Artificial Intelligence

Open Elective

Module 3: Symbolic Reasoning Under Uncertainty CH7

Dr. Santhi Natarajan
Associate Professor
Dept of AI and ML
BMSIT, Bangalore

Uncertainty

- With FOL or propositional logic based knowledge representation, we might write A→B, which means if A is true then B is true

- But consider a situation where we are not sure about whether A is true or not then we cannot express this statement, this situation is called uncertainty.

- So to represent uncertain knowledge, where we are not sure about the predicates, we need uncertain reasoning or probabilistic reasoning.

Probabilistic Reasoning

- Probabilistic reasoning is a way of knowledge representation where we apply the concept of probability to indicate the uncertainty in knowledge.

- In probabilistic reasoning, we combine probability theory with logic to handle the uncertainty.

- We use probability in probabilistic reasoning because it provides a way to handle the uncertainty that is the result of someone's laziness and ignorance.

- In the real world, there are lots of scenarios, where the certainty of something is not confirmed, such as
  - "It will rain today,"
  - "behavior of someone for some situations,"
  - "A match between two teams or two players."

Need for Probabilistic Reasoning

Need of probabilistic reasoning in AI:

- When there are unpredictable outcomes.
- When specifications or possibilities of predicates becomes too large to handle.
- When an unknown error occurs during an experiment.

In probabilistic reasoning, there are two ways to solve problems with uncertain knowledge:

- Bayes' rule
- Bayesian Statistics

Probability

Probability

Probability can be defined as a chance that an uncertain event will occur. It is the numerical measure of the likelihood that an event will occur. The value of probability always remains between 0 and 1 that represent ideal uncertainties.

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

P(A) = 0, indicates total uncertainty in an event A.

P(A) =1, indicates total certainty in an event A.

We can find the probability of an uncertain event by using the below formula.

$$\textbf{Probability of occurrence} = \frac{\text{Number of desired outcomes}}{\text{Total number of outcomes}}$$

Probability

- P(¬A) = probability of a not happening event.
- P(¬A) + P(A) = 1.
- **Event:** Each possible outcome of a variable is called an event.
- **Sample space:** The collection of all possible events is called sample space.
- **Random variables:** Random variables are used to represent the events and objects in the real world.
- **Prior probability:** The prior probability of an event is probability computed before observing new information.
- **Posterior Probability:** The probability that is calculated after all evidence or information has taken into account. It is a combination of prior probability and new information.

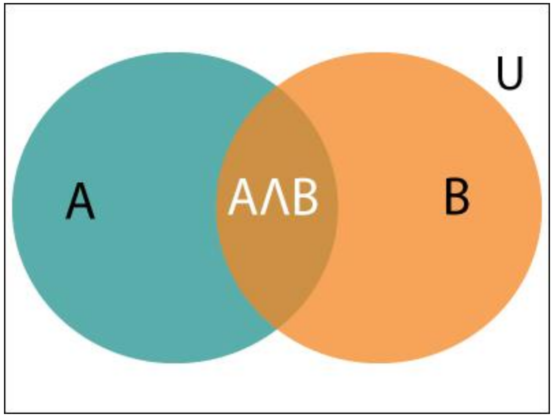
Probability

Conditional probability

- Conditional probability is a probability of occurring an event when another event has already happened.

- Let's suppose, we want to calculate the event A when event B has already occurred, "the probability of A under the conditions of B", it can be written as:

$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$



Where P(A∧B)= Joint probability of a and B

P(B)= Marginal probability of B.

If the probability of A is given and we need to find the probability of B, then it will be given as:

$$P(B|A) = \frac{P(A \wedge B)}{P(A)}$$

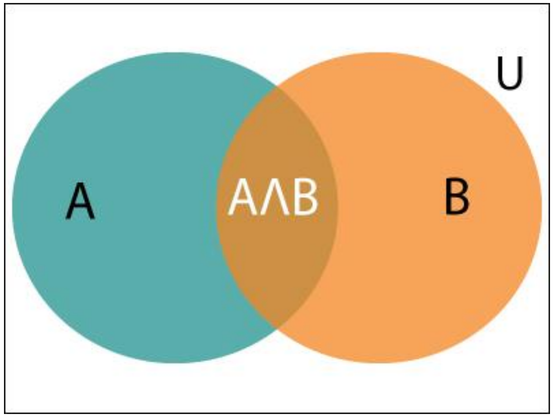
Probability

Conditional probability

- Conditional probability is a probability of occurring an event when another event has already happened.

- Let's suppose, we want to calculate the event A when event B has already occurred, "the probability of A under the conditions of B", it can be written as:

$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$



Where P(A∧B)= Joint probability of a and B

P(B)= Marginal probability of B.

