Unit-5

Knowledge Inference

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," "behaviour 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.

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**

Bayes' Theorem

- $P(H \mid E)$

Hypothesis Evidence

- Truthiness is depend on how many evidence supporting the hypothesis.
- $P(H_i \mid E)$ – What is probability of hypothesis that evidence is available.
- $P(E \mid H_i)$ – Probability of evidence being present for a particular ith hypothesis.
- $P(H_i)$ – Probability for ith hypothesis to true (Whatever evidence available or not)
- Baye's theorem:

$$P(H_i \mid E) = P(E|H_i)*P(H_i) / \sum P(E|H_n)*P(H_n) \text{ where } n=1 \text{ to } k.$$

Baye's Theorem

BAYE'S THEOREM: Describes the probability of an event, based on prior knowledge of conditions that might be related to the event.

↳ In Probability theory it relates the Conditional probability & marginal probabilities of two random events.

↳ Calculate $P(B|A)$ with knowledge of $P(A|B)$.

↳ $$P(H|E) = \frac{\text{no. of time H and E}}{\text{no. of times E}}$$

$$P(H|E) = \frac{P(H \cap E)}{P(E)}$$ } Prob. of H when E is true.

Baye's Theorem

BAYE'S THEOREM: Describes the probability of an event, based on prior knowledge of conditions that might be related to the event.

↳ In Probability theory it relates the Conditional probability & marginal probabilities of two random events.

↳ Calculate P(B|A) with Knowledge of P(A|B).

↳ $P(H|E) = \frac{\text{no. of time H and E}}{\text{no. of times E}}$

$P(H|E) = \frac{P(H \cap E)}{P(E)}$ } Prob. of H when E is true.

$P(A \cap B) = P(A|B) \cdot P(B)$ — (i)
$P(A \cap B) = P(B|A) \cdot P(A)$ — (ii)
} From (i) and (ii) L.H.S are equal.

↳ $P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$

So, $\boxed{P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}}$ Baye's theorem formula.

→ Posterior (Prob. of A when B is true)

→ marginal Prob. (Prob. of evidence)

Likelihood (Prob. of evidence)

→ Prior Prob (Prob. of hypothesis)

Baye's Theorem

Baye's theorem Example 1:

Ques 1:- what is the probability that Person has disease dengue with neck pain?

↳ Given:- → 80% of time dengue causes neck pain. $P(a|b) = 0.8$

↳ $P(\text{dengue}) = 1/30{,}000$ ($P(b) \Rightarrow 1/30{,}000$

↳ $P(\text{neckpain}) = .02$ $\quad P(a) = .02$

a = Proposistion that Person has neck pain

b = Person has dengue.

$P(b|a) = ?$

$$P(b|a) = \frac{P(a|b) . P(b)}{P(a)}$$

$$= \frac{0.8 . 1/30000}{0.02} = \boxed{0.00133}$$

Baye’s Theorem

- Application of Baye’s theorem in AI

➢ Next step is calculated based on prior step
  - Robot
  - Automatic Machine

➢ Forecasting
  - Weather

Probabilistic Reasoning

- 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.
- $0 \leq P(A) \leq 1$, 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.

Probabilistic Reasoning

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

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

- 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.

Probabilistic Reasoning

- **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($A \wedge B$)= 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:

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

Joint Probability Distribution

- Because events are rarely isolated from other events, we may want to define a joint probability distribution, or $P(X_1, X_2, .., X_n)$.
- Each X_i is a vector of probabilities for values of variable X_i.
- The joint probability distribution is an n-dimensional array of combinations of probabilities.

	Wet	~Wet
Rain	**0.6**	**0.4**
~Rain	0.4	0.6

Probabilistic Reasoning

- **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 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 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.

Bayes' Theorem

- **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: $P(A \land B) = P(A|B)\,P(B)$
- Similarly, the probability of event B with known event A: $P(A \land B) = P(B|A)\,P(A)$
- Equating right hand side of both the equations, we will get:

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

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

Bayes' Theorem

- 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.
- In the equation (a), in general, we can write P (B) = P(A)*P(B|Ai), hence the Bayes' rule can be written as: $P(A_i|B) = \frac{P(A_i)*P(B|A_i)}{\sum_{i=1}^{k} P(A_i)*P(B|A_i)}$

Bayes' Theorem

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

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:

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

Applying Bayes' rule

- **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

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

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

Applying Bayes' rule

- **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.**
- **Solution:**

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

- 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:

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

Example

- We wish to know probability that John has malaria, given that he has a slightly unusual symptom: a high fever.
- We have 4 kinds of information
  a) probability that a person has malaria regardless of symptoms
  b) probability that a person has the symptom of fever given that he has malaria
  c) probability that a person has symptom of fever, given that he does NOT have malaria
  d) John has high fever
- H = John has malaria
- E = John has a high fever

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

Suppose P(H) = 0.0001, P(E|H) = 0.75, P(E|~H) = 0.14

Then P(E) = 0.75 * 0.0001 + 0.14 * 0.9999 = 0.14006

and P(H|E) = (0.75 * 0.0001) / 0.14006 = 0.0005354

On the other hand, if John did not have a fever, his probability of having malaria would be

$$P(H|{\sim}E) = \frac{P({\sim}E|H) * P(H)}{P({\sim}E)} = \frac{(1-0.75)(0.0001)}{(1-0.14006)} = 0.000029$$

