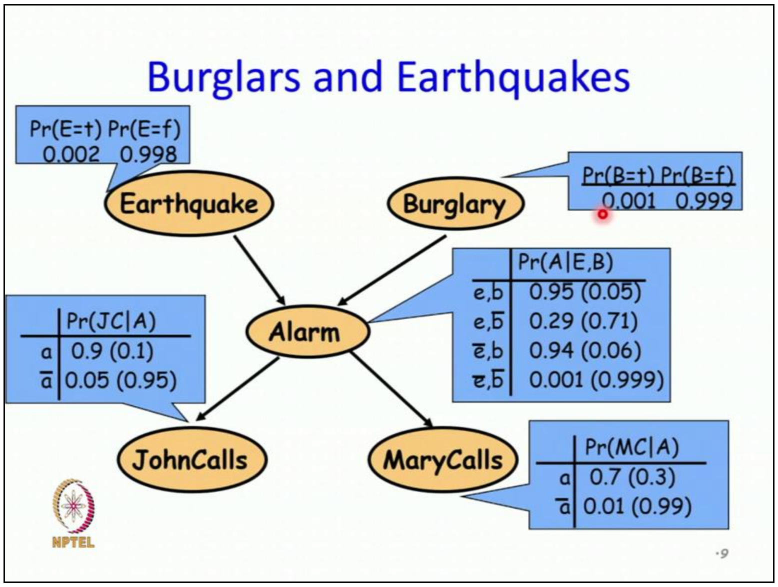
Artificial Intelligence
Prof. Mausam
Department of Computer Science and Engineering
Indian Institute of Technology - Delhi

Lecture-54
Bayesian Networks Factorization, Part – 2

(Refer Slide Time: 00:18)



So, this is our graph. Now, let us see in how many parameters can I represent each node? Remember, that we have said that for every node will represent probability that random variable is given all the parents. So, for earthquake, how many parameters do I need? Just one because earthquake has no parent so, I simply want to represent probability of earthquake and probability of not earthquake but they sum to 1. So, basically I need one parameter is same for Burglary.

So we need two parameters for earthquake and burglary, ok. Now let us think about Alarm, probability of alarm so, we have to set probability of alarm given all the parents what are the parents? Earthquake and Burglary. So, how many parameters do I need? 2 parameters, people said two, OK what are the two parameters? What is your name? Mayur and the blue shirt behind you Chinnah, yeah Chinnah. Why do you say we need 2 parameters, what are the two parameters?

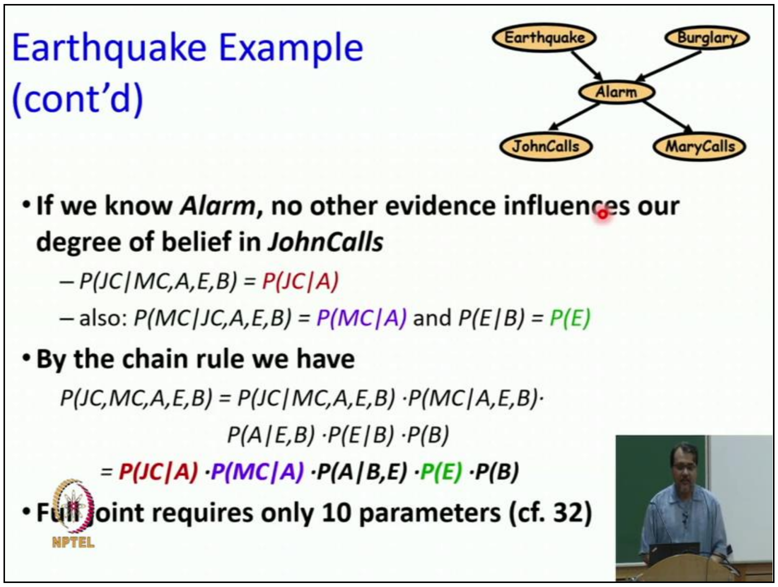
Earthquake and not earthquake, Burglary not burglary, so how many such possibilities are there for you said? Two right very important, so what are the possible States for my parents? 00,01 10,11 and for all such possible states there is a different probability of when the alarm is going to go off. So, we will end up having this kind of a conditional probability table with the divisibility alarm, it will be a probability of alarm given earthquake, burglary, if both earthquake happened and at the same time burglary happens 0.95, if only earthquake happens when only happens 0.29 and so on. So, we need 4 parameters.

