Uncertain knowledge

↳ Probability

↳ Probabilistic logic / Probabilistic reasoning

→ **Probabilistic reasoning** is using logic and probability to handle uncertain situations.

→ The aim of a probabilistic logic / probabilistic reasoning is to combine the capacity of probability theory to handle uncertainity with the capacity of deductive logic to exploit structure of formal argument.

→ In situations where " the relevant world is random" or appears to be random because of poor representation" or "not random but our program can not access large database", probabilistic reasoning is to be applied.

→ One has to apply probabilistic reasoning in deciding about the next card to play in a game of cards or in diagnosing the illness from the symptoms. These are random world.

→ Uncertainities can arise from an inability to predict outcomes due to unreliable, vague, incomplete or inconsistent knowledge

Probability

→ The chance that something will happen.

↳ real number in the range 0 to 1.

→ $P(A) = 0$ indicates total uncertainity i.e. there is no chance that a particular event A will occur.

→ $P(A) = 1$ indicates total certainity i.e. event A is certain to occur.

* Random variables

↳ variables that would occur in KB.
↳ These values might represent the possible outcomes of one experiment or potential values of a quantity whose value is uncertain.

For e.g.
The possible outcomes for one fair coin toss can be described using the following random variables:

$$X = \begin{cases} head, \\ tail \end{cases}$$

Random variable's domain
↳ Boolean → $x = \{T, F\}$
↳ Discrete → Distinct values that a random variable have. $x = \{d_1, d_2, d_3, d_4\}$
↳ Continuous → $x = \{1, 2, 3, \ldots \infty\}$

- Types of probability

1. Prior probability / unconditional probability

The prior / unconditional probability $P(a)$ is the probability ~~pea~~ of a to be true.

E.g.
$P(a) = 0.7$
↳ probability of a to be true = 0.7

$P(weather = sunny) = 0.72$
$P(weather = rain) = 0.1$

2. Posterior probability / conditional probability

The posterior / conditional probability $P(a|b)$ is the probability of a to be true given the condition b.

e.g.

$P(\text{cavity}/\text{toothache}) = 0.8$ means if a patient have toothache and no other information is yet available, the the probability of patient's having the cavity is 0.8.

$$\left\{\begin{array}{l} P(a/b) = \dfrac{P(a \wedge b)}{P(b)} \;;\; P(b) \neq 0 \\ P(a \wedge b) = P(a/b) \cdot P(b) = P(b/a) \cdot P(a) \end{array}\right\}$$

Q. In a group of 100 sports car buyers, 40 bought alarm system, 30 purchased bucket seats, and 20 purchased an alarm system and bucket seats. If a car buyer chosen at random bought an alarm system. what is the probability they also bought bucket seats?

Soln

let us define,

A = purchasing alarm system , B = purchasing bucket seats

Then,

$P(A) = 40/100 = 0.4$

$P(B) = 30/100 = 0.3$

$P(A \wedge B) = 20/100 = 0.2$

$$P(B/A) = \frac{P(A \wedge B)}{P(A)} = \frac{0.2}{0.4} = 0.5$$

∴ The probability that a buyer bought bucket seats given that they purchased an alarm system is 50%.

Probability Distribution

A probability distribution is a table or a mathematical function that provides the probabilities of occurrence of different possible outcomes in an experiment.

→ Probability distribution gives values for all possible assignments.

E.g.

If P(weather = sunny) = 0.72, P(weather = rain) = 0.1, P(weather = cloudy) = 0.08 & P(weather = snow) = 0.1 then

$$P(\text{weather}) = (0.72, 0.1, 0.08, 0.1)$$

If two coins are tossed simultaneously, then the probability distribution of getting head !

→ If two coins are being tossed, there can be either 0 heads, 1 head or 2 heads.

Sample space S = {HH, HT, TH, TT}

The prob. distribution of getting heads can be shown as:

x	0	1	2
$p(x)$	$1/4$	$1/4$	$1/4$

If a coin is tossed, the probability distribution of getting tails.

x	0	1
$p(x)$	$1/2$	$1/2$

→ {If a coin is tossed, either 0 tail or 1 tail can be obtained}

Joint Probability Distribution

Joint probability is a statistical measure that calculates the likelihood of two events occurring together and at the same point in time.
Joint probability is the probability of event Y occurring at the same time that event X occurs.
For e.g.
If we have two random variables: weather and cavity with set of domain for weather = {rainy, sunny, cloudy, snow} and cavity = {True, false} the P(weather, cavity) = 4×2 matrix of values

Weather =	sunny	rainy	cloudy	snow
cavity = true	0.144	0.02	0.016	0.02
cavity = false	0.576	0.08	0.064	0.08

Inference using full joint probability distribution

We use the full joint distribution as the knowledge base from which answers to all questions may be derived. The probability of a proposition is equal to the sum of the probabilities of the atomic events in which it holds.

$$P(a) = \sum P(e_i)$$

Therefore, given a full joint distribution that specifies the probabilities of all the atomic events, one can compute the probability of any proposition.

