KNOWLEDGE INFERENCE

UNIT-3

Knowledge Representation

KNOWLEDGE REPRESENTATION AND INFERENCE

- What is knowledge
- What is knowledge representation (KR)
- Knowledge representation languages
- Approaches to KR
  - Semantic networks
  - Frames
  - Predicate Logic
  - Production Rules

KNOWLEDGE

- Knowledge is the collection of facts, inference rules etc. which can be used for a particular purpose.
- Knowledge requires the use of data and information.
- It combines relationships, correlations, dependencieswith data and information.
- The basic components of knowledge are:

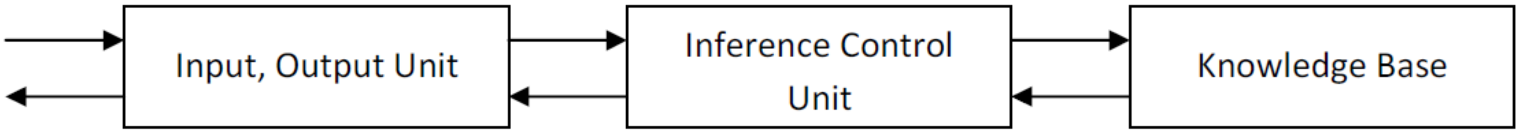
  1) A set of collected data

  2) A form of belief or hypothesis

  3) A kind of information.

KNOWLEDGE BASED SYSTEMS

Knowledge based systems get their power from the expert knowledge that has been coded into facts, rules, heuristics and procedures. The knowledge is stored in a knowledge base separate from the control and inferencing components. Knowledge is important and essential for knowledge based intelligent behaviour.



knowledge is information about objects, concepts and relationships that are assumed to exist in a particular area of interest.

KNOWLEDGE REPRESENTATION (KR)

- Knowledge representation is the most important ingredient for developing an AI.
- layer between information accessible from outside world and high level thinking processes.
- Knowledge representation refers to the data structures techniques and organizing notations that are used in AI.
- These include semantic networks, frames, logic, production rules and conceptual graphs.

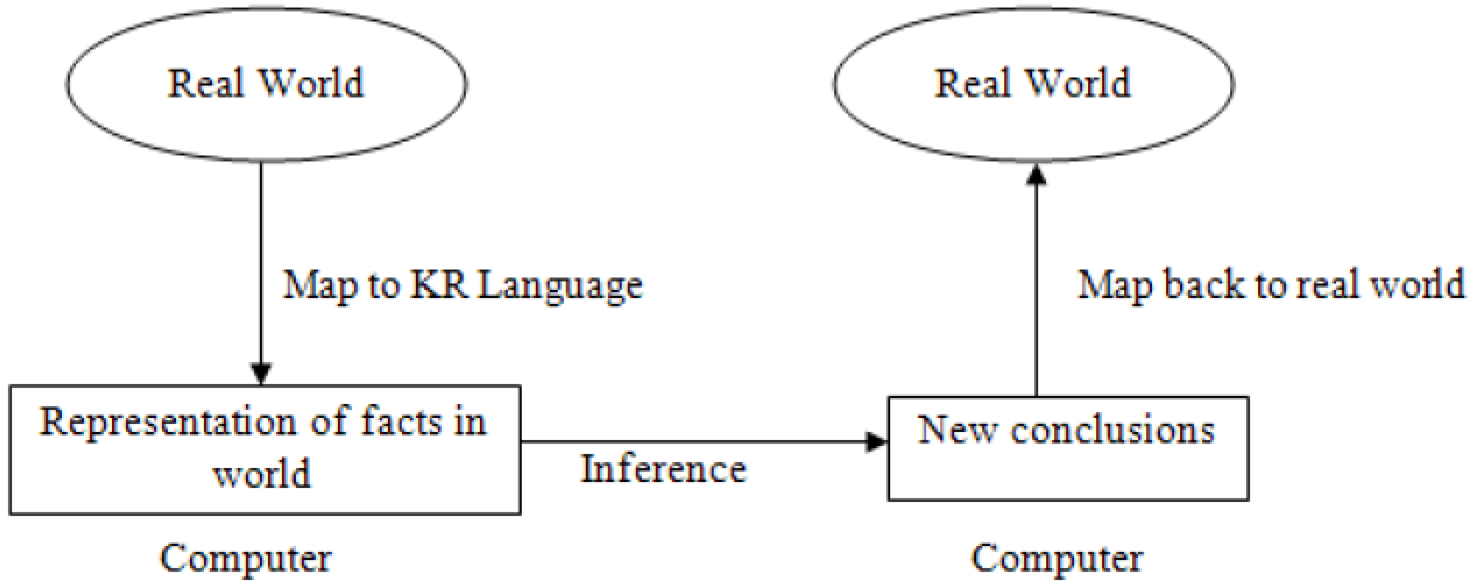
Properties for knowledge Representation

The following properties should be possessed by a knowledge representation system.

- a. **Representational Adequacy:** It is the ability to represent the required knowledge.
- b. **Inferential Adequacy:** It is the ability to manipulate the knowledge represented to produce new knowledge corresponding to that inferred from the original.
- c. **Inferential Efficiency**: The ability to direct the inferential mechanisms into the most productive directions by storing appropriate guides.
- d. **Acquisitional Efficiency:** The ability to acquire new knowledge using automatic methods wherever possible rather than reliance on human intervention.

Syntax and semantics for Knowledge Representation

Knowledge representation languages should have precise syntax and semantics. You must know exactly what an expression means in terms of objects in the real world. Suppose we have decided that "red 1" refers to a dark red colour, "car1" is my car, car2 is another. Syntax of language will tell you which of the following is legal: red1 (car1), red1 car1, car1 (red1), red1 (car1 & car2)?



Types of Knowledge Representation

- Knowledge can be represented in different ways.
- Different knowledge representation techniques are

  a. Logic

  b. Semantic Network

  c. Frame

  d. Production rules

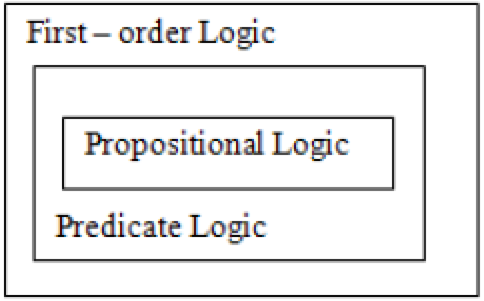
  e. Conceptual Dependency

  f. Script

First Order Predicate Logic

- **FOPL** has a well defined syntax and semantics,
- It is concerned with truth preserving inference.

Problems : time, beliefs and uncertainty are difficult to represent

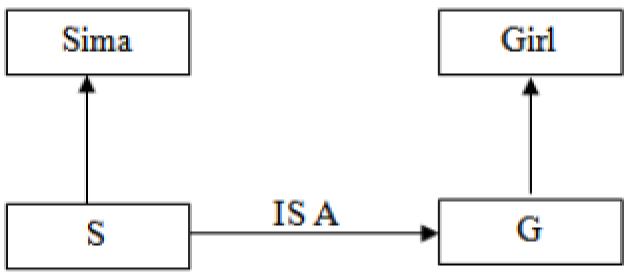


Semantic Network

- Represent factual knowledge about classes of objects and their properties
- Not formal systems.
- Basic inference mechanism: inheritance of properties

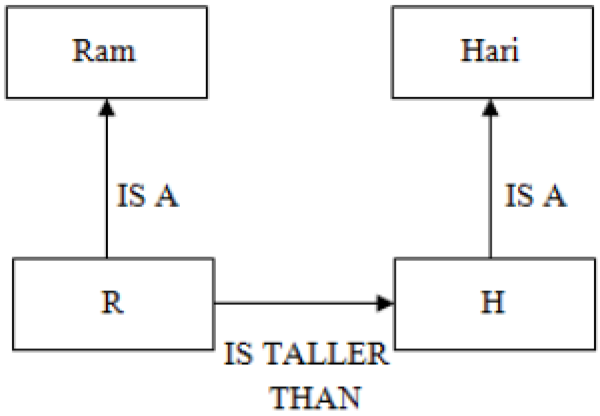
Problems: quantifiers, representing disjunction and negation

1. Suppose we have to represent the sentence "Sima is a girl".

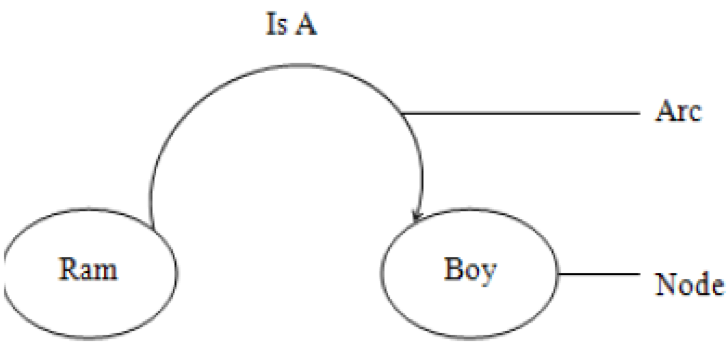


Figure

2. Ram is taller than Hari



For example: Ram is a boy.



Semantic Nets

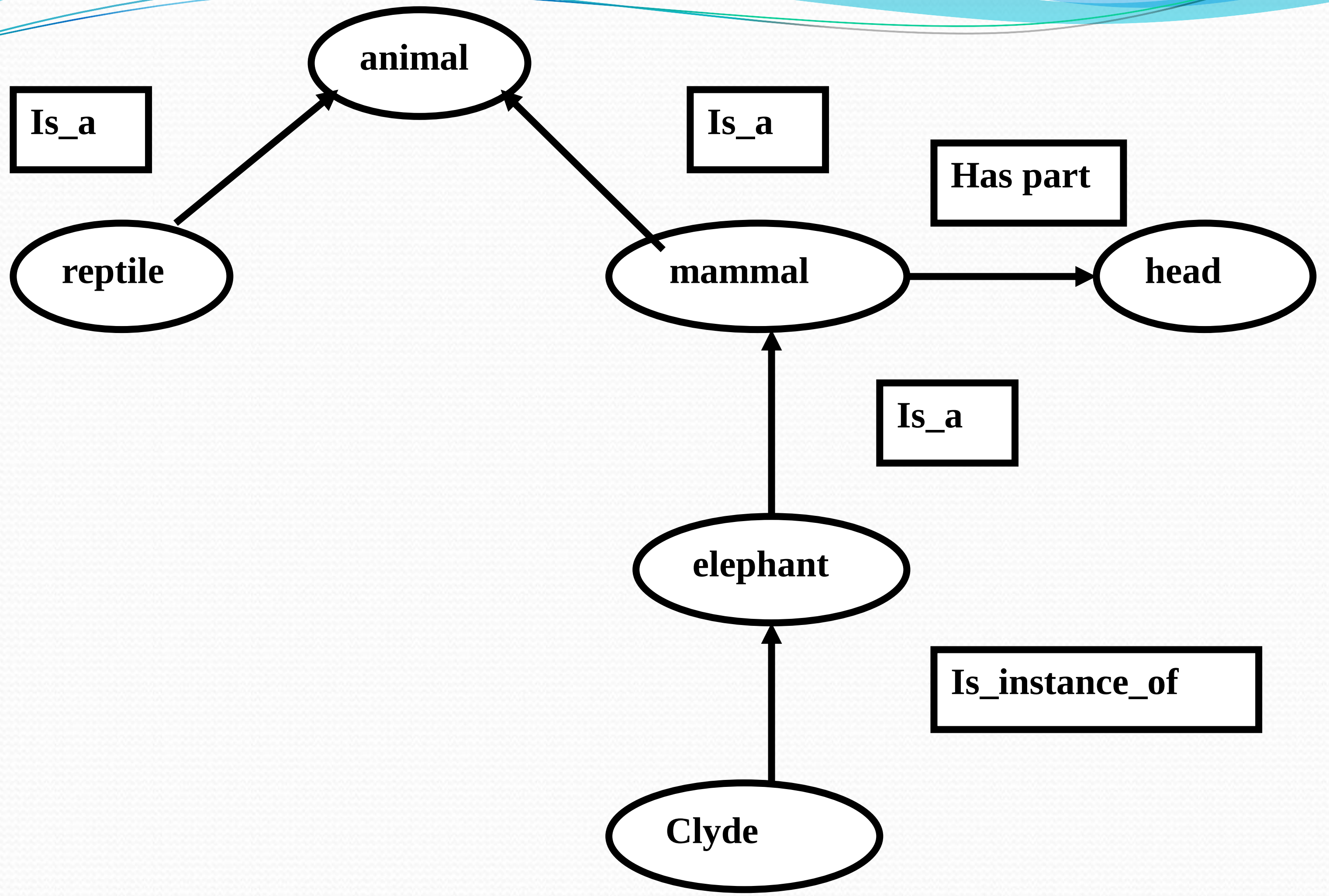
A semantic net is represented as a **graph**, where the nodes in the graph represent concepts,

and the arcs represent binary relationships between concepts.

