Probabilistic Reasoning in Artificial Intelligence

Uncertainty

- Knowledge representation with first-order logic and propositional logic is based on certainty, means we are sure about the predicates. For example, we can say that A→B, which means if A is true then B is true, but what about the situation where we are not sure about whether A is true or not then we cannot express this statement, this situation comes under uncertainty.
- So to characterize uncertain knowledge, where we are not sure about the predicates, we need uncertain reasoning or probabilistic reasoning.

Causes of uncertainty:

Following are some leading causes of uncertainty to occur in the real world.

1. Information occurred from unreliable sources.
2. Experimental Errors
3. Equipment fault
4. Temperature variation
5. Climate change.

Need of Probabilistic Reasoning in AI

- **Unpredictable outcomes**
- **Predicates are too large to handle**
- **Unknown error occurs**

In probabilistic reasoning, there are two methods to solve difficulties with uncertain knowledge:

  - **Bayes' rule**
  - **Bayesian Statistics**

- **Probability:** Probability can be defined as chance of occurrence of an uncertain event. It is the numerical measure of the likelihood that an event will occur. The value of probability always remains between 0 and 1.

  $0 \leq P(X) \leq 1$, where $P(X)$ is the probability of an event X.

- $P(X) = 0$, indicates total uncertainty in an event X.
- $P(X) = 1$, indicates total certainty in an event X.

Bayes' Theorem in Artificial Intelligence

Bayes' theorem determines the probability of an event with uncertain knowledge.

It is a way to calculate the value of P(B|A) with the knowledge of P(A|B).

- **Example**: If cancer corresponds to one's age then by using Bayes' theorem, we can determine the probability of cancer more accurately with the help of age.

Bayes' theorem can be derived using product rule and conditional probability of event A with known event B:

As from product rule we can write: **$P(A \wedge B) = P(A|B)\, P(B)$**

Similarly, the probability of event B with known event A: **$P(A \wedge B) = P(B|A)\, P(A)$** , here $P(A \wedge B)$ is joint probability.

- Equating right hand side of both the equations, we will get:

$$P(A|B) = \frac{P(B|A)\, P(A)}{P(B)} \quad \text{....(a)}$$

- The above equation (a) is called as **Bayes' rule** or **Bayes' theorem**. This equation is basic of most modern AI systems for **probabilistic inference**.
- Here, P(A|B) is known as **posterior**, which we need to calculate and it will be read as Probability of hypothesis A with occurrence of an evidence B.

- P(B|A) - likelihood, in which we consider that hypothesis is true, then we compute the likelihood of evidence.
- P(A) - **prior probability**, probability of hypothesis before seeing the evidence
- P(B) - **marginal probability** of an evidence.
- In the equation (a), in general, we can write P (B) = P(A)*P(B|Ai), hence the Bayes' rule can be written as:

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

- Where A_1, A_2,........, A_n is a set of mutually exclusive and exhaustive events.

- **Question: From a standard deck of playing cards, a single card is drawn. The probability that the card is king is 4/52, then calculate posterior probability P(King|Face), which means the drawn face card is a king card.**
- **Solution:**

$$P(king|face) = \frac{P(Face|king) * P(King)}{P(Face)} \quad ....(i)$$

- P(king): probability that the card is King= 4/52= 1/13
- P(face): probability that a card is a face card= 3/13
- P(Face|King): probability of face card when we assume it is a king = 1
- Putting all values in equation (i) we will get:

$$P(king|face) = \frac{1 * (\frac{1}{13})}{(\frac{3}{13})} = \mathbf{1/3},$$ **it is a probability that a face card is a king card.**