E.g.
The full joint distribution is the following 2×2×2 table.

	toothache		¬toothache	
	Gum Problem	¬Gum Problem	Gum Problem	¬Gum Problem
Cavity	0.108	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

$P(\text{cavity or toothache}) = 0.108 + 0.012 + 0.072 + 0.08 + 0.016 + 0.064$
$= 0.28$

Marginalization or summing out

Distribution over Y can be obtained by summing out all the other variables from any joint distribution containing Y. This process is called marginalization.

$P(Y) = \sum P(Y, z)$
↳ No. of variables where Y seems to be true.

E.g.
From above table

$P(\text{cavity}) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2$
$P(\neg GP) = 0.012 + 0.064 + 0.008 + 0.576 = 0.66$
$P(\neg \text{toothache}) = 0.072 + 0.008 + 0.144 + 0.576 = 0.8$
$P(\text{cavity}, \neg\text{toothache}) = 0.072 + 0.008 = 0.08$

$P(Y) = \sum P(Y, z)$
$P(Y, z) = P(Y/z) \cdot P(z)$
Therefore, for any set of variables Y & Z:
$P(Y) = \sum P(Y/z) \cdot P(z)$
↳ This rule is the Conditioning rule.

Calculating conditional probability,

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)}$$

$$= \frac{(0.016 + 0.064)}{(0.108 + 0.012 + 0.016 + 0.064)}$$

$$= 0.4$$

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)}$$

$$= \frac{(0.108 + 0.012)}{(0.108 + 0.012 + 0.016 + 0.064)}$$

$$= 0.6$$

Independence

A and B are independent iff

$$P(A \mid B) = P(A) \text{ or } P(B \mid A) = P(B) \text{ or } P(A, B) = P(A) \cdot P(B)$$

e.g.

$P(\text{toothache, Gum problem, cavity, weather})$
$= P(\text{toothache, Gum problem, cavity}) \cdot P(\text{weather})$

Here, weather is independent of other three variables.

Imp

Baye's Theorem / Rule

Baye's theorem is a way to apply conditional probability for prediction. Conditional probability is the probability of an event happening, given that it has some relationship to one or more other events.
Mathematically, Baye's theorem is defined as:

$$P(b/a) = \frac{P(a/b) * P(b)}{P(a)}$$

Proof:

We know that,

$$P(b/a) = \frac{P(b \wedge a)}{P(a)}$$

$$P(b \wedge a) = P(b/a) * P(a) \quad \text{—— (I)}$$

Similarly,

$$P(a/b) = \frac{P(a \wedge b)}{P(b)}$$

$$P(a \wedge b) = P(a/b) * P(b) \quad \text{—— (II)}$$

from eq^n (I) & (II)

$$P(b/a) * P(a) = P(a/b) * P(b) \quad [P(a \wedge b) \approx P(b \wedge a)]$$

$$P(b/a) = \frac{P(a/b) * P(b)}{P(a)}$$

→ Bayes theorem provides a way to revise existing predictions or theories (update probabilities) given new or additional evidence. This, in turn, makes the predictions more accurate.

Applications of Bayes Theorem

1. Medical science :
Baye's rule is used for predicting a particular disease based on the symptoms and physical condition of the patient.

2. Weather forcasting:
Baye's rule is a powerful algorithm for predictive modeling weather forcast.

3. Robotics :
Baye's rule is used to calculate the probability of a robot's next steps given the steps the robot has already executed.

4. Finance :
Baye's theorem can be used to rate the risk of lending money to potential borrowers.

Q: A doctor knows that the disease meningitis causes the patient to have a stiff neck 50% of the time. The doctor also knows that the probability that a patient has meningitis is 1/50,000, and the probability that any patient has a stiff neck is 1/20. Now find the probability that a patient with stiff neck has meningitis.

