[Chapter 1: Introduction]
Artificial Intelligence (CSC 355)

Arjun Singh Saud

Central Department of Computer Science & Information Technology
Tribhuvan University

What is Artificial Intelligence?

Artificial Intelligence is a branch of *Science* which deals with helping machines to find solutions to complex problems in a more human-like fashion. This generally involves borrowing characteristics from human intelligence, and applying them as algorithms in a computer friendly way. A more or less flexible or efficient approach can be taken depending on the requirements established, which influences how artificial the intelligent behavior appears.

AI is generally associated with *Computer Science*, but it has many important links with other fields such as *Maths*, *Psychology*, *Cognition*, *Biology* and *Philosophy*, among many others. Our ability to combine knowledge from all these fields will ultimately benefit our progress in the quest of creating an intelligent artificial being.

Different definitions of AI are given by different books/writers. These definitions can be divided into two dimensions.

| Systems that think like humans	Systems that think rationally
Systems that act like humans	Systems that act rationally

Top dimension is concerned with thought processes and reasoning, where as bottom dimension addresses the behavior.

The definition on the left measures the success in terms of fidelity of human performance, whereas definitions on the right measure an ideal concept of intelligence, which is called rationality.

Human-centered approaches must be an empirical science, involving hypothesis and experimental confirmation. A rationalist approach involves a combination of mathematics and engineering.

Acting Humanly: The Turing Test Approach

The Turing test, proposed by Alan Turing (1950) was designed to convince the people that whether a particular machine can think or not. He suggested a test based on indistinguishability from undeniably intelligent entities- human beings.
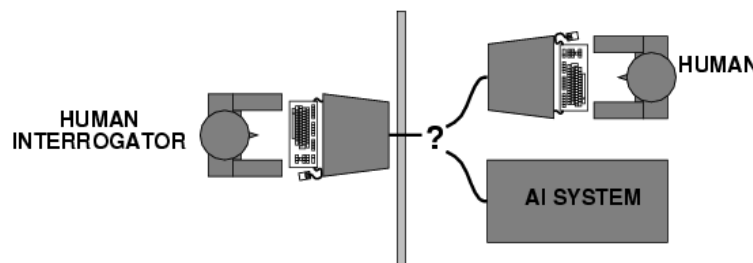
The computer passes the test if a human interrogator after posing some written questions, can not tell whether the written response come from human or not.

To pass a Turing test, a computer must have following capabilities:

- ➢ Natural Language Processing: Must be able to communicate in English successfully
- ➢ Knowledge representation: To store what it knows and hears.
- ➢ Automated reasoning: Answer the Questions based on the stored information.
- ➢ Machine learning: Must be able to adapt in new circumstances.

Turing test avoid the physical interaction with human interrogator. Physical simulation of human beings is not necessary for testing the intelligence.

The total Turing test includes video signals and manipulation capability so that the interrogator can test the subject's perceptual abilities and object manipulation ability. To pass the total Turing test computer must have following additional capabilities:

- ➢ Computer Vision: To perceive objects
- ➢ Robotics: To manipulate objects and move

Thinking Humanly: Cognitive modeling approach

Make the machines with mind. Cognition means the action or process of acquiring knowledge and understanding through thought, experience and senses. To make a machine that think like human brain, scientific theories of internal brain activities (cognitive model) are required. Two ways of doing this is:

- − Predicting and testing human behavior (cognitive science)
- − Identification from neurological data (Cognitive neuroscience)

Once we have precise theory of mind, it is possible to express the theory as a computer program. But unfortunately until up to now there is no precise theory about thinking process of human brain. Therefore it is not possible to make the machines that think like human brain

Thinking rationally: The laws of thought approach

Aristotle was one of the first who attempt to codify the right thinking that is irrefutable reasoning process. He gave Syllogisms that always yielded correct conclusion when correct premises are given.

For example:

Ram is a man

Man is mortal

⇨ Ram is mortal

Let
p(x) → x is man
q(x) → x is mortal

then above statement can be written as
p(x) => q(x)   Man is mortal
p(Ram)   Ram is man

Then from modus ponens q(x) ia also true. That is Ram is mortal

This study initiated the field of logic. The logicist tradition in AI hopes to create intelligent systems using logic programming.

Problems:
It is not easy to take informal knowledge and state in the formal terms required by logical notation, particularly when knowledge is not 100% certain.

Solving problem principally is different from doing it in practice. Even problems with certain dozens of fact may exhaust the computational resources of any computer unless it has some guidance as which reasoning step to try first.

Acting Rationally: The rational Agent approach:
Agent is something that acts. Computer agent is expected to have following attributes:
- ❖ Autonomous control
- ❖ Perceiving their environment
- ❖ Persisting over a prolonged period of time
- ❖ Adapting to change
- ❖ And capable of taking on another's goal

Rational behavior means doing the right thing. The right thing is that which is expected to maximize goal achievement, given the available information. Rational Agent is one that acts so as to achieve the best outcome or, when there is uncertainty, the best expected outcome. In this approach the emphasis is given to correct inferences.

One way to act rationally is to reason logically to the conclusion and act on that conclusion. On the other hand there are also some ways of acting rationally that can not be said to involve inference. For Example, recoiling from a host stove is a reflex action that is usually more successful than a slower action taken after careful deliberation.

Advantages:
- ✓ It is more general than laws of thought approach, because correct inference is just one of several mechanisms for achieving rationality.

✓ It is more amenable to scientific development than are approaches based on human behavior or human thought because the standard of rationality is clearly defined and completely general.

Applications of AI

Game playing

You can buy machines that can play master level chess for a few hundred dollars. There is some AI in them, but they play well against people mainly through brute force computation--looking at hundreds of thousands of positions. To beat a world champion by brute force and known reliable heuristics requires being able to look at 200 million positions per second.

Speech recognition

In the 1990s, computer speech recognition reached a practical level for limited purposes. Thus United Airlines has replaced its keyboard tree for flight information by a system using speech recognition of flight numbers and city names. It is quite convenient. On the the other hand, while it is possible to instruct some computers using speech, most users have gone back to the keyboard and the mouse as still more convenient.

Understanding natural language

Just getting a sequence of words into a computer is not enough. Parsing sentences is not enough either. The computer has to be provided with an understanding of the domain the text is about, and this is presently possible only for very limited domains.

Computer vision

The world is composed of three-dimensional objects, but the inputs to the human eye and computers' TV cameras are two dimensional. Some useful programs can work solely in two dimensions, but full computer vision requires partial three-dimensional information that is not just a set of two-dimensional views. At present there are only limited ways of representing three-dimensional information directly, and they are not as good as what humans evidently use.

Expert systems

A ``knowledge engineer'' interviews experts in a certain domain and tries to embody their knowledge in a computer program for carrying out some task. How well this works depends on whether the intellectual mechanisms required for the task are within the present state of AI. When this turned out not to be so, there were many disappointing results. One of the first expert systems was MYCIN in 1974, which diagnosed bacterial infections of the blood and suggested treatments. It did better than medical students or practicing doctors, provided its limitations were observed. Namely, its ontology included bacteria, symptoms, and treatments and did not include patients, doctors, hospitals, death, recovery, and events occurring in time. Its interactions depended on a single patient being considered. Since the experts consulted by the knowledge engineers knew about patients, doctors, death, recovery, etc., it is clear that the knowledge engineers forced what the experts told them into a predetermined framework. In the present state of AI,

---

this has to be true. The usefulness of current expert systems depends on their users having common sense.

Heuristic classification

One of the most feasible kinds of expert system given the present knowledge of AI is to put some information in one of a fixed set of categories using several sources of information. An example is advising whether to accept a proposed credit card purchase. Information is available about the owner of the credit card, his record of payment and also about the item he is buying and about the establishment from which he is buying it (e.g., about whether there have been previous credit card frauds at this establishment).

