Artificial Intelligence

Unit-7 Statistical Reasoning

Artificial Intelligence 01CE0702

Marwadi University

Department of Computer Engineering

Shilpa Singhal

Probability

- The word 'Probability' means the chance of occurring of a particular event.
- It is generally possible to predict the future of an event quantitatively with a certain probability of being correct.
- The probability is used in such cases where the outcome of the trial is uncertain.
- Probability Definition:
  - The probability of hap[illegible]d by P(A), is defined as

$$P(A) = \frac{\text{number of cases favourable to A}}{\text{number of possible outcomes}}$$

  - Thus, if an event can happen in m ways and fails to occur in n ways and m+n ways is equally likely to occur then the probability of happening of the event A is given by

$$P(A) = \frac{m}{m+n}$$

Probability

- Important Terms related to Probability:
  - Trial and Event: The performance of an experiment is called a trial, and the set of its outcomes is termed an event.
    - Example: Tossing a coin and getting head is a trial. Then the event is {HT, TH, HH}
  - Random Experiment: It is an experiment in which all the possible outcomes of the experiment are known in advance. But the exact outcomes of any specific performance are not known in advance.
    - Example:
      - Tossing a Coin
      - Rolling a die
  - Outcome: The result of a random experiment is called an Outcome.
    - Example: 1. Tossing a coin is an experiment and getting head is called an outcome.

Probability

- Sample Space: The set of all possible outcomes of an experiment is called sample space and is denoted by S.
  - Example: When a die is thrown, sample space is S = {1, 2, 3, 4, 5, 6} It consists of six outcomes 1, 2, 3, 4, 5, 6

Probability

- # Independent Events:
  - Events A and B are said to be independent if the occurrence of any one event does not affect the occurrence of any other event.

$$P(A \cap B) = P(A)\,P(B).$$

- # Example:
  - A coin is tossed thrice,
    and all 8 outcomes are equally likely
  - A: "The first throw results in heads."
  - B: "The last throw results in Tails.“
  - Prove that event A and B are independent.
  - Solution:

Sample Space: [HHH, HHT, HTH, THH, TTT, TTH, THT, HTT]

A: [HHH, HHT, HTH, HTT]

B: [HHT, TTT, THT, HTT]

A∩B: [HHT, HTT]

$P(A) = \frac{4}{8} = \frac{1}{2}$ $\qquad$ $P(B) = \frac{4}{8} = \frac{1}{2}$

$P(A \cap B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

Probability

- Dependent Event:
  - Events are said to be dependent if occurrence of one affect the occurrence of other events.
- Conditional Probability
  - The conditional probability of an event B is the probability that the event will occur given the knowledge that an event A has already occurred.
  - This probability is written P(B|A), notation for the probability of B given A.
  - In the case where events A and B are independent, the conditional probability of event B given event A is simply the probability of event B, that is P(B).
  - If events A and B are not independent, then the probability of the intersection of A and B (the probability that both events occur) is defined by

    $$P(A \text{ and } B) = P(A)P(B|A).$$
  - From this definition, the conditional probability P(B|A) is easily obtained by dividing by P(A):

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

Marwadi University

Probability

- Conditional Probability Example
  - In a card game, suppose a player needs to draw two cards of the same suit in order to win. Of the 52 cards, there are 13 cards in each suit.
  - Suppose first the player draws a heart. Now the player wishes to draw a second heart. Since one heart has already been chosen, there are now 12 hearts remaining in a deck of 51 cards.
  - So the conditional probability P(Draw second heart|First card a heart) = 12/51.

Marwadi University

Probability

- Probability density function (PDF):
  - Also known as density of a continuous random variable, is a function whose value at any given sample (or point) in the sample space (the set of possible values taken by the random variable) can be interpreted as providing a relative likelihood that the value of the random variable would equal that sample.
  - In other words, while the absolute likelihood for a continuous random variable to take on any particular value is 0 (since there is an infinite set of possible values to begin with), the value of the PDF at two different samples can be used to infer, in any particular draw of the random variable, how much more likely it is that the random variable would equal one sample compared to the other sample.

Probability

- Probability distribution is the mathematical function that gives the probabilities of occurrence of different possible outcomes for an experiment.
- A discrete probability distribution is made up of discrete variables.
  - Specifically, if a random variable is discrete, then it will have a discrete probability distribution.
- The following are examples of discrete probability distributions commonly used in statistics:
  - Binomial distribution.
  - Geometric Distribution
  - Hypergeometric distribution.
  - Multinomial Distribution.
  - Negative binomial distribution.
  - Poisson distribution.

Probability

- Poisson distribution:
  - It is a tool that helps to predict the probability of certain events happening when you know how often the event has occurred.
  - It gives us the probability of a given number of events happening in a fixed interval of time.
- Binomial distribution
  - It can be thought of as simply the probability of a SUCCESS or FAILURE outcome in an experiment or survey that is repeated multiple times.
  - The binomial is a type of distribution that has two possible outcomes (the prefix "bi" means two, or twice).
  - For example, a coin toss has only two possible outcomes: heads or tails and taking a test could have two possible outcomes: pass or fail.

Marwadi University

Probability

- Geometric distribution:
  - It represents the number of failures before you get a success in a series of Bernoulli trials. This discrete probability distribution is represented by the probability density function:

    $f(x) = (1 - p)^{(x - 1)} * p$

Here,

P: Probability of success

X: event number

Marwadi University

Bayes Theorem

- Naive Bayes classifiers are a collection of classification algorithms based on Bayes’ Theorem.
- It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other.
- To start with, let us consider a dataset.
    - Consider a fictional dataset that describes the weather conditions for playing a game of golf.
    - Given the weather conditions, each tuple classifies the conditions as fit(“Yes”) or unfit(“No”) for playing golf.

Marwadi University

Bayes Theorem

Outlook	Temperature	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

Marwadi University

Bayes Theorem

- The fundamental Naive Bayes assumption is that each feature makes an:
  - independent
  - Equal
- With relation to our dataset, this concept can be understood as:
  - We assume that no pair of features are dependent.
    - For example, the temperature being ‘Hot’ has nothing to do with the humidity or the outlook being ‘Rainy’ has no effect on the winds.
    - Hence, the features are assumed to be independent.
  - Secondly, each feature is given the same weight(or importance).
    - For example, knowing only temperature and humidity alone can’t predict the outcome accurately.
    - None of the attributes is irrelevant and assumed to be contributing equally to the outcome.