Ans

Let s be the proposition that the patient has a stiff neck and m be the proposition that the patient has meningitis.

Here, we are given

$P(S/m) = 0.5$

$P(m) = 1/50,000$

$P(s) = 1/20$

$P(m/s) = ?$

Now using Baye's rule

$$P(m/s) = \frac{P(s/m) * P(m)}{P(s)} = \frac{0.5 * 1/50,000}{1/20} = 0.0002$$

Hence, the probability that a patient with a stiff neck has meningitis is 0.0002.

Q. 1% of women at age fourty who participates in routine screening have breast cancer 80% of women with breast cancer will get positive mammographies. 9.6% of women without breast cancer will also get positive mammographies. A women in this age group has a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

Soln

Let, B be the proposition that women has breast cancer.
B' be the proposition that women without breast cancer.
+ be the proposition that women getting positive mammographies.

Here, we are given

$P(B) = 0.01$

$P(+/B) = 0.8$

$P(+/B') = 0.096$

$P(B/+) = ?$

We have,

$$P(B/+) = \frac{P(+/B) * P(B)}{P(+)}$$

Here $P(B)$, $P(+/B)$ & $P(+/B')$ are known. $P(+)$ is needed to find $P(B/+)$.

$$P(+) = P(+/B) * P(B) + P(+/B') * P(B')$$
$$= (0.8 * 0.01) + (0.096 * 0.99)$$
$$= 0.1030$$

$$\therefore P(B/+) = \frac{0.8 * 0.01}{0.1030} = 0.07767$$

T.U. 074

Q. Consider in Nepal, 51% of adults are males and rests are females, Consider one adult is randomly selected for a survey of drinking alcohol. It is found that 15% of males drink alcohol where as 2% of female drink alcohol. Now find the probability that the selected adult is male.

Soln

let M be the adult males.
F be the adult females.
A be the adult who is drinking alcohol.

Here, we are given,

$P(M) = 0.51$

$P(F) = 0.49$

$P(A/M) = 0.15$

$P(A/F) = 0.02$

$P(M/A) = ?$

Now,

$P(A) = P(M) * P(A/M) + P(F) * P(A/F)$

$= 0.51 * 0.15 + 0.49 * 0.02$

$= 0.0863$

By using Baye's rule

$$P(M/A) = \frac{P(A/M) * P(M)}{P(A)}$$

$$= \frac{0.15 * 0.51}{0.0863}$$

$= 0.8864$

Q. Two different suppliers, A and B, provid a manufacturer with the same part. All supplies of this part are kept in a large bin. In the past, 5% of the parts supplied by A and 9% of the parts supplied by B have been defective. A supplies four times as many parts as B. suppose you reach into the ~~been~~ bin and select a part, and find it is non-defective. What is the prob. that it was supplied by A ?

Sol^n

Let,

A is the parts supplied by A.
B is the parts supplied by B.
D is the non-defective parts.

Given,

$P(D/A) = 0.095 \quad (1-0.05)$

$P(D/B) = 0.91$ $(1-0.08)$

$P(A) = 0.8$

$P(B) = 0.2$

$P(A/D) = ?$

Now,

$P(D) = P(D/A) * P(A) + P(D/B) * P(B)$

$= 0.95 * 0.8 + 0.91 * 0.2$

$= 0.942$

By using Baye's rule

$$P(A/D) = \frac{P(D/A) * P(A)}{P(D)} = \frac{0.95 * 0.8}{0.942} = 0.8068$$

Q. Manju is getting married tomorrow, at an outdoor ceremony in the desert. In recent years, it has rained only 5 days each year. Unfortunately, the weatherman has predicated rain for tomorrow. When it actually rains, the weatherman correctly forecasts rain 90% of the time. When it doesn't rain, he incorrectly forecasts 10% of the time. What is the probability that it will rain on the day of Manju's wedding?

Sol.

let,

Event A_1 → It rains on Manju's wedding.

Event A_2 → It does not rain on Manju's wedding.

Event B → The weatherman predicts rain.

Given,

$P(A_1) = \frac{5}{365} = 0.0136985 \approx 0.014$

$P(A_2) = \frac{360}{365} = 0.9863015 \approx 0.986$

$P(B/A_1) = 0.9$

$P(B/A_2) = 0.1$

$P(A_1/B) = ?$

Now,

$P(B) = P(A_1) * P(B/A_1) + P(A_2) * P(B/A_2)$

$= 0.014 * 0.9 + 0.986 * 0.1$

$= 0.1112$

Using Baye's rule

$P(A_1/B) = \frac{P(A_1) * P(B/A_1)}{P(B)} = \frac{0.014 * 0.9}{0.112} = 0.111$

$\therefore$ Prob. of rain on the day of Manju's wedding, given a forecast of rain by the weatherman is 0.111.