How many parameters for John? Two parameters for John and two parameters for Mary. So now notice that we have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 parameters. We have 10 parameters to represent this Bayesian Network. And what I am going to show you now is that that has the full power for representing the full joint distribution. That is the beauty right. So, for that we have to ask some very basic Independence questions.

Let us quickly say is burglary and earthquake are independent of each other, by knowing earthquake, burglary does not change although people can say that in some ways it may be dependent also earthquake happened burglars as both. This is a really good time people have gone crazy right now. Let me go in a burglar the house, possible. But it is possible there is some small effect but we are not modeling that. They are saying that they are independent events.

Then, by knowing alarm does earthquake give me additional information on whether John is going to call? Important question I already know that alarm has gone off or it has not gone off. By additional knowing whether earthquake occurred or it did not occur, does that changed my probability distribution about whether John is going to call? John only depends on alarm if alarm is already known to me. I have idea of how long is John going to behave for any additional knowledge about earthquake and burglary, does not give me information.

(Refer Slide Time: 04:23)



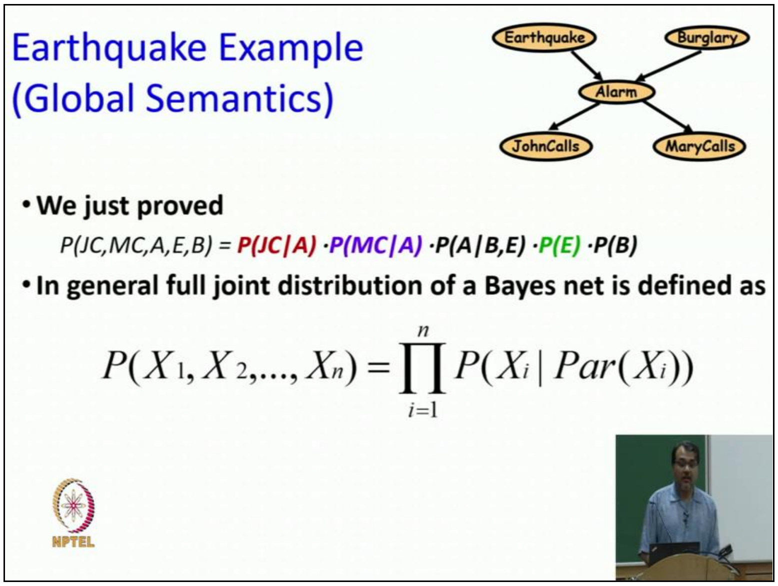
So we can say some conditional independence is here. We can say that John given Mary alarm earthquake, burglary only depends on alarm. If there is no alarm all the other things do not give me any additional information. Similarly, for Mary, we can also say that burglary and earthquake because they are independent of each other. Now, we are given all these conditional dependences now, it is very easy to use the chain rule. So what is the chain rule?

John, Mary, earthquake, burglary and alarm is a joint distribution. You can always write down without anything as P of P times your P given B, P of A given E, B, P of M given A, E, B and P of J given A, E, B. But P of A given B, E is P of E, M given A is P of A and J given MA, E, B given A. So, therefore, alternatively this would become because of the conditional dependences, probability of earthquake and probability of burglar alarm given earthquake from a burglary, probability of john given alarm and probability of Mary given alarm, end product of square.

Anyone think about the graph. What have you done? We have simply multiplied all the CPT, CPT are conditional probability table given. So for a given joint, suppose, I want to know what is the probability of John not Mary, no alarm earthquake, no burglary. There was a earthquake 0.002, no burglary 0.999, alarm no alarm. No alarm given earthquake, no burglary would be 0.71. John given alarm would be 0.05 not Mary given no alarm would be 0.99.

