Bayesian Networks

Yunwen Lei

Southern University of Science and Technology

leiyw@sustc.edu.cn

November 21, 2018

Review of Probability

- Joint probability
  - the probability of two events happening together, denoted by $P(AB)$ or $P(A, B)$
  - $P(AB) = P(B)P(A|B) = P(A)P(B|A)$
- Conditional probability
  - $P(A|B) = P(AB)/P(B)$
  - $P(B|A) = P(AB)/P(A)$
  - Chain rule:

$$P(A_1, \ldots, A_n) = P(A_n|A_1, \ldots, A_{n-1})P(A_{n-1}|A_1, \ldots, A_{n-2}) \cdots P(A_2|A_1)P(A_1)$$

  - Bayesian theorem:

$$P(A|B) = P(B|A)P(A)/P(B)$$

Review of Probability Theorem

- Independence

$$P(A|B) = P(A) \quad \text{or} \quad P(B|A) = P(B) \quad \text{or} \quad P(AB) = P(A)P(B)$$

- Total probability formula: Let $A_1, \ldots, A_n$ be a partition of sample space, then

$$P(B) = \sum_{i=1}^{n} P(A_i B) = \sum_{i=1}^{n} P(A_i)P(B|A_i)$$

- Bayesian formula (complete form):

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

Bayesian Network: definition

Definition: Bayesian network is a directed graph without cycles with each node annotating some information on probability. To be detailed:

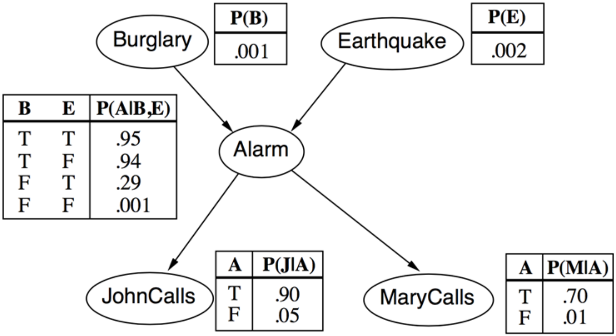
- Each node corresponds to a random variable (either discrete or continuous)
- If there is an arrow from node X to node Y, we call X is a parent of Y
- Each node X_i has a conditional probability distribution $P(X_i|\text{Parent}(X_i))$ to quantify the influence of its parent to the node

Conditional probability table (CFT) is used to represent the conditional probability distribution of X_i conditioned at its parent. Each row of CFT quantifies the probability of node conditioned the occurrence of other events. We always ignore the case with the node being false since it can be induced from the case with the node being true

Bayesian Network: sample

- To prevent thieves from entering, the home has installed a new burglar alarm
- Alarm reliable for detecting thieves & occasionally respond to earthquakes
- Two neighbors called John and Mary. Call you if they hear an alarm
- John sometimes interpret the phone ringing sounds as an alarm
- Mary likes to listen to music loudly, and sometimes can not hear the alarm

Please construct a Bayesian network and CFT.



Bayesian Network: Calculation of Fully Joint Probability

- A Bayesian network can essentially represent any fully joint probability distribution, but is simpler than a fully joint probability distribution.

  If there are n random variables, then with a fully joint probability distribution, the required space is $O(2^n)$. But for Bayesian networks, if the number of parent nodes of each child node does not exceed k, then the required space is $O(n2^k)$, which is linearly increasing for the problem size n.

- Calculation of fully joint probabilities in a Bayesian network:

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | \mathrm{parent}(x_i))$$

Bayesian Network vs Fully Joint Probability

- We use B, E, A, J and M to represent Burglary, Earthquake, Alarm, JohnCalls and MaryCalls
- The following table is a fully joint probability distribution with size 2^n

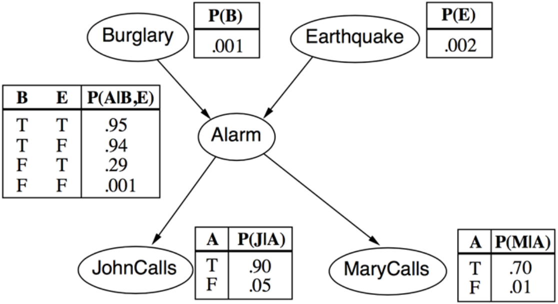
	ABE	A¬BE	AB¬E	A¬B¬E	¬ABE	¬A¬BE	¬AB¬E	¬A¬B¬E
JM	?	?	?	?	?	?	?	?
¬JM	?	?	?	?	?	?	?	?
J¬M	?	?	?	?	?	?	?	?
¬J¬M	?	?	?	?	?	?	?	?

- In addition to the large space, it is very difficult to calculate the probability of the 32 cases.
- Bayesian network separates independent events, and identify the conditional probability relationship between nodes.