Nodes represent objects, attributes and values

Links represent attributes and relationships between nodes

Labels attached to links: the name of the corresponding attribute or relation


animal
Is_a
Is_a
Has part
reptile
mammal
head
Is_a
elephant
Is_instance_of
Clyde

Production rules

Production based systems: (rule based system)

a set of if-then rules - typically state that if certain conditions hold, then some action should be taken.

If -then relation:

If high_temperature then prescribe aspirin

Production systems use a working memory - represents the facts (as semantic nets or frames) that are currently believed to hold.

- If :condtn-1 & condition -2 condition -3

  Then :take action -4

  IF:the temperature is greater than 200 degrees and water level is low

  THEN: open the safety valve.
- Each rule represents a small knowledge relating to the given domain to expertise.
- A no. of relative rules collectively may correspond to a chain of inferences which leads from some initially known facts to some useful conclusions.

- Inference is accomplished by a process of chaining through rules in forward or backward direction,until conclusion or failure is reached.
- Selection of rules is determined by matching current facts against the domain knowledge or variables in rules and choosing among a candidate set of rules.
- Inference process is carried out in an iterative mode with the user providing parameters needed to complete the rule chaining process.

Sl. No.	Scheme	Advantages	Disadvantages
1	Production rules	• Simple syntax • Easy to understand • Simple interpreter • Highly Modular • Easy to add or modify	• Hard to follow Hierarchies • Inefficient for large systems • Poor at representing structured descriptive knowledge.

Frames(frame based system)

- Proposed in 1968 by Marvin Minsky
- All the information relevant to a particular concept is stored in a single complex entity, called a frame.
- Frames support inheritance.
- Frames can be viewed as a structural representation of semantic nets.

Example

```
Mammal
   subclass:      Animal
   warm_blooded:  yes

Elephant
   subclass:      Mammal
   * colour:      grey
   * size:        large

Clyde
   instance:      Elephant
   color:         pink
   owner:         Fred
```

Syntax of a frame

(<frame name>
(<slot1> (<facet1> <value 1>……….<value n_1>)
(<facet2> <value1>…………<value n_2>)
.
.
.
.
.
(<facet n> <value1>…………. <value n_n>))
(<slot 2> (<facet1> <value 1>……<value n_1>)
(<facet2><value2>…………<value n_2>)
.
.
))

1) Create a frame of the person Ram who is a doctor. He is of 40. His wife name is Sita. They have two children Babu and Gita. They live in 100 kps street in the city of Delhi in India. The zip code is 756005.

```
(Ram
        (PROFESSION (VALUE Doctor))
        (AGE (VALUE 40))
        (WIFE (VALUE Sita))
        (CHILDREN (VALUE Bubu, Gita))
        (ADDRESS
                (STREET (VALUE 100 kps))
                (CITY(VALUE Delhi))
                (COUNTRY(VALUE India))
                (ZIP (VALUE 756005))))
```

Components of a Frame Entity

Name - correspond to a node in a semantic net

Attributes* or *slots filled with particular values

E.G. in the frame for Clyde, *instance* is the name of a slot, and *elephant* is the value of the slot.

- Names of slots correspond to the links in semantic nets
- Values of slots correspond to nodes.

Hence **each slot can be another frame**.

Example

```
Size:
  instance:        Slot
  single_valued:   yes
  range:           Size-set

Owner:
  instance:        Slot
  single_valued:   no
  range:           Person

Fred:
  instance:        Person
  occupation:      Elephant-breeder
```

Inheritance

If a slot is not defined for a given frame, we look at the parent-class slot with the same name

Simple if single parent-class

several parent classes :

multiple inheritance problem

(e.g., Clyde is both an elephant and a circus-animal)

Choose which parent to inherit from first.

Applications

- Classifying new instances of familiar entities (objects/events/places/tasks)
- Anticipating the attributes of such instances
- Inferring the presence and properties of their parts or participants

Sl. No.	Scheme	Advantages	Disadvantages
	Frame	• Expressive Power • Easy to set up slots for new properties and relations • Easy to create specialized procedures	• Difficult to program • Difficult for inference • Lack of inexpensive software

Inferential Knowledge

The knowledge, which can use inference mechanism to use this knowledge is called inferential knowledge. The inheritance property is a very powerful form of inferential knowledge. The inference procedures implement the standard logic rules of inference. There are two types of inference procedures like forward inference and backward inference. Forward inference moves from start state to goal state whereas backward inference moves from goal state to start state. In this type of knowledge several symbols are generally used like ∀ (universal quantifier), ∃(existential quantifier), → (arrow indicator) etc.

For example: All cats have tails

$$\forall X: \text{cat}(x) \rightarrow \text{has tail}(x)$$

Advantages:

1) A set of strict rules are defined which can be used to derive more facts.
2) Truths of new statements can be verified.
3) It gives guarantee about the correctness.
4) Many inference procedures available to implement standard rules of logic.

Inference

- Inference engine performs 2 major tasks:

1) examines existing facts and rules and adds new facts when possible

2) decides the order in which inferences are made.

- Infer means " to derive as a conclusion from facts or premises".
- There are 2 common rules for deriving new facts from rules and known facts. These are
    1. Modus Ponens
    2. Modus Tollens

MODUS PONENS

- most common inference strategy
- simple ,reasoning based on it is easily understood.
- The rule states that when A is known to be true and if a rule states " If A then B "

  it is valid to conclude that B is true

MODUS TOLLENS

- When B is false rule If A, then B then A is false.

E.g:

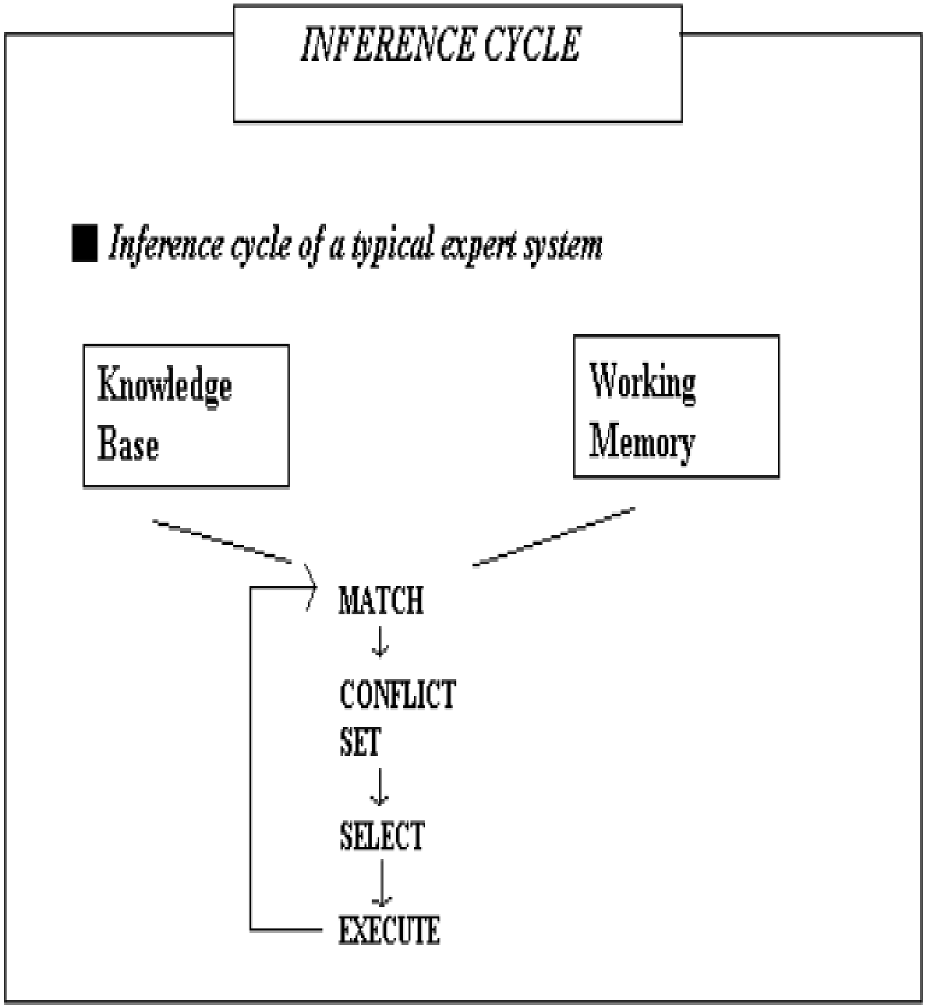
Rule : IF Ahmet's CAR IS DIRTY

THEN Ahmet HAS BEEN DRIVING OUTSIDE ANKARA

Given fact : Ahmet has not been outside Ankara.

New rule : Ahmet car is not dirty.

- This conclusion seems quite obvious but cannot be reached by most expert systems.
- Because they use modus ponens for deriving new facts from rules.



There are two problems addressed by the inference engine:

1) It must have a way to decide where to start. Rules and facts reside in a static knowledge base. There must be a way for the reasoning process to begin.
2) The inference engine must resolve conflicts that occur when alternative links of reasoning emerge. The system may reach a point where there are more than a few rules ready to fire. The inference engine must choose which rule to examine next.

TYPES OF INFERENCE

- There are two inferencing methods. These are Forward and Backward Chaining

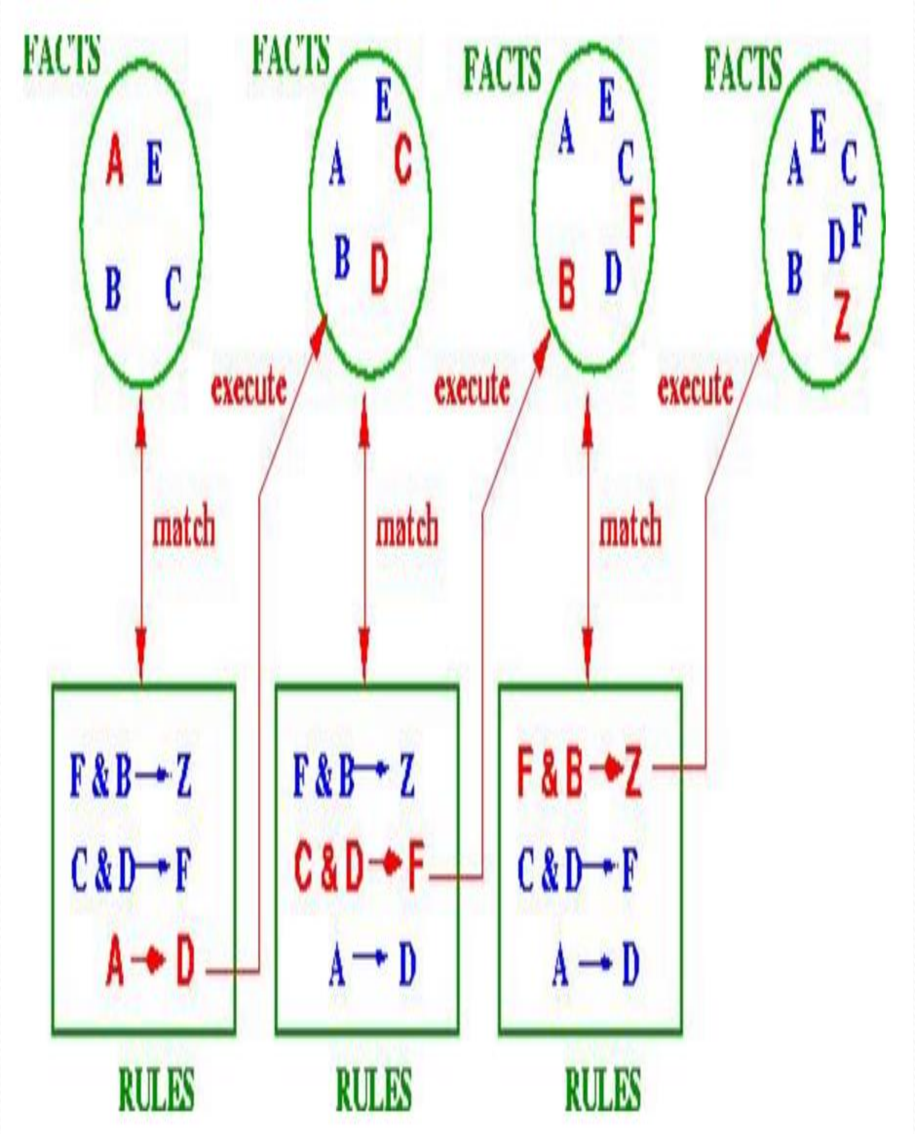
Forward chaining is one of the two main methods of reasoning when using an inference engine and can be described logically as repeated application of *modus ponens*. Forward chaining is a popular implementation strategy for expert systems, business and production rule systems. The opposite of forward chaining is backward chaining.