I will multiply all these five numbers and that give me the joint probability for this particular state. So, what I have done it have shown you with an example and this is true in general.

(Refer Slide Time: 06:35)



That is semantics of a Bayesian network is that the full joint distribution is nothing more than a product of all the conditional probability. Product for all I, where I is node, probability of XI given all the parents of XI and because in a cyclic directed graph, because it is cyclic, you can start in the technological order you can start from the parents of everything. There will always be some nodes which has no parent.

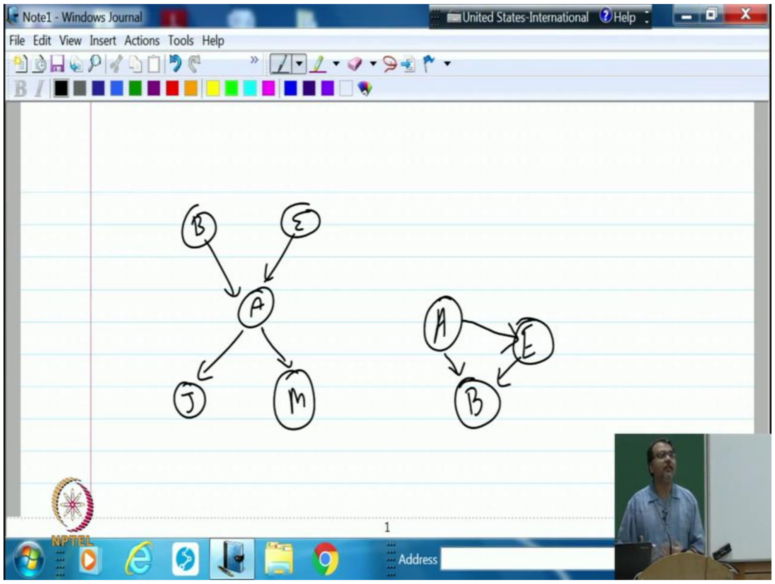
Now, you have already defined it, there will be at most one node which has a parent which would be something already defined and you can keep using those probability values in the right CPT to prove all the notifications. This is nothing but a factorization of the joint distribution. In 10 parameters, I have been able to represent the full joint distribution, which would have otherwise; taken 31 parameters because there are 5 variables each variable can be true or false. 32 States they all sum to 1. Then 31 parameters.

Something that requires 31 parameters has not been done in 10 parameters because they were conditional dependences in the domain that you are able to exploit. Any questions, at this point? Yes. How we will be determining the probability? That is called learning and we will talk about that in 4th class from now. Ok. So first we are defining what Bayesian network is, then, we will ask queries for Bayesian Network and make inferences.

Given the confirmation network we will try to figure out what is the probability that if alarm went off and burglary happened. What is the probability that John Corbett never made a call. These are sets of questions we will answer and once when John has called what is the probability that burglary has not happened and questions like that. Then, we will talk about how do we learn, we will talk about this in the next lectures this will get clearer and clearer.

Is it possible that a Bayesian network can have a earth storm? Earthquake is possible, I believe it is possible. And it is possible if we choose a bad order of making the Bayesian network.

(Refer Slide Time: 09:20)