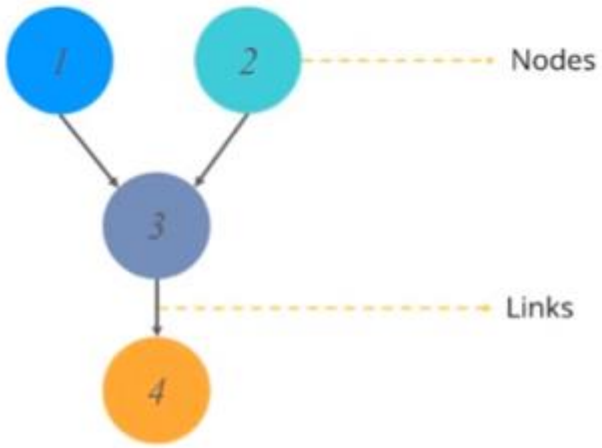
Which is much smaller.

Belief Networks

- A belief network (Bayes net) represents the dependence between variables.
- Components of a belief network graph:
- Nodes
  - These represent variables
- Links
  - X points to Y if X has a direct influence on Y
- Conditional probability tables
  - Each node has a CPT that quantifies the effects the parents have on the node
- The graph has no directed cycles

Bayesian Network

- Bayesian network falls under the category of probabilistic Graphical Modelling (PGM) technique that is used to compute uncertainties by using the concept of probability.
- Bayesian network is known as **belief network** or **casual network**
- Bayesian network is based on the concept of probability. It can be represented by using a directed acyclic graph.
- Directed acyclic graph is used to represent Bayesian network and like any other statistical group. DAG contains a set of nodes and links. where the link denote the relationship between the nodes.



Bayesian Network

- DAG models the uncertainty of an event occurring based on conditional probability distribution of each random variable.
- Conditional probability table is used to represent this distribution of each variable in the network.
- **Joint Probability** is a measure of two events happening at the same time, i.e., p(A and B). The probability of the intersection of A and B may be written p(A ∩ B).

Conditional Probability of an event B is the probability that the event will occur given that an event A has already occurred.

p(B|A): probability of event B occurring, given that event A occurs.
If A and B are **dependent** events : $P(B|A) = P(A \text{ and } B) / P(A)$
If A and B are **independent** events: $P(B|A) = P(B)$

Bayesian Belief Network

Bayesian Belief N/w in AI: It defines probabilistic independencies and dependencies among the Variables in the N/w.

- "It is a probabilistic graphical model which represents a set of Variables and their Conditional dependencies using a directed acyclic graph." → (DAG).
- Built from Probability distribution.
- Consists of
  1. DAG.
  2. Table of Conditional Probabilities.
- Node: Corresponds to a Random Variable → Continuous / Discrete
- Arc / Directed arrows: rep. Casual relationship or Conditional Probabilities among random Variables.

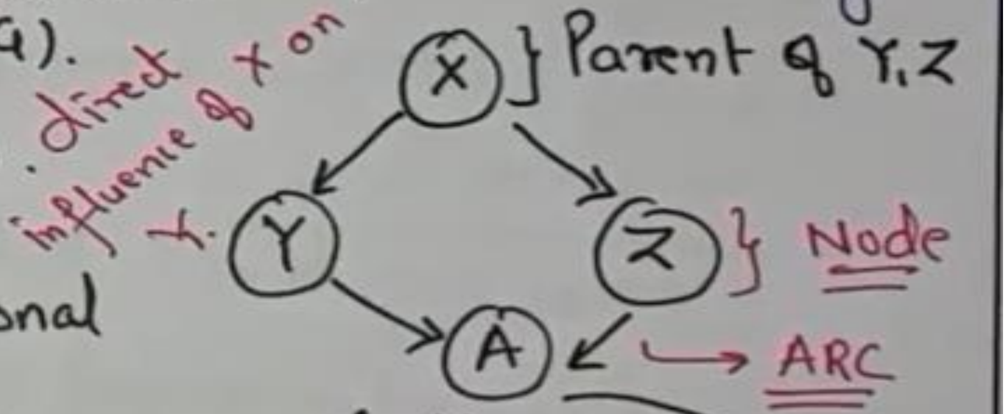
Bayesian Belief Network

Bayesian Belief N/W in AI: It defines probabilistic independencies and dependencies among the Variables in the N/W.

↳ "It is a probabilistic graphical model which represents a set of Variables and their Conditional dependencies using a directed acyclic graph." → (DAG).

↳ Built from Probability distribution.

↳ Consists of
- ① DAG.
- ② Table of Conditional Probabilities.



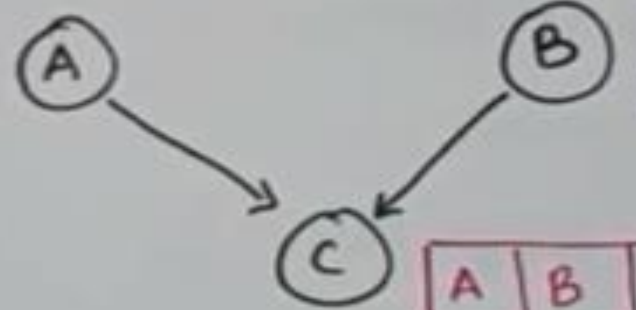
↳ Node: Corresponds to a Random Variable → Continuous / → Discrete

↳ Arc/Directed arrows: rep. Casual relationship or Conditional probabilities among random Variables.