1.U 068 Q3. After your yearly checkup, the doctor has bad news and good news. The bad news is that you tested positive for a serious disease, and the test is 99 % accurate (i.e. the probability of testing positive given that you have the disease is 0.99, as the probability of testing negative if you don't have disease). The good news is that is a rare disease, striking only one in 10,000 people?

a) Why it is good news that the disease is rare?

b) What are the chances that you actually have the disease?

Sol^n

let, T^+ : test is positive , T^- : test is negative.
D : has disease $\bar{D}$: doesn't have disease

Here, we are given,

$P(T^+/D) = 0.99$

$P(T^-/D) = 0.99$

$P(D) = \frac{1}{10,000} = 0.0001$

a) As probability of having disease is very small i.e. 0.0001, so it is good news that the disease is rare.

b) we can infer,

$P(\bar{D}) = 1 - 0.0001 = 0.9999$

$P(T^+/\bar{D}) = 1 - 0.99 = 0.01$

$P(D/T^+) = ?$

Now,

$P(T^+) = P(T^+/D) * P(D) + P(T^+/\bar{D}) * P(\bar{D})$

$= 0.99 * 0.0001 + 0.01 * 0.9999$

$= 0.010098$

Now, using Baye's rule

$$P(D/T^+) = \frac{P(T^+/D) * P(D)}{P(T^+)}$$

$$= \frac{0.99 * 0.0001}{0.010098}$$

$$= 0.09804.$$

Vimp.

Bayesian Network (Belief N/w / Causal nets / Baye's N/w)

Bayesian networks are a type of probabilistic graphical model that uses Bayesian inference for probability computations.

→ Bayesian network is a probabilistic graphical model that represents a set of random variables and their conditional dependencies via a directed acyclic graph.

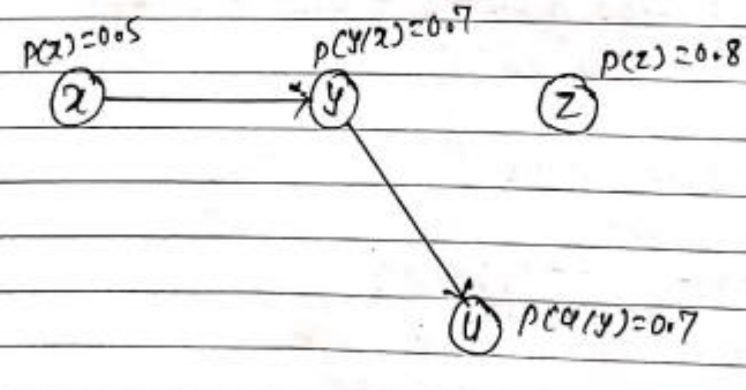
→ Nodes in the graph represent the random variables and the directed edges between nodes represent conditional dependencies.

→ The edge exists between nodes iff there exists conditional probability i.e. a link from x to y means y is dependent of x.

→ Each nodes are labelled with probability.

For eg.

$P(x) = 0.5$
$P(y/x) = 0.7$
$P(z) = 0.8$
$P(u/y) = 0.47$



Bayesian Graph.

Probabilistic Reasoning over time

Probabilistic reasoning over time involves using probabilistic models to handle uncertainty and make predictions or inferences about systems that evolve over time. These models are particularly useful in dynamic environments where the state of the system changes, and the exact knowledge of the system's state may not be available.

A **Hidden Markov Model (HMM)** is a statistical model used to represent systems with hidden states that evolve over time. It is widely used in areas like speech recognition, bioinformatics, and time-series analysis. Below is a detailed exploration of HMMs.

Mathematical Representation

An HMM is defined by the tuple: $\lambda = (\pi, A, B)$

Where:

- π: Initial state distribution.
- A: State transition probabilities.
- B: Emission probabilities.