Foundations of AI

Different fields have contributed to AI in the form of ideas, viewpoints and techniques.

*Philosophy***:**
Logic, reasoning, mind as a physical system, foundations of learning, language and rationality.

*Mathematics***:**
Formal representation and proof algorithms, computation, undecidability, intractability, probability.

*Psychology***:**
adaptation, phenomena of perception and motor control.

*Economics***:**
formal theory of rational decisions, game theory.

*Linguistics***:**
Knowledge representation, grammar

*Neuroscience***:**
Physical substrate for mental activities

*Control theory***:**
Homeostatic systems, stability, optimal agent design

Brief history of AI

What happened after WWII?

− 1943: Warren Mc Culloch and Walter Pitts: a model of artificial boolean neurons to perform computations.
  − First steps toward connectionist computation and learning (Hebbian learning).
  − Marvin Minsky and Dann Edmonds (1951) constructed the first neural network computer

− 1950: Alan Turing's "Computing Machinery and Intelligence"
  − First complete vision of AI.

The birth of AI (1956)

- Dartmouth Workshop bringing together top minds on automata theory, neural nets and the study of intelligence.
  − Allen Newell and Herbert Simon: The logic theorist (first nonnumeric thinking program used for theorem proving)
  − For the next 20 years the field was dominated by these participants.

Great expectations (1952-1969)

− Newell and Simon introduced the General Problem Solver.
  − Imitation of human problem-solving
− Arthur Samuel (1952-)investigated game playing (checkers) with great success.
− John McCarthy(1958-) :
  − Inventor of Lisp (second-oldest high-level language)
  − Logic oriented, Advice Taker (separation between knowledge and reasoning)

− Marvin Minsky (1958 -)
  − Introduction of microworlds that appear to require intelligence to solve: e.g. blocks-world.
  − Anti-logic orientation, society of the mind.

Collapse in AI research (1966 - 1973)

− Progress was slower than expected.
  − Unrealistic predictions.
− Some systems lacked scalability.
  − Combinatorial explosion in search.
− Fundamental limitations on techniques and representations.
  − Minsky and Papert (1969) Perceptrons.

AI revival through knowledge-based systems (1969-1970)

− General-purpose vs. domain specific
  - E.g. the DENDRAL project (Buchanan et al. 1969)
        First successful knowledge intensive system.

- Expert systems
    - MYCIN to diagnose blood infections (Feigenbaum et al.)
        - Introduction of uncertainty in reasoning.

- Increase in knowledge representation research.
    - Logic, frames, semantic nets, …

AI becomes an industry (1980 - present)
- R1 at DEC (McDermott, 1982)
- Fifth generation project in Japan (1981)
- American response …

    Puts an end to the AI winter.

Connectionist revival (1986 - present): (Return of Neural Network)
- Parallel distributed processing (RumelHart and McClelland, 1986); backprop.

AI becomes a science (1987 - present)

- In speech recognition: hidden markov models
- In neural networks
- In uncertain reasoning and expert systems: Bayesian network formalism

The emergence of intelligent agents (1995 - present)
- The whole agent problem: "How does an agent act/behave embedded in real environments with continuous sensory inputs"

Knowledge and its importance

Knowledge is information that helps us to solve problems in particular domain, to predict what will happen next and to explain why and how something has happened? Knowledge contains facts, procedures and judgmental rules. Knowledge is important in AI for making intelligent machines. Key issues confronting the designer of AI system are:

Knowledge acquisition: Gathering the knowledge from the problem domain to solve the AI problem.

Knowledge representation: Expressing the identified knowledge into some knowledge representation language such as propositional logic, predicate logic etc.

Knowledge manipulation: Large volume of knowledge has no meaning until up to it is processed to deduce the hidden aspects of it. Knowledge is manipulated to draw conclusions from knowledgebase.

[chapter 2: Problem Solving]
Artificial Intelligence (CSC 355)

Arjun Singh Saud

Central Department of Computer Science & Information Technology
Tribhuvan University

Problem Solving:

Problem solving, particularly in artificial intelligence, may be characterized as a systematic search through a range of possible actions in order to reach some predefined goal or solution. Problem-solving methods divide into special purpose and general purpose. A special-purpose method is tailor-made for a particular problem and often exploits very specific features of the situation in which the problem is embedded. In contrast, a general-purpose method is applicable to a wide variety of problems. One general-purpose technique used in AI is means-end analysis—a step-by-step, or incremental, reduction of the difference between the current state and the final goal.

Four general steps in problem solving:
- Goal formulation
  - What are the successful world states
- Problem formulation
  - What actions and states to consider given the goal
- Search
  - Determine the possible sequence of actions that lead to the states of known values and then choosing the best sequence.
- Execute
  - Give the solution perform the actions.

Problem formulation:

A problem is defined by:

- An initial state: State from which agent start
- Successor function: Description of possible actions available to the agent.
- Goal test: Determine whether the given state is goal state or not
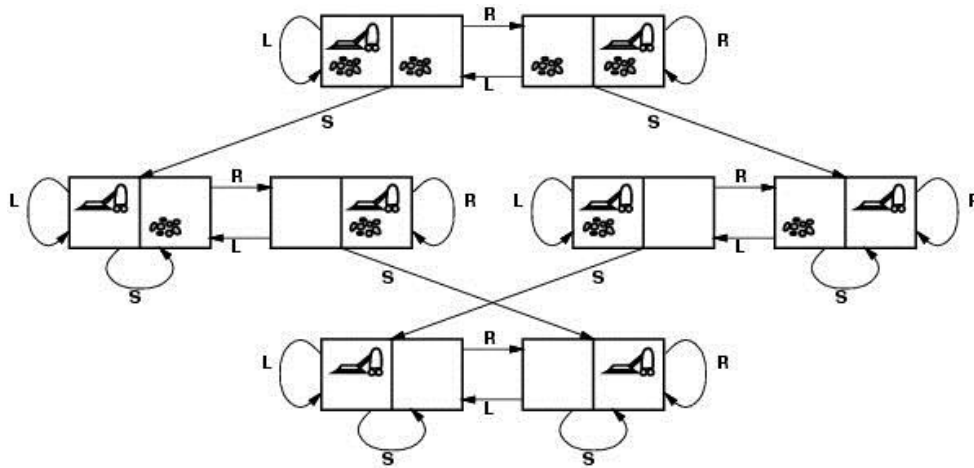- Path cost: Sum of cost of each path from initial state to the given state.

A solution is a sequence of actions from initial to goal state. Optimal solution has the lowest path cost.

State Space representation

The state space is commonly defined as a directed graph in which each node is a state and each arc represents the application of an operator transforming a state to a successor state.

   A **solution** is a path from the initial state to a goal state.