Bayesian Belief Network

To propagate belief in Bayesian N/W, initial "Directed Acyclic Graph" is converted into an undirected graph in which the arcs can be used to transmit probabilities in dir^n of evidence.

T	·002
F	·998

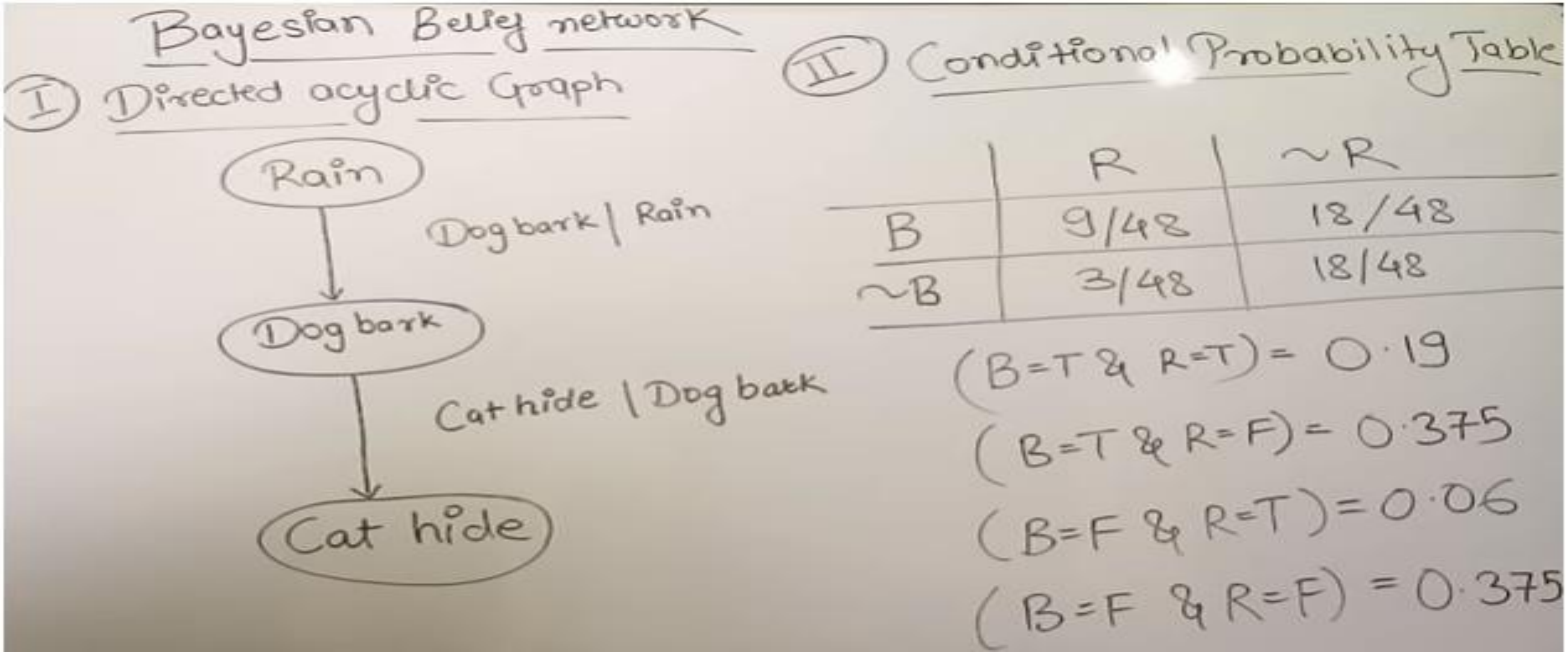


T	0·001
F	·999

A	B	P(C=T)	C=F
T	T	–	–
T	F	–	–
F	T	–	–
F	F	–	–

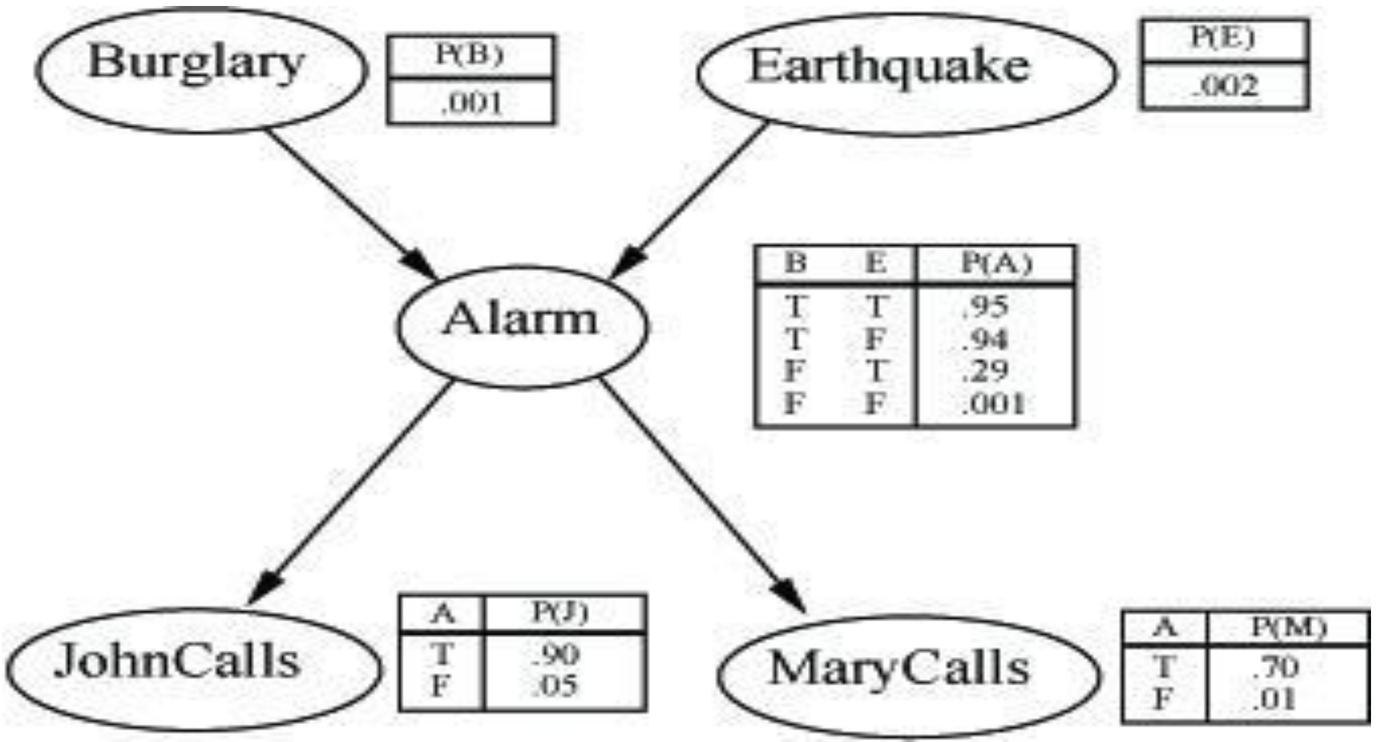
Bayesian Network Example

- Convenient for representing probabilistic relation between multiple events.



Bayesian Network Example2

- I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?
- Variables: Burglar, Earthquake, Alarm, JohnCalls, MaryCalls Network topology reflects “causal” knowledge:

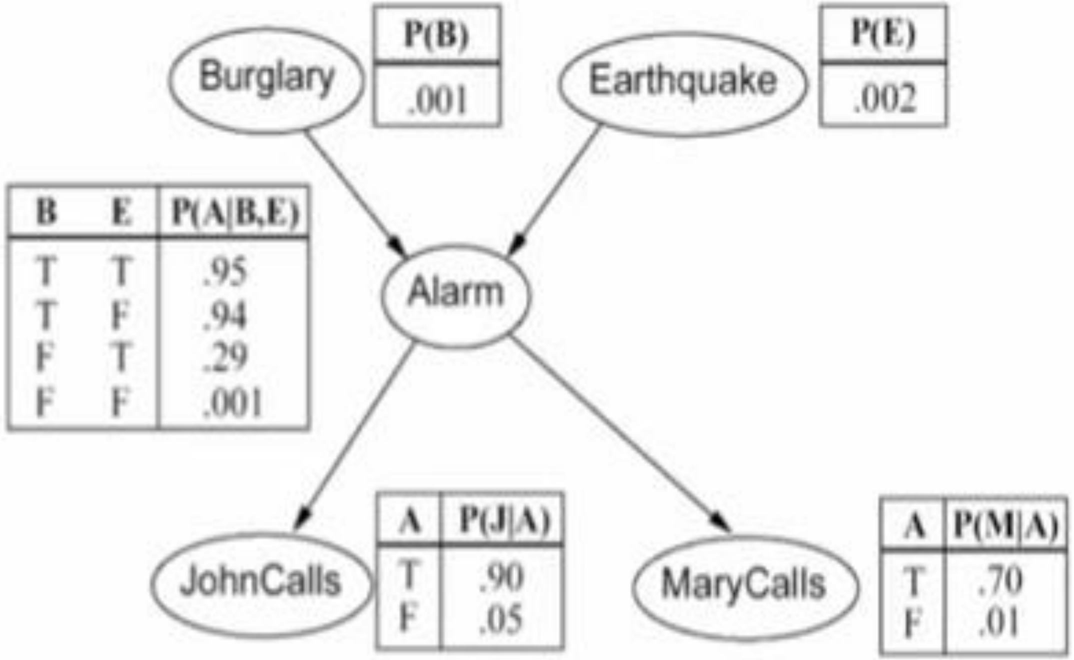


Bayesian Network Example2

- You have a new burglar alarm installed at home.
- It is fairly reliable at detecting burglary, but also sometimes responds to minor earthquakes.
- You have two neighbors, John and Merry , who promised to call you at work when they hear the alarm.
- John always calls when he hears the alarm, but sometimes confuses telephone ringing with the alarm and calls too.
- Merry likes loud music and sometimes misses the alarm.
- Given the evidence of who has or has not called, we would like to estimate the probability of a burglary.