Concepts in HMM

1. **States (Hidden States):**
    - Represent the underlying system that we cannot directly observe.
    - Examples:
        - In speech recognition: phonemes.
        - In weather modeling: "Rainy" or "Sunny."
2. **Observations:**
    - Observable data or evidence generated by the hidden states.
    - Examples:
        - In speech recognition: sound waves (audio signals).
        - In weather modeling: temperature readings.
3. **Transition Probabilities (A):**
    - Represent the probability of moving from one hidden state to another.
4. **Emission Probabilities (B):**
    - Represent the probability of an observation being generated from a hidden state.
5. **Initial Probabilities (π):**
    - Represent the probabilities of starting in each hidden state.

Example: Weather Prediction

Hidden States:

- Rainy, Sunny

Observations:

- Walk, Shop, Clean

Probabilities:

- **Initial Probabilities (π):**
  - P(Rainy) = 0.6, P(Sunny) = 0.4

- **Transition Probabilities (A):**
  - P(Rainy → Rainy) = 0.7, P(Rainy → Sunny) = 0.3
  - P(Sunny → Rainy) = 0.4, P(Sunny → Sunny) = 0.6
- **Emission Probabilities (B):**
  - P(Walk | Rainy) = 0.1, P(Shop | Rainy) = 0.4, P(Clean | Rainy) = 0.5
  - P(Walk | Sunny) = 0.6, P(Shop | Sunny) = 0.3, P(Clean | Sunny) = 0.1

Given a sequence of observations (e.g., "Walk, Shop"), the HMM can infer the most likely weather sequence (e.g., "Sunny, Rainy").

Object Based knowledge Representation system

- Frames
- semantic Network
- scripts
- Conceptual Dependencies
- OAV triplets

Frames

A frame is a data structure containing typical knowledge about concept or objects. A frame represents knowledge about real world things (or entities).

Each frame contains frame name and slots. slots Contains attributes defining frame object and values associated with the attributes. Values may be:

- a default values
- an inherited value from a higher frame
- a procedure
- a specific value

When the slots of a frame are all filled, the frame is said to be instantiated. Empty frames are sometimes called object prototypes.

E.g.

Employee

Name	Ram
Address	ktm
Salary	15,000/-
Tax	15% of salary
Gender	M
Department	

Department

DID	001
DName	HR
DLocation	Lalitpur

Semantic Network

Semantic networks can
- show natural relationship between objets/concepts.
- be used to represent declarative/descriptive knowledge.

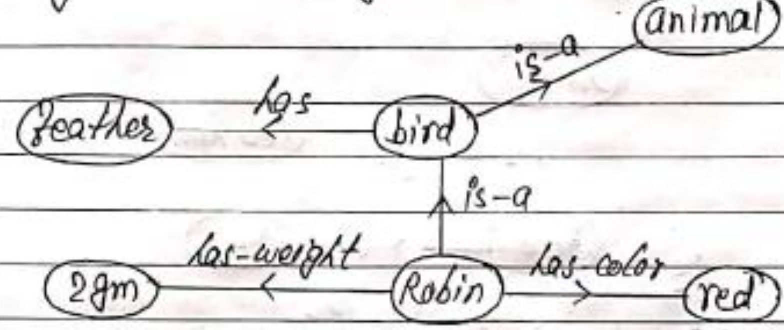
Knowledge is represented as a collection of concepts; represented by nodes. Thus, semantic networks are constructed using nodes (vertices) linked by directional lines called arcs (edges).

A node represents:
- physical object
- concept
- event
- attributes
- attribute values etc.

