Uncertainty in AI

Sukanta Ghosh

Need of Considering Probability

- Take a self-driving car for example — you can set the goal to get from A to B in an efficient and safe manner that follows all laws.
- But what happens if the traffic gets worse than expected, maybe because of an accident ahead? Sudden bad weather?
- Random events like a ball bouncing in the street, or a piece of trash flying straight into the car's camera?
- Self driving cars work on different sensors, which are not perfect always.

Causes of Uncertainty

- Information occurred from unreliable sources.
- Experimental Errors
- Equipment fault
- Temperature variation
- Climate change.

History of dealing with uncertainty

- Fuzzy logic was for a while a contender for the best approach to handle uncertain and imprecise information.
- Probability has turned out to be the best approach for reasoning under uncertainty, and almost all current AI applications are based, to at least some degree, on probabilities.

Why probability matters

- What are the chances of getting three of a king in poker? (about 1 in 47)
- What are the chances of winning in the lottery? (very small)

- Probability is used to quantify and compare risks in everyday life.
- What are the chances of crashing your car if you exceed the speed limit?

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.

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.

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.

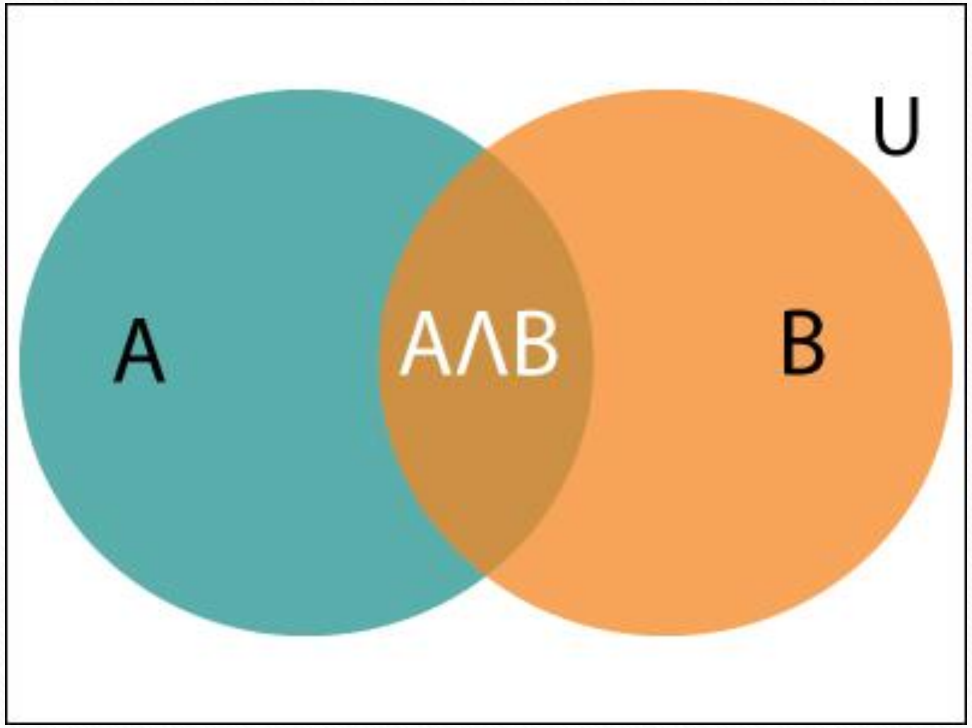
Conditional Probability

- 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.

Venn diagram - Conditional probability



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.

Derivation of Formula

- $P(A \land B) = P(A|B)\,P(B)$ or
- $P(A \land B) = P(B|A)\,P(A)$ { Similarly }
- Equating right hand side of both the equations, we will get

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

Bayes' theorem

- 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.

Bayes' theorem

- $P(B|A)$ 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.

Bayes' theorem

- Bayes' rule can be written as:

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

Applying Bayes' rule

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

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

Example

- **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.