Bayesian Network Example2

What is the probability that the alarm has sounded but neither a burglary nor an earthquake has occurred, and both John and Merry call?



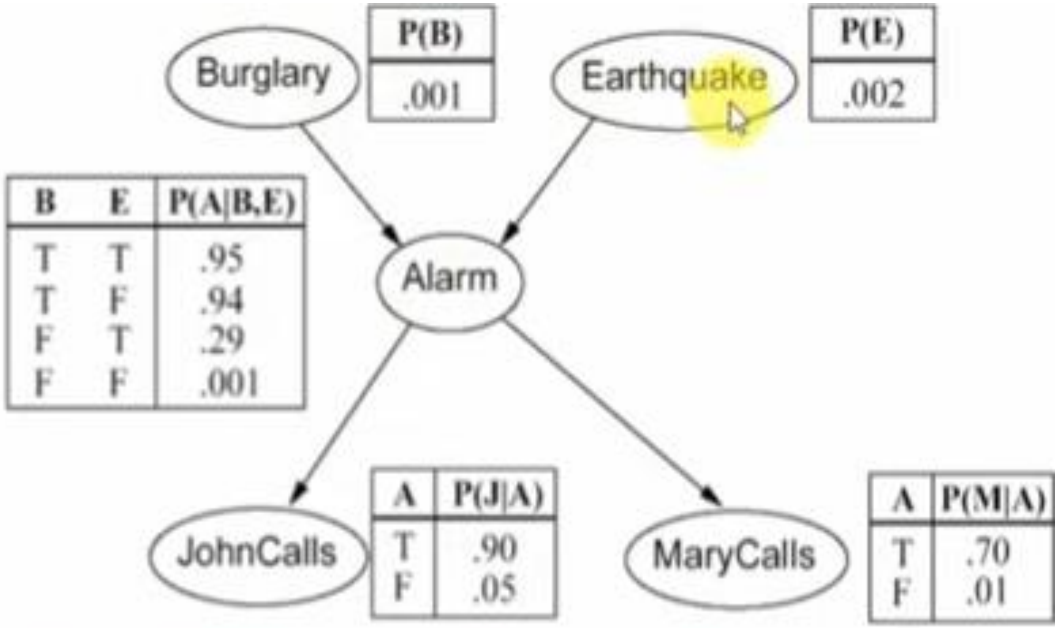
$$P(j \wedge m \wedge a \wedge \neg b \wedge \neg e) = P(j \mid a)\, P(m \mid a)\, P(a \mid \neg b, \neg e)\, P(\neg b)\, P(\neg e)$$

$$= 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998$$

$$= 0.00062$$

Bayesian Network Example2

2. What is the probability that John call?



Solution:

$$P(j) = P(j \mid a)\, P(a) + P(j \mid \neg a)\, P(\neg a)$$

$$= P(j|a)\{P(a|b,e)*P(b,e)+P(a|\neg b,e)*P(\neg b,e)+P(a|b,\neg e)*P(b,\neg e)+P(a|\neg b,\neg e)*P(\neg b,\neg e)\}$$

$$+ P(j|\neg a)\{P(\neg a|b,e)*P(b,e)+P(\neg a|\neg b,e)*P(\neg b,e)+P(\neg a|b,\neg e)*P(b,\neg e)+P(\neg a|\neg b,\neg e)*P(\neg b,\neg e)\}$$