State Space representation of Vacuum World Problem:

States?? two locations with or without dirt: 2 x 2^2=8 states.
Initial state?? Any state can be initial
Actions?? {*Left*, *Right*, *Suck*}
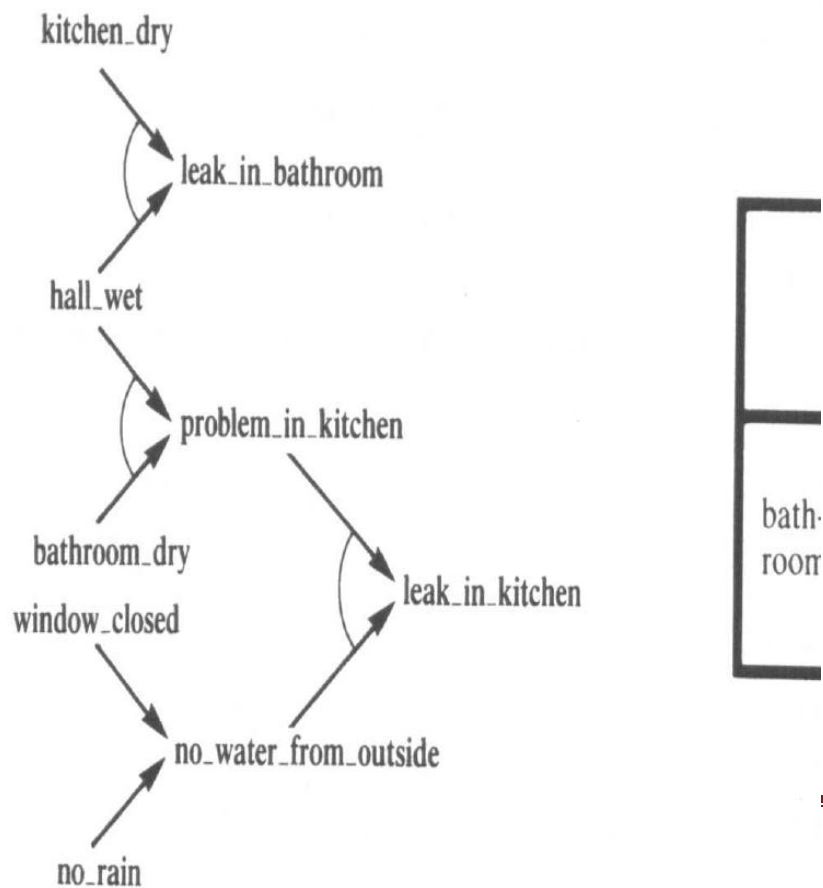Goal test?? Check whether squares are clean.
Path cost?? Number of actions to reach goal.

For following topics refer Russell and Norvig's Chapter 3 from pages 87-96.
     Problem Types: Toy Problems & Real World Problems
     Well Defined Problems

<u>**Water Leakage Problem:**</u>

If

    hall _wet and kitchen_dry
then

    leak_in_bathroom
If

    hall_wet and bathroom_dry
then

    problem_in_kitchen
If

    window_closed or no_rain
then

    no_water_from_outside

Production System:

A **production system** (or **production rule system**) is a computer program typically used to provide some form of artificial intelligence, which consists primarily of a set of rules about behavior. These rules, termed **productions**, are a basic representation found useful in automated planning, expert systems and action selection. A production system provides the mechanism necessary to execute productions in order to achieve some goal for the system.

Productions consist of two parts: a sensory precondition (or "IF" statement) and an action (or "THEN"). If a production's precondition matches the current state of the world, then the production is said to be *triggered*. If a production's action is executed, it is said to have *fired*. A production system also contains a database, sometimes called worcking memory, which maintains data about current state or knowledge, and a rule interpreter. The rule interpreter must provide a mechanism for prioritizing productions when more than one is triggered.

The underlying idea of production systems is to represent knowledge in the form of condition-action pairs called production rules:

If the condition C is satisfied then the action A is appropriate.

Types of production rules

Situation-action rules
    If it is raining then open the umbrella.
Inference rules
    If Cesar is a man then Cesar is a person

Production system is also called *ruled-based system*

**Architecture of Production System:**

Short Term Memory:
- Contains the description of the current state.

Set of Production Rules:
- Set of condition-action pairs and defines a single chunk of problem solving knowledge.

Interpreter:
- A mechanism to examine the short term memory and to determine which rules to fire (According to some strategies such as DFS, BFS, Priority, first-encounter etc)

Rule Base
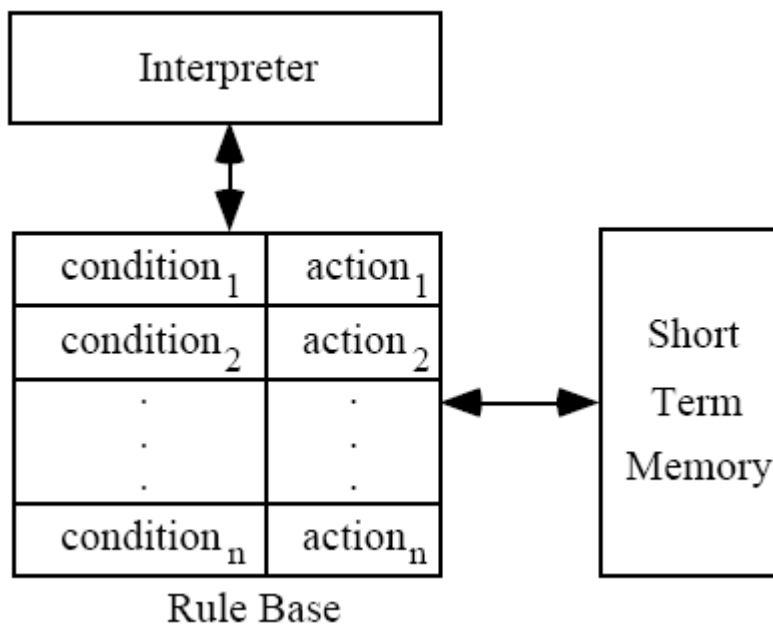
The execution of a production system can be defined as a series of recognize-act cycles: *Match* –memory contain matched against condition of production rules, this produces a subset of production called **conflict set**. *Conflict resolution* –one of the production in the conflict set is then selected, *Apply the rule.*

Consider an example:

Problem: Sorting a string composed of letters a, b & c.
Short Term Memory: cbaca
Production Set:

1. ba → ab
2. ca → ac.
3. cb → bc

Interpreter: Choose one rule according to some strategy.

Iteration #	Memory	Conflict Set	Rule fired
0	cbaca	1, 2, 3	1
1	cabca	2	2
2	acbca	2, 3	2
3	acbac	1, 3	1
4	acabc	2	2
5	aacbc	3	3
6	aabcc	ø	halt

Production System: The water jug problem

Problem:

There are two jugs, a 4-gallon one and a 3-gallon one. Neither jug has any measuring markers on it. There is a pump that can be used to fill the jugs with water.
How can you get exactly n (0, 1, 2, 3, 4) gallons of water into one of the two jugs ?

Solution Paradigm:

- build a simple production system for solving this problem.
- represent the problem by using the state space paradigm.

State = (x, y); where: x represents the number of gallons in the 4-gallon jug; y represents the number of gallons in the 3-gallon jug. $x \in \{0, 1, 2, 3, 4\}$ and $y \in \{0, 1, 2, 3\}$.

The initial state represents the initial content of the two jugs.

For instance, it may be (2, 3), meaning that the 4-gallon jug contains 2 gallons of water and the 3-gallon jug contains three gallons of water.

The goal state is the desired content of the two jugs.

The left hand side of a production rule indicates the state in which the rule is applicable and the right hand side indicates the state resulting after the application of the rule.

For instance;

(x, y) such that $x < 4 \rightarrow (4, y)$ represents the production
If the 4-gallon jug is not full then fill it from the pump.

The rule base contains the following production rules:

1. (x, y) such that $x < 4 \rightarrow (4, y)$; Fill the 4-gallon jug from pump
2. (x, y) such that $y < 3 \rightarrow (x, 3)$; Fill the 3-gallon jug from pump
3. (x, y) such that $x > 0 \rightarrow (0, y)$; Empty the 4-gallon jug on the ground
4. (x, y) such that $y > 0 \rightarrow (x, 0)$; Empty the 3-gallon jug on the ground
5. (x, y) such that $x + y \geq 4, x < 4, y > 0 \rightarrow (4, y - (4 - x))$
    ; Completely fill the 4-gallon jug from the 3-gallon jug
6. (x, y) such that $x + y \geq 3, x > 0, y < 3 \rightarrow (x - (3 - y), 3)$
    ; Completely fill the 3-gallon jug from the 4-gallon jug
7. (x, y) such that $x + y \leq 4, y > 0 \rightarrow (x+y, 0)$
    ; Empty the 3-gallon jug into the 4-gallon jug
8. (x, y) such that $x + y \leq 3, x > 0 \rightarrow (0, x + y)$
    ; Empty the 4-gallon jug into the 3-gallon jug

The short term memory contains the current state (x, y).