Example

- **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.**

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

- 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.**

Application 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 Belief Network in AI

- 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.

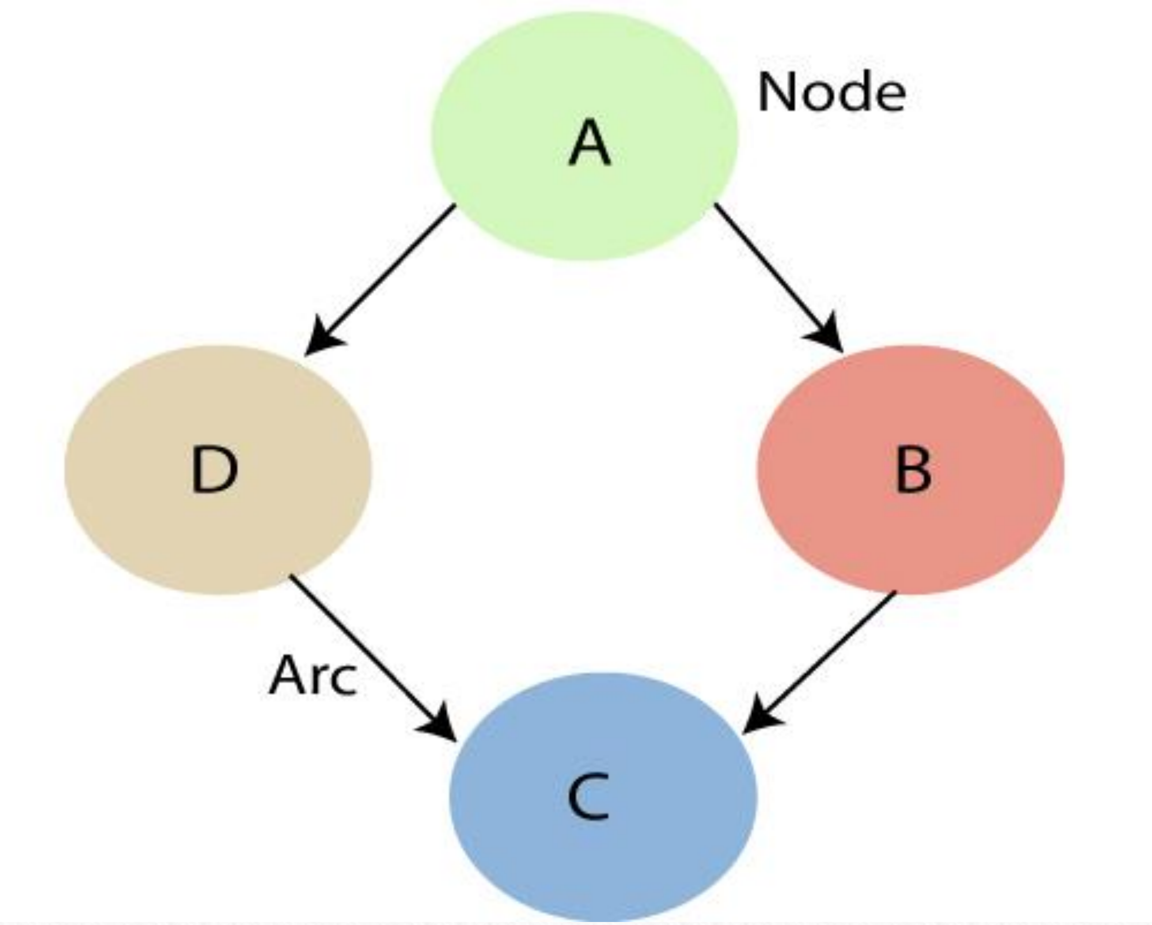
- Real world applications are probabilistic in nature, and to represent the relationship between multiple events, we need a Bayesian network.
- It can also be used in various tasks including **prediction, anomaly detection, diagnostics, automated insight, reasoning, time series prediction**, and **decision making under uncertainty**.