Bayes Theorem

- Bayes' Theorem finds the probability of an event occurring given the probability of another event that has already occurred. Bayes' theorem is stated mathematically as the following equation:

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

- where A and B are events and P(B) ?
  - Basically, we are trying to find probability of event A, given the event B is true. Event B is also termed as evidence.
  - P(A) is the priori of A (the prior probability, i.e. Probability of event before evidence is seen). The evidence is an attribute value of an unknown instance(here, it is event B).
  - P(A|B) is a posteriori probability of B, i.e. probability of event after evidence is seen.

Bayes Theorem

- Now, with regards to our dataset, we can apply Bayes' theorem in following way: $P(y|X) = \frac{P(X|y)P(y)}{P(X)}$
- where, y is class variable and X is a dependent feature vector (of size n) where: $X = (x_1, x_2, x_3, \ldots.., x_n)$
- Just to clear, an example of a feature vector and corresponding class variable can be: (refer 1st row of dataset)
  - X = (Rainy, Hot, High, False)
  - y = No
  - So basically, P(y|X) here means, the probability of "Not playing golf" given that the weather conditions are "Rainy outlook", "Temperature is hot", "high humidity" and "no wind".

Marwadi University

Bayes Theorem

- Hence, we reach to the result: $P(y|x_1,...,x_n) = \frac{P(x_1|y)P(x_2|y)...P(x_n|y)P(y)}{P(x_1)P(x_2)...P(x_n)}$
- which can be expressed as: $P(y|x_1,...,x_n) = \frac{P(y)\prod_{i=1}^{n} P(x_i|y)}{P(x_1)P(x_2)...P(x_n)}$
- Now, as the denominator remains constant for a given input, we can remove that term: $P(y|x_1,...,x_n) \propto P(y)\prod_{i=1}^{n} P(x_i|y)$
- Now, we need to create a classifier model. For this, we find the probability of given set of inputs for all possible values of the class variable y and pick up the output with maximum probability. This can be expressed mathematically as: $y = argmax_y P(y)\prod_{i=1}^{n} P(x_i|y)$

Bayes Theorem

- So, finally, we are left with the task of calculating P(y) and $P(x_i \mid y)$.
- Please note that P(y) is also called class probability and $P(x_i \mid y)$ is called conditional probability.
- Let us try to apply the above formula manually on our weather dataset.
- For this, we need to do some precomputations on our dataset.
- We need to find $P(x_i \mid y_j)$ for each x_i in X and y_j in y.

Bayes Theorem

So, in the figure above, we have calculated $P(x_i \mid y_j)$ for each x_i in X and y_j in y manually in the tables 1-4. For example, probability of playing golf given that the temperature is cool, i.e P(temp. = cool | play golf = Yes) = 3/9.

Also, we need to find class probabilities (P(y)) which has been calculated in the table 5. For example, P(play golf = Yes) = 9/14.

Outlook

	Yes	No	P(yes)	P(no)
Sunny	2	3	2/9	3/5
Overcast	4	0	4/9	0/5
Rainy	3	2	3/9	2/5
Total	9	5	100%	100%

Temperature

	Yes	No	P(yes)	P(no)
Hot	2	2	2/9	2/5
Mild	4	2	4/9	2/5
Cool	3	1	3/9	1/5
Total	9	5	100%	100%

Humidity

	Yes	No	P(yes)	P(no)
High	3	4	3/9	4/5
Normal	6	1	6/9	1/5
Total	9	5	100%	100%

Wind

	Yes	No	P(yes)	P(no)
False	6	2	6/9	2/5
True	3	3	3/9	3/5
Total	9	5	100%	100%

Play		P(Yes)/P(No)
Yes	9	9/14
No	5	5/14
Total	14	100%

Marwadi University

Bayes Theorem

- Let us test it on a new set of features (let us call it today):
- today = (Sunny, Hot, Normal, False)

$$P(Yes|today) = \frac{P(SunnyOutlook|Yes)P(HotTemperature|Yes)P(NormalHumidity|Yes)P(NoWind|Yes)P(Yes)}{P(today)}$$

$$P(No|today) = \frac{P(SunnyOutlook|No)P(HotTemperature|No)P(NormalHumidity|No)P(NoWind|No)P(No)}{P(today)}$$

Since, P(today) is common in both probabilities, we can ignore P(today) and find proportional probabilities as:

$$P(Yes|today) \propto \frac{2}{9}.\frac{2}{9}.\frac{6}{9}.\frac{6}{9}.\frac{9}{14} \approx 0.0141$$

$$P(No|today) \propto \frac{3}{5}.\frac{2}{5}.\frac{1}{5}.\frac{2}{5}.\frac{5}{14} \approx 0.0068$$

$$P(Yes|today) + P(No|today) = 1$$

These numbers can be converted into a probability by making the sum equal to 1 (normalization):

$$P(Yes|today) = \frac{0.0141}{0.0141+0.0068} = 0.67$$

$$P(No|today) = \frac{0.0068}{0.0141+0.0068} = 0.33$$

$$P(Yes|today) > P(No|today)$$

Prediction that golf would be played is ‘Yes’.

Bayesian Belief Network

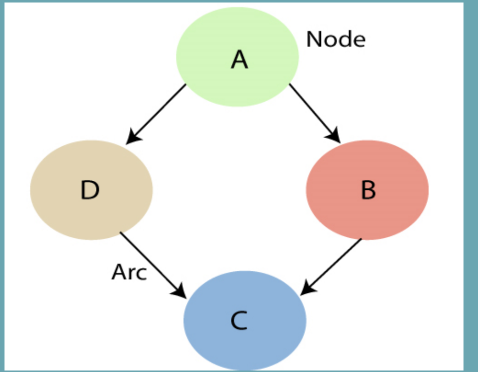
- Bayesian belief network is key computer technology for dealing with probabilistic events and to solve a problem which has uncertainty.
- We can define a Bayesian network as:
  - "A Bayesian network is a probabilistic graphical model which represents a set of variables and their conditional dependencies using a directed acyclic graph."
- It is also called a Bayes network, belief network, decision network, or Bayesian model.
- Bayesian networks are probabilistic, because these networks are built from a probability distribution, and also use probability theory for prediction and anomaly detection.