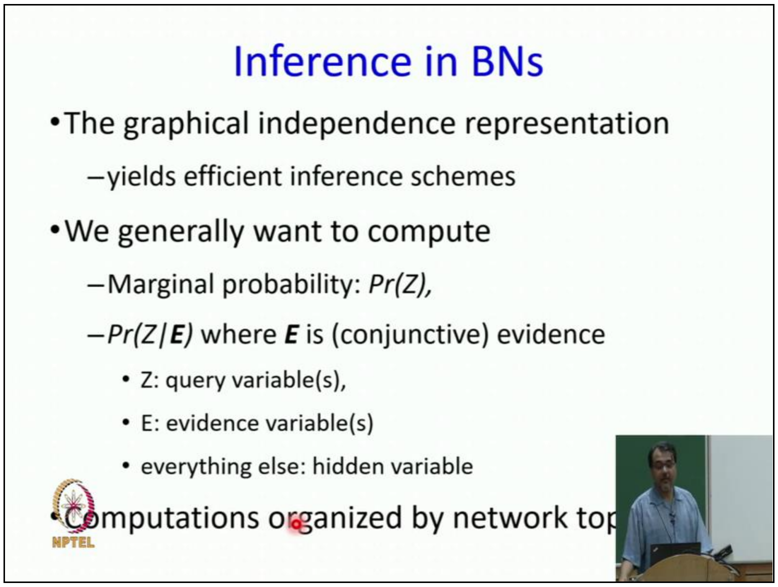


Here is a Simple example, suppose, we start with making in edge from alarm. Then does alarm influence earthquake. It is very weird to think in the noncausal direction. So, therefore it becomes confusing but does alarm influence earthquake. Yes, because alarm goes off then it increases the probability of earthquake, if alarm does not goes off, it decrease the Earthquake. However, if I know Burglary and Alarm and I want to figure out about burglary in any conditional independence happening here, it is a hard word?

The reason it is a hard question is alarm going up increase the probability of burglary, alarm going up and earthquake also happening, decrease the probability of operation. In fact, in this weird is order both earthquake and alarm actually influence burglary. So, if you if you take a very different order of variables, then you can get very different notations of networks.

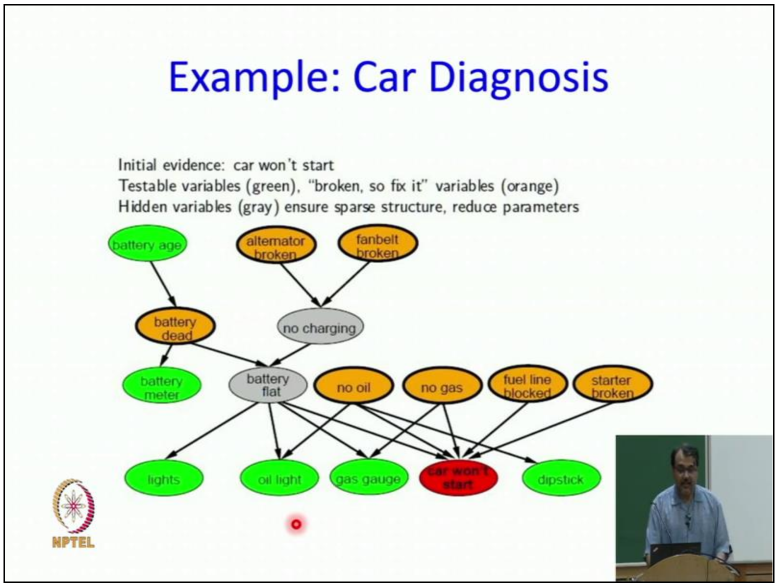
And then influences. So, therefore the general principle here is that it is intuitive to think about these edges in the causal order. But probabilistically, theoretically, mathematically it is not needed. They can work even if these edges are not in the causal order except that thinking about human is resistant. So, now what we are going to do is we are going to use the Bayesian network think about, what are the conditional dependence assumptions it is making?

(Refer Slide Time: 10:55)