Bayesian Network

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

Directed Acyclic Graph

- **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.


Node
A
D
B
Arc
C

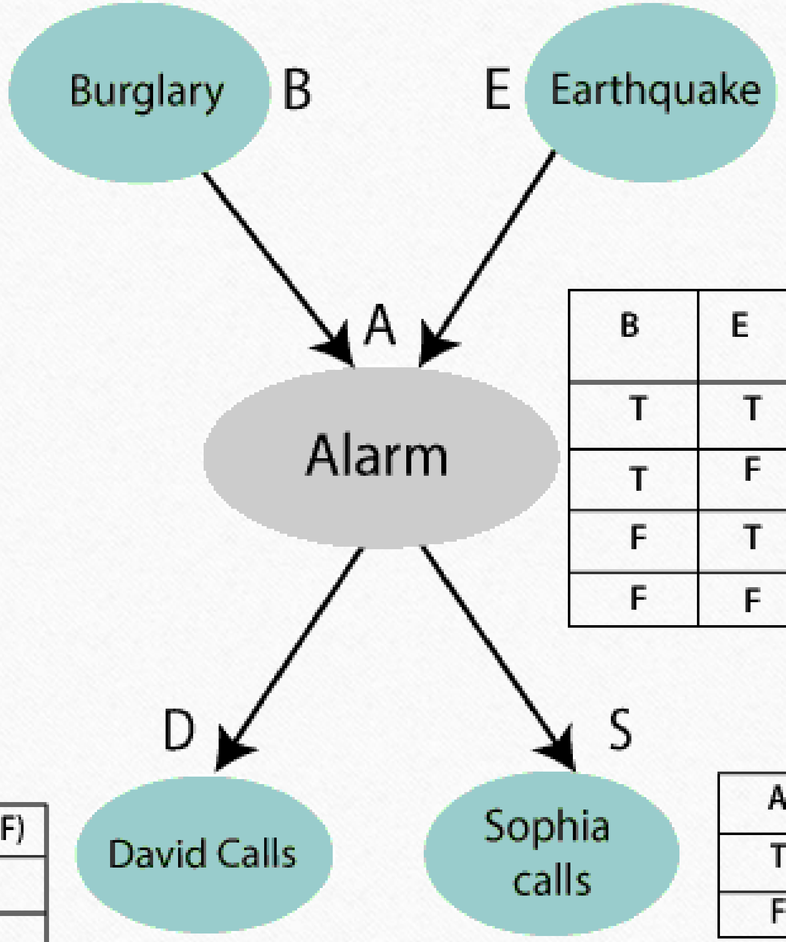
- These links represent that one node directly influence the other node, and if there is no directed link that means that nodes are independent with each other
  - **In the above diagram, A, B, C, and D are random variables represented by the nodes of the network graph.**
  - **If we are considering 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.**

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.
- **$P[x_1, x_2, x_3,....., x_n]$**, it can be written as the following way in terms of the joint probability distribution.
- **$= P[x_1 | x_2, x_3,....., x_n]P[x_2, x_3,....., x_n]$**
- **$= P[x_1 | x_2, x_3,....., x_n]P[x_2 | x_3,....., x_n]....P[x_{n-1} | x_n]P[x_n]$.**
- In general for each variable Xi, we can write the equation as:
- $P(X_i | X_{i-1},........., X_1) = P(X_i | Parents(X_i))$

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

Discussion points !!

- Why probability matters?
- Why quantifying uncertainty matters?

Probabilistic forecasts

- Consider the following four probabilistic forecasts and outcomes. What can we conclude based on the outcome about the correctness of the forecasts? Can we conclude that the probability given by the forecast was indeed the correct probability (choose "right"), that the forecast was wrong (choose "wrong"), or can we conclude neither way (choose "cannot be concluded")?