Bayesian Belief Network

- 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 can be used for building models from data and experts opinions, and it consists of two parts:
  - Directed Acyclic Graph
  - Table of conditional probabilities.
- The generalized form of Bayesian network that represents and solve decision problems under uncertain knowledge is known as an Influence diagram.

Bayesian Belief Network

- Bayesian network graph is made up of nodes and Arcs, where:
- Each node corresponds to the random variables, and a variable can be continuous or discrete.
- Arc represent the causal relationship or conditional probabilities between random variables. These directed links or arrows connect the pair of nodes in the graph.
- These links represent that one node directly influence the other node, and if there is no directed link that means that nodes are independent with each other
    - In the above diagram, A, B, C, and D are random variables represented by the nodes of the network graph.
    - If we are considering node B, which is connected with node A by a directed arrow, then node A is called the parent of Node B.
    - Node C is independent of node A.



Bayesian Belief Network

- The Bayesian network has mainly two components:
  - Causal Component
  - Actual numbers
- Each node in the Bayesian network has condition probability distribution $P(X_i \mid Parent(X_i))$, which determines the effect of the parent on that node.
- Bayesian network is based on Joint probability distribution and conditional probability.

Marwadi University

Statistical Learning

- Statistical learning plays a key role in many areas of science, finance and industry. A few examples are already considered in Lesson 1. Some more examples of the learning problems are:
    - Predict whether a patient, hospitalized due to a heart attack, will have a second heart attack. The prediction is to be based on demographic, diet and clinical measurements for that patient.
    - Predict the price of a stock in 6 months from now, on the basis of company performance measures and economic data.
    - Estimate the amount of glucose in the blood of a diabetic person, from the infrared absorption spectrum of that person’s blood.
    - Identify the risk factors for prostate cancer, based on clinical and demographic variables.

Marwadi University

Statistical Learning

- Statistical learning problem is learning from the data.
- In a typical scenario, we have an outcome measurement, usually quantitative (such as a stock price) or categorical (such as heart attack/no heart attack), that we wish to predict based on a set of features (such as diet and clinical measurements).
- We have a Training Set which is used to observe the outcome and feature measurements for a set of objects. Using this data we build a Prediction Model, or a Statistical Learner , which enables us to predict the outcome for a set of new unseen objects.

Statistical Learning

- Elements of Statistical Learning
  - Supervised Learning
    - Clustering
    - Classification
    - Reinforcement Learning
  - Unsupervised Learning
    - Clustering

Probabilistic Graphical Model

- A graphical model or probabilistic graphical model (PGM) or structured probabilistic model is a probabilistic model for which a graph expresses the conditional dependence structure between random variables.
- They are commonly used in probability theory, statistics—particularly Bayesian statistics—and machine learning.
- Generally, probabilistic graphical models use a graph-based representation as the foundation for encoding a distribution over a multi-dimensional space and a graph that is a compact or factorized representation of a set of independences that hold in the specific distribution.

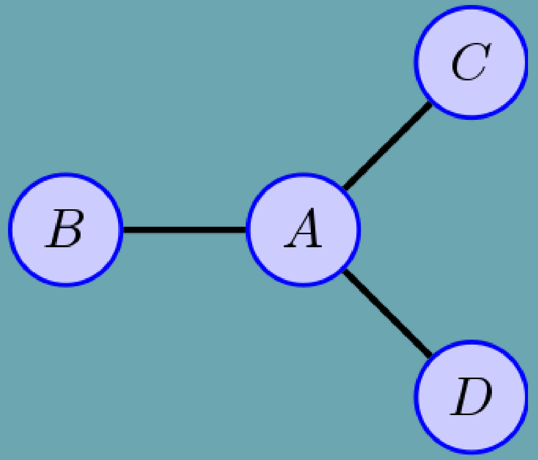
Probabilistic Graphical Model

- Undirected Graphical Model
  - The undirected graph shown may have one of several interpretations; the common feature is that the presence of an edge implies some sort of dependence between the corresponding random variables. From this graph we might deduce that B , C , D are all mutually independent, once A is known, or (equivalently in this case) that

$$P[A,B,C,D] = f_{AB}[A,B] \cdot f_{AC}[A,C] \cdot f_{AD}[A,D]$$

or some non-negative functions

$$f_{AB}, f_{AC}, f_{AD}$$



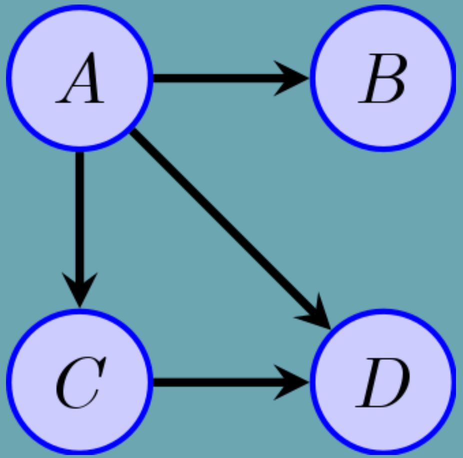
Probabilistic Graphical Model

- Directed graphical model
  - If the network structure of the model is a directed acyclic graph, the model represents a factorization of the joint probability of all random variables. More precisely, if the events are X1, … , X n then the joint probability satisfies

$$P[X_1, \ldots, X_n] = \prod_{i=1}^{n} P[X_i | \mathrm{pa}(X_i)]$$

  - where pa(X i) is the set of parents of node X i (nodes with edges directed towards Xi).
  - In other words, the joint distribution factors into a product of conditional distributions.
  - For example, the directed acyclic graph shown in the Figure this factorization would be

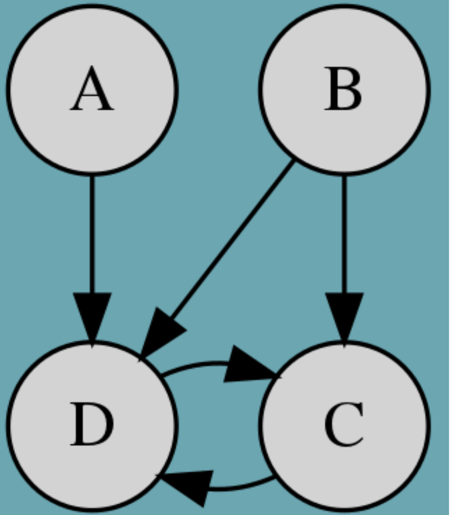
$$P[A, B, C, D] = P[A] \cdot P[B|A] \cdot P[C|A] \cdot P[D|A, C]$$