Forward chaining starts with the available data and uses inference rules to extract more data (from an end user, for example) until a goal is reached. An inference engine using forward chaining searches the inference rules until it finds one where the antecedent (**If** clause) is known to be true. When such a rule is found, the engine can conclude, or infer, the consequent (**Then** clause), resulting in the addition of new information to its data.[1]

FORWARD CHAINING

Problem: Does situation Z exists or not ?

- The first rule that fires is A->D because A is already in the database. Next we infer D.
- Existence of C and D causes second rule to fire and as a consequence F is inferred and placed in the database.
- This in turn, causes the third rule F&B->Zto fire, placing Z in the database.
- This technique is called forward chaining.



A very simple Forward chaining Algorithm

- Given m facts F1,F2,....,Fm? N RULES

  R1,R2,......Rn

- repeat

  for i ?- 1to n do

if one or more current facts match the antecedent of Ri then

1) add the new fact(s) define by the consequent

2) flag the rule that has been fired

3) increase m until no new facts have been produced.

EXAMPLE

Rule 1

IF the car overheats , THEN the car will stall.

Rule 2

IF the car stalls

THEN it will cost me money

AND I will be late getting home

Now, the question is

How do you arrive at conclusion that this situation will cost money and cause you to be late ?

The condition that triggers the chain of events is the car overheating

BACKWARD CHAINING

Backward Chaining (Example 1)

Rule 1

IF the car is not tuned AND the battery is weak

THEN not enough current will reach the starter.

Rule 2

IF not enough current reaches the starter

THEN the car will not start.

Given facts:

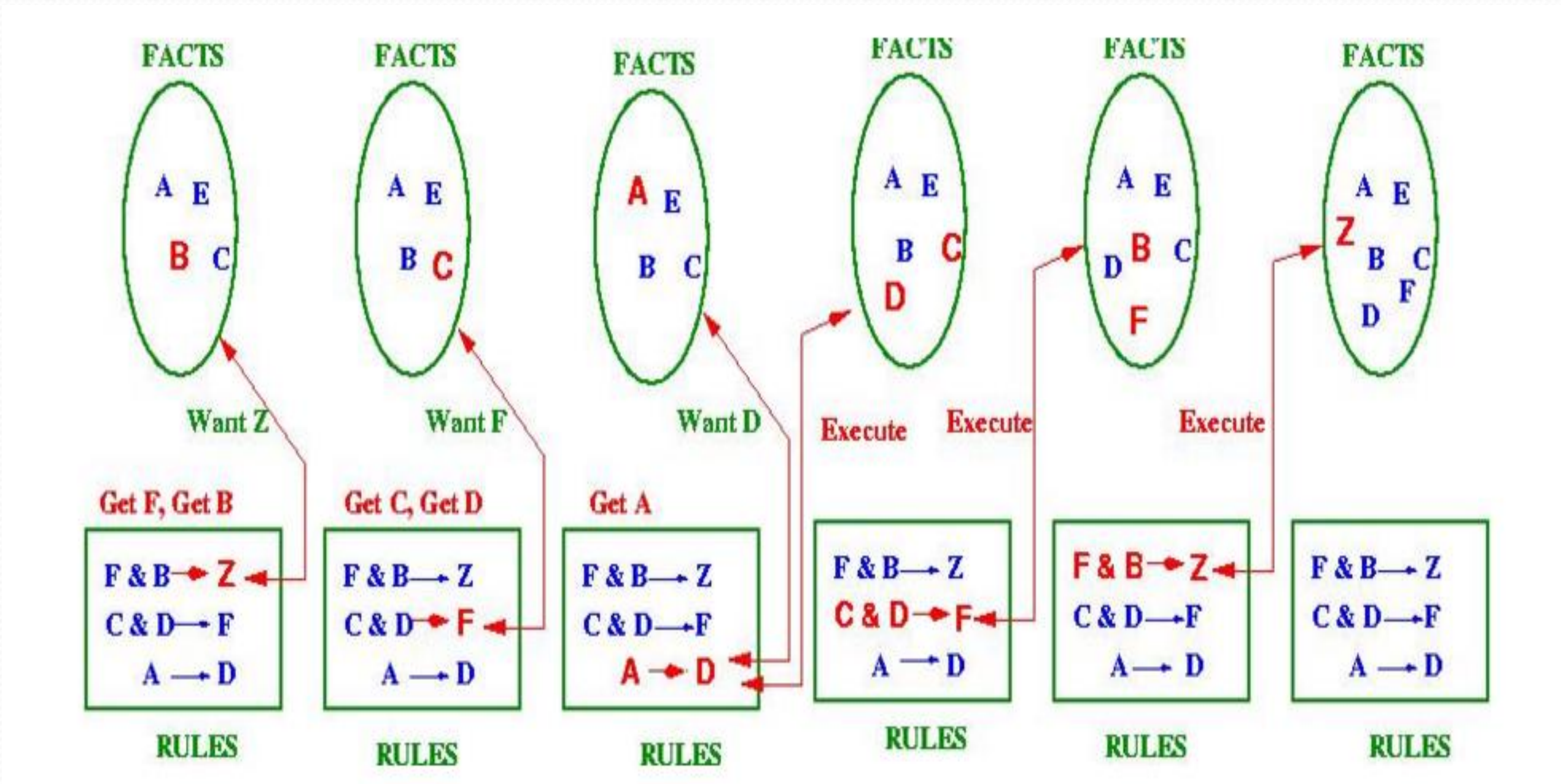
The car is not tuned

The battery is weak.

Now, the question is

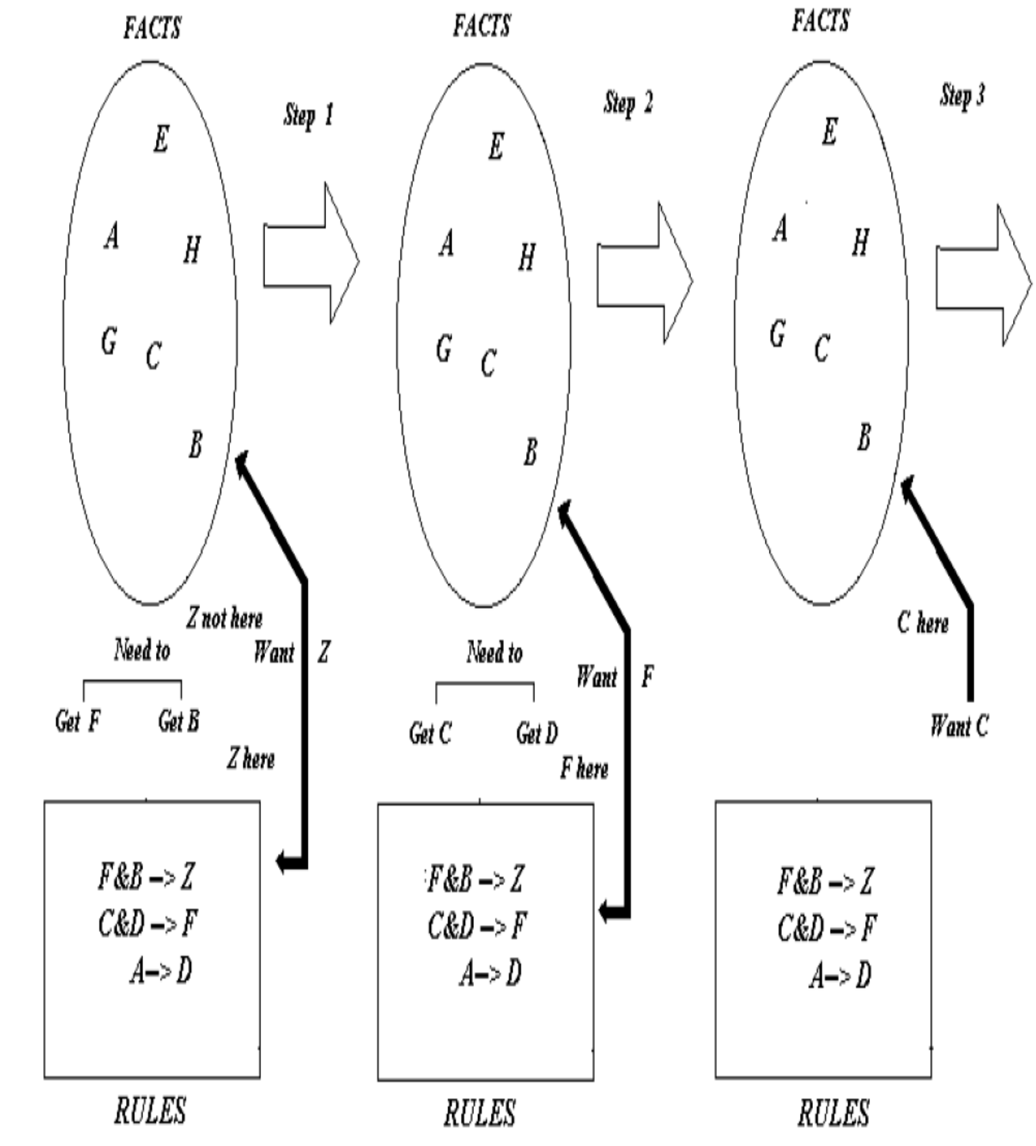
How would you arrive at the conditions that have resulted inthe car failing to start?

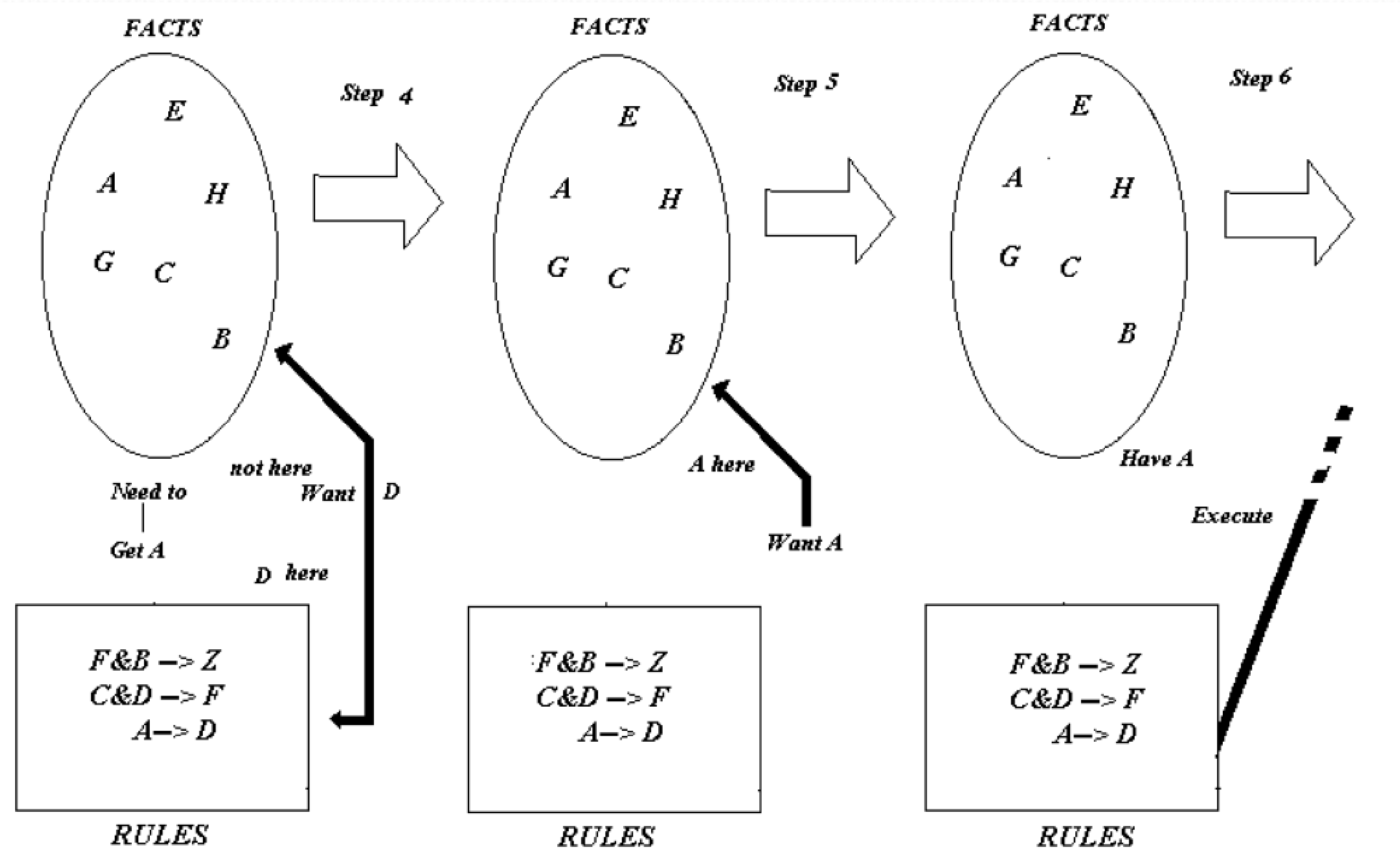
- In such a situation backward chaining might be more cost effective.
- With this inference method the system starts with what it wants to prove, e.g., that situation Z exists, and only executes rules that are relevant to establishing it.
- Figure following shows how backward chaining would work using the rules from the forward chaining example.