Let us consider the initial situation (0, 0) and the goal situation (n, 2)

short term memory :   (0, 0)
1. Match: 1, 2.   2. Conflict resolution: select rule 2.   3. Apply the rule
short term memory becomes (0, 3)
1. Match: 1, 4, 7   2. Conflict resolution: select rule 7   3. Apply the rule
short term memory becomes (3, 0)
1. Match:1, 2, 3, 6, 8   2. Conflict resolution: select rule 2   3. Apply the rule
short term memory becomes (3, 3)
1. Match: 1, 3, 4, 5   2. Conflict resolution: select rule 5   3. Apply the rule
short term memory becomes (4, 2)   Goal achieved

The sequence of the applied rules:
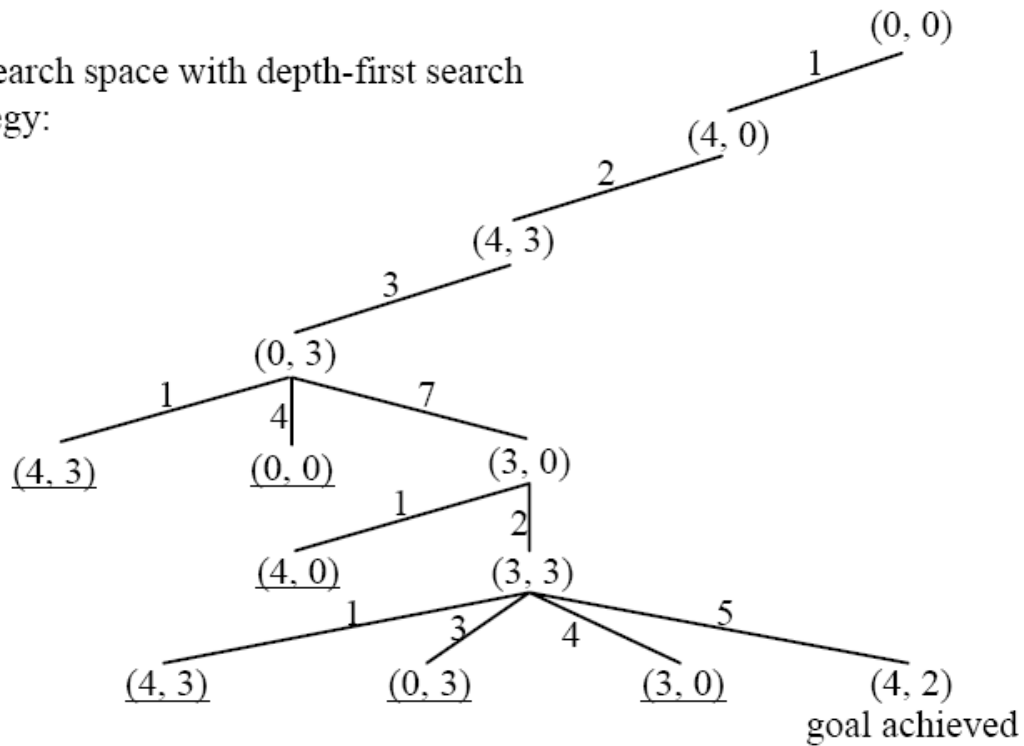   Fill the 3-gallon jug from pump
    Empty the 3-gallon jug into the 4-gallon jug
    Fill the 3-gallon jug from pump
    Fill the 4-gallon jug from the 3-gallon jug

The Water Jug Problem: Representation

the search space with depth-first search strategy:

```
                                                              (0, 0)
                                                         1
                                                  (4, 0)
                                             2
                                      (4, 3)
                                 3
                          (0, 3)
                    1         4        7
              (4, 3)      (0, 0)      (3, 0)
                                   1         2
                              (4, 0)      (3, 3)
                          1        3      4        5
                    (4, 3)      (0, 3)      (3, 0)      (4, 2)
                                                      goal achieved
```

Constraint Satisfaction Problem:

A **Constraint Satisfaction Problem** is characterized by:

- a *set of variables* {x1, x2, .., xn},
- for each variable xi a *domain* Di with the possible values for that variable, and
- a set of *constraints*, i.e. relations, that are assumed to hold between the values of the variables. [These relations can be given intentionally, i.e. as a formula, or extensionally, i.e. as a set, or procedurally, i.e. with an appropriate generating or recognizing function.] We will only consider constraints involving one or two variables.

The constraint satisfaction problem is to find, for each i from 1 to n, a value in Di for xi so that all constraints are satisfied. Means that, we must find a value for each of the variables that satisfies all of the constraints.

A CS problem can easily be stated as a sentence in first order logic, of the form:

```
(exist x1)..(exist xn) (D1(x1) & .. Dn(xn) => C1..Cm)
```

A CS problem is usually represented as an undirected graph, called *Constraint Graph* where the nodes are the variables and the edges are the binary constraints. Unary constraints can be disposed of by just redefining the domains to contain only the values that satisfy all the unary constraints. Higher order constraints are represented by hyperarcs. In the following we restrict our attention to the case of unary and binary constraints.

Formally, a constraint satisfaction problem is defined as a triple $\langle X, D, C \rangle$, where *X* is a set of variables, *D* is a domain of values, and *C* is a set of constraints. Every constraint is in turn a pair $\langle t, R \rangle$, where *t* is a tuple of variables and *R* is a set of tuples of values; all these tuples having the same number of elements; as a result *R* is a relation. An evaluation of the variables is a function from variables to values, $v : X \rightarrow D$. Such an evaluation satisfies a constraint $\langle (x_1, \ldots, x_n), R \rangle$ if $(v(x_1), \ldots, v(x_n)) \in R$. A solution is an evaluation that satisfies all constraints.

Constraints

- A constraint is a relation between a **local** collection of variables.
- The constraint restricts the values that these variables can simultaneously have.
- For example, **all-diff(X1, X2, X3)**. This constraint says that X1, X2, and X3 must take on different values. Say that {1,2,3} is the set of values for each of these variables then:
  X1=1, X2=2, X3=3 OK X1=1, X2=1,X3=3 NO

The constraints are the key component in expressing a problem as a CSP.
- The constraints are determined by how the variables and the set of values are chosen.
- Each constraint consists of;
  1. A set of variables it is over.
  2. A specification of the sets of assignments to those variables that satisfy the constraint.
- The idea is that we break the problem up into a set of distinct conditions each of which have to be satisfied for the problem to be solved.

Example: In N-Queens: Place N queens on an NxN chess board so that queen can attack any other queen.

- No queen can attack any other queen.
- Given any two queens Qi and Qj they cannot attack each other.
- Now we translate each of these individual conditions into a separate constraint.
  - Qi cannot attack Qj(i $\neq$ j)
    - Qi is a queen to be placed in column i, Qj is a queen to be placed in column j.

- The value of Qi and Qj are the rows the queens are to be placed in.
- Note the translation is dependent on the representation we chose.
- **Queens can attack each other**,
    1. *Vertically*, if they are in the same column---this is impossible as Qi and Qj are placed in different columns.
    2. *Horizontally*, if they are in the same row---we need the constraint Qi≠Qj.
    3. *Along a diagonal*, they cannot be the same number of columns apart as they are rows apart: we need the constraint |i-j| ≠|Qi-Qj| (| | is absolute value)
- **Representing the Constraints;**
    1. Between every pair of variables (Qi,Qj) (i ≠j), we have a constraint Cij.
    2. For each Cij, an assignment of values to the variables Qi= A and Qj= B, satisfies this constraint if and only if;

        A ≠B
        | A-B| ≠|i-j|
- **<u>Solutions:</u>**
    o A solution to the N-Queens problem will be any assignment of values to the variables $Q_1,\ldots,Q_N$ that satisfies all of the constraints.
    o Constraints can be over any collection of variables. In N-Queens we only need binary constraints---constraints over pairs of variables.