If the probability of A is given and we need to find the probability of B, then it will be given as:

$$P(B|A) = \frac{P(A \wedge B)}{P(A)}$$

Probability

- In a class, there are 70% of the students who like English and 40% of the students who likes English and mathematics, and then what is the percent of students those who like English also like mathematics?
- **Solution:**
- Let, A is an event that a student likes Mathematics
- B is an event that a student likes English.
- **Hence, 57% are the students who like English also like Mathematics.**

$$P(A|B) = \frac{P(A \wedge B)}{P(B)} = \frac{0.4}{0.7} = 57\%$$

Bayesian Probability

Bayes' theorem

- Bayes' theorem is also known as **Bayes' rule, Bayes' law**, or **Bayesian reasoning**, which determines the probability of an event with uncertain knowledge

.

- In probability theory, it relates the conditional probability and marginal probabilities of two random events.

- Bayes' theorem was named after the British mathematician **Thomas Bayes**. The **Bayesian inference** is an application of Bayes' theorem, which is fundamental to Bayesian statistics.

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

- Bayes' theorem allows updating the probability prediction of an event by observing new information of the real world.

- **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

Bayesian Probability

Bayes' theorem

- $P(A \wedge B) = P(A|B)\,P(B)$

- $P(A \wedge B) = P(B|A)\,P(A)$

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

- The above equation (a) is called as **Bayes' rule** or **Bayes' theorem**. This equation is basic of most modern AI systems for **probabilistic inference**.

Bayesian Probability

Bayes' theorem

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

- P(A|B) is known as **posterior**, which we need to calculate, and it will be read as Probability of hypothesis A when we have occurred an evidence B.

- P(B|A) is called the likelihood, in which we consider that hypothesis is true, then we calculate the probability of evidence.

- P(A) is called the **prior probability**, probability of hypothesis before considering the evidence

- P(B) is called **marginal probability**, pure 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 $A_1, A_2, A_3, \ldots\ldots, A_n$ is a set of mutually exclusive and exhaustive events.

Bayesian Probability

Bayes' theorem

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

- P(A|B) is known as **posterior**, which we need to calculate, and it will be read as Probability of hypothesis A when we have occurred an evidence B.

- P(B|A) is called the likelihood, in which we consider that hypothesis is true, then we calculate the probability of evidence.

- P(A) is called the **prior probability**, probability of hypothesis before considering the evidence

- P(B) is called **marginal probability**, pure 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 $A_1, A_2, A_3, \ldots\ldots, A_n$ is a set of mutually exclusive and exhaustive events.

Bayesian Probability

Bayes' theorem

- Bayes' rule allows us to compute the single term P(B|A) in terms of P(A|B), P(*B*), and P(A).

- This is very useful in cases where we have a good probability of these three terms and want to determine the fourth one.

- Suppose we want to perceive the effect of some unknown cause, and want to compute that cause, then the Bayes' rule becomes:

$$P(\text{cause} \mid \text{effect}) = \frac{P(\text{effect} \mid \text{cause})\, P(\text{cause})}{P(\text{effect})}$$

Bayesian Probability

Bayes' theorem

Example-1:

Question: what is the probability that a patient has diseases meningitis with a stiff neck?

Given Data:

- A doctor is aware that disease meningitis causes a patient to have a stiff neck, and it occurs 80% of the time. He is also aware of some more facts, which are given as follows:
- The Known probability that a patient has meningitis disease is 1/30,000.
- The Known probability that a patient has a stiff neck is 2%.

Solution

- Let a be the proposition that patient has stiff neck and b be the proposition that patient has meningitis. , so we can calculate the following as:
- P(a|b) = 0.8
- P(b) = 1/30000
- P(a)= .02

$$P(b|a) = \frac{P(a|b)P(b)}{P(a)} = \frac{0.8*(\frac{1}{30000})}{0.02} = 0.001333333.$$

Hence, we can assume that 1 patient out of 750 patients has meningitis disease with a stiff neck.

Bayesian Probability

Bayes' theorem

Example-2:

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.

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

Solution

- 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:

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

Bayesian Probability

Bayes' theorem

Following are some applications of Bayes' theorem:

- It is used to calculate the next step of the robot when the already executed step is given.
- Bayes' theorem is helpful in weather forecasting.
- It can solve the Monty Hall problem.