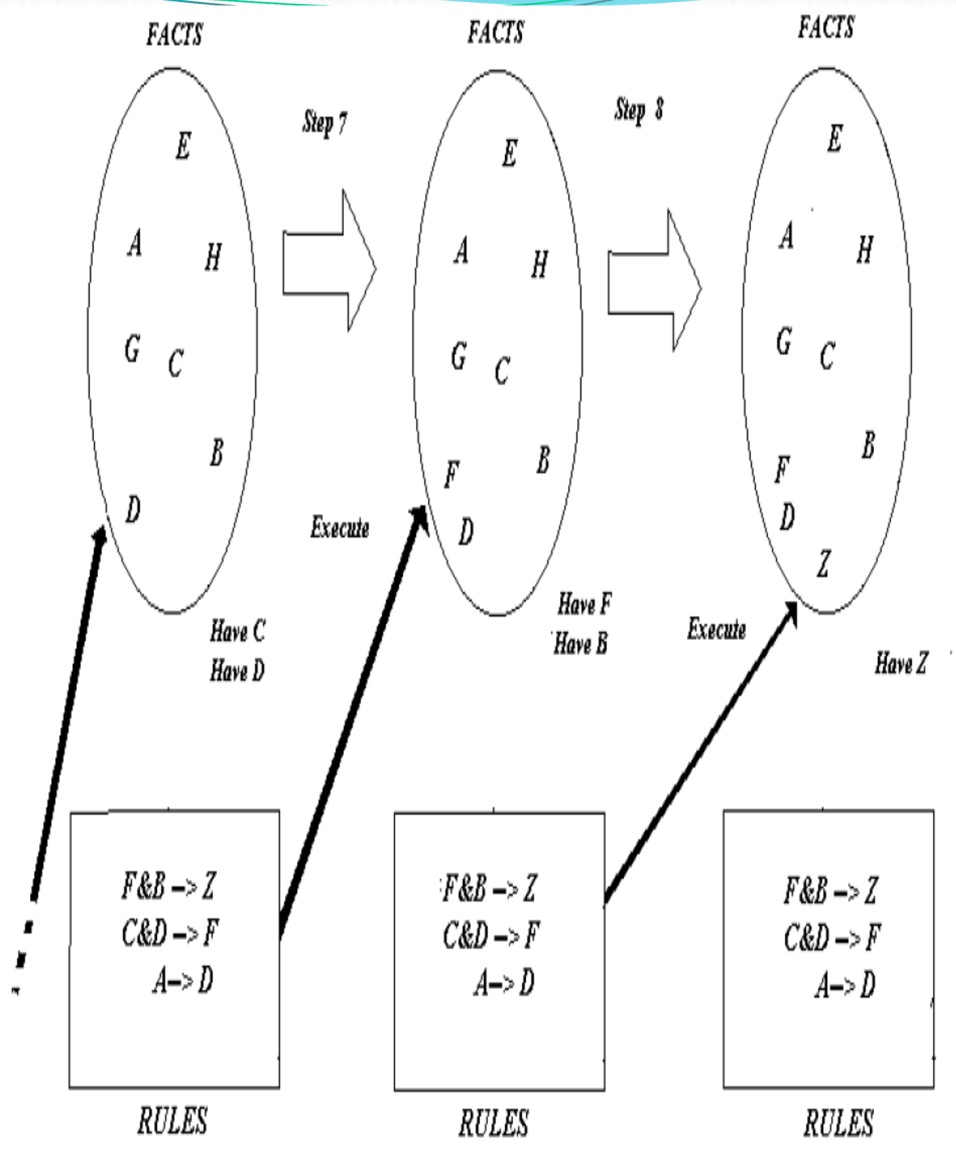
FACTS
A E B C
FACTS
A E B C
FACTS
A E B C
FACTS
A E B C D
FACTS
A E D B C F
FACTS
A E Z B C F D
Want Z
Want F
Want D
Execute
Execute
Execute
Get F, Get B
Get C, Get D
Get A
F & B → Z
C & D → F
A → D
RULES
RULES
RULES
RULES
RULES
RULES

- step 1 the system is told to establish (if it can) that situation Z exists, It first checks the data base for Z, and when that fails, searches for rules that conclude Z, i.e., have Z on the right side of the arrow. It finds the rule F&B->Z, and decides that it must establish F and B in order to conclude Z.
- step 2 the system tries to establish F, first checking the data base and then finding a rule that concludes F. From this rule, C&D->F, the system decides it must establish C and D to conclude F.
- steps 3 through 5 the system finds C in the data base but decides it must establish A before it can conclude D. It then finds A in the data base.




FACTS
Step 4
E
A
H
G
C
B
not here
Need to
Want
D
Get A
D here
F&B --> Z
C&D --> F
A--> D
RULES
FACTS
Step 5
E
A
H
G
C
B
A here
Want A
F&B --> Z
C&D --> F
A--> D
RULES
FACTS
Step 6
E
A
H
G
C
B
Have A
Execute
F&B --> Z
C&D --> F
A--> D
RULES

- steps 6 through 8 the system executes the third rule to establish D, then executes the second rule to establish the original goal, Z.
- The inference chain created here is identical to the one created by forward chaining.
- The difference in two approaches hinges on the method in which data and rules are searched.



What is Reasoning ?

- Reasoning is the act of deriving a conclusion from certain premises using a given methodology.
- Reasoning is a process of thinking; reasoning is logically arguing; reasoning is drawing inference.

Example :

If we know : *Robins are birds.*

All birds have wings

Then if we ask: *Do robins have wings?*

To answer this question - some reasoning must go.

Uncertainty in Reasoning

- The world is an uncertain place; often the Knowledge is imperfect which causes uncertainty. Therefore reasoning must be able to operate under uncertainty.

- AI systems must have ability to reason under conditions of uncertainty.

Uncertainties		Desired action
‡ Incompleteness Knowledge	:	Compensate for lack of knowledge
‡ Inconsistencies Knowledge	:	Resolve ambiguities and contradictions
‡ Changing Knowledge	:	Update the knowledge base over time

Approaches to Reasoning

There are three different approaches to reasoning under uncertainties.

- Symbolic reasoning
- Statistical reasoning
- Fuzzy logic reasoning

Fuzzy reasoning

- Human reasoning capabilities are divided into three areas:
  - ‡ **Mathematical Reasoning** – axioms, definitions, theorems, proofs
  - ‡ **Logical Reasoning** – deductive, inductive, abductive
  - ‡ **Non-Logical Reasoning** – linguistic , language

The **IQ** (Intelligence quotient) is the summation of mathematical reasoning skill and the logical reasoning.

The **EQ** (Emotional Quotient) depends mostly on non-logical reasoning capabilities.

Note : The Logical Reasoning is of our concern in AI

Logical Reasoning

Logic is a language for reasoning. It is a collection of rules called Logic arguments, we use when doing logical reasoning.

Logic reasoning is the process of drawing conclusions from premises using rules of inference.

Fuzzy Logic Systems(*FLS*)

- Produce definite output in response to incomplete, ambiguous, distorted, or inaccurate *fuzzy input*.

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.
- 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

- The fuzzy logic works on the levels of possibilities of input to achieve the definite output.

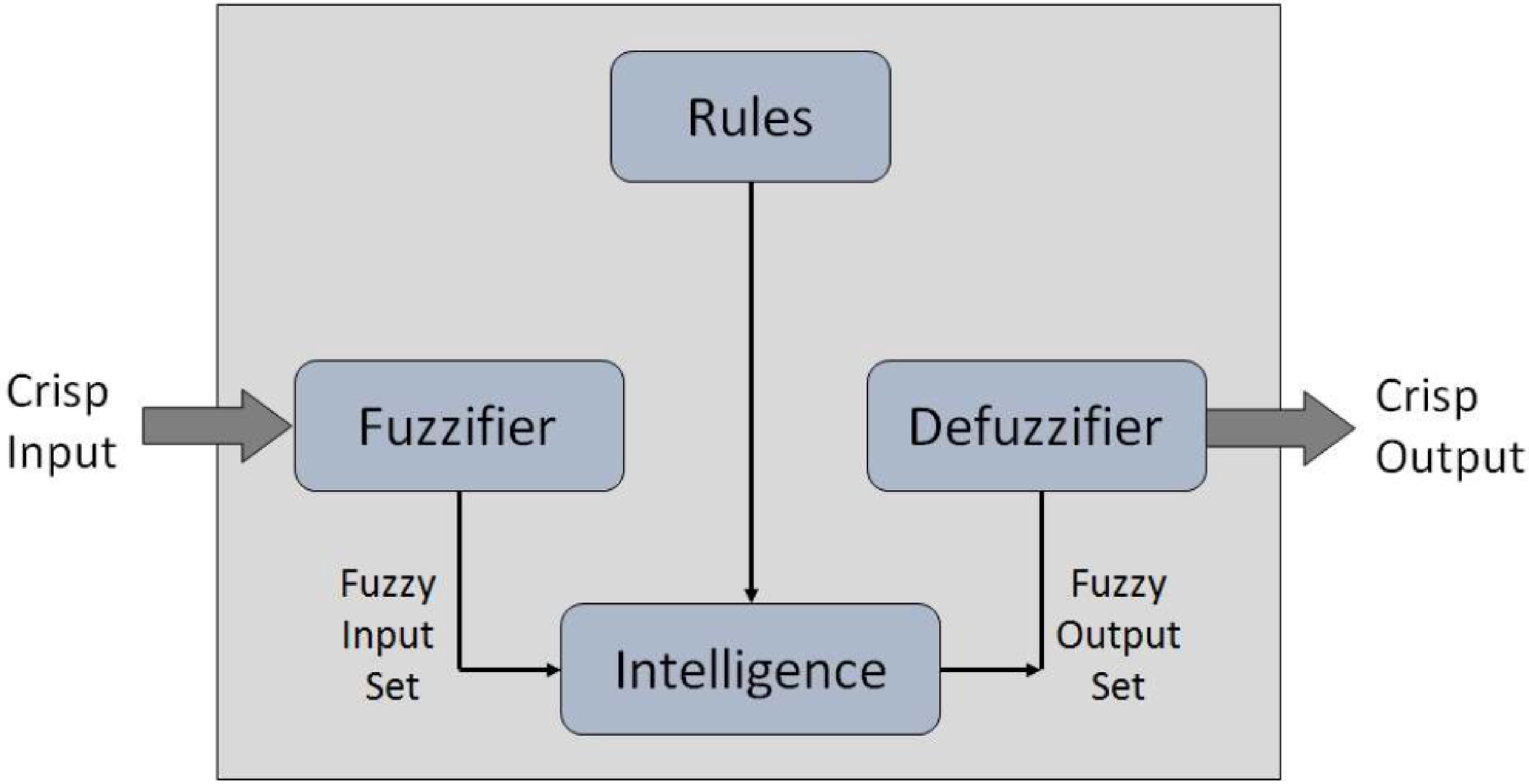
Implementation

- 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



- It has four main parts as shown –
  1. Fuzzification Module
  2. Knowledge Base
  3. Inference Engine
  4. Defuzzification Module

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

Knowledge Base – It stores IF-THEN rules provided by experts.