Probabilistic Graphical Model

- Cyclic Directed Graphical Models
  - The next figure depicts a graphical model with a cycle. This may be interpreted in terms of each variable 'depending' on the values of its parents in some manner. The particular graph shown suggests a joint probability density that factors as

$$P[A, B, C, D] = P[A] \cdot P[B] \cdot P[C, D|A, B]$$



Fuzzy Logic

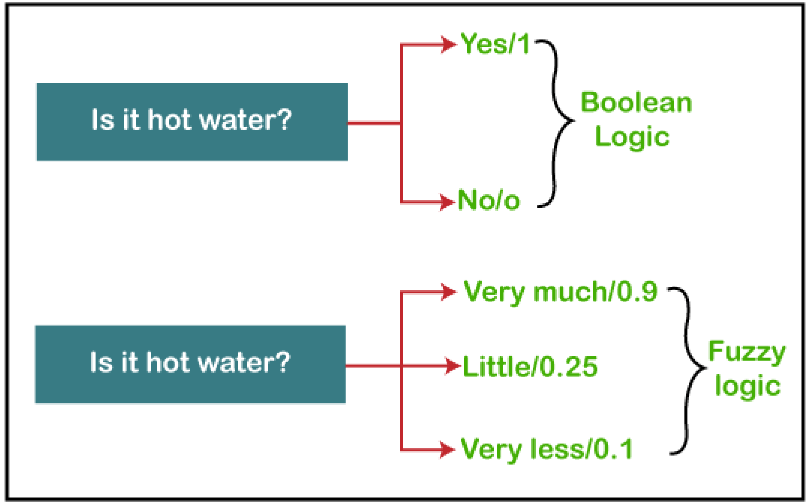
- What is Fuzzy Logic?
  - Fuzzy Logic (FL) is a method of reasoning that resembles human reasoning.
  - The approach of FL imitates the way of decision making in humans that involves all intermediate possibilities between digital values YES and NO.
  - The conventional logic block that a computer can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

What is Fuzzy Logic?

- The **'Fuzzy'** word means the things that are not clear or are vague. Sometimes, we cannot decide in real life that the given problem or statement is either true or false. At that time, this concept provides many values between the true and false and gives the flexibility to find the best solution to that problem.

Example of Fuzzy Logic as comparing to Boolean Logic



- Fuzzy logic contains the multiple logical values and these values are the truth values of a variable or problem between 0 and 1.
- This concept was introduced by **Lofti Zadeh** in **1965** based on the **Fuzzy Set Theory**. This concept provides the possibilities which are not given by computers, but similar to the range of possibilities generated by humans.
- In the Boolean system, only two possibilities (0 and 1) exist, where 1 denotes the absolute truth value and 0 denotes the absolute false value. But in the fuzzy system, there are multiple possibilities present between the 0 and 1, which are partially false and partially true.

Fuzzy Sets and Membership function

- The concept of a set is fundamental to mathematics. Crisp set theory is governed by a logic that uses one of only two values: true or false.
- This logic cannot represent vague concepts, and therefore fails to give the answers on the inconsistencies.
- In fuzzy set theory, an element is with a certain degree of membership. Thus, a proposition is not either true or false, but may be partly true (or partly false) to any degree.
- This degree is usually taken as a real number in the interval [0,1].

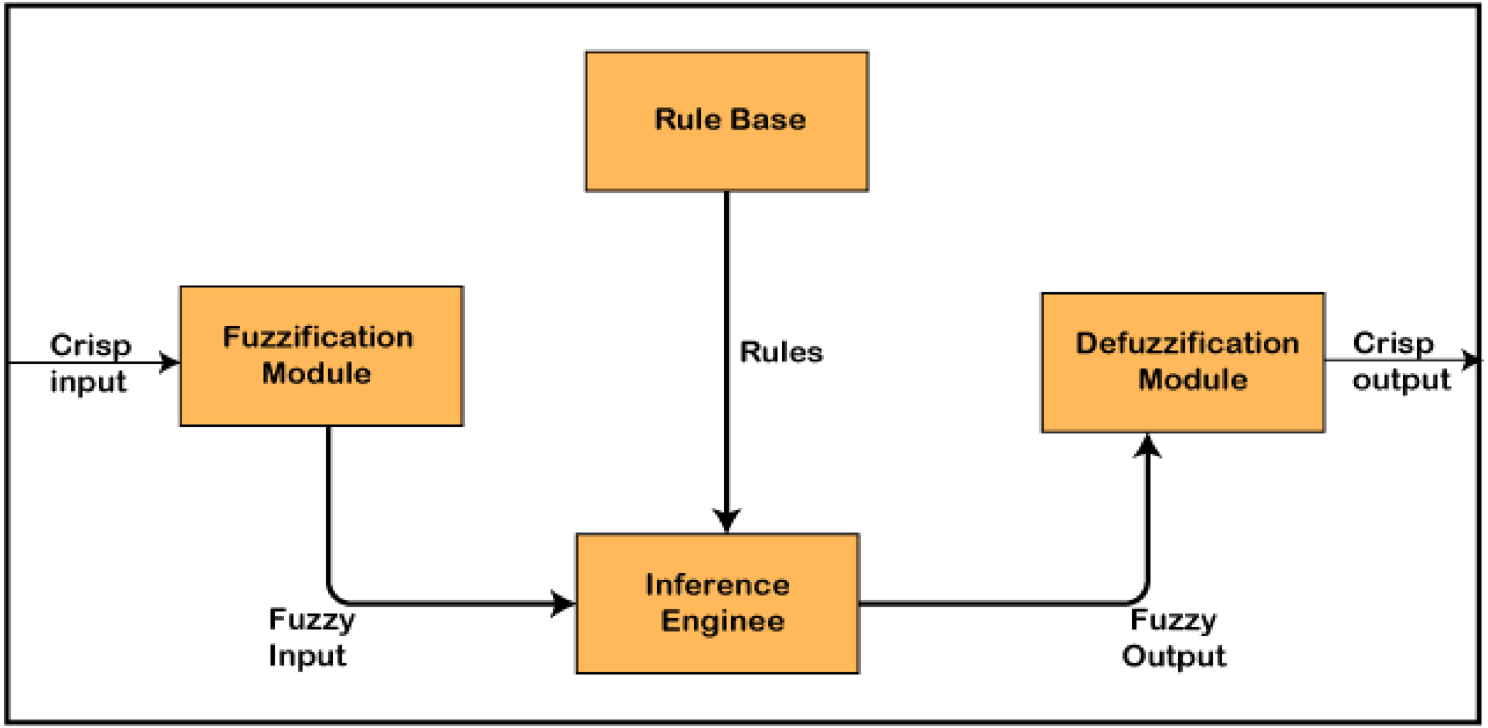
The classical example in fuzzy sets is tall men. The elements of the fuzzy set “tall men” are all men, but their degrees of membership depend on their height.