Probabilistic forecasts

- The weather forecast says it's going to rain with 90% probability tomorrow but the day turns out to be all sun and no rain.
- Right
- Wrong
- Cannot be concluded

Probabilistic forecasts

- The weather forecast says it's going to rain with 0% probability tomorrow but the day turns out to be rainy.
- Right
- Wrong
- Cannot be concluded

Probabilistic forecasts

- Suppose you monitor a weather forecaster for a long time. You only consider the days for which the forecast gives 80% chance of rain. You find that in the long run, on the average it rains on three out of every five days.
- Right
- Wrong
- Cannot be concluded

Probabilistic forecasts

- In the United States presidential election 2016, a well-known political forecast blog, Five-Thirty-Eight, gave Clinton a 71.4% chance of winning (vs Trump's 28.6%). However, contrary to the prediction, Donald Trump was elected the 45th president of the United States.

- Right

- Wrong

- Cannot be concluded

Fuzzy Logic

What is Fuzzy Logic?

- Fuzzy Logic (FL) is a method of reasoning that resembles human reasoning.
- The approach of FL imitates the way of decision making in humans that involves all intermediate possibilities between digital values YES and NO.
- The conventional logic block that a computer can understand takes precise input and produces a definite output as TRUE or FALSE, which is equivalent to human's YES or NO.

What is Fuzzy Logic?

- The inventor of fuzzy logic, Lotfi Zadeh, observed that unlike computers, the human decision making includes a range of possibilities between YES and NO, such as −
- CERTAINLY YES
- POSSIBLY YES
- CANNOT SAY
- POSSIBLY NO
- CERTAINLY NO