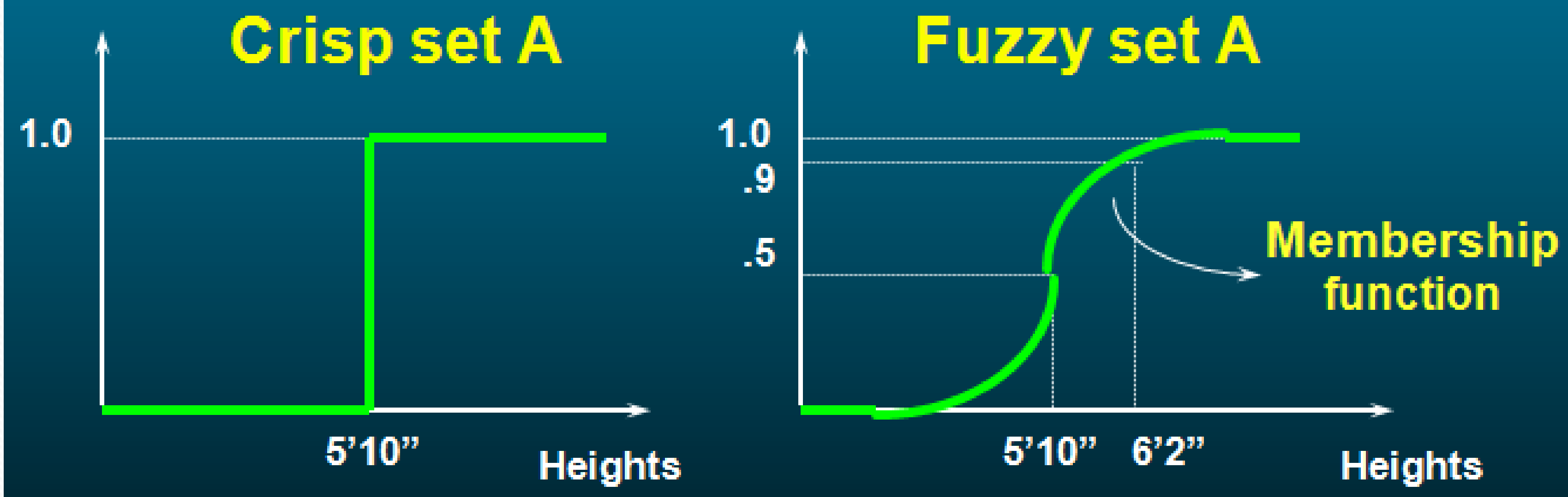
Inference Engine – It simulates the human reasoning process by making fuzzy inference on the inputs and IF-THEN rules.

Defuzzification Module – It transforms the fuzzy set obtained by the inference engine into a crisp value.

Fuzzy Sets

Sets with fuzzy boundaries

A = Set of tall people



Membership Function

- Membership functions allow 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***

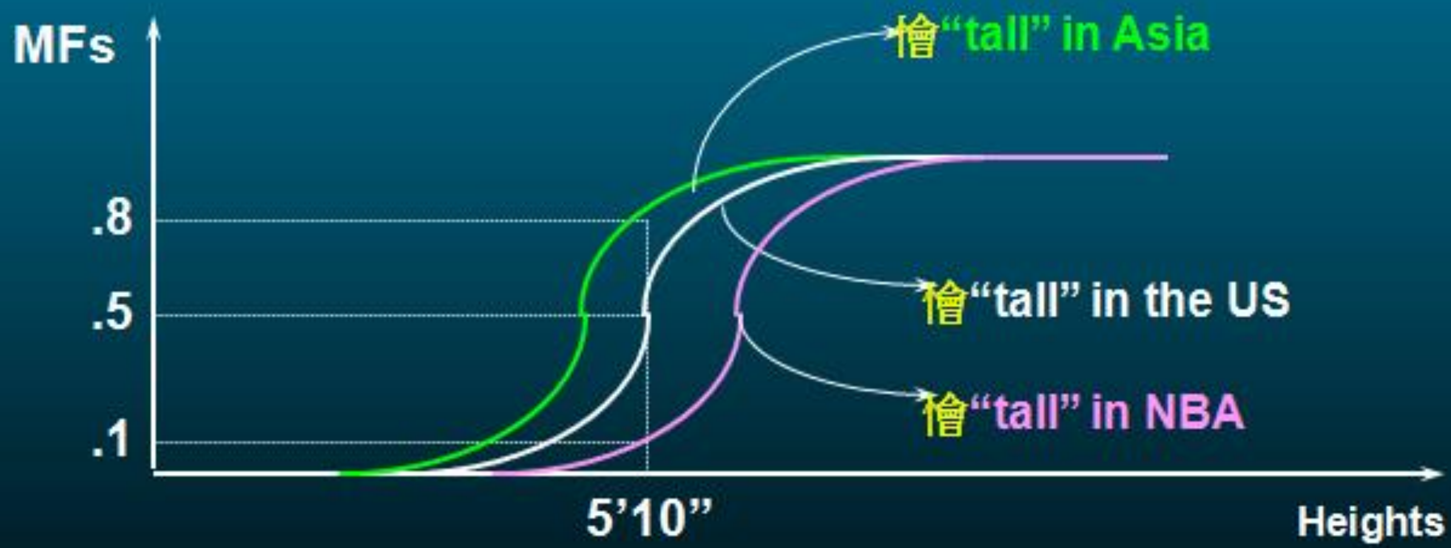
$$\mu A : X \rightarrow [0,1].$$

  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*.
- x axis represents the universe of discourse.
- y axis represents the degrees of membership in the [0, 1] interval.
- There can be multiple membership functions applicable to fuzzify a numerical value.

Membership Functions (MFs)

Characteristics of MFs:

- Subjective measures
- Not probability functions



Fuzzy Sets

Formal definition:

A fuzzy set *A* in *X* is expressed as a set of ordered pairs:

$$A = \{(x, \underbrace{\mu_A(x)}_{\text{Membership function (MF)}}) \mid x \in X\}$$

Fuzzy set

Membership function (MF)

Universe or universe of discourse

A fuzzy set is totally characterized by a membership function (MF).

Fuzzy Sets with Discrete Universes

Fuzzy set C = "desirable city to live in"

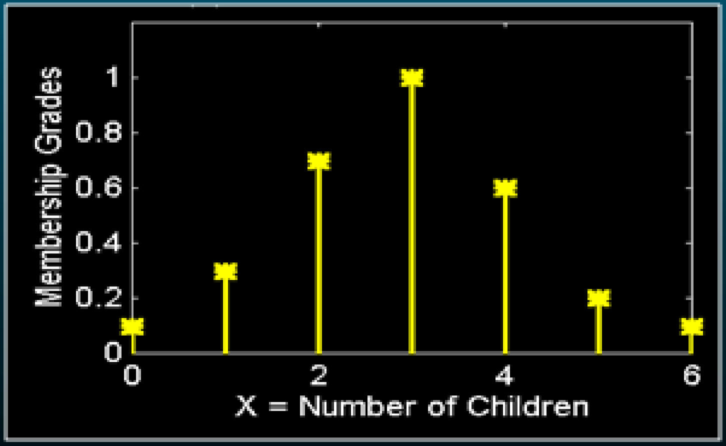
X = {SF, Boston, LA} (discrete and nonordered)

C = {(SF, 0.9), (Boston, 0.8), (LA, 0.6)}

Fuzzy set A = "sensible number of children"

X = {0, 1, 2, 3, 4, 5, 6} (discrete universe)

A = {(0, .1), (1, .3), (2, .7), (3, 1), (4, .6), (5, .2), (6, .1)}

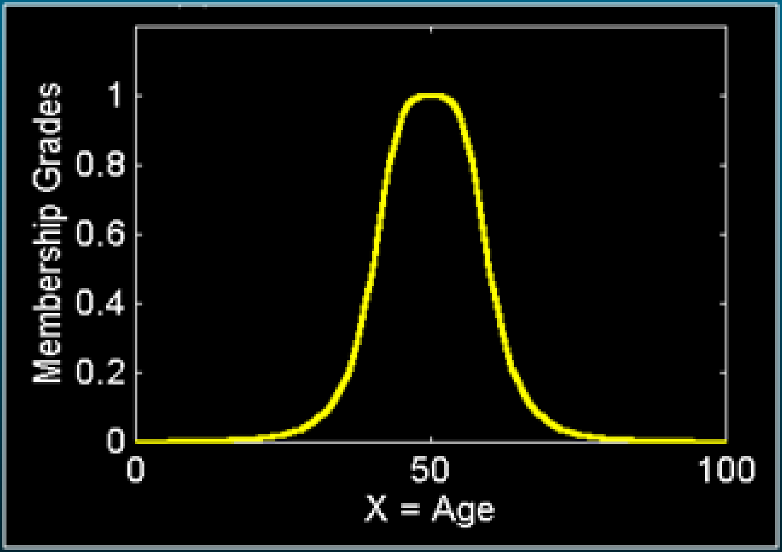


Fuzzy Sets with Cont. Universes

Fuzzy set B = "about 50 years old"

- X = Set of positive real numbers (continuous)
- B = {(x, μ_B(x)) | x in X}

$$\mu_B(x) = \frac{1}{1 + \left(\frac{x-50}{10}\right)^2}$$



Alternative Notation

A fuzzy set A can be alternatively denoted as follows:

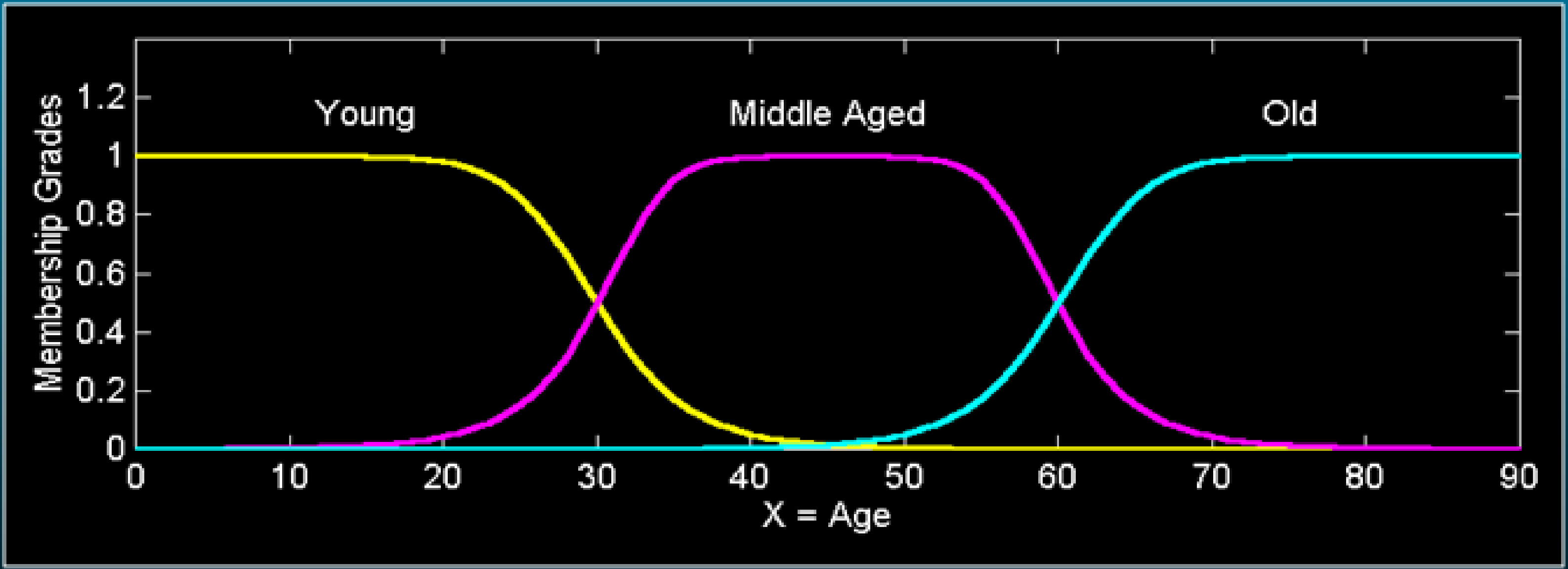
X is discrete → $A = \sum_{x_i \in X} \mu_A(x_i) / x_i$

X is continuous → $A = \int_X \mu_A(x) / x$

Note that Σ and integral signs stand for the union of membership grades; "/" stands for a marker and does not imply division.