Name	Height in cm	Degree of Membership	
		Crisp	Fuzzy
John	208	1	1
Tom	181	1	0.8
Bob	152	0	0.0
Mike	198	1	0.9
Billy	158	0	0.4

Architecture of a Fuzzy Logic System

In the architecture of the **Fuzzy Logic** system, each component plays an important role. The architecture consists of the different four components which are given below.

1. Rule Base
2. Fuzzification
3. Inference Engine
4. Defuzzification


Rule Base
Rules
Crisp
input
Fuzzification
Module
Fuzzy
Input
Inference
Enginee
Fuzzy
Output
Defuzzification
Module
Crisp
output

1. Rule Base

Rule Base is a component used for storing the set of rules and the If-Then conditions given by the experts are used for controlling the decision-making systems.

There are so many updates that come in the Fuzzy theory recently, which offers effective methods for designing and tuning of fuzzy controllers. These updates or developments decreases the number of fuzzy set of rules.

2. Fuzzification

Fuzzification is a module or component for transforming the system inputs, i.e., it converts the crisp number into fuzzy steps. The crisp numbers are those inputs which are measured by the sensors and then fuzzification passed them into the control systems for further processing. This component divides the input signals into following five states in any Fuzzy Logic system:

- Large Positive (LP)
- Medium Positive (MP)
- Small (S)
- Medium Negative (MN)
- Large negative (LN)

3. Inference Engine

- This component is a main component in any Fuzzy Logic system (FLS), because all the information is processed in the Inference Engine. It allows users to find the matching degree between the current fuzzy input and the rules. After the matching degree, this system determines which rule is to be added according to the given input field. When all rules are fired, then they are combined for developing the control actions.

4. Defuzzification

- **Defuzzification** is a module or component, which takes the fuzzy set inputs generated by the **Inference Engine**, and then transforms them into a crisp value. It is the last step in the process of a fuzzy logic system. The crisp value is a type of value which is acceptable by the user. Various techniques are present to do this, but the user has to select the best one for reducing the errors.

Membership Function

The membership function is a function which represents the graph of fuzzy sets, and allows users to quantify the linguistic term. It is a graph which is used for mapping each element of x to the value between 0 and 1.

- This function is also known as indicator or characteristics function.
- This function of Membership was introduced in the first papers of fuzzy set by **Zadeh**. For the Fuzzy set B, the membership function for X is defined as: $\mu B:X \rightarrow [0,1]$. In this function X, each element of set B is mapped to the value between 0 and 1. This is called a degree of membership or membership value.

Fuzzy Logic

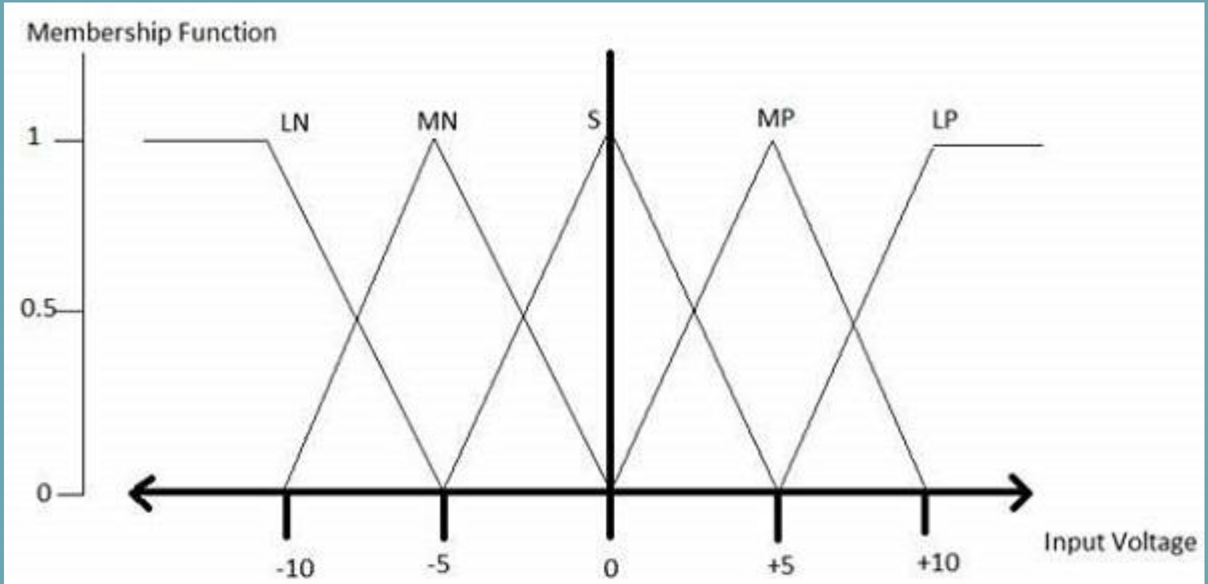
- Why Fuzzy Logic?
- Fuzzy logic is useful for commercial and practical purposes.
  - It can control machines and consumer products.
  - It may not give accurate reasoning, but acceptable reasoning.
  - Fuzzy logic helps to deal with the uncertainty in engineering.

Fuzzy Logic

- 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

- Fuzzification Module – It transforms the system inputs, which are crisp numbers, into fuzzy sets.