More Examples: Map Coloring Problem (Discussed in class)

Refer Russell and Norvig's Chapter 5 from pages 165-169. Also have a brief look on page 172-173 for forward checking that we have discussed in class.

[Chapter 3: Searching]

Artificial Intelligence (CSC 355)

Arjun Singh Saud

Central Department of Computer Science & Information Technology
Tribhuvan University

Problem Solving by Searching

Four general steps in problem solving:
- Goal formulation
    - What are the successful world states
- Problem formulation
    - What actions and states to consider given the goal
- Search
    - Determine the possible sequence of actions that lead to the states of known values and then choosing the best sequence.
- Execute
    - Give the solution perform the actions.

Problem formulation

A problem is defined by:

- An initial state: State from which agent start
- Successor function: Description of possible actions available to the agent.
- Goal test: Determine whether the given state is goal state or not
- Path cost: Sum of cost of each path from initial state to the given state.

A solution is a sequence of actions from initial to goal state. Optimal solution has the lowest path cost.

State Space representation

The state space is commonly defined as a directed graph in which each node is a state and each arc represents the application of an operator transforming a state to a successor state.

A **solution** is a path from the initial state to a goal state.

State Space representation of Vacuum World Problem:

States?? two locations with or without dirt: 2 x 2^2=8 states.
Initial state?? Any state can be initial
Actions?? {*Left*, *Right*, *Suck*}
Goal test?? Check whether squares are clean.
Path cost?? Number of actions to reach goal.

A search problem

Figure below contains a representation of a map. The nodes represent cities, and the links represent direct road connections between cities. The number associated to a link represents the length of the corresponding road.

The search problem is to find a path from a city S to a city G
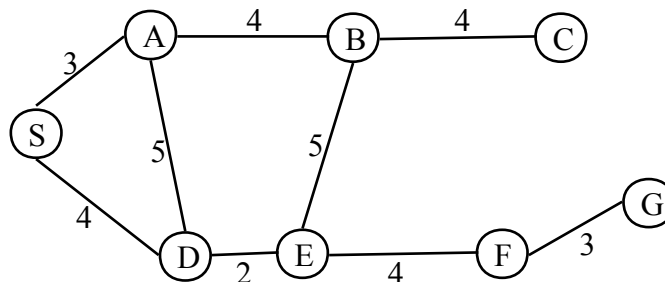
Figure : A graph representation of a map

This problem will be used to illustrate some search methods.

There are two broad classes of search methods:
 - uninformed (or blind) search methods;
 - heuristically informed search methods.

In the case of the uninformed search methods the order in which potential solution paths are considered is arbitrary, using no domain-specific information to judge where the solution is likely to lie.

In the case of the heuristically informed search methods one uses domain-dependent (heuristic) information in order to search the space more efficiently.

Measuring problem Solving Performance

We will evaluate the performance of a search algorithm in four ways

- **Completeness**
    - An algorithm is said to be complete if it definitely finds solution to the problem, if exist.
- **Time Complexity**
    - How long (worst or average case) does it take to find a solution? Usually measured in terms of the **number of nodes expanded**
- **Space Complexity**
    - How much space is used by the algorithm? Usually measured in terms of the **maximum number of nodes in memory at a time**
- **Optimality/Admissibility**
    - If a solution is found, is it guaranteed to be an optimal one? For example, is it the one with minimum cost?

Uninformed search

Breadth First Search(BFS)

Looks for the goal node among all the nodes at a given level before using the children of those nodes to

Expand *shallowest* unexpanded node

fringe is implemented as a FIFO queue

Note: Fringe is a generated but unexpanded node.
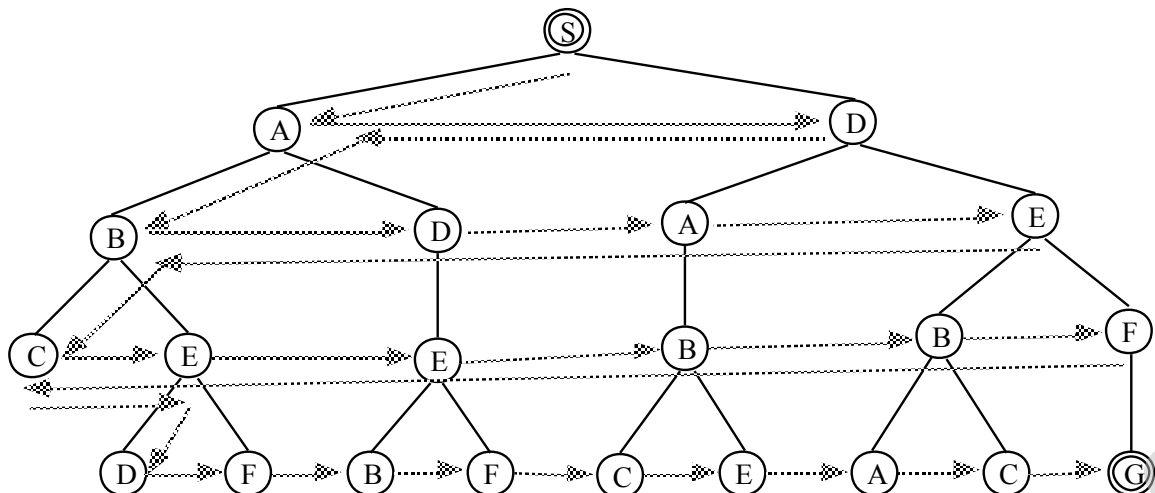
BFS evaluation

Completeness:
- *Does it always find a solution if one exists?*
- YES
    - If shallowest goal node is at some finite depth d and If *b* is finite

Time complexity:
- Assume a state space where every state has *b* successors.
    - root has *b* successors, each node at the next level has again b successors (total b^2), …
    - Assume solution is at depth *d*
    - Worst case; expand all except the last node at depth *d*
    - Total no. of nodes generated:
$$b + b^2 + b^3 + \dots\dots\dots\dots\dots b^d + (b^{d+1} - b) = \mathbf{O(b^{d+1})}$$

Space complexity:
- Each node that is generated must remain in memory
- Total no. of nodes in memory:
$$1 + b + b^2 + b^3 + \dots\dots\dots\dots\dots b^d + (b^{d+1} - b) = \mathbf{O(b^{d+1})}$$

Optimal (i.e., admissible):
- if all paths have the same cost. Otherwise, not optimal but finds solution with shortest path length (shallowest solution). If each path does not have same path cost shallowest solution may not be optimal

Two lessons:
- Memory requirements are a bigger problem than its execution time.
- Exponential complexity search problems cannot be solved by uninformed search methods for any but the smallest instances.