Fuzzy Partition

Fuzzy partitions formed by the linguistic values "young", "middle aged", and "old":



lingmf.m

Set-Theoretic Operations

Subset:

$$A \subseteq B \Leftrightarrow \mu_A \le \mu_B$$

Complement:

$$\overline{A} = X - A \Leftrightarrow \mu_{\overline{A}}(x) = 1 - \mu_A(x)$$

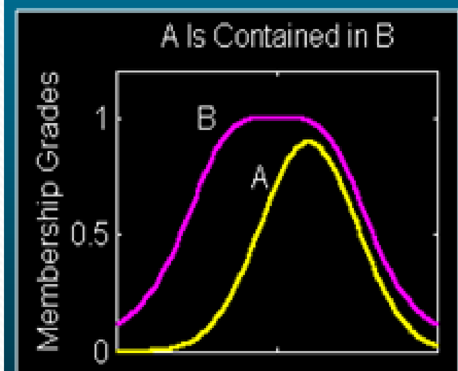
Union:

$$C = A \cup B \Leftrightarrow \mu_c(x) = \max(\mu_A(x), \mu_B(x)) = \mu_A(x) \vee \mu_B(x)$$

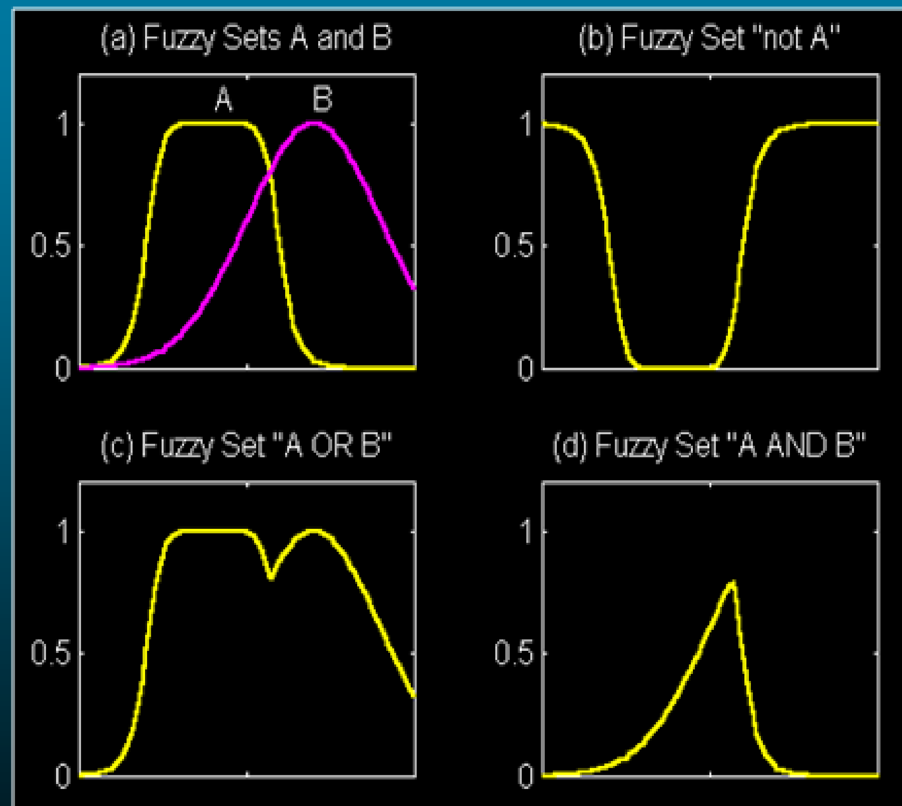
Intersection:

$$C = A \cap B \Leftrightarrow \mu_c(x) = \min(\mu_A(x), \mu_B(x)) = \mu_A(x) \wedge \mu_B(x)$$

Set-Theoretic Operations



subset.m



fuzsetop.m

MF Formulation

Triangular MF:

$$trimf(x;a,b,c) = \max\left(\min\left(\frac{x-a}{b-a},\frac{c-x}{c-b}\right),0\right)$$

Trapezoidal MF:

$$trapmf(x;a,b,c,d) = \max\left(\min\left(\frac{x-a}{b-a},1,\frac{d-x}{d-c}\right),0\right)$$

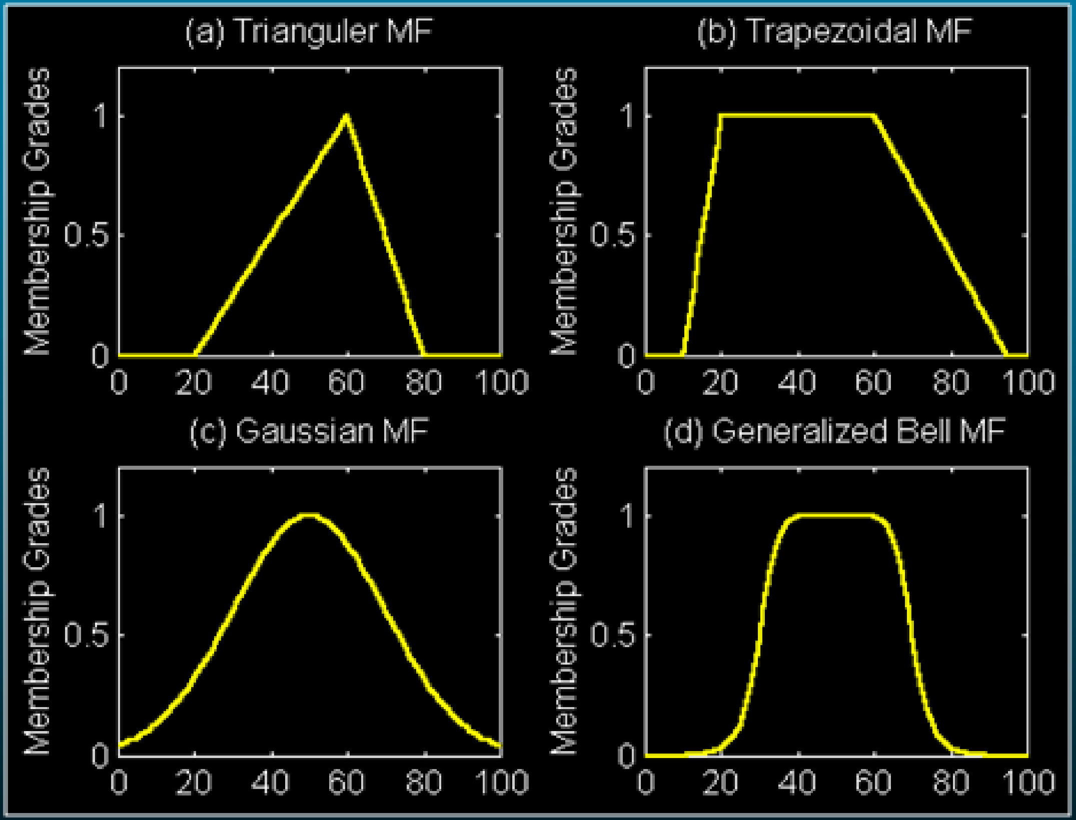
Gaussian MF:

$$gaussmf(x;a,b,c) = e^{-\frac{1}{2}\left(\frac{x-c}{\sigma}\right)^2}$$

Generalized bell MF:

$$gbellmf(x;a,b,c) = \frac{1}{1+\left|\frac{x-c}{b}\right|^{2b}}$$

MF Formulation



disp_mf.m

Application Areas of Fuzzy Logic

Automotive Systems

- Automatic Gearboxes
- Four-Wheel Steering
- Vehicle environment control

Consumer Electronic Goods

- Hi-Fi Systems
- Photocopiers
- Still and Video Cameras
- Television

Domestic Goods

- Microwave Ovens
- Refrigerators
- Toasters
- Vacuum Cleaners
- Washing Machines

Environment Control

- Air Conditioners/Dryers/Heaters
- Humidifiers

Advantages of FLSs

- Mathematical concepts within fuzzy reasoning are very simple.
- 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.

Probability

- **Probabilities :**

  Usually, are descriptions of the likelihood of some event occurring (ranging from **0** to **1**).

- **Event :**

  One or more outcomes of a probability experiment .

- **Probability Experiment :**

  Process which leads to well-defined results call outcomes.

- **Sample Space :**

  Set of all possible outcomes of a probability experiment.

- **Independent Events :**

  Two events, E_1 and E_2, are independent if the fact that E_1 occurs does not affect the probability of E_2 occurring.

- **Mutually Exclusive Events :**

  Events E_1, E_2, ..., E_n are said to be mutually exclusive if the occurrence of any one of them automatically implies the non-occurrence of the remaining $n - 1$ events.

- **Classical Probability :**

  Also called a priori theory of probability.

  The probability of event **A** = no of possible outcomes **f** divided by the total no of possible outcomes **n** ; ie., **P(A) = f / n.**

  Assumption: All possible outcomes are equal likely.

- **Conditional Probability :**

  The probability of some event **A**, given the occurrence of some other event **B**. Conditional probability is written **P(A|B)**, and read as "the probability of **A**, given **B** ".

- **Joint probability :**

  The probability of two events in conjunction. It is the probability of both events together. The joint probability of **A** and **B** is written **P(A ∩ B)** ; also written as **P(A, B)**.

Marginal Probability :

The probability of one event, regardless of the other event. The marginal probability of **A** is written **P(A)**, and the marginal probability of **B** is written **P(B)**.

- **Probability :** The Probabilities are numeric values between **0** and **1** (both inclusive) that represent ideal uncertainties (not beliefs).

  - **Probability of event A is P(A)**

$$P(A) = \frac{\text{instances of the event A}}{\text{total instances}}$$

$P(A) = 0$	indicates total uncertainty in **A**,
$P(A) = 1$	indicates total certainty and
$0 < P(A) < 1$	values in between tells degree of uncertainty

Probability Rules :

- ‡ All probabilities are between **0** and **1** inclusive $0 <= P(E) <= 1$.
- ‡ The sum of all the probabilities in the sample space is **1**.
- ‡ The probability of an event which must occur is **1**.
- ‡ The probability of the sample space is **1**.
- ‡ The probability of any event which is not in the sample space is zero.
- ‡ The probability of an event not occurring is $P(E') = 1 - P(E)$

Example 1 : A single 6-sided die is rolled.

What is the probability of each outcome?

What is the probability of rolling an even number?

What is the probability of rolling an odd number?

The possible outcomes of this experiment are 1, 2, 3, 4, 5, 6.

The Probabilities are :

P(1)	=	No of ways to roll 1 / total no of sides	=	1/6
P(2)	=	No of ways to roll 2 / total no of sides	=	1/6
P(3)	=	No of ways to roll 3 / total no of sides	=	1/6
P(4)	=	No of ways to roll 4 / total no of sides	=	1/6
P(5)	=	No of ways to roll 5 / total no of sides	=	1/6
P(6)	=	No of ways to roll 6 / total no of sides	=	1/6
P(even)	=	ways to roll even no / total no of sides	=	3/6 = 1/2
P(odd)	=	ways to roll odd no / total no of sides	=	3/6 = 1/2

Example 2 : Roll two dices

Each dice shows one of 6 possible numbers;

Total unique rolls is 6 x 6 = 36;

List of the joint possibilities for the two dices are:

(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

Roll two dices;

The rolls that add up to 4 are ((1,3), (2,2), (3,1)).

The probability of rolling dices such that total of 4 is 3/36 = 1/12

and the chance of it being true is (1/12) x 100 = 8.3%.

Conditional probability $P(A|B)$

A conditional probability is the probability of an event given that another event has occurred.

Example : Roll two dices.

What is the probability that the total of two dice will be greater than 8 given that the first die is a 6 ?

First List of the **joint possibilities** for the two dices are:

(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
(6, 1)	(6, 2)	**(6, 3)**	**(6, 4)**	**(6, 5)**	**(6, 6)**

There are 6 outcomes for which the first die is a 6, and of these, there are 4 outcomes that total more than 8 are (6,3; 6,4; 6,5; 6,6).

The probability of a total > 8 given that first die is 6 is therefore 4/6 = 2/3 .

This probability is written as: $\underbrace{P(\text{total}>8}_{\textbf{event}} \mid \underbrace{\text{1st die} = 6)}_{\textbf{condition}} = 2/3$

Read as "The probability that the total is > 8 given that die one is 6 is 2/3."

Summary of symbols & notations

$A \cup B$	(A union B)	'Either A or B occurs or both occur'
$A \cap B$	(A intersection B)	'Both A and B occur'
$A \subseteq B$	(A is a subset of B)	'If A occurs, so does B'
A'	$\bar{A}$	'Event A does not occur'
Φ	(the empty set)	An impossible event
S	(the sample space)	An event that is certain to occur
$A \cap B = \Phi$		Mutually exclusive Events
$P(A)$		Probability that event A occurs
$P(B)$		Probability that event B occurs
$P(A \cup B)$		Probability that event A or event B occurs
$P(A \cap B)$		Probability that event A and event B occur
$P(A \cap B) = P(A) . P(B)$		Independent events
$P(A \cap B) = 0$		Mutually exclusive Events
$P(A \cup B) = P(A) + P(B) - P(AB)$		Addition rule;
$P(A \cup B) = P(A) + P(B) - P(A) . P(B)$		Addition rule; independent events
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$		
$P(A \cup B) = P(A) + P(B) - P(B\|A).P(A)$		
$P(A \cup B) = P(A) + P(B)$		Addition rule; mutually exclusive Events
$A\|B$	(A given B)	"Event A will occur given that event B has occurred"

A|B (A given B)	"Event A will occur given that event B has occurred"
P(A|B)	Conditional probability that event A will occur given that event B has occurred already
P(B|A)	Conditional probability that event B will occur given that event A has occurred already
P(A ∩ B) = P(A|B).P(B) or P(A ∩ B) = P(B|A).P(A)	Multiplication rule
P(A ∩ B) = P(A) . P(B)	Multiplication rule; independent events; ie probability of joint events A and B
P(A|B) = P(A ∩ B) / P(B)	Rule to determine a conditional probability from unconditional probabilities.