$$= 0.90 * 0.00252 + 0.05 * 0.9974 = 0.0521$$

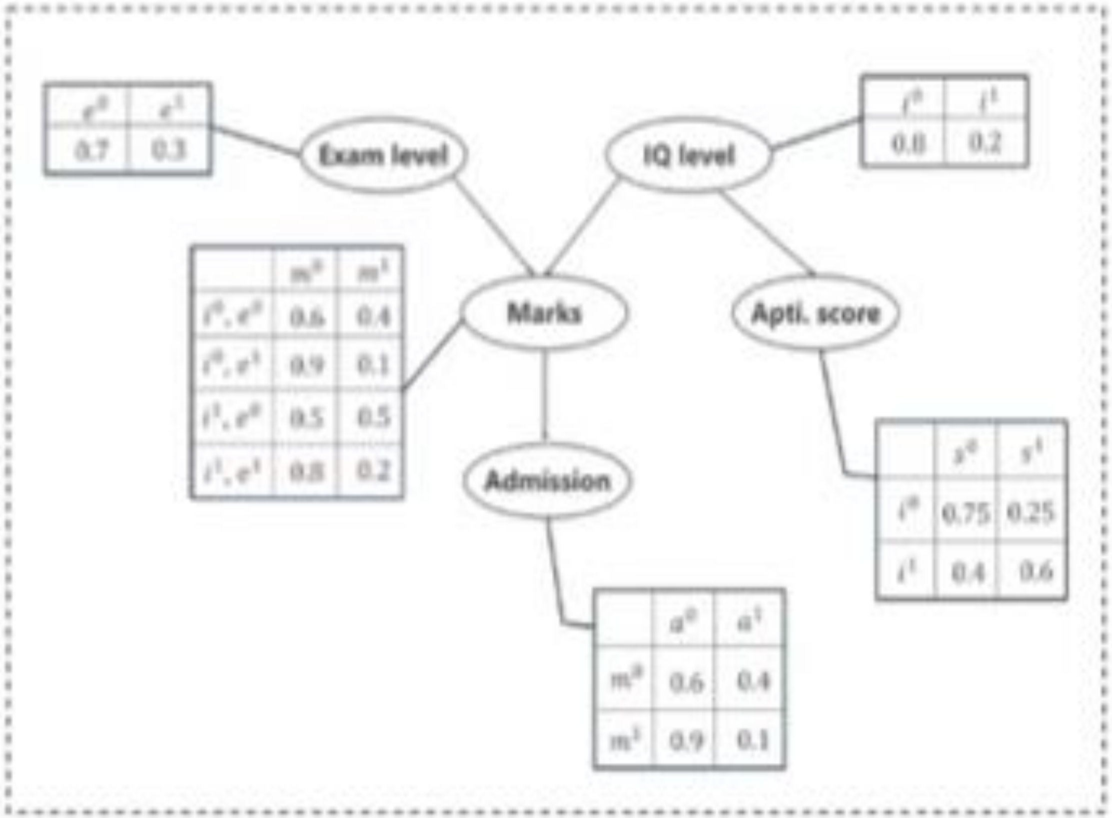
Bayesian Network Example

- There should not be a cycle.

Create a Bayesian Network that will model the marks (m) of a student on his examination.

The marks will depend on:

- ***Exam level** (e): (difficult, easy)*
- ***IQ** of the student (i): (high, low)*
- *Marks -> **admitted (a)** to a university*
- *The IQ -> **aptitude score (s)** of the student*

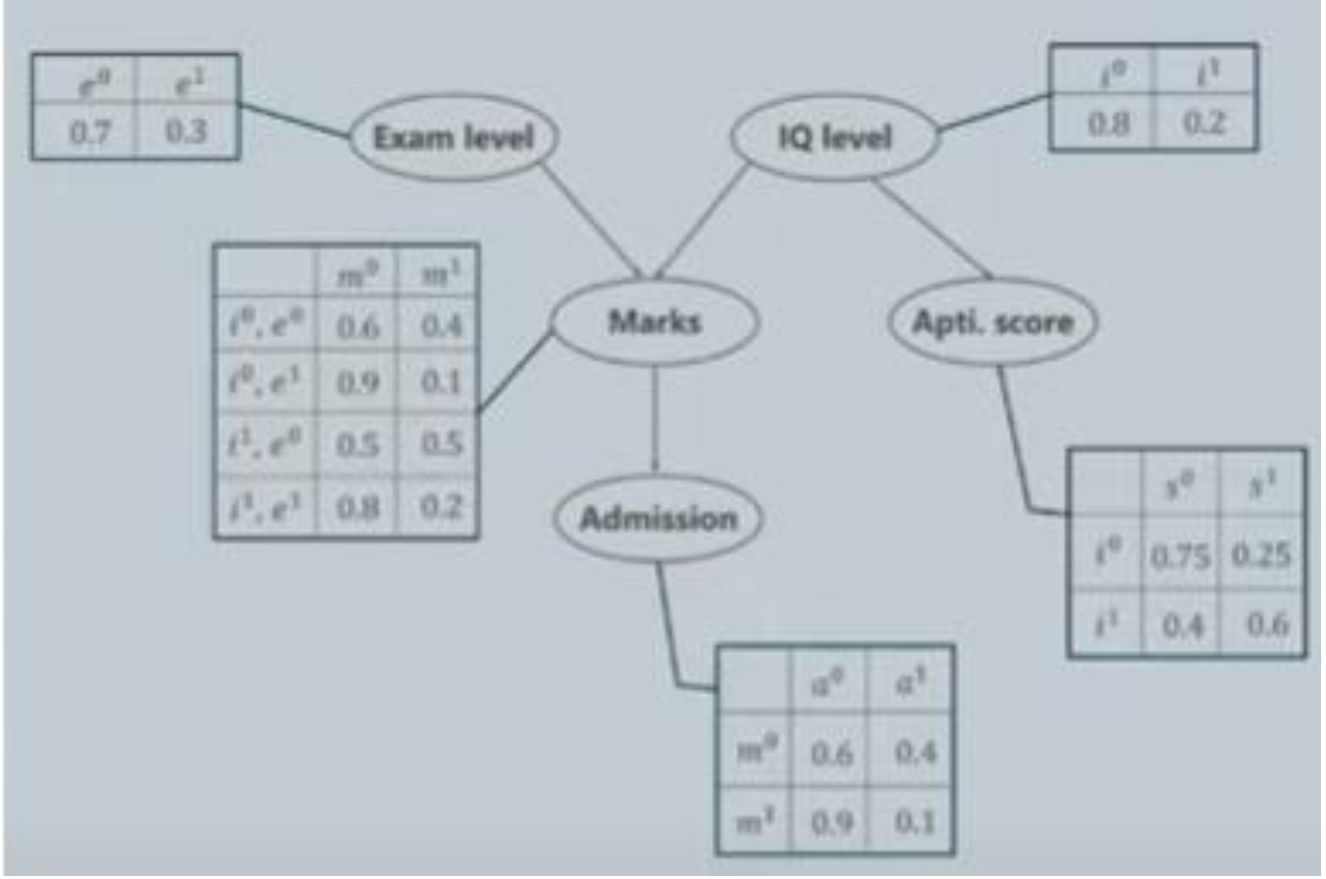


Bayesian Network Example

Factorizing Joint Probability Distribution:

p(a, m, i, e, s) = p(a | m) p(m | i, e) p(i) p(e) p(s | i)