LP	x is Large Positive
MP	x is Medium Positive
S	x is Small
MN	x is Medium Negative
LN	x is Large Negative

Applications of Fuzzy Logic System

- 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

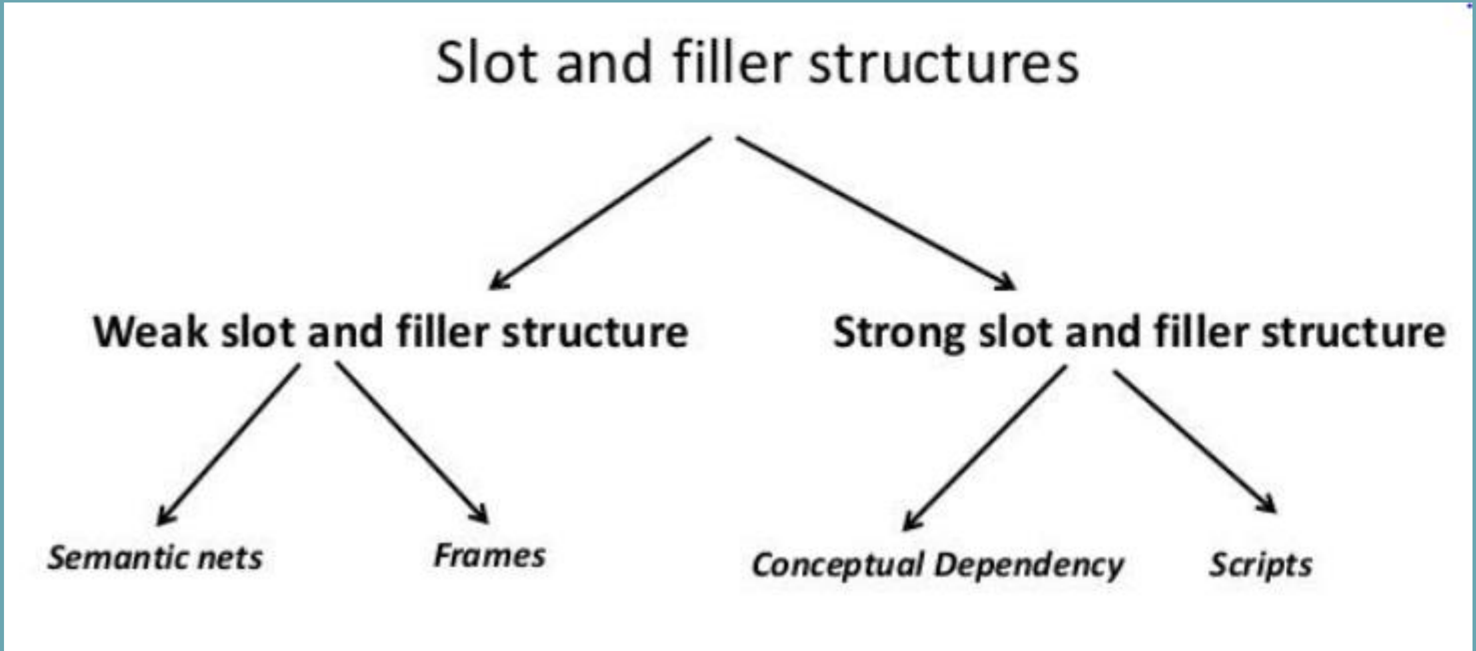
Fuzzy Logic Systems

- 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.

Weak Slot-and-Filler Structures, Strong Slot-and-Filler Structures

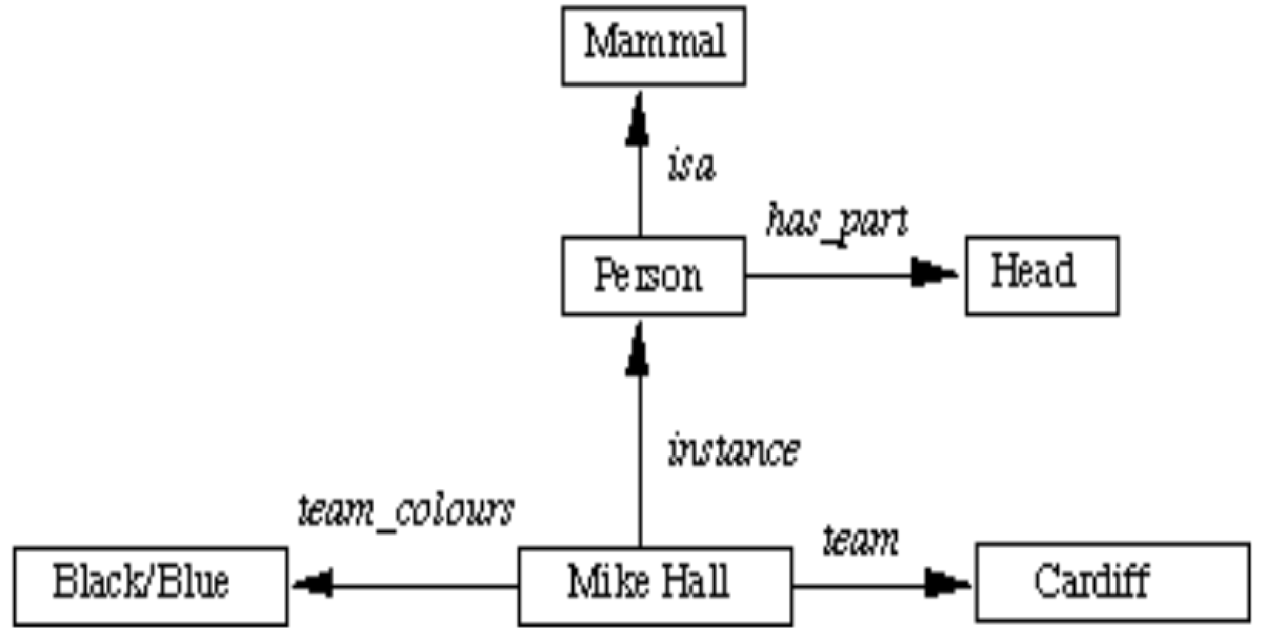
- Weak slot and filler structure
  - The knowledge in slot and filler systems consists of structures as a set of entities and their attributes
  - This structure is called a weak slot and filler structure.
- Strong slot and filler structure:
  - But later we used strong slot and filler structures.
  - It represent links between objects according to more rigid rules.