Bayesian Belief Network

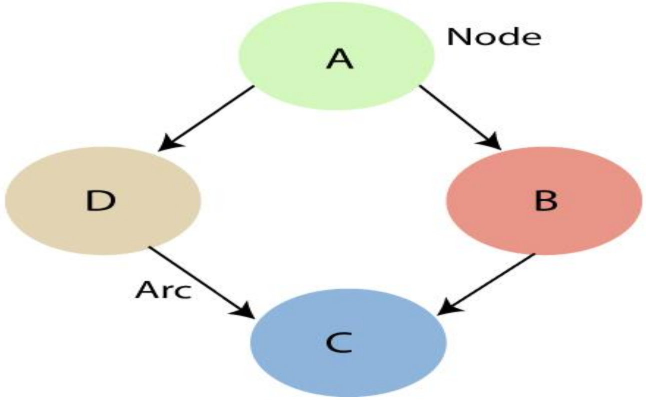
Bayesian Network can be used for building models from data and experts opinions, and it consists of two parts:

- **Directed Acyclic Graph**
- **Table of conditional probabilities**

The generalized form of Bayesian network that represents and solve decision problems under uncertain knowledge is known as an **Influence diagram**.

Note: It is used to represent conditional dependencies.

A Bayesian network graph is made up of nodes and Arcs (directed links), where:



- Each **node** corresponds to the random variables, and a variable can be **continuous** or **discrete**.
- **Arc or directed arrows** represent the causal relationship or conditional probabilities between random variables. These directed links or arrows connect the pair of nodes in the graph. These links represent that one node directly influence the other node, and if there is no directed link that means that nodes are independent with each other
- **Note: The Bayesian network graph does not contain any cyclic graph. Hence, it is known as a directed acyclic graph or DAG.**
- The Bayesian network has mainly two components:
  1. **Causal Component**
  2. **Actual numbers**
- Each node in the Bayesian network has condition probability distribution **$P(X_i \mid Parent(X_i))$**, which determines the effect of the parent on that node.
- Bayesian network is based on Joint probability distribution and conditional probability.

- **Example:** Harry installed a new burglar alarm at his home to detect burglary. The alarm reliably responds at detecting a burglary but also responds for minor earthquakes. Harry has two neighbors David and Sophia, who have taken a responsibility to inform Harry at work when they hear the alarm. David always calls Harry when he hears the alarm, but sometimes he got confused with the phone ringing and calls at that time too. On the other hand, Sophia likes to listen to high music, so sometimes she misses to hear the alarm. Here we would like to compute the probability of Burglary Alarm.

- **Problem: Calculate the probability that alarm has sounded, but there is neither a burglary, nor an earthquake occurred, and David and Sophia both called the Harry.**

Note: List of all events occurring in this network:

Burglary (B)

Earthquake(E)

Alarm(A)

David Calls(D)

Sophia calls(S)

- From the formula of joint distribution, we can write the problem statement in the form of probability distribution:

$$P(S, D, A, \neg B, \neg E) = P(S|A) * P(D|A) * P(A|\neg B \wedge \neg E) * P(\neg B) * P(\neg E)$$

- **Example:** Harry installed a new burglar alarm at his home to detect burglary. The alarm reliably responds at detecting a burglary but also responds for minor earthquakes. Harry has two neighbors David and Sophia, who have taken a responsibility to inform Harry at work when they hear the alarm. David always calls Harry when he hears the alarm, but sometimes he got confused with the phone ringing and calls at that time too. On the other hand, Sophia likes to listen to high music, so sometimes she misses to hear the alarm. Here we would like to compute the probability of Burglary Alarm.

- **Problem: Calculate the probability that alarm has sounded, but there is neither a burglary, nor an earthquake occurred, and David and Sophia both called the Harry.**

Note: List of all events occurring in this network:

Burglary (B)

Earthquake(E)