Bayesian Probability

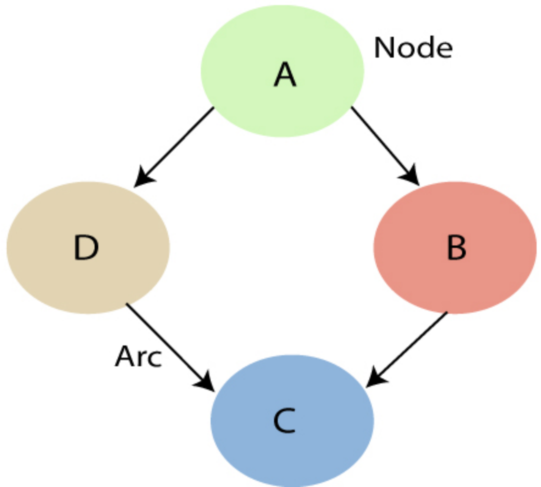
Bayesian Belief Network in artificial intelligence

- Bayesian belief network is key computer technology for dealing with probabilistic events and to solve a problem which has uncertainty. We can define a Bayesian network as:
- "A Bayesian network is a probabilistic graphical model which represents a set of variables and their conditional dependencies using a directed acyclic graph."
- It is also called a Bayes network, belief network, decision network, or Bayesian model.
- Bayesian networks are probabilistic, because these networks are built from a **probability distribution**, and also use probability theory for prediction and anomaly detection.
- Real world applications are probabilistic in nature, and to represent the relationship between multiple events, we need a Bayesian network.
- It can also be used in various tasks including **prediction, anomaly detection, diagnostics, automated insight, reasoning, time series prediction**, and **decision making under uncertainty**.

Bayesian Probability

Bayesian Belief Network in artificial intelligence

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



Bayesian Probability

Bayesian Belief Network in artificial intelligence

- 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

- In the above diagram, A, B, C, and D are random variables represented by the nodes of the network graph.

- If we are considering node B, which is connected with node A by a directed arrow, then node A is called the parent of Node B.

- Node C is independent of node A.

- The Bayesian network graph does not contain any cyclic graph. Hence, it is known as a **directed acyclic graph or DAG**.

Bayesian Probability

Bayesian Belief Network in artificial intelligence

The Bayesian network has mainly two components:

Causal Component

Actual numbers

Each node in the Bayesian network has condition probability distribution $\mathbf{P(X_i \mid Parent(X_i))}$, which determines the effect of the parent on that node.

Bayesian network is based on Joint probability distribution and conditional probability.

Bayesian Probability

Bayesian Belief Network in artificial intelligence

Joint probability distribution:

If we have variables x1, x2, x3,....., xn, then the probabilities of a different combination of x1, x2, x3.. xn, are known as Joint probability distribution.

$P[x_1, x_2, x_3, \ldots, x_n]$, it can be written as the following way in terms of the joint probability distribution.

$= P[x_1 | x_2, x_3, \ldots, x_n] P[x_2, x_3, \ldots, x_n]$

$= P[x_1 | x_2, x_3, \ldots, x_n] P[x_2 | x_3, \ldots, x_n] \ldots P[x_{n-1} | x_n] P[x_n].$

In general for each variable Xi, we can write the equation as:

$P(X_i | X_{i-1}, \ldots, X_1) = P(X_i | Parents(X_i))$

Bayesian Probability

Bayesian Belief Network in artificial intelligence

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.

Bayesian Probability

Solution:

List of all events occurring in this network:
Burglary (B)
Earthquake(E)
Alarm(A)
David Calls(D)
Sophia calls(S)

We can write the events of problem statement in the form of probability: **P[D, S, A, B, E]**, can rewrite the above probability statement using joint probability distribution:

P[D, S, A, B, E]= P[D | S, A, B, E]. P[S, A, B, E]

=P[D | S, A, B, E]. P[S | A, B, E]. P[A, B, E]