Weak Slot-and-Filler Structures



Weak Slot-and-Filler Structures

- Semantic Net



Weak Slot-and-Filler Structures

- Semantic Net

Ex- *John gave Mary the book.*

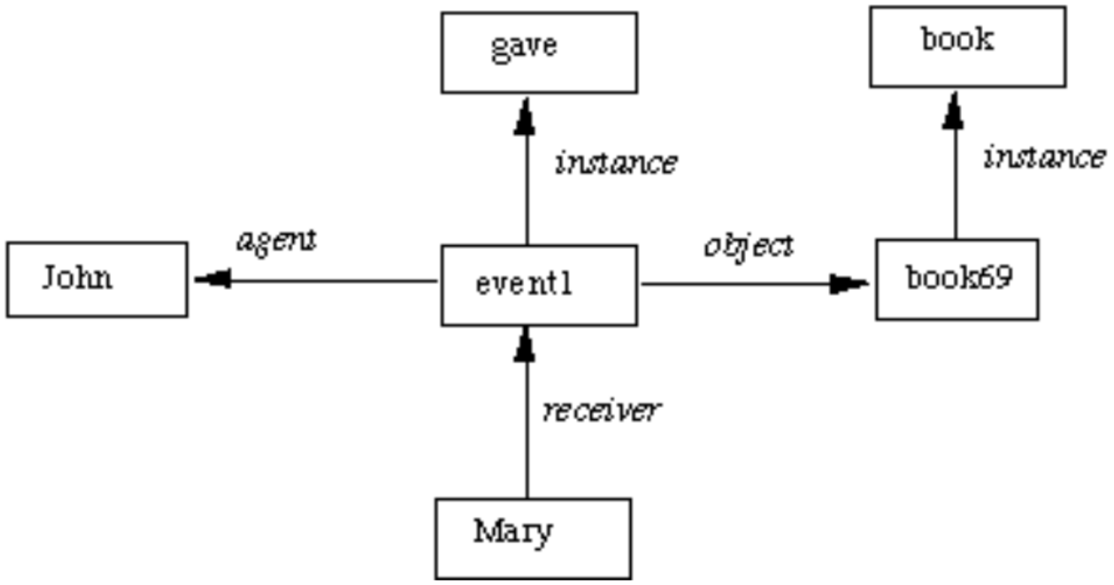


Fig. A Semantic Network for a Sentence

Weak Slot-and-Filler Structures

- Semantic Net
  - Semantic networks became popular in artificial intelligence and natural language processing only because it represents knowledge or supports reasoning.
  - These act as another alternative for predicate logic in a form of knowledge representation.
  - Semantic nets consist of nodes, links and link label.

Weak Slot-and-Filler Structures

- Frames
  - A frame is a record like structure which consists of a collection of attributes and its values to describe an entity in the world.
  - Frames are the AI data structure which divides knowledge into substructures by representing stereotypes situations. It consists of a collection of slots and slot values.
  - These slots may be of any type and sizes. Slots have names and values which are called facets.
  - Facets:
    - The various aspects of a slot is known as Facets. Facets are features of frames which enable us to put constraints on the frames.

Weak Slot-and-Filler Structures

- Frames

Ex- Frame for a book

Slots	Filters
Title	Artificial Intelligence
Author	Peter Norvig
Edition	Third Edition
Year	1996
Page	1152

Let's suppose we are taking an entity, Peter. Peter is an engineer as a profession, and his age is 25, he lives in city London, and the country is England. So following is the frame representation for this:

Slots	Filter
Name	Peter
Profession	Doctor
Age	25
Marital status	Single
Weight	78

Marwadi University

Weak Slot-and-Filler Structures

- Frames

A Simplified Frame System

Person	
isa :	Mammal
cardinality :	6,000,000,000
* handed :	Right
Adult-Male	
isa :	Person
cardinality :	2,000,000,000
* height :	5-10
ML-Baseball-Player	
isa :	Adult-Male
cardinality :	624
*height :	6-1
* bats :	equal to handed
* batting-average :	.252
* team :	
* uniform-color :	
Fielder	
isa :	ML-Baseball-Player
cardinality :	376
*batting-average :	.262
Pee-Wee-Reese	
instance :	Fielder
height :	5-10
bats :	Right
batting-average :	.309
team :	Brooklyn-Dodgers
uniform-color :	Blue
ML-Baseball-Team	
isa:	Team
cardinality :	26
* team-size :	24
* manager :	

Weak Slot-and-Filler Structures

- Frames
  - Advantages
    - The frame knowledge representation makes the programming
    - easier by grouping the related data.
    - The frame representation is comparably flexible and used by many applications in AI.
    - It is very easy to add slots for new attribute and relations.
    - It is easy to include default data and to search for missing values.
    - Frame representation is easy to understand and visualize.

Marwadi University

Weak Slot-and-Filler Structures

- Frames
  - Disadvantages
    - In frame system inference mechanism is not be easily processed.
    - Inference mechanism cannot be smoothly proceeded by frame representation.
    - Frame representation has a much generalized approach.

Strong Slot-and-Filler Structures

- Strong Slot and Filler Structures typically:
  - Represent links between objects according to more rigid rules.
  - Specific notions of what types of object and relations between them are provided.
  - Represent knowledge about common situations.

Marwadi University

Strong Slot-and-Filler Structures

- Conceptual Dependency (CD)
  - Conceptual Dependency originally developed to represent knowledge acquired from natural language input.
  - The goals of this theory are:
    - To help in the drawing of inference from sentences.
    - To be independent of the words used in the original input.
  - CD provides:
    - a structure into which nodes representing information can be placed
    - a specific set of primitives
    - at a given level of granularity.
  - Sentences are represented as a series of diagrams depicting actions using both abstract and real physical situations.
    - The agent and the objects are represented
    - The actions are built up from a set of primitive acts which can be modified by tense.

Strong Slot-and-Filler Structures

- Conceptual Dependency (CD)

Six primitive conceptual categories provide *building blocks* which are the set of allowable dependencies in the concepts in a sentence:

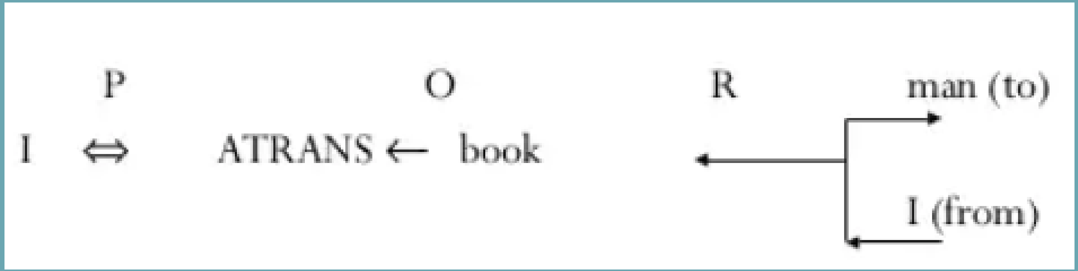
PP
-- Real world objects.
ACT
-- Real world actions.
PA
-- Attributes of objects.
AA
-- Attributes of actions.
T
-- Times.
LOC
-- Locations.