Bayesian theory

Bayes' Theorem

- Bayesian view of probability is related to degree of belief.

- Bayes' theorem is also known as Bayes' rule or Bayes' law, or called Bayesian reasoning.

- The probability of an event **A** conditional on another event **B** ie **P(A|B)** is generally different from probability of **B** conditional on **A** ie **P(B|A)**.
  - There is a definite relationship between the two, P(A|B) and P(B|A), and Bayes' theorem is the statement of that relationship.
  - Bayes theorem is a way to calculate P(A|B) from a knowledge of P(B|A).
  - Bayes' Theorem is a result that allows new information to be used to update the conditional probability of an event.

Bayes' Theorem

Let S be a sample space.

Let A_1, A_2, ... , A_n be a set of mutually exclusive events from S.

Let B be any event from the same S, such that $P(B) > 0$.

Then Bayes' Theorem describes following two probabilities :

$$P(A_k|B) = \frac{P(A_k \cap B)}{P(A_1 \cap B) + P(A_2 \cap B) + - - - - + P(A_n \cap B)}$$

and

by invoking the fact $P(A_k \cap B) = P(A_k).P(B|A_k)$ the probability

$$P(A_k|B) = \frac{P(A_k).P(B|A_k)}{P(A_1).P(B|A_1) + P(A_2).P(B|A_2) + - - - - + P(A_n).P(B|A_n)}$$

Applying Bayes' Theorem :

Bayes' theorem is applied while following conditions exist.

- the sample space **S** is partitioned into a set of mutually exclusive events **{A1, A2, ... , An }.**

- within **S**, there exists an event **B**, for which **P(B) > 0.**

- the goal is to compute a conditional probability of the form : **P(Ak|B)**.

- you know at least one of the two sets of probabilities described below
  - $P(A_k \cap B)$ for each A_k
  - $P(A_k)$ and $P(B|A_k)$ for each A_k

Example 1: Applying Bayes' Theorem

Problem : Marie's marriage is tomorrow.

- in recent years, each year it has rained only 5 days.
- the weatherman has predicted rain for tomorrow.
- when it actually rains, the weatherman correctly forecasts rain 90% of the time.
- when it doesn't rain, the weatherman incorrectly forecasts rain 10% of the time.

The question : What is the probability that it will rain on the day of Marie's wedding?

Solution : The sample space is defined by two mutually exclusive events – "it rains" or "it does not rain". Additionally, a third event occurs when the "weatherman predicts rain".

The events and probabilities are stated below.

- Event A1 : rains on Marie's wedding.
- Event A2 : does not rain on Marie's wedding
- Event B : weatherman predicts rain.
- P(A1) = 5/365 =0.0136985 [Rains 5 days in a year.]
- P(A2) = 360/365 = 0.9863014 [Does not rain 360 days in a year.]

- $P(B|A1) = 0.9$ [When it rains, the weatherman predicts rain 90% time.]
- $P(B|A2) = 0.1$ [When it does not rain, weatherman predicts rain 10% time.]

We want to know **P(A1|B)**, the probability that it will rain on the day of Marie's wedding, given a forecast for rain by the weatherman.

The answer can be determined from Bayes' theorem, shown below.

$$P(A1|B) = \frac{P(A1).P(B|A1)}{P(A1).P(B|A1)+P(A2).P(B|A2)} = \frac{(0.014)(0.9)}{[(0.014)(0.9)+(0.986)(0.1)]}$$

$$= 0.111$$

So, despite the weatherman's prediction, there is a good chance that Marie will not get rain on at her wedding.

Thus Bayes theorem is used to calculate conditional probabilities.

Bayesian Networks

Why This Matters

- Bayesian networks have been the most important contribution to the field of AI in the last 10 years
- Provide a way to represent knowledge in an uncertain domain and a way to reason about this knowledge
- Many applications: medicine, factories, help desks, spam filtering, etc.

Bayesian Networks and Certainty Factors

A Bayesian network (or a belief network) is a probabilistic graphical model that represents a set of variables and their probabilistic independencies.

For example, a Bayesian network could represent the **probabilistic relationships** between diseases and symptoms. Given symptoms, the network can be used to compute the probabilities of the presence of various diseases.

Bayesian Networks are also called : Bayes nets, Bayesian Belief Networks (BBNs) or simply Belief Networks. Causal Probabilistic Networks (CPNs).

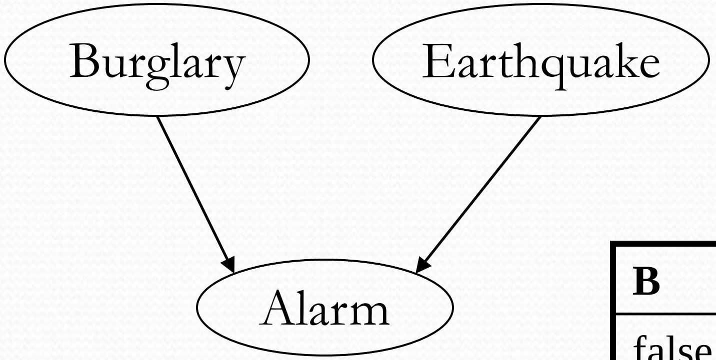
A Bayesian network consists of :

- a set of nodes and a set of directed edges between nodes.
- the edges reflect cause-effect relations within the domain.
- The effects are not completely deterministic (e.g. disease -> symptom).
- the strength of an effect is modeled as a probability.

A Bayesian Network

B	P(B)
false	0.999
true	0.001

E	P(E)
false	0.998
true	0.002

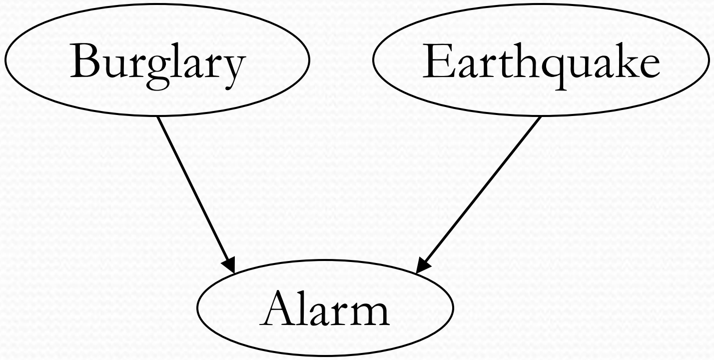


A Bayesian network is made up of two parts:

1. A directed acyclic graph
2. A set of parameters

B	E	A	P(A|B,E)
false	false	false	0.999
false	false	true	0.001
false	true	false	0.71
false	true	true	0.29
true	false	false	0.06
true	false	true	0.94
true	true	false	0.05
true	true	true	0.95

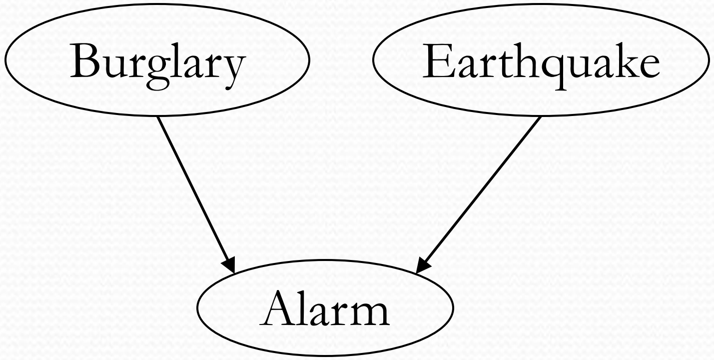
A Directed Acyclic Graph



1. A directed acyclic graph:

- The nodes are random variables (which can be discrete or continuous)
- Arrows connect pairs of nodes (X is a parent of Y if there is an arrow from node X to node Y).

A Directed Acyclic Graph



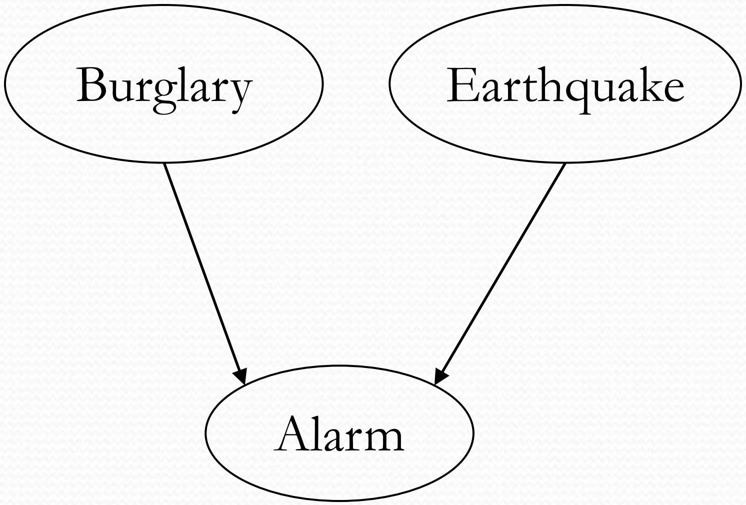
- Intuitively, an arrow from node X to node Y means X has a direct influence on Y (we can say X has a casual effect on Y)
- Easy for a domain expert to determine these relationships
- The absence/presence of arrows will be made more precise later on

A Set of Parameters

B	P(B)
false	0.999
true	0.001

E	P(E)
false	0.998
true	0.002

B	E	A	P(A|B,E)
false	false	false	0.999
false	false	true	0.001
false	true	false	0.71
false	true	true	0.29
true	false	false	0.06
true	false	true	0.94
true	true	false	0.05
true	true	true	0.95



A Set of Parameters

Conditional Probability Distribution for Alarm

B	E	A	P(A|B,E)
false	false	false	0.999
false	false	true	0.001
false	true	false	0.71
false	true	true	0.29
true	false	false	0.06
true	false	true	0.94
true	true	false	0.05
true	true	true	0.95

Stores the probability distribution for Alarm given the values of Burglary and Earthquake

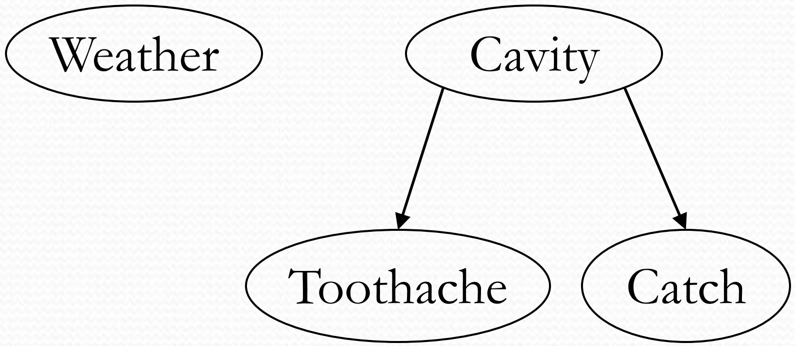
For a given combination of values of the parents (B and E in this example), the entries for P(A=true|B,E) and P(A=false|B,E) must add up to 1 eg. P(A=true|B=false,E=false) + P(A=false|B=false,E=false)=1

Semantics of Bayesian Networks

Two ways to view Bayes nets:

1. A representation of a joint probability distribution
2. An encoding of a collection of conditional independence statements

Bayesian Network Example



P(A|B,C) = P(A|C)
I(ToothAche,Catch|Cavity)

- Weather is independent of the other variables, I(Weather,Cavity) or P(Weather) = P(Weather|Cavity) = P(Weather|Catch) = P(Weather|Toothache)
- Toothache and Catch are conditionally independent given Cavity
- I(Toothache,Catch|Cavity) meaning P(Toothache|Catch,Cavity) = P(Toothache|Cavity)

Certainty Factors

- Another method of dealing with uncertainty uses certainty factors, originally developed for the MYCIN expert system.

