we do not talk about the probability of an event, but we talk about the degree of truth of an event. As I had given an example previously, then, that if we have a person and we want to find out the probability that he is fat, then, we say that that fatness is a Boolean, that he is either fat or he is not fat. And then, we talk about the probability that X or Y is fat. But when we talk about fuzzy logic, we say that no, fat is not a Boolean, it is a set which has fuzzy contours, so, at some part in the set for some members, well, if they are definitely fat, for some members, well, they are fuzzy.

(Refer Slide Time: 39:18)

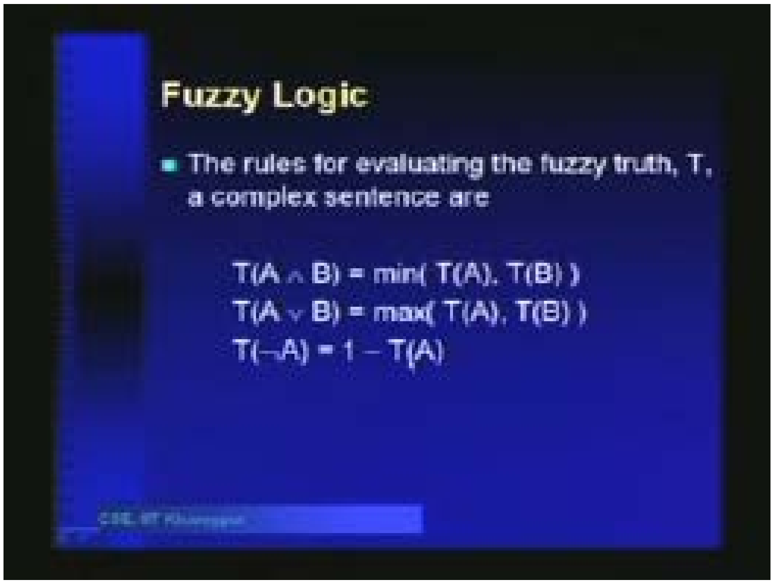


So, fatness is actually a graded truth. How fat- now, depending on how fat, we can have other kinds of things, like for example, suppose we want to make a rule based system that if a person is fat, then, he is more likely to have a cardiac problem. Now, given that X is fat, now then, it depends on how fat. If he is very fat, then, he has more chance of having a cardiac problem. If he is less fat, then, he has less chance of having the problem. If he is not fat at all, then, he has negligible chance of this.

So, again, the outcome of whether he has a cardiac problem or not, or the likeliness or the possibility of that person having a cardiac problem depends on how fat he is. In this logic, we deal with fuzzy set theory, which is a means of specifying how well an object satisfies a vague description. Truth is a value between 0 and one. Uncertainty stems from lack of evidence.

But given the dimensions of a man, concluding whether he is fat has no uncertainty involved, but it has fuzziness involved. So, if you know the person, if you have the dimensions of the person, then, there is no uncertainty about that person is being fat or not. It is only about how fat we are going to rate him. Then, the rules for evaluating the fuzzy truth of a complex sentence are as follows. When we have 2 events A and B, then, the truth of A and B is the minimum of the truths of A and the truths of B.

(Refer Slide Time: 41:29)



If A is fat and B is slow, then, the truth that a person is fat and slow is given by the minimum of how fat he is and how slow he is. Now, we can debate for hours about the choice of the mean function for this people found was- okay, I will come back to that. Let us first see the other rules. If you have A or B, then, it is the maximum of the 2. So, you see here, we are not looking at anything like conditionality or anything. There is no such thing, because there is no uncertainty involved in it. It is just the degree of truth that we are talking about.

So, there is no conditionals are anything like that and we just deal with the min and max operators, so, the analysis itself becomes much simpler than that of Bayes networks. And the truth of not A is equal to 1 minus the truth of A. So, these are the 3 basic rules for analysis on fuzzy sets. Every event A, B, C, whatever they are, all sets. For example, fat is a set. If you if define the continuum of weights of all persons of the world, then, a subset of that is fat. Tall is another subset. These are all sets and we are working on the sets only, so, it is a kind of set theoretic operation, except that we associate a truth with the set and perform all the deductions based on this min max and complement operators.

People did find that this kind of analysis was simplifying things a lot, and they found also that in those kinds of simple situations where you do not have too much of dependence between events, that the kind of deduction that you do with fuzzy logic, approximates what you would do with a comprehensive Bayes analysis. So, people found that there is actually an approximation of the Bayes analysis that you do obtain through fuzzy analysis. But in cases where the system is large, you have many events and there is lot of interdependencies.

This has failed also (unclear words). It is not good for everything also. It has been used to some extent in different kinds of control systems, particularly in the kinds of control systems where you have continuous variables and you want to specify the gradation of control through a fuzzy set. You want to say that if the blast furnace becomes very hot, then, do this. If the blast furnace is somewhat hot, then, you do this. So, that hot is a fuzzy set and different rules that you want to apply on controlling the blast furnace can be specified in terms of fuzzy rules.

And the degree of control exercised by all these different controlling mechanism are also graded depending on the value of the hot, or the truth of the hot- how hot, right? The critics of fuzzy logic believe that the same kind of control could also be achieved if you just perform the simple discritization of the continuous system, and then, you use normal reasoning. And people think that it is more, because of the sophistication in the control system itself that lead to the success of fuzzy control. Things- you know, you have fuzzy control washing machine and things like that being advertised.

So, the critics of fuzzy logic believe that it is more because of the design of the systems that they are successful, rather than the success of the logic which is driving this and that more simpler logic, which just discritizes the system and performs elementary reasoning, would have done just as well. But it still open to debate- fuzzy logic is still a part of research and well, in future, we will know that whether this is really very useful or not. With that- yes? (Student speaking). Yes (Student speaking). Yes, you have problems like that in fuzzy logic as well. (Student speaking). Which formula? The third one? (Student speaking). Yes. (Student speaking). Hmm. (Student speaking). T of A and T of not A, yes. (Student speaking). Yes. (Student speaking). No, see, that is what is the meaning of a fuzzy set.

See, in the border areas, the truth and its negation both can have certain values. It is not going to be 0, it is not a Boolean that you are talking about; it is not a Boolean, so, it is not that if it is A, then, it is not not A, it can be- a person can be fat and he can also be not fat because you are in the fuzzy region, so, that is why we have this (unclear words). With this, we come to the end of that the chapter on reasoning under uncertainty and the next topic and the possibly the last topic that we will pick us is on learning and machine learning.