DEPTH2	NODES	TIME	MEMORY
2	1100	0.11 seconds	1 megabyte
4	111100	11 seconds	106 megabytes
6	107	19 minutes	10 gigabytes
8	109	31 hours	1 terabyte
10	1011	129 days	101 terabytes
12	1013	35 years	10 petabytes
14	1015	3523 years	1 exabyte

Depth First Search(DFS)

Looks for the goal node among all the children of the current node before using the sibling of this node

Expand *deepest* unexpanded node

Fringe is implemented as a LIFO queue (=stack)
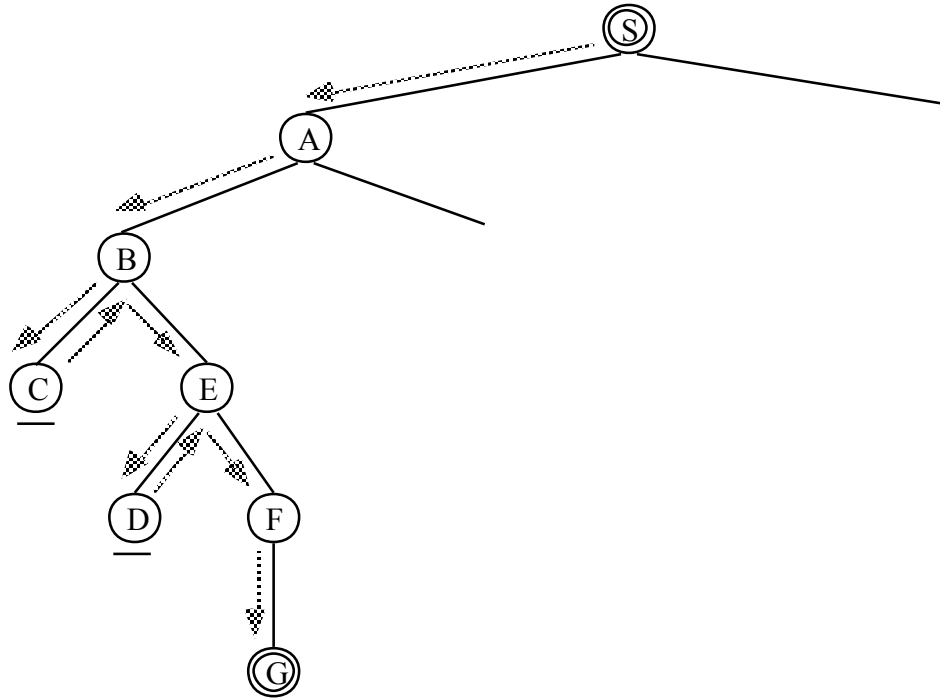
DFS evaluation

Completeness;
- *Does it always find a solution if one exists?*
- NO
    - If search space is infinite and search space contains loops then DFS may not find solution.

Time complexity;
- Let m is the maximum depth of the search tree. In the worst case Solution may exist at depth m.
- root has *b* successors, each node at the next level has again b successors (total b^2), …
- Worst case; expand all except the last node at depth *m*
- Total no. of nodes generated:
$$b + b^2 + b^3 + \ldots\ldots\ldots\ldots\ldots\ldots b^m = \mathbf{O(b^m)}$$

Space complexity:
- It needs to store only a single path from the root node to a leaf node, along with remaining unexpanded sibling nodes for each node on the path.
- Total no. of nodes in memory:

$$1+ b + b + b + \ldots\ldots\ldots\ldots\ldots.. b \quad m \text{ times} = \textbf{O(bm)}$$

Optimal (i.e., admissible):
- DFS expand deepest node first, if expands entire let sub-tree even if right sub-tree contains goal nodes at levels 2 or 3. Thus we can say DFS may not always give optimal solution.

Depth-limited search

It is DF-search with depth limit *l*.
- i.e. nodes at depth *l is assumed to* have no successors.
- Problem knowledge can be used to guess the value of *l*

Solves the infinite-path problem of DFS. Yet it introduces another source of problem if we are unable to find good guess of *l*. Let d is the depth of shallowest solution.

If $l < d$ then incompleteness results.

If $l > d$ then not optimal.

Time complexity: $O(b^l)$
Space complexity: $O(bl)$

Challenge!!

Derive Time complexity and Space Complexity

Iterative deepening search

- A general strategy to find best depth limit *l*.
  - Goals is found at depth *d*, the depth of the shallowest goal-node.

- Often used in combination with DF-search
- Combines benefits of DFS and BFS
- It does this by gradually increasing the limit—first 0, then 1, then 2, and so on until the goal is found. This will occur when depth d is reached: that is depth of shallowest goal node.

Limit = 0

Limit = 1

Limit = 2

Limit = 3

ID search evaluation

Completeness:
- − YES (no infinite paths)

Time complexity:
- − Algorithm seems costly due to repeated generation of certain states.
- − Node generation:

  level d: once

  level d-1: 2

  level d-2: 3

  …

  level 2: d-1

  level 1: d

- − Total no. of nodes generated:

  $d.b + (d-1).b^2 + (d-2).b^3 + \ldots\ldots\ldots\ldots\ldots.+1.b^d = \mathbf{O(b^d)}$

Space complexity:
- − It needs to store only a single path from the root node to a leaf node, along with remaining unexpanded sibling nodes for each node on the path.
- − Total no. of nodes in memory:

  $1 + b + b + b + \ldots\ldots\ldots\ldots\ldots.. b \quad d$ times $= \mathbf{O(bd)}$

Optimality:
- − YES if path cost is non-decreasing function of the depth of the node.

Note::-

Notice that BFS generates some nodes at depth d+1, whereas IDS does not. The result is that IDS is actually faster than BFS , despite the repeated generation of node.

Example:

Num. of nodes generated for b=10 and d=5 solution at far right

N(IDS) = 50 + 400 + 3000 + 20000 + 100000 = 123450

N(BFS) = 10 + 100 + 1000 + 10000 + 100000 + 999990 = 1111100

Recommended reading!!!

> Practice the problems for calculating time and sapace requirement of search algorithm that are done in class and also other problems from book.

Bidirectional search
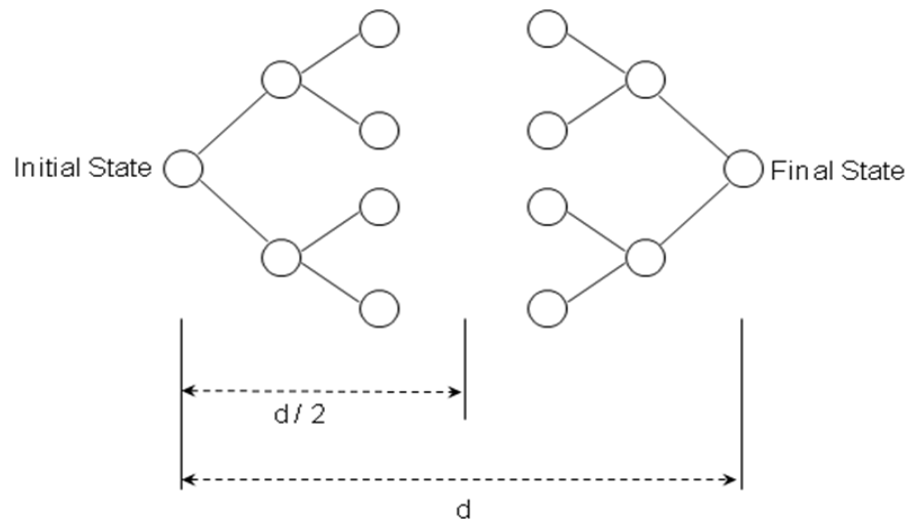
Bidirectional search is a graph search algorithm that finds a shortest path from an initial vertex to a goal vertex in a directed graph. It runs two simultaneous searches. one forward from the initial state and one backward from the goal, stopping when the two meet in the middle. The reason for this approach is that in many cases it is faster: for instance, in a simplified model of search problem complexity in which both searches expand a tree with branching factor b, and the distance from start to goal is d, each of the two searches has complexity $O(b^{d/2})$ and the sum of these two search times is much less than the $O(b^d)$, complexity that would result from a single search from the beginning to the goal. Even if theoretically bidirectional search seems effective than unidirectional search practically the case may be different because finding out where two search algorithms (forward and backward search) intersects needs additional data structure to be maintained and addition implementation steps to identify whether two search algorithms intersects or not.

Drawbacks of uniformed search :

- Criterion to choose next node to expand depends only on a global criterion: level.
- Does not exploit the structure of the problem.
- One may prefer to use a more flexible rule, that takes advantage of what is being discovered on the way, and hunches about what can be a good move.
- Very often, we can select which rule to apply by comparing the current state and the desired state

Heuristic Search:

Heuristic Search Uses domain-dependent (heuristic) information in order to search the space more efficiently.

