ARTIFICIAL INTELLIGENCE

Prof. Sneha Saini, Assistant Professor
IT & Computer Science

CHAPTER-7

Statistical Reasoning

Probabilistic reasoning in Artificial intelligence

Till now, we have learned knowledge representation using first-order logic and propositional logic with certainty, which means we were sure about the predicates.
With this knowledge representation, we might write $A \rightarrow B$, 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.

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

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." These are probable sentences for which we can assume that it will happen but not sure about it, so here we use 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**

As probabilistic reasoning uses probability and related terms, so before understanding probabilistic reasoning, let's understand some common terms:

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.

- 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.

- $0 \leq P(A) \leq 1$, where P(A) is the probability of an event A.
- $P(\neg A)$ = probability of a not happening event.
- $P(\neg A) + P(A) = 1$.

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

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.

Conditional probability:

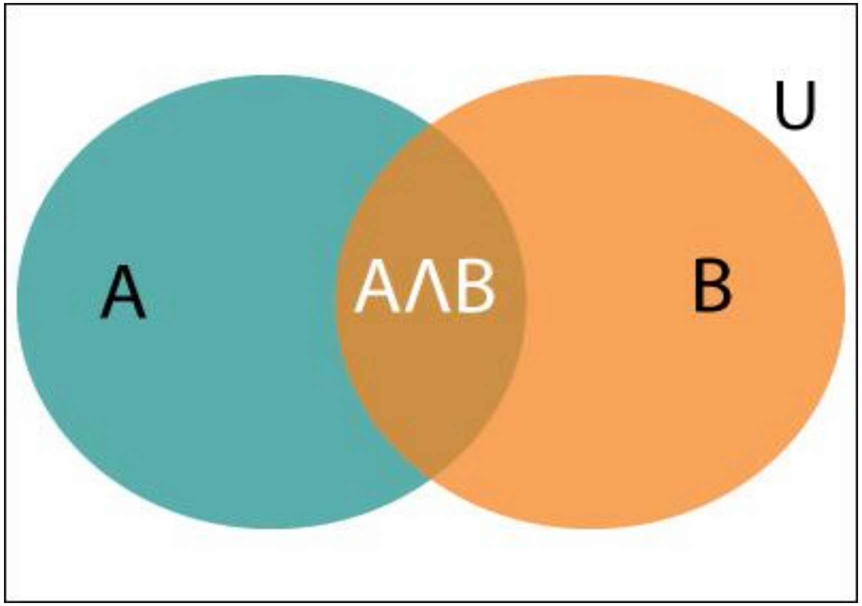
Conditional probability is a probability of occurring an event when another event has already happened.

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

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:

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

It can be explained by using the below Venn diagram, where B is occurred event, so sample space will be reduced to set B, and now we can only calculate event A when event B is already occurred by dividing the probability of **P(A⋀*B*) by P(B)**.



Example:
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.

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

Hence, 57% are the students who like English also like Mathematics.

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.

As from product rule we can write:

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

Similarly, the probability of event B with known event A:

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

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**.

It shows the simple relationship between joint and conditional probabilities. Here,

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.

Applying Bayes' rule:

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:

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

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%.

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.

Application of Bayes' theorem in Artificial intelligence:

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.

Dempster - Shafer Theory (DST)

DST is a mathematical theory of evidence based on belief functions and plausible reasoning. It is used to combine separate pieces of information (evidence) to calculate the probability of an event.

DST is a mathematical theory of evidence based on belief functions and plausible reasoning. It is used to combine separate pieces of information (evidence) to calculate the probability of an event.

Application of Bayes' theorem in Artificial intelligence:

DST offers an alternative to traditional probabilistic theory for the mathematical representation of uncertainty.DST can be regarded as, a more general approach to represent uncertainty than the Bayesian approach.
Bayesian methods are sometimes inappropriate

Example : Let A represent the proposition "Moore is attractive". Then the axioms of probability insist that P(A) + P(¬A) = 1.

Now suppose that Andrew does not even know who "Moore" is, then
We cannot say that Andrew believes the proposition if he has no idea what it means.
Also, it is not fair to say that he disbelieves the proposition.It would therefore be meaningful to denote Andrew's belief B of

B(A) and B(¬A) as both being 0.

Dempster-Shafer Model

- The idea is to allocate a number between 0 and 1 to indicate a degree of belief on a proposal as in the probability framework.However, it is not considered a probability but a belief mass. The distribution of masses is called basic belief assignment.
- In other words, in this formalism a degree of belief (referred as mass) is represented as a belief function rather than a Bayesian probability distribution.

Example: Belief assignment

Suppose a system has five members, say five independent states, and exactly one of which is actual. If the original set is called S, | S | = 5, then the set of all subsets (the power set) is called 2S.

If each possible subset as a binary vector (describing any member is present or not by writing 1 or 0), then 25 subsets are possible, ranging from the empty subset (0, 0, 0, 0, 0) to the "everything" subset (1, 1, 1, 1, 1).The "empty" subset represents a "contradiction", which is not true in any state, and is thus assigned a mass of one ;

The remaining masses are normalized so that their total is 1.

- The "empty" subset represents a "contradiction", which is not true in any state, and is thus assigned a mass of one ;

The remaining masses are normalized so that their total is 1.

The "everything" subset is labeled as "unknown"; it represents the state where all elements are present one , in the sense that you cannot tell which is actual.

Belief and Plausibility

Shafer's framework allows for belief about propositions to be represented as intervals, bounded by two values, belief (or support) and plausibility:

belief ≤ plausibility

Belief in a hypothesis is constituted by the sum of the masses of all sets enclosed by it (i.e. the sum of the masses of all subsets of the hypothesis). It is the amount of belief that directly supports a given hypothesis at least in part, forming a lower bound.

DIGITAL LEARNING CONTENT

Parul® University

www.paruluniversity.ac.in