  For example, It is more easy to estimate the probability of the thief making the alarm sound, the probability of the earthquake making the alarm sound. Sometimes can be a parameter of the alarm.

Calculation of Fully Joint Probability in Bayesian Networks

P(JMABE)
=P(J | A)P(M | A)P(A | BE)P(B)P(E)
=0.9*0.7*0.95*0.001*0.002
= 0.0001197
P(JMAB¬E)
=P(J | A)P(M | A)P(A | B¬E)P(B)P(¬E)
=0.9*0.7*0.94*0.001*0.998
= 0.0005910156
P(JM¬AB¬E)
=P(J | ¬A)P(M | ¬A)P(¬A | B¬E)P(B)P(¬E)
=P(J | ¬A)P(M | ¬A)(1-P(A | B¬E))P(B)(1-P(E))
=0.05*0.01*(1-0.94)*0.001*(1-0.002)
=0.05*0.01*0.06*0.001*0.998
=0.00000002994



Precise Inference of Bayesian Networks: Enumeration

- The basic task of any probabilistic inference system is to calculate the posterior probability distribution of a set of query variables given a particular observed event.
- Example: When JohnCalls and MaryCalls are known to be true, what is the probability of a thief?

$$P(B|J = true, M = true) =?$$

- Here $P(B|J = true, M = true)$ is the posterior probability. The variable B is the query variable, the evidence variable set is $\{J, M\}$, the hidden variable set is $\{A, E\}$, and the specific events observed are $J =$ true and $M =$ true (the so-called specific event is to give assignments to the evidence variable).

Precise Inference of Bayesian Networks: Enumeration

- P(B | J=true,M=true)

= <P(B=true,J=true,M=true), P(B=false, J=true,M=true)>

P(B=true,J=true,M=true)

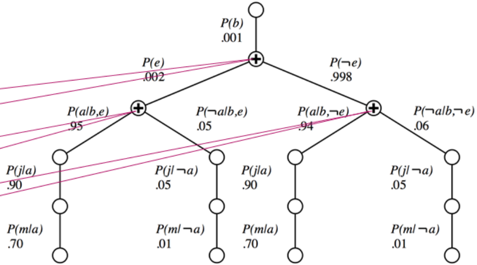
=P(B=true,J=true,M=true,E=true)

+P(B=true,J=true,M=true,E=false)

=P(B=true,J=true,M=true,E=true,A=true)

+P(B=true,J=true,M=true,E=true,A=flase)

+P(B=true,J=true,M=true,E=false,A=true)
+P(B=true,J=true,M=true,E=false,A=false)



There are calculations for repeated redundancy in the enumeration method. For example, $P(J|A)$ and $P(M|A)$ appeared twice. $P(J|A), P(M|A)$ also appeared twice.

Assignment

- $P(B = \text{false}, J = \text{true}, M = \text{true}) = ?$
- According to the previous example, please draw the structure graph to calculate the above probability

Approximate Inference in Bayesian Network: Monte Carlo Sampling

Why approximate inference?

- When the Bayesian network is single-connected, the scale of the problem grows linearly. But when it is connected, it will grow exponentially.
- The precise reasoning of the Bayesian network problem is NP-hard.
- It is impractical to use precise inference for large-scale multi-connectivity problems.

Application of Monte Carlo

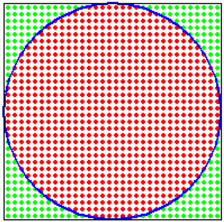
- Some models contain uncertain random factors
- Some models are difficult to quantify, or the cost of quantification is too high

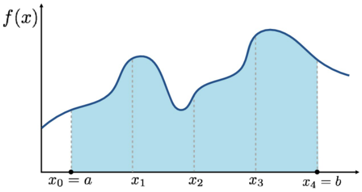
Procedure of Monte-Carlo

- The problem can be expressed by probability, or it can be used to estimate a certain feature, or a parameter which helps to obtain the solution to the problem.
- Simulate a process using random numbers or pseudo-random numbers, and count the probability of an event occurring through multiple experiments. The more number of experiments, the more accurate the statistics.

Monte Carlo

- It refers to algorithms using random numbers (pseudo random variables) to solve computation problems
- Is it a name of a person? No, it is the name of a city famous for gambling
- Classic applications
  - computation of π
  - Monte Carlo integration
- Development attributed to the increment on the computing power





Simple Sampling

```
def Sample(p):
    return p>random.uniform(0.0, 1.0)
```

Solution

- P(B=false, J=true,M=true) = ?
- Struture graph: P(B=false,J=true,M=true)

=P(B=false,J=true,M=true,E=true)

+P(B=false,J=true,M=true,E=false)

=P(B=false,J=true,M=true,E=true,A=true)

+P(B=false,J=true,M=true,E=true,A=flase)

+P(B=false,J=true,M=true,E=false,A=true)

+P(B=false,J=true,M=true,E=false,A=false)