Difficulties with Bayesian Method

- The Bayesian method is useful in medicine / geology because we are determining the probability of a specific event (disease / location of mineral deposit), given certain symptoms / analyses.
- The problem is with the difficulty / impossibility of determining the probabilities of these givens – symptoms / analyses.
- Evidence tends to accumulate over time.

Belief and Disbelief

- Consider the statement:

  "The probability that I have a disease plus the probability that I do not have the disease equals one."

- Now, consider an alternate form of the statement:

  "The probability that I have a disease is one minus the probability that I don't have it."

Belief and Disbelief

- It was found that physicians were reluctant to state their knowledge in the form:

  "The probability that I have a disease is one minus the probability that I don't have it."

- Symbolically, $P(H|E) = 1 - P(H'|E)$, where E represents evidence

Likelihood of Belief / Disbelief

- The reluctance by the physicians stems from the likelihood of belief / disbelief – not in the probabilities.
- The equation, $P(H|E) = 1 - P(H'|E)$, implies a cause-and-effect relationship between E and H.
- The equation implies a cause-and-effect relationship between E and H' if there is a cause-and-effect between E and H.

Measures of Belief and Disbelief

- The certainty factor, CF, is a way of combining belief and disbelief into a single number.

- This has two uses:
  1. The certainty factor can be used to rank hypotheses in order of importance.
  2. The certainty factor indicates the net belief in a hypothesis based on some evidence.

Certainty Factor Values

- Positive CF – evidence supports the hypothesis
- CF = 1 – evidence definitely proves the hypothesis
- CF = 0 – there is no evidence or the belief and disbelief completely cancel each other.
- Negative CF – evidence favors negation of the hypothesis – more reason to disbelieve the hypothesis than believe it

Difficulties with Certainty Factors

- In MYCIN, which was very successful in diagnosis, there were difficulties with theoretical foundations of certain factors.
- There was some basis for the CF values in probability theory and confirmation theory, but the CF values were partly ad hoc.
- Also, the CF values could be the opposite of conditional probabilities.

Dempster-Shafer Theory

- The Dempster-Shafer Theory is a method of inexact reasoning.
- It is based on the work of Dempster who attempted to model uncertainty by a range of probabilities rather than a single probabilistic number.
- The Dempster-Shafer theory assumes that there is a fixed set of mutually exclusive and exhaustive elements called environment and symbolized by the Greek letter s:

$$: = \{=_1, , _2, \cdots, , _N\}$$

- The environment is another term for the universe of discourse in set theory.
- Consider the following:

1 = {rowboat, sailboat, destroyer, aircraft carrier}

- These are all mutually exclusive elements
- Consider the question:

  "What are the military boats?"

- The answer would be a subset of T :

  { {$_3$, , $_4$} = {destroyer, aircraft carrier}

- Consider the question:

  "What boat is powered by oars?"

- The answer would also be a subset of T :

  {{$_1$} = {rowboat}

  This set is called a singleton because it contains only one element.

Dempster-Shafer

- Each of these subsets of E is a possible answer to the question, but there can be only one correct answer.
- Consider each subset an implied proposition:
  - The correct answer is: $\{T_1, , _2, , _3)$
  - The correct answer is: $\{T_1, , _3\}$
- All subsets of the environment can be drawn as a hierarchical lattice with h at the top and the null set s at the bottom

Dempster-Shafer

- An environment is called a frame of discernment when its elements may be interpreted as possible answers and only one answer is correct.
- If the answer is not in the frame, the frame must be enlarged to accommodate the additional knowledge of element..

Dempster-Shafer

1. Mass Functions and Ignorance

   In Bayesian theory, the posterior probability changes as evidence is acquired. In Dempster-Shafer theory, the belief in evidence may vary.

   We talk about the degree of belief in evidence as analogous to the mass of a physical object – evidence measures the amount of mass.

Dempster-Shafer

- Dempster-Shafer does not force belief to be assigned to ignorance – any belief not assigned to a subset is considered no belief (or non-belief) and just associated with the environment.
- Every set in the power set of the environment which has mass > 0 is a focal element.
- Every mass can be thought of as a function:

$$m: \mathscr{P}((\ ) \rightarrow [0, 1]$$

Dempster-Shafer

- Combining Evidence

  Dempster's rule combines mass to produce a new mass that represents the consensus of the original, possibly conflicting evidence

  The lower bound is called the support; the upper bound is called the plausibility; the belief measure is the total belief of a set and all its subsets.

Dempster-Shafer

1. The moving mass analogy is helpful to understanding the support and plausibility.
   - The support is the mass assigned to a set and all its subsets
   - Mass of a set can move freely into its subsets
   - Mass in a set cannot move into its supersets
   - Moving mass from a set into its subsets can only contribute to the plausibility of the subset, not its support.
   - Mass in the environment can move into any subset.

Dempster –Shafer Theory(DST)

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 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 **$P(A) + P(\neg A) = 1$**.

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 **$B(\neg A)$** as both being **0**.
- ‡ Certainty factors do not allow this.

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.
  - .. it is not considered a probability but a belief mass.
- The distribution of masses is called basic belief assignment.

KEY TERMS

Belief: In Dempster-Shafer theory, the level of representing the confidence that a proposition lies in a focal element or any subset of it.

Body of Evidence: In Dempster-Shafer theory, a series of focal elements and associated mass values.

Focal Element: In Dempster-Shafer theory, a set of hypotheses with positive mass value in a body of evidence.

Frame of Discernment: In Dempster-Shafer theory, the set of all hypotheses considered.

Dempster-Shafer Theory: General methodology, also known as the theory of belief functions, its rudiments are closely associated with uncertain reasoning.

Ignorance: In Dempster-Shafer theory, the level of mass value not discernible among the hypotheses.

Mass Value: In Dempster-Shafer theory, the level of exact belief in a focal element.

Non-Specificity: In Dempster-Shafer theory, the weighted average of the focal elements' mass values in a body of evidence, viewed as a species of a higher uncertainty type, encapsulated by the term ambiguity.

Plausibility: In Dempster-Shafer theory, the extent to which we fail to disbelieve a proposition lies in a focal element.

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 2^S.

- If each possible subset as a binary vector (describing any member is present or not by writing **1** or **0**), then 2^5 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 "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.

Note : Given a set S, the power set of S, written 2^S, is the set of all subsets of S, including the empty set and S.

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.

Plausibility is 1 minus the sum of the masses of all sets whose intersection with the hypothesis is empty. It is an upper bound on the possibility that the hypothesis could possibly happen, up to that value, because there is only so much evidence that contradicts that hypothesis.

Example :

A proposition say **"the cat in the box is dead."**

Suppose we have **belief of 0.5** and **plausibility of 0.8** for the proposition.

- Evidence to state strongly, that proposition is true with confidence **0.5**.
- Evidence contrary to hypothesis ("the cat is alive") has confidence **0.2**.
- Remaining mass of **0.3** (the gap between the **0.5** supporting evidence and the **0.2** contrary evidence) is "**indeterminate**," meaning that the cat could either be dead or alive. This interval represents the level of uncertainty based on the evidence in the system.

Hypothesis	Mass	belief	plausibility
Null (neither alive nor dead)	0	0	0
Alive	0.2	0.2	0.5
Dead	0.5	0.5	0.8
Either (alive or dead)	0.3	1.0	1.0

- Null hypothesis is set to **zero** by definition, corresponds to "no solution".
- Orthogonal hypotheses "Alive" and "Dead" have probabilities of **0.2** and **0.5**, respectively. This could correspond to "Live/Dead Cat Detector" signals, which have respective reliabilities of **0.2** and **0.5**.

- All-encompassing "Either" hypothesis (simply acknowledges there is a cat in the box) picks up the slack so that the sum of the masses is **1**.
- Belief for the "Alive" and "Dead" hypotheses matches their corresponding masses because they have no subsets;
- Belief for "Either" consists of the sum of all three masses (Either, Alive, and Dead) because "Alive" and "Dead" are each subsets of "Either".
- "Alive" plausibility is **1- m** (Death) and "Dead" plausibility is **1- m** (Alive).
- "Either" plausibility sums **m(Alive) + m(Dead) + m(Either)**.
- Universal hypothesis ("Either") will always have **100%** belief and plausibility; it acts as a checksum of sorts.

Dempster-Shafer Calculus

- The Dempster-Shafer (DS) Theory, requires a Universe of Discourse **U** (or Frame of Judgment) consisting of mutually exclusive alternatives, corresponding to an attribute value domain. For instance, in satellite image classification the set **U** may consist of all possible classes of interest.

- Each subset $S \subseteq U$ is assigned a basic probability **m(S)**, a belief **Bel(S)**, and a plausibility **Pls(S)** so that

  m(S), Bel(S), Pls(S) $\in [0, 1]$ and **Pls(S)** $\geq$ **Bel(S)** where

  - **m** represents the strength of an evidence, is the basic probability; e.g., a group of pixels belong to certain class, may be assigned value **m**.
  - **Bel(S)** summarizes all the reasons to believe **S**.
  - **Pls(S)** expresses how much one should believe in **S** if all currently unknown facts were to support **S**.

- The true belief in **S** is somewhere in the belief interval **[Bel(S), Pls(S)]**.

- The basic probability assignment **m** is defined as function $m : 2^U \rightarrow [0,1]$, where $m(\emptyset) = 0$ and sum of **m** over all subsets of **U** is **1** (i.e., $\sum_{s \subseteq U} m(s) = 1$).

- For a given basic probability assignment **m**, the belief **Bel** of a subset **A** of **U** is the sum of **m(B)** for all subsets **B** of **A**, and the plausibility **Pls** of a subset **A** of **U** is $Pls(A) = 1 - Bel(A')$ (5) where **A'** is complement of **A** in **U**.

- **<u>Advantages of DST</u>**
- The difficult problem of specifying priors can be avoided
- In addition to uncertainty, also ignorance can be expressed
- It is straightforward to express pieces of evidence with different levels of abstraction
- Dempster's combination rule can be used to combine pieces of evidence
- **<u>Disadvantages of DST</u>**
- Potential computational complexity problems
- It lacks a well-established decision theory

IMPORTANT QUESTIONS

PART-B

1. Define knowledge?
2. What is meant by knowledge representation?
3. List out the properties of knowledge representation?
4. List out the various knowledge representation schemes ***(or)*** types of knowledge?
5. What is meant by first order logic (FOL)?
6. Define semantic network?
7. Define Production based systems (rule based system)?
8. What are the advantages & disadvantages of Production based systems?
9. Define Frames? Give the structure of Frame?
10. What is meant by Frame based system?

11. What are the advantages & disadvantages of frame based systems?

12. Define inference?

13. Define modus ponens?

14. What is meant by Inference cycle?

15. What are the types of inference ***(or)*** inference technique***(or)*** inference strategy?

16. Define forward chaining (Goal driven Reasoning)? Give example.

17. Define backward chaining (Data driven Reasoning)? Give example.

18. What is meant by fuzzy logic (fuzzy reasoning)? **

19. Define fuzzy logic system (FLS)?

20. Define member function (MF)?

21. What are the applications of fuzzy logic?

22. What are the advantages & disadvantages of fuzzy logic?

23. Define probability?

24. Define conditional probability?

25. Define certainty factor?

26. Define Bayesian ***(or)***Bayes's theorem***(or)*** Bayesian reasoning ***(or)***Bayes's ***(or)*** Bayes's law? **

27. Define Bayesian network ***(or)***Bayes nets***(or)*** Bayesian Belief network(BBNs) ***(or)*** Belief network***(or)*** Casual probabilistic Networks(CPNs)? **

28. Define Dempster-Shafer Theory (DST)?

PART-A

1. Explain in detail fuzzy logic***(or)*** fuzzy reasoning ***(or)*** fuzzy logic system (FLS) with example? ******
2. Explain in detail Bayesian ***(or)***Bayes's theorem***(or)*** Bayesian reasoning ***(or)***Bayes's ***(or)*** Bayes's law with example? ******
3. Explain Dempster-Shafer theory (DST) with example? *****
4. Define inference? Explain the various inference strategies (technique) with example? ****
5. Explain frame based system with example? ****
6. Explain Production based systems (rule based system **or** rule value approach) with example? ****
7. Explain in detail Bayesian network ***(or)***Bayes nets***(or)*** Bayesian Belief network(BBNs) ***(or)*** Belief network***(or)*** Casual probabilistic Networks(CPNs) with example ? ****
8. Explain in detail knowledge representation with example?

FIRST PEFERENCE

- FUZZY REASONING****
- BAYESIAN THEORY****
- DEMPSTER-SHAFER THEORY(DST)*****
- INFERENCE-FORWARD & BACKWARD CHAINING***
- FRAME BASED SYSTEM****
- PROUCTION BASED SYSTEM**
- BAYESIAN NETWORK***