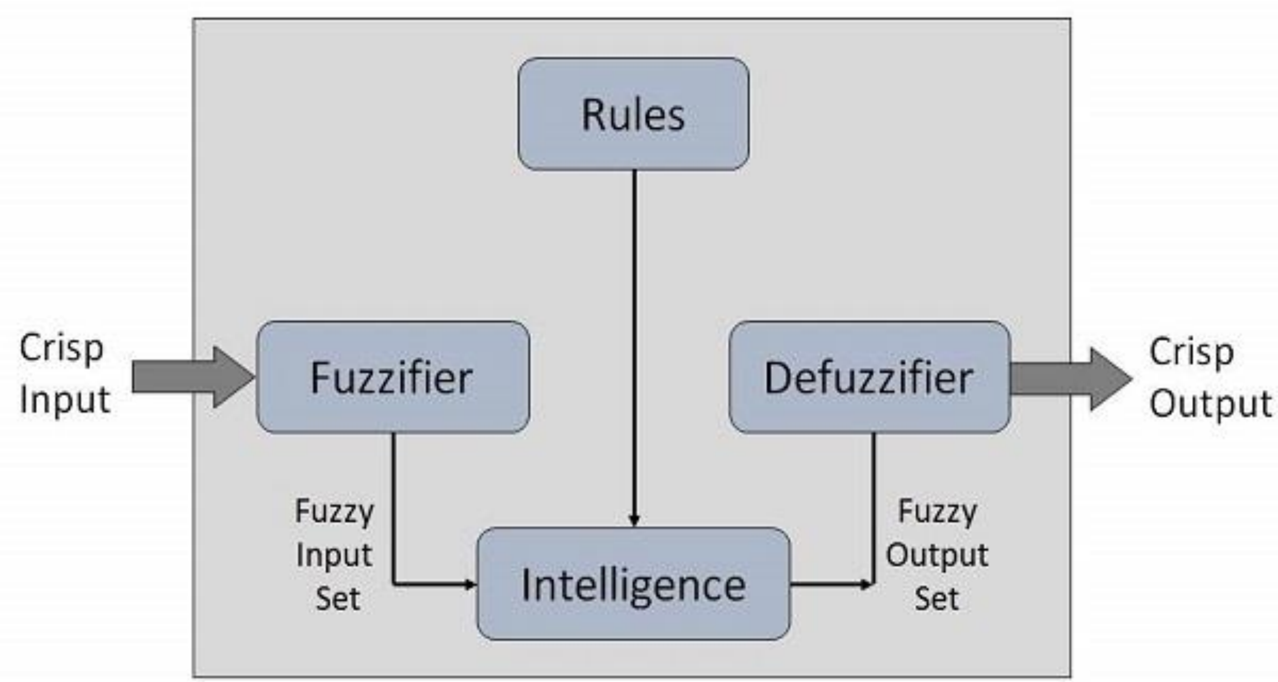
Implementation

- It can be implemented in systems with various sizes and capabilities ranging from small micro-controllers to large, networked, workstation-based control systems.
- It can be implemented in hardware, software, or a combination of both.

Why Fuzzy Logic?

- Fuzzy logic is useful for commercial and practical purposes.
- It can control machines and consumer products.
- It may not give accurate reasoning, but acceptable reasoning.
- Fuzzy logic helps to deal with the uncertainty in engineering.

Fuzzy Logic Systems Architecture



Fuzzy Logic Systems Architecture

- Fuzzification Module − It transforms the system inputs, which are crisp numbers, into fuzzy sets. It splits the input signal into five steps such as −
- LP x is Large Positive
- MP x is Medium Positive
- S x is Small
- MN x is Medium Negative
- LN x is Large Negative