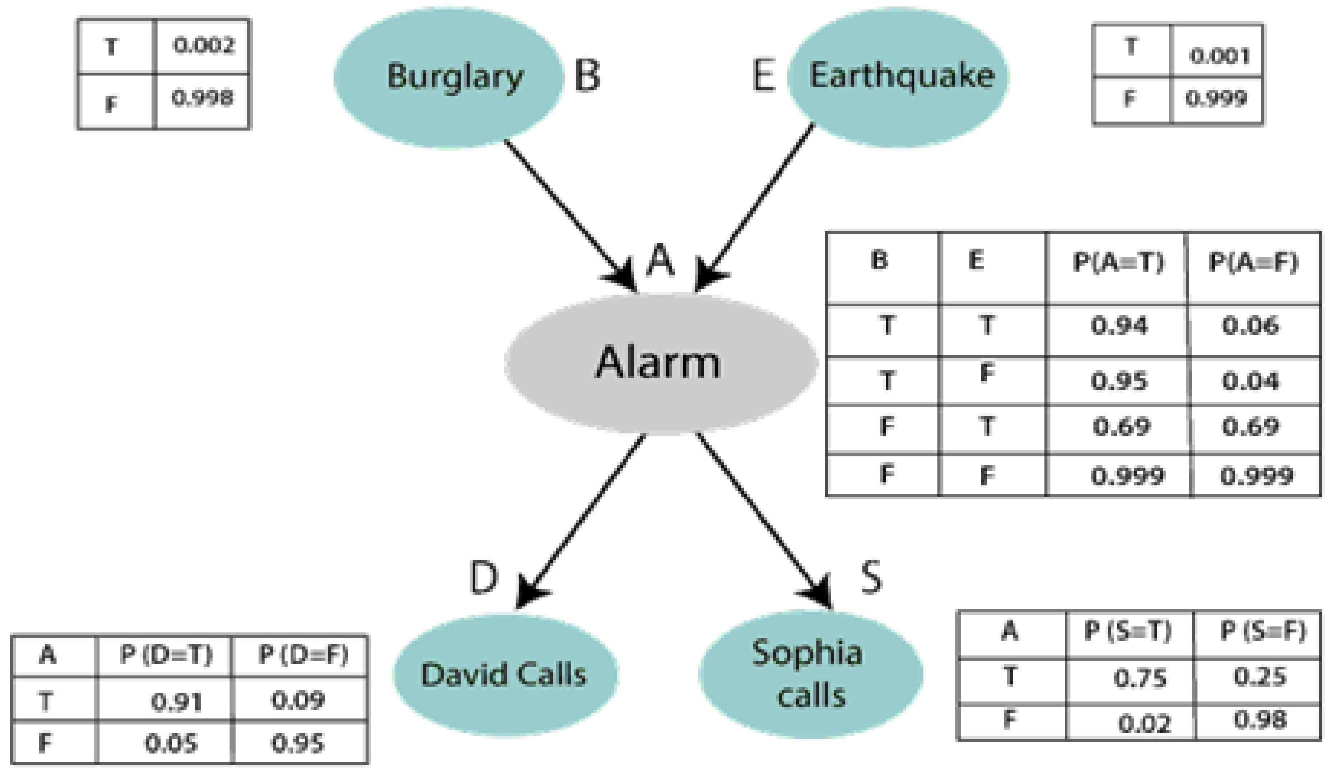
Alarm(A)

David Calls(D)

Sophia calls(S)

- From the formula of joint distribution, we can write the problem statement in the form of probability distribution:

$$P(S, D, A, \neg B, \neg E) = P(S|A) * P(D|A) * P(A|\neg B \wedge \neg E) * P(\neg B) * P(\neg E)$$



T	0.002
F	0.998

T	0.001
F	0.999

B	E	P(A=T)	P(A=F)
T	T	0.94	0.06
T	F	0.95	0.04
F	T	0.69	0.69
F	F	0.999	0.999

A	P (D=T)	P (D=F)
T	0.91	0.09
F	0.05	0.95

A	P (S=T)	P (S=F)
T	0.75	0.25
F	0.02	0.98

P(S, D, A, ¬B, ¬E) = P (S|A) *P (D|A)*P (A|¬B ^ ¬E) *P (¬B) *P (¬E).

= 0.75* 0.91* 0.001* 0.998*0.999

= 0.00068045.

Hence, a Bayesian network can answer any query about the domain by using Joint distribution.