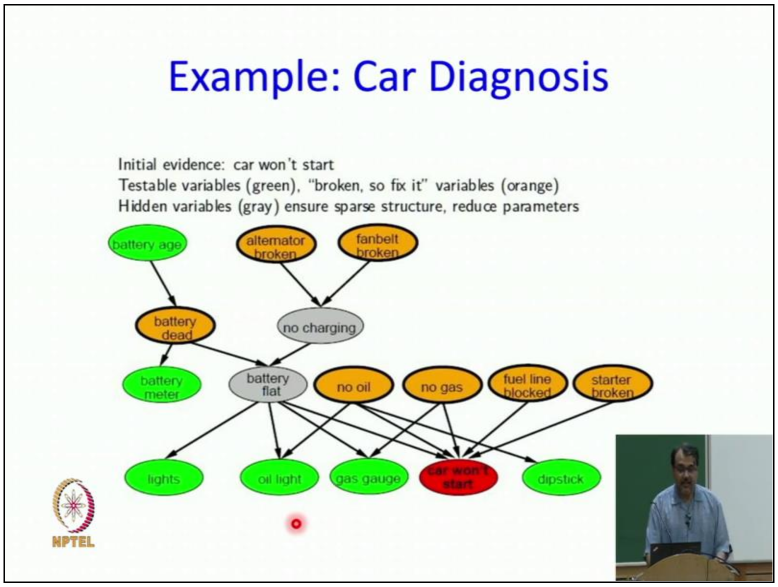
And after that we will start talking about how do we do inferences? So, inferences like suppose, I want to make this inference, ah let us say I want to know that John has called, Mary has called, What is the probability that burglary actually happened? That is the question that and inference question, right? We want to compute some probabilities of some variable given some other variable I mean these kinds of Bayesian networks are incredibly common.

(Refer Slide Time: 11:20)



For example, this is a car diagnosis Bayesian networks. It basically says that battery age I can observe, battery metre I can observe. Whether battery is dead or not, I cannot observe. I can observe that. You know there is oil in light is on or not, but I also observe car is not starting. So what is the likely cause of this. This is exactly the Diagnosis Centre in logic also. Accept we now will have probability distribution of what is the more likely cause of a certain event and what is a less likely cause of a certain.

(Refer Slide Time: 12:01)



Similarly car insurance, believe it or not a lot of car insurance in figuring out. How much should we give you insurance or not? How much your Premium should be depends on a lot of factors and all these factors in a complicated way and are dependent on each other. So what is your age?

If you are very old person, then you might get a senior citizen discount, but if you are too old then you may not be able to have good reflexes for your premium should go up.

If you have a good social economic status, that will be one thing. What kind of mileage given the car if your price is too old and they want to increase your fast, but you also want to decrease your premium because now your car is cheaper, it will cost you less money et cetera et cetera. So, all these factors which are all dependent on each other; your driving skill depends on your age, your driving quality depends on your driving skill, whether you are going to have an accident or not depends upon how good a driver you are, whether you have airbags or not?

Or do you have anti lock system or not? Some of these things you can observe for example, you can observe the vehicle here. You can observe the age, you might not be able to observe the driving quality, right? You should know whether there is antiseptic, anyway the point is that all the factors probabilistically come together into determining how much should be the liability.

(Refer Slide Time: 13:18)

And this Bayesian network has had a huge number application. In medical diagnosis, in the; what is the probability of the disease given symptoms? In computational Biology, in natural language processing what is the past history of whether something a word is a verb or noun given the sentence document classification given the document. Is it of class sentiment positives sentiments or negative sentiments? In image processing, given the image is it of is there a boy in there or not?

In robotics, given the sensor readings, how far is the actual obstacle? There are so many cases where you have to ask these questions and there answer has to be probabilistic because the world is probabilistic because either we cannot because observe or because sensors make mistakes because whatever and for all such situations, some kind of probabilistic graphical model become the basis in Bayesian network. It is the first fundamental Probabilistic model.

So what we are doing is actually incredibly important. Even though deep rule networks have become much more important in terms of real practical applications, Bayesian Network and their power is believed to be coming back in a better while. Also for some better accuracy, sometimes we combine the two. Like the best model for sequence labeling is a Bi-Elestian CRM. Bi-Elestian is user network; a CRM is a probabilistic graphical model.

Deep network does something and then the probability graphical models does something and together they make the best decision. I think that Bayesian network is something that is introductorily important for all of us to study and the next class we will talk about the D-separation, the general principles of conditional dependence and then will get to how to make probabilistic inferences using a given example. We can stop with this.