- p(a | m): CP of student admit -> marks
- p(m | i, d): CP of the student's marks -> (IQ & exam level)
- p(i): Probability -> IQ level
- p(d): Probability -> exam level
- p(s | i) CP of aptitude scores ->IQ level

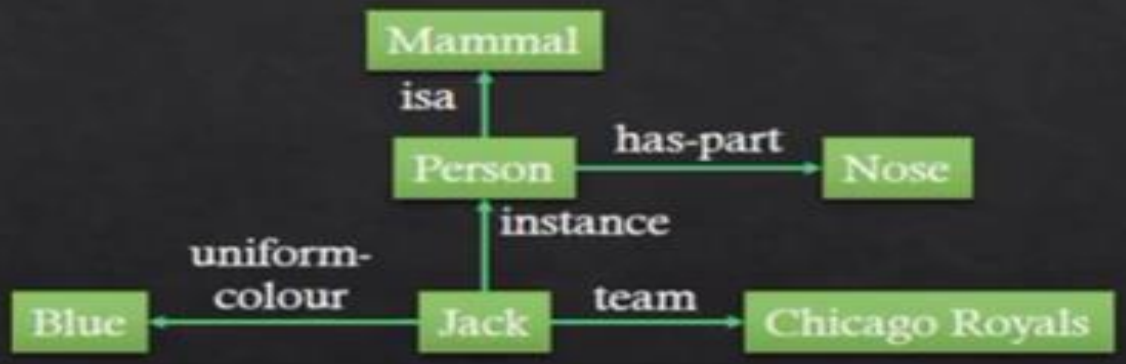


Semantic Nets

Semantic Nets

- Semantic Nets – way of representing knowledge and are based on property inheritance.
- How to represent English statements using semantic nets?
- Represent knowledge with the help of property inheritance we take the help of structures called slot and filler structures.

- Slot and filler structures are the devices to support Property Inheritance using isa and instance links.



- <u>Arcs</u>: Represent relationships among the nodes
- <u>Nodes</u>: classes, or objects, or values of attributes of an object.
- A slot is an attribute-value pair and a filler is the actual value that the slot can take.

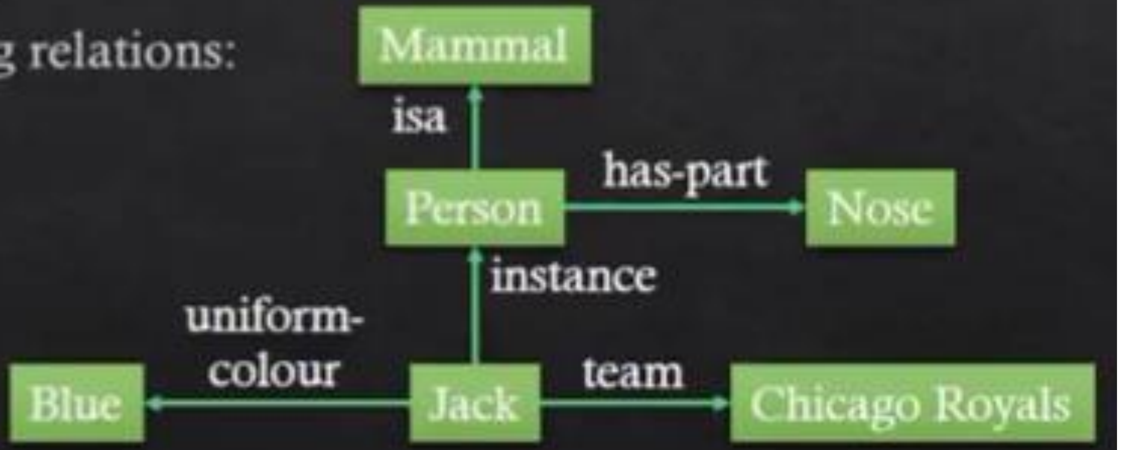
Semantic Nets

- Advantages of Slot-and-Filler structures
  - They support inheritance,
  - They support monotonic as well as non-monotonic reasoning effectively.
  - They make it easy to describe properties of relations.
  - Since they follow Object-Oriented approach, there is modularity

- Slot and filler structures are of two types:
  - Weak Slot-and-Filler Structures: Very little importance is given to the specific knowledge the structure should contain
  - Strong Slot-and-Filler structures: Specific commitments are made to the context of the representation.