Ways of using heuristic information:

• Deciding which node to expand next, instead of doing the expansion in a strictly breadth-first or depth-first order;

• In the course of expanding a node, deciding which successor or successors to generate, instead of blindly generating all possible successors at one time;

• Deciding that certain nodes should be discarded, or *pruned*, from the search space.

Best-First Search

Idea: use an *evaluation function f(n)* that gives an indication of which node to expand next for each node.

- usually gives an estimate to the goal.
- the node with the lowest value is expanded first.

A key component of *f(n)* is a heuristic function, *h(n),* which is a additional knowledge of the problem.

There is a whole family of best-first search strategies, each with a different evaluation function.

Typically, strategies use estimates of the cost of reaching the goal and try to minimize it.

Special cases: based on the evaluation function.
- Greedy best-first search
- A*search

Greedy Best First Search

Greedy best first search expands the node that seems to be closest to the goal node. Heuristic function is used to estimate which node is closest to the goal node. It doest not always gives optimal solution. Optimality of the solution depends upon how good our heuristic is. Once the greedy best first search takes a node, it never looks backward.
Heuristic used by Greedy Best First Search is:

$f(n) = h(n)$

Where h(n) is the estimated cost of the path from node n to the goal node.

For example consider the following graph
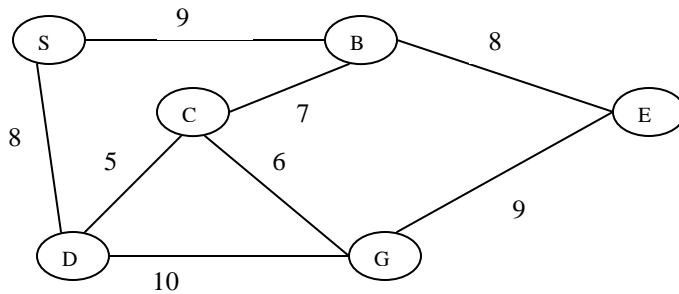
Straight Line distances to node G (goal node) from other nodes is given below:

S→ G = 12
B→ G = 4
E→ G = 7
D→ G = 5
C→ G = 3

Let H(n)= Straight Line distance
Now Greedy Search operation is done as below:

Step 1
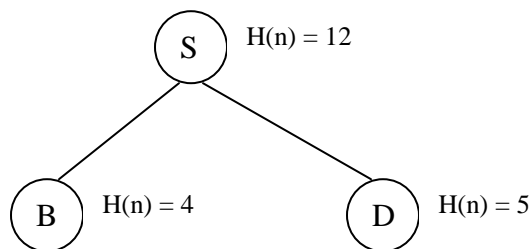
Step 2:
Now the turn is of
node B
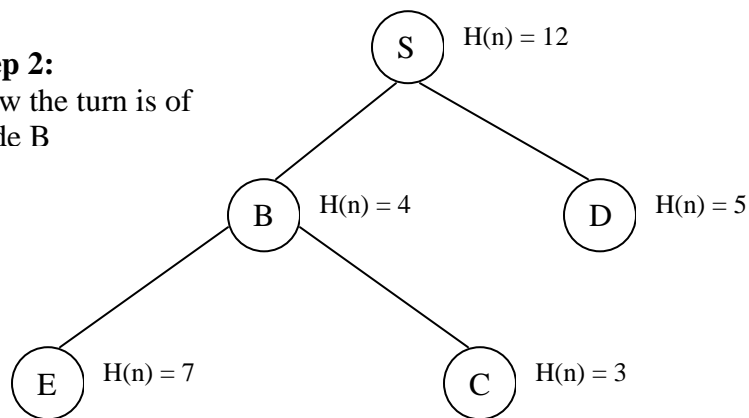
Step 3:
Now the turn is
of Node C
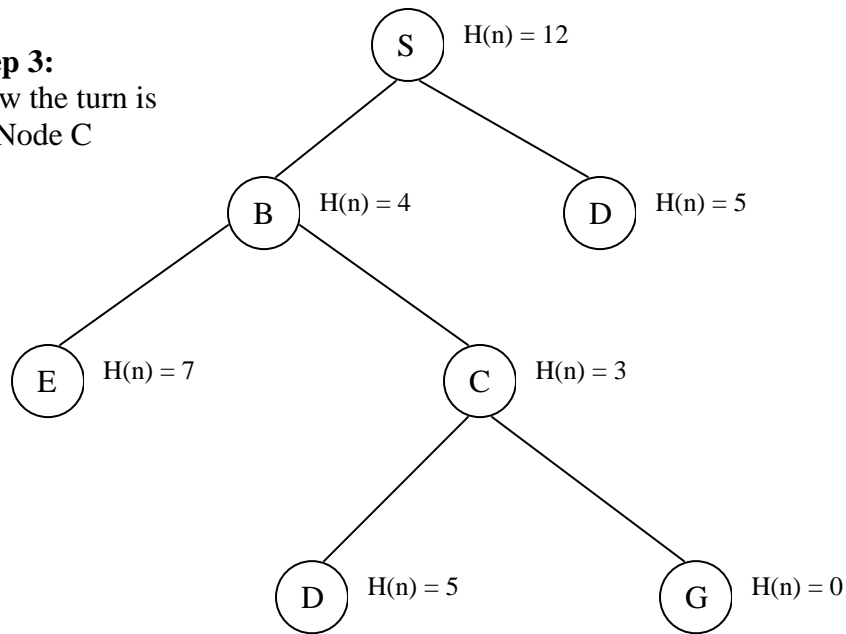
Step 4:
Now the turn id
of node G
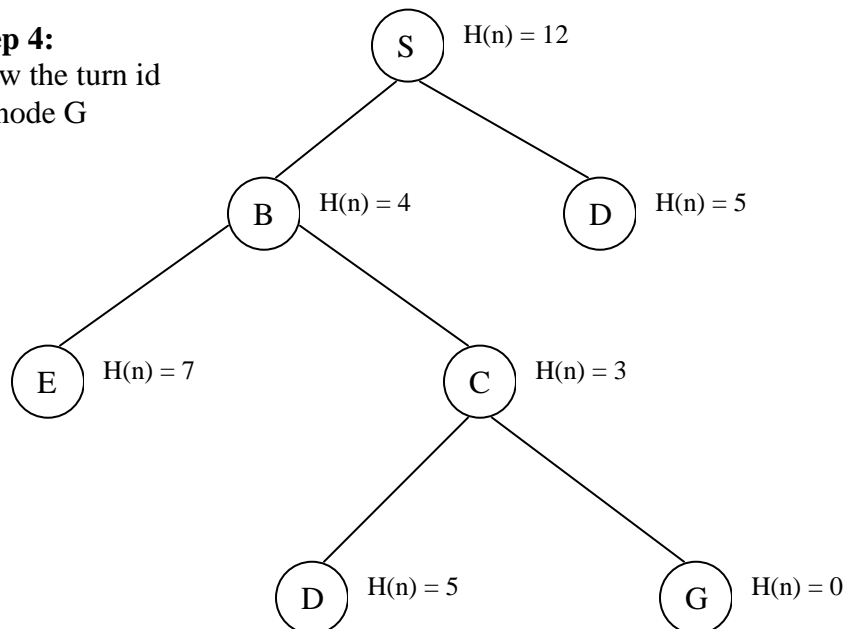
Evaluation Greedy Search

Completeness:
- While minimizing the value of heuristic greedy may oscilate between two nodes. Thus it is not complete.

Optimality:
- Greedy search is not optimal. Same as DFS.

Time complexity:

> - In worst case Greedy search is same as DFS therefore it's time complexity is $O(b^m)$.

Space Complexity:
> - Space complexity of DFS is $O(b^m)$. No nodes can be deleted from memory.