ATRANS
-- Transfer of an abstract relationship. *e.g. give.*
PTRANS
-- Transfer of the physical location of an object. *e.g. go.*
PROPEL
-- Application of a physical force to an object. *e.g. push.*
MTRANS
-- Transfer of mental information. *e.g. tell.*
MBUILD
-- Construct new information from old. *e.g. decide.*
SPEAK
-- Utter a sound. *e.g. say.*

ATTEND
-- Focus a sense on a stimulus. *e.g. listen, watch.*
MOVE
-- Movement of a body part by owner. *e.g. punch, kick.*
GRASP
-- Actor grasping an object. *e.g. clutch.*
INGEST
-- Actor ingesting an object. *e.g. eat.*
EXPEL
-- Actor getting rid of an object from body. *e.g. ????.*

Strong Slot-and-Filler Structures

- Conceptual Dependency (CD)



O – for the object case relation

R – for the recipient case relation

P – for past tense

D - destination

It should be noted that this representation is same for different saying with same meaning. For example

- I gave the man a book,
- The man got book from me,
- The book was given to man by me etc.

Strong Slot-and-Filler Structures

- Conceptual Dependency (CD)

John ran

$$\text{John} \overset{P}{\Leftrightarrow} \text{PTRANS}$$

John pushed the bike

$$\text{John} \Leftrightarrow \text{PROPEL} \overset{o}{\leftarrow} \text{bike}$$

John is Doctor

$$\text{John} \leftrightarrow \text{doctor}$$

Marwadi University

Strong Slot-and-Filler Structures

- Script
  - A script is a structure that prescribes a set of circumstances which could be expected to follow on from one another.
  - It is similar to a thought sequence or a chain of situations which could be anticipated.
  - It could be considered to consist of a number of slots or frames but with more specialized roles.
- Scripts are beneficial because:
  - Events tend to occur in known runs or patterns.
  - Causal relationships between events exist.
  - Entry conditions exist which allow an event to take place
  - Prerequisites exist upon events taking place. E.g. when a student progresses through a degree scheme or when a purchaser buys a house.

Strong Slot-and-Filler Structures

- # Components of Script are:
  - Entry Conditions
    - these must be satisfied before events in the script can occur.
  - Results
    - Conditions that will be true after events in script occur.
  - Props
    - Slots representing objects involved in events.
  - Roles
    - Persons involved in the events.
  - Track
    - Variations on the script. Different tracks may share components of the same script.
  - Scenes
    - The sequence of events that occur. Events are represented in conceptual dependency form.

Strong Slot-and-Filler Structures

- Scripts are useful in describing certain situations such as robbing a bank. This might involve:
  - Getting a gun.
  - Hold up a bank.
  - Escape with the money.

Here the *Props* might be

- Gun, *G*.
- Loot, *L*.
- Bag, *B*
- Get away car, *C*.

The *Roles* might be:

- Robber, *S*.
- Cashier, *M*.
- Bank Manager, *O*.
- Policeman, *P*.

The *Entry Conditions* might be:

- *S* is poor.
- *S* is destitute.

The *Results* might be:

- *S* has more money.
- *O* is angry.
- *M* is in a state of shock.
- *P* is shot.

Strong Slot-and-Filler Structures

- Script

Script: ROBBERY *Track: Successful Snatch*

Props:
G = Gun,
L = Loot,
B= Bag,
C = Get away car.

Roles:
R = Robber,
M = Cashier,
O = Bank Manager,
P = Policeman.

Entry Conditions:
R is poor.
R is destitute.

Results:
R has more money.
O is angry.
M is in a state of shock.
P is shot.

Scene 1: Getting a gun

R PTRANS R into Gun Shop
R MBUILD R choice of G
R MTRANS choice.
R ATRANS buys G

(go to scene 2)

Strong Slot-and-Filler Structures

- Script

Scene 2 Holding up the bank

R PTRANS R into bank
R ATTEND eyes M, O and P
R MOVE R to M position
R GRASP G
R MOVE G to point to M
R MTRANS "Give me the money or ELSE" to M
P MTRANS "Hold it Hands Up" to R
R PROPEL shoots G
P INGEST bullet from G
M ATRANS L to M
M ATRANS L puts in bag, B
M PTRANS exit
O ATRANS raises the alarm

(go to scene 3)

Scene 3: The getaway

M PTRANS C

Marwadi University

Strong Slot-and-Filler Structures

- Script
- Advantages of Scripts:
  - Ability to predict events.
  - A single coherent interpretation may be build up from a collection of observations.
- Disadvantages:
  - Less general than frames.
  - May not be suitable to represent all kinds of knowledge.

Strong Slot-and-Filler Structures

- # CYC
  - What is CYC?
    - An ambitious attempt to form a very large knowledge base aimed at capturing commonsense reasoning.
    - Initial goals to capture knowledge from a hundred randomly selected articles in the EnCYClopedia Britannica.
    - Both Implicit and Explicit knowledge encoded.
    - Emphasis on study of underlying information
  - Example: Suppose we read that Wellington learned of Napoleon's death
  - Then we (humans) can conclude Napoleon never new that Wellington had died.

Statistical Reasoning

- Uncertainty:
  - We have learned knowledge representation using first-order logic and propositional logic with certainty, which means we were sure about the predicates.
  - With this knowledge representation, we might write A→B, which means if A is true then B is true, but consider a situation where we are not sure about whether A is true or not then we cannot express this statement, this situation is called uncertainty.
- So to represent uncertain knowledge, where we are not sure about the predicates, we need uncertain reasoning or probabilistic reasoning.
- Causes of uncertainty:
  - Information occurred from unreliable sources.
  - Experimental Errors
  - Equipment fault
  - Temperature variation
  - Climate change.