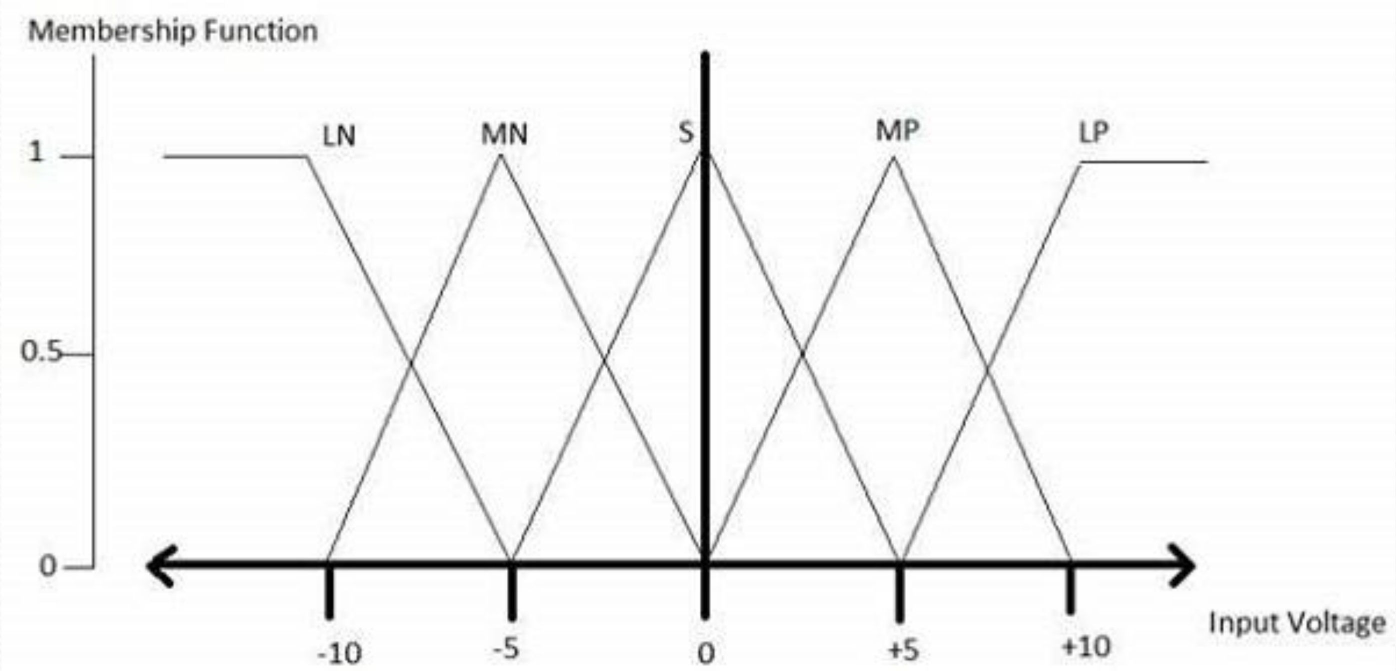
Fuzzy Logic Systems Architecture

- **Knowledge Base** – It stores IF-THEN rules provided by experts.
- **Inference Engine** – It simulates the human reasoning process by making fuzzy inference on the inputs and IF-THEN rules.
- **Defuzzification Module** – It transforms the fuzzy set obtained by the inference engine into a crisp value.

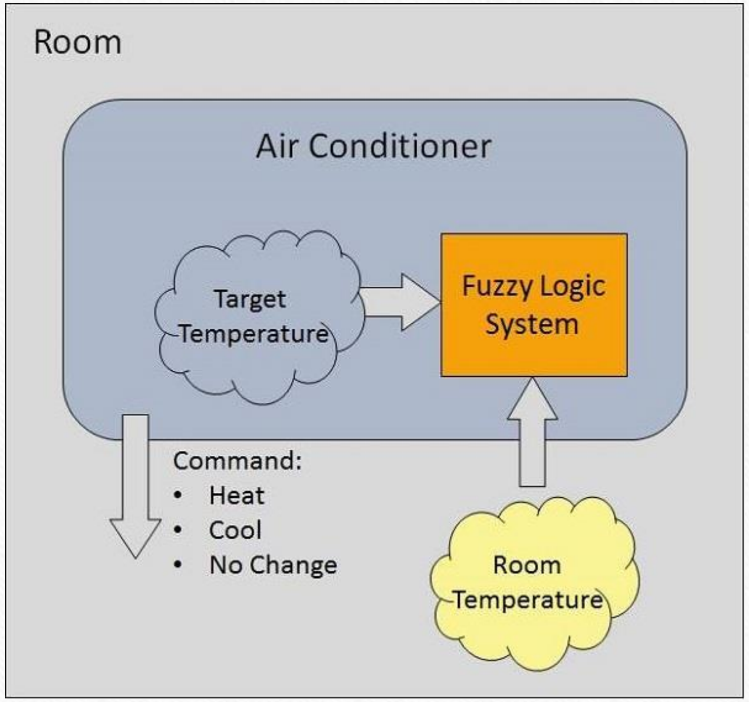
Membership Function

- Membership functions allow you to quantify linguistic term and represent a fuzzy set graphically.
- A **membership function** for a fuzzy *set A* on the universe of discourse X is defined as $\mu_A: X \rightarrow [0,1]$.
- Here, each element of X is mapped to a value between 0 and 1.
- It is called **membership value** or **degree of membership**.
- It quantifies the degree of membership of the element in X to the fuzzy set A.

All membership functions for **LP, MP, S, MN,** and **LN** are shown as below



Example of a Fuzzy Logic System



Application Areas of Fuzzy Logic

- Automotive Systems
- Consumer Electronic Goods
- Domestic Goods
- Environment Control

Advantages of FLSs

- Mathematical concepts within fuzzy reasoning are very simple.
- You can modify a FLS by just adding or deleting rules due to flexibility of fuzzy logic.
- Fuzzy logic Systems can take imprecise, distorted, noisy input information.
- FLSs are easy to construct and understand.
- Fuzzy logic is a solution to complex problems in all fields of life, including medicine, as it resembles human reasoning and decision making.

Disadvantages of FLSs

- There is no systematic approach to fuzzy system designing.
- They are understandable only when simple.
- They are suitable for the problems which do not need high accuracy.

“

Thank You

”