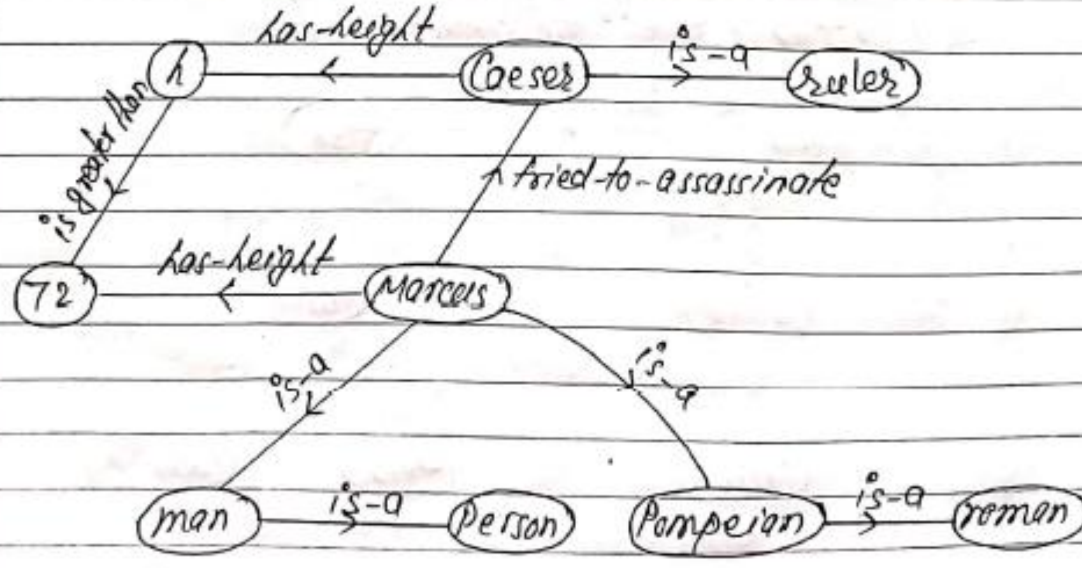
An arc represents relationship betn nodes. Relationship types are:
- Is-a : represent class/instance relationships
- Has-a : identify property relationship.

E.g.

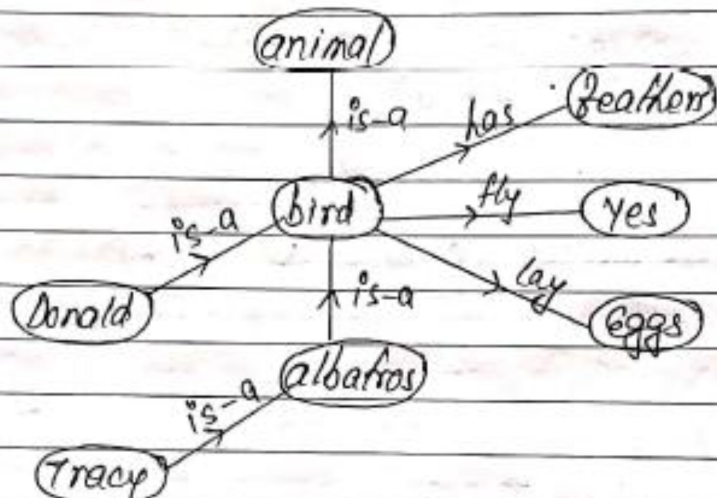
① All birds are animal.
Birds have feathers.
Robin is a bird.
Robin is red in color.
Weight of Robin is 2gm.



② All men are person.
All pompeians are roman.
Marcus, is a men, who is pompeian.
Caeser is ruler.
Marcus tried to assassinate caeser.
Marcus has height 72.
Height of marcus is less than the height of caeser.