A * Search

Like all informed search algorithms, it first searches the routes that appear to be most likely to lead towards the goal. What sets A* apart from a greedy best-first search is that it also takes the distance already traveled into account (the g(n) part of the heuristic is the cost from the start, and not simply the local cost from the previously expanded node).The algorithm traverses various paths from start to goal. Evaluation function used by A* search algorithm is

$f(n) = h(n) + g(n)$
Where:
- g(n): the actual shortest distance traveled from initial node to current node
- h(n): the estimated (or "heuristic") distance from current node to goal
- f(n): the sum of g(n) and h(n)

For example consider the following graph
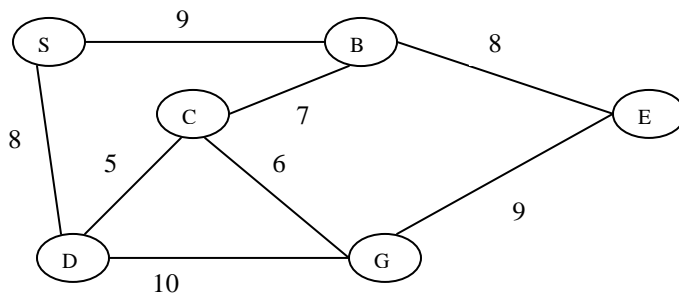
Straight Line distances to node G (goal node) from other nodes is given below:
S→ G = 12
B→ G = 4
E→ G = 7
D→ G = 6
C→ G = 3
Labels in the graph shows actual distance.
Let H(n)= Straight Line distance

Now A* Search operation is done as below

Step 1

S   f(n)= 12

9   8

B   f(n) = 13   D   f(n) = 14
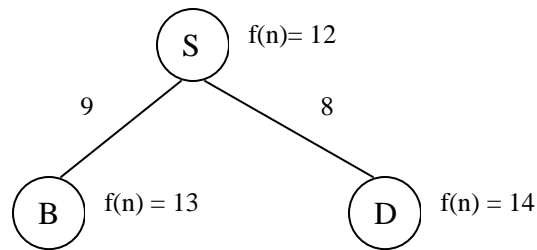
f(B) = 9 + 4 =13
f(D) = 8 + 6 =14

Step 2:
Now the turn is of
node B

S   f(n) = 12

9   8

B   f(n) = 4   D   f(n) = 5

8   7

E   C

f(n) = 24

f(n) = 19
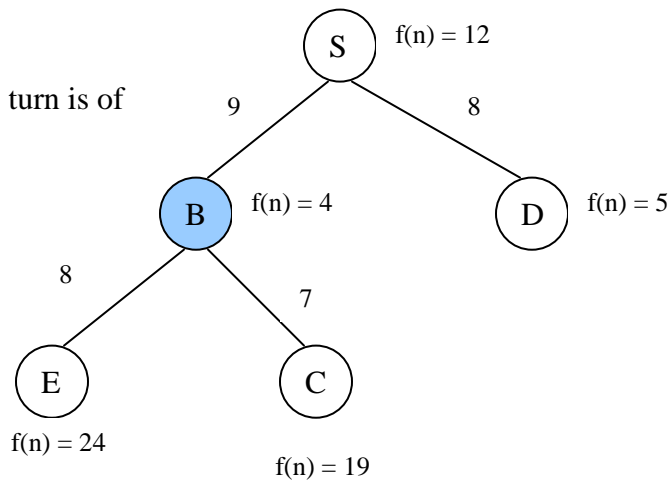
f(E) =17 + 7 =24
f(C) =16 + 3 =19

Step 3:
Now the turn is of
node D

S   H(n) = 12

9   8

B   H(n) = 4   D   H(n) = 5

8   7   10   5

E   C   G   C

f(n) = 24

f(n) = 19   f(n) = 18   f(n) = 16
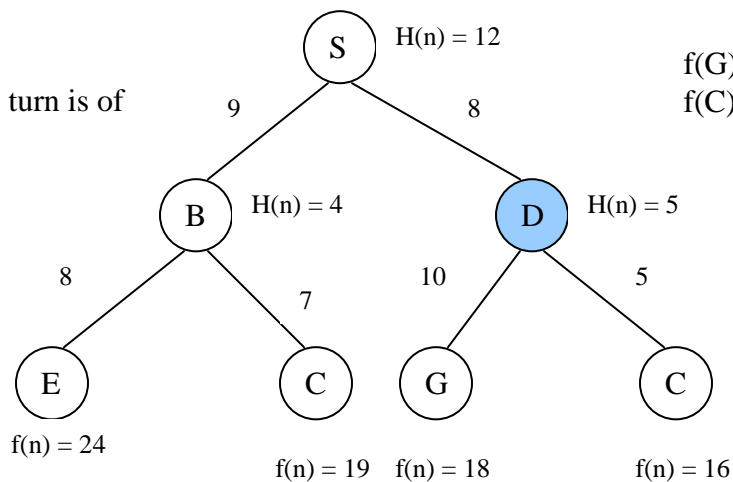
f(G) =18 + 0 =18
f(C) =13 + 3 =16

Step 4:
Now the turn is of node C

H(n) = 12

f(G) = 19 + 0 = 19
f(C) = 20 + 4 = 24

H(n) = 4

H(n) = 5

9

8

8

7

10

5

f(n) = 24

f(n) = 19

f(n) = 18

f(n) = 16

6

7

f(n) = 19

f(n) = 24

Step 4:
Now the turn is of node G

S   H(n) = 12

f(G) =19 + 0 =19
f(C) =20 + 4 =24

9   8

B   H(n) = 4   D   H(n) = 5

8   7   10   5

E   C   G   C

f(n) = 24

f(n) = 19   f(n) = 18   f(n) = 16

6   7

G   B

f(n) = 19   f(n) = 24
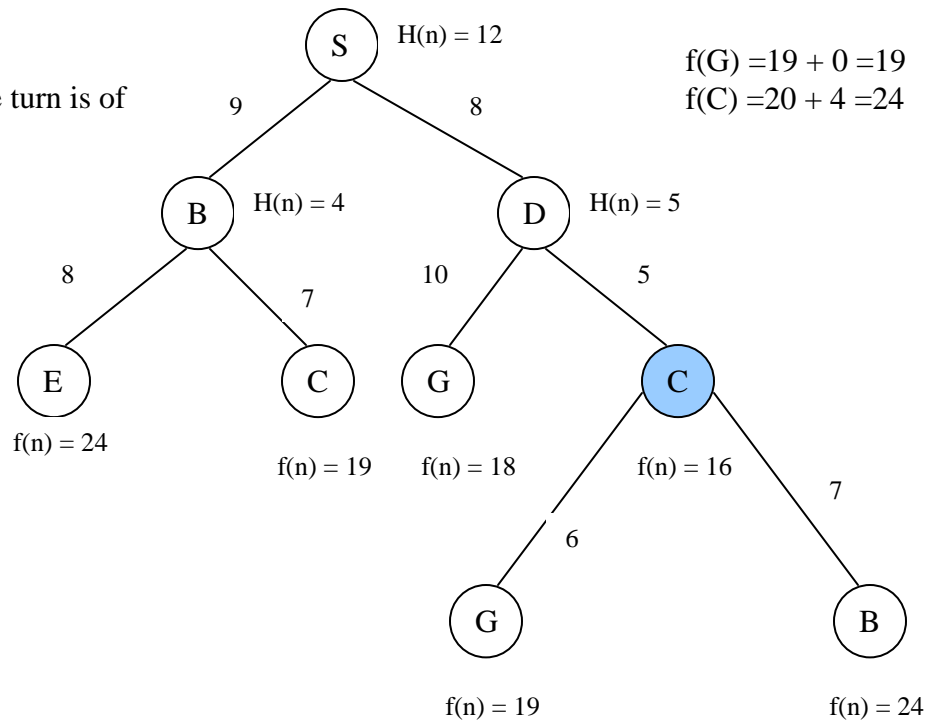