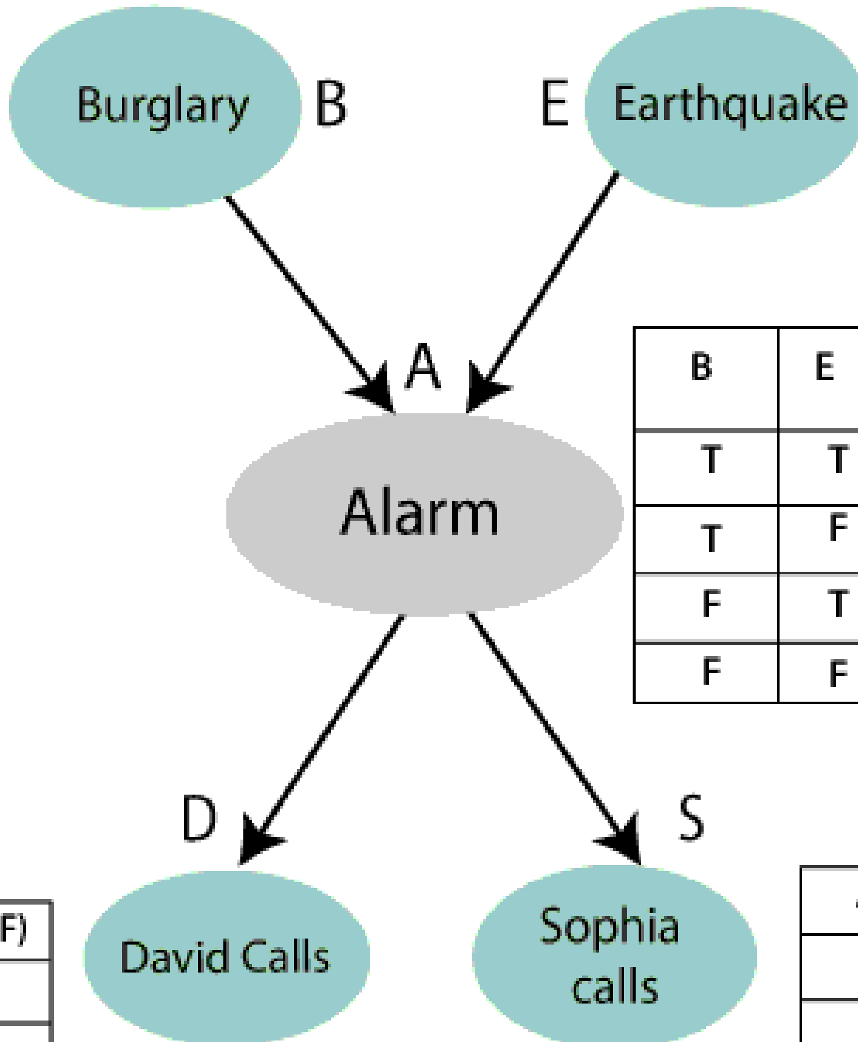
= P [D| A]. P [S| A, B, E]. P[A, B, E]

= P[D | A]. P[S | A]. P[A| B, E]. P[B, E]

= P[D | A]. P[S | A]. P[A| B, E]. P[B |E]. P[E]

Bayesian Probability

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

Bayesian Probability

- The network structure is showing that burglary and earthquake is the parent node of the alarm and directly affecting the probability of alarm's going off, but David and Sophia's calls depend on alarm probability.

- The network is representing that our assumptions do not directly perceive the burglary and also do not notice the minor earthquake, and they also not confer before calling.

- The conditional distributions for each node are given as conditional probabilities table or CPT.

- Each row in the CPT must be sum to 1 because all the entries in the table represent an exhaustive set of cases for the variable.

- In CPT, a boolean variable with k boolean parents contains 2^K probabilities. Hence, if there are two parents, then CPT will contain 4 probability values

Bayesian Probability

- Let's take the observed probability for the Burglary and earthquake component:
- P(B= True) = 0.002, which is the probability of burglary.
- P(B= False)= 0.998, which is the probability of no burglary.
- P(E= True)= 0.001, which is the probability of a minor earthquake
- P(E= False)= 0.999, Which is the probability that an earthquake not occurred.

Conditional probability table for Alarm A:

The Conditional probability of Alarm A depends on Burglar and earthquake:

B	E	P(A= True)	P(A= False)
True	True	0.94	0.06
True	False	0.95	0.04
False	True	0.31	0.69
False	False	0.001	0.999

Bayesian Probability

Conditional probability table for David Calls:
The Conditional probability of David that he will call depends on the probability of Alarm.

A	P(D= True)	P(D= False)
True	0.91	0.09
False	0.05	0.95

Conditional probability table for Sophia Calls:
The Conditional probability of Sophia that she calls is depending on its Parent Node "Alarm."

A	P(S= True)	P(S= False)
True	0.75	0.25
False	0.02	0.98

Bayesian Probability

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

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.

Bayesian Probability

The semantics of Bayesian Network:
There are two ways to understand the semantics of the Bayesian network, which is given below:

1. **To understand the network as the representation of the Joint probability distribution.**

It is helpful to understand how to construct the network.

2. To understand the network as an encoding of a collection of conditional independence statements.

It is helpful in designing inference procedure.

Dempster Shafer (D-S) Theory

- Provides a numerical method to represent and reason about uncertainty.
- "Absence of evidence is not an evidence of absence".
- Provides a way to combine evidence from two or more sources and to draw conclusions from them.

Dempster Shafer (D-S) Theory

- **Frame of Discernment - ** Sample space of DS theory denoted by Ω
- **Propositions** - Subsets of frame of discernment
- Probability values are assigned to the propositions.
- **Basic Probability Assignments** - Probability values assigned to the propositions denoted by **m**.
- **Focal Elements** - Propositions with non-zero probability assignment.
- **Core** – Union of focal elements.