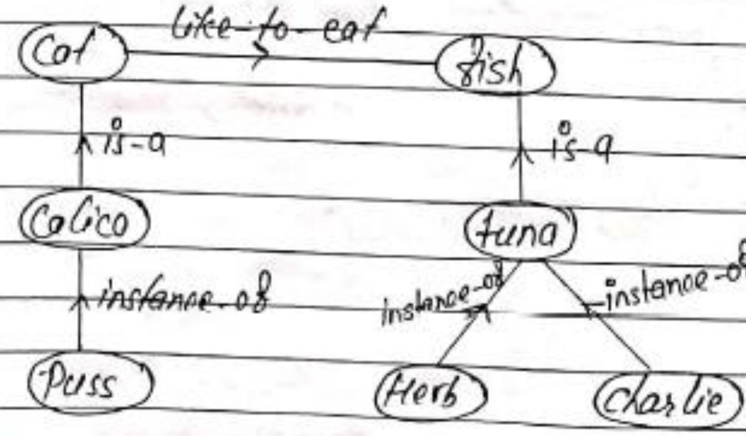


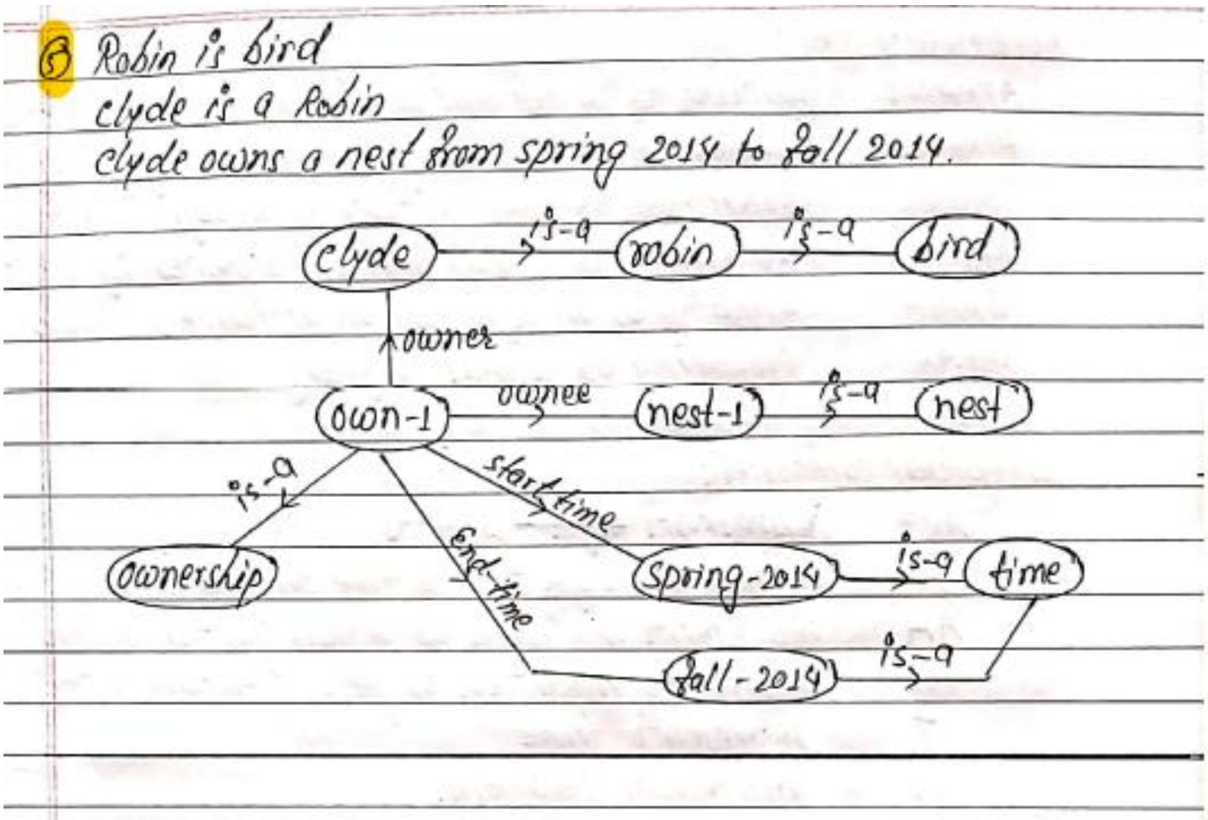
(3) Birds are animals.
Birds have feathers, fly and lay eggs.
Albatros is a bird.
Donald is a bird.
Tracy is an albatros.



(4) Puss is a calico.
Herb is a tuna.
charlie is a tuna.
All tunas are fishes.
All calicos are cats
All cats like to eat all kinds of fishes.

{is / instance-of}





Rule-based knowledge representation

In artificial intelligence (AI), **rule-based knowledge representation** is a method of encoding knowledge in the form of explicit **if-then rules**. These rules describe the relationships between facts, actions, and outcomes in a domain of interest. Rule-based systems are commonly used in **expert systems**, **decision support systems**, and **logic programming**.

1. Components of Rule-Based Systems

- **Knowledge Base:** Contains the rules in the form of **if-then statements** (also called production rules).

  Example: IF a person has fever AND a sore throat THEN the person might have the flu.

- **Inference Engine:** The reasoning mechanism that applies the rules in the knowledge base to the given data or facts to derive conclusions. There are two main inference strategies:
  - **Forward Chaining:** Starts with known facts and applies rules to derive new facts until a goal is reached.
  - **Backward Chaining:** Starts with a goal and works backward by looking for rules that can support that goal.

- **Working Memory:** A dynamic area that contains the facts or data currently under consideration.

Characteristics of Rule-Based Systems

- **Symbolic Representation:** Rules represent knowledge symbolically, making it human-readable and interpretable.
- **Modular:** Rules can often be added, removed, or modified without impacting other rules.
- **Deterministic or Non-Deterministic:** Depending on the domain, rules can lead to deterministic (predictable) or probabilistic (uncertain) outcomes.

ONTOLOGICAL BASED REPRESENTATION (SEE ON CHAPTER SEVEN)