Semantic Nets

- Semantics: Meaning of words
- Net: Network.
- In Semantic Nets, the meaning of a concept comes from the ways in which it is connected to other concepts.
- Information is represented as a set of nodes connected to each other by a set of labelled arcs, which represent relationships among nodes.
- In our example, we have represented the following relations:
  1. isa(Person, Mammal)
  2. has-part(Person, Nose)
  3. instance(Jack, Person)
  4. team(Jack, Chicago-Royals)
  5. uniform-color(Jack, Blue)

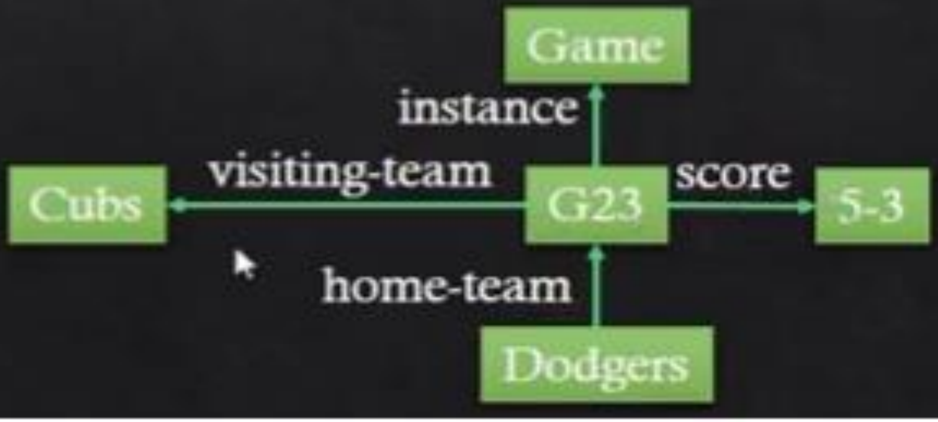


- But, from the same network, we could use inheritance to derive the additional relation:
  6. has-part(Jack, Nose)

Representing N-place Predicates using Semantic Nets

- Single Place Predicate:
  - Marcus is a man
    - man(Marcus)
    - instance(Marcus, Man)

- N-Place Predicate:
  - score(Cubs, Dodgers, 5-3)
    - three place predicate
      - visiting team, Cubs
      - home team, Dodgers
      - Score
    - Represent the game as a base class
    - G23 as instance of Game, with attributes:
      - Visiting team
      - Home team
      - Score
    - Game can be a class, instance can be G23



Representing Declarative Statement using Semantic Nets

- A declarative statement is a statement that states a fact.
  - Prabha likes riding a bike.
  - Vikram exchanges good books.
  - Pratima sang a classical song.
- Consider the statement: John gave the book to Mary
- Organise, the information into base class, derived class, their attributes, the values of those attributes, objects, their attributes and their values.

Representing Declarative Statement using Semantic Nets

- Consider the statement: John gave the book to Mary
- The event is an instance of the act of giving.
  - Give: class node
  - EV1: object of class give
  - John is the doer or the agent of the instance of giving.
  - The receiver of the instance of giving is Mary: beneficiary. The value of beneficiary attribute is Mary.
  - the item given is an instance of a class of books. That means we'll have to create a class node called book and its instance called say BK23 and this instance is the filler of the attribute item

Frame Based System

Frame Representation

- When an agent encounters a new situation, it will need to retrieve information to act rationally in that situation. This information is likely to be multi-faceted and hierarchical. One way of structuring the knowledge is in terms of frames.
- These are frameworks consisting of slots.
- Each slot containing information in various representations like logical sentences, production rule, frame etc.
- Each framework represents a stereotypical object or situation.
- Whenever an agent encounters an object or situation which fits the stereotype, agent retrieve the framework and change some default information or fill the blank information.

Frame Representation

- Some of the information is procedural, so that when a blank is filled in with certain values, a procedure must be carried out.
- In this way frame dictate how to act rationally in situations.
- Types of Information store in slot

➢ Information for choosing the frame

➢ Information about relationships between frames.

➢ Procedures to be carried out.

➢ Default information.

➢ Blank slots.

Frame Representation

- Information for choosing the frame
  - It may also be information about situations or descriptors for the stereotype the frame represents.
  - E.g. name ,id
- Information about relationships between frames.
  - Two frames should never be considered at the same time whether this frame is a generalisation or specialisation of another frame.
- Procedures to be carried out.
  - It is rational action an agent should do in a situation where a particular value for a slot has been identified.
- Default information.
  - It’s the values when certain information required for the frame is missing.
  - Default information is used in choosing actions until more specific information is found.
- Blank slots.
  - These are flagged to be left blank unless required for a particular task.

FRAMES

- **FRAMES** :- means of representing common sense knowledge. Knowledge is organized into small packets called “Frames”. All frames of a given situation constitute the system.
- A frame can be defined as a structure that has slots for various objects & a collection of frames consist of expectation for a given situation.
- Frame are used to represent two types of knowledge viz. declarative/factual and procedural, declarative & procedural Frames: -
- A frame that merely contains description about objects is call a declarative type/factual situational frame.

Name : Computer Centre	
A/c	Stationary cupboard
Computer	Dumb terminals
Printer	

Name of the frame

Slots in the frame

- Frames which have procedural knowledge embedded in it are called action procedure frames. The action frame has the following slots.
- Actor slot which holds information @ who is performing the activity.
- Source Slot hold information from where the action has to begin.
- Destination slot holds information about the place where action has to end.
- Task slot This generates the necessary sub frames required to perform the operation.

Name : Cleaning the ict of carburetor

Actor

Expert

Object

Source Destination

Scooter | Scooter

Task 1 Task 2

Task 3

Remove Carburetor | Clean Nozzle | Fix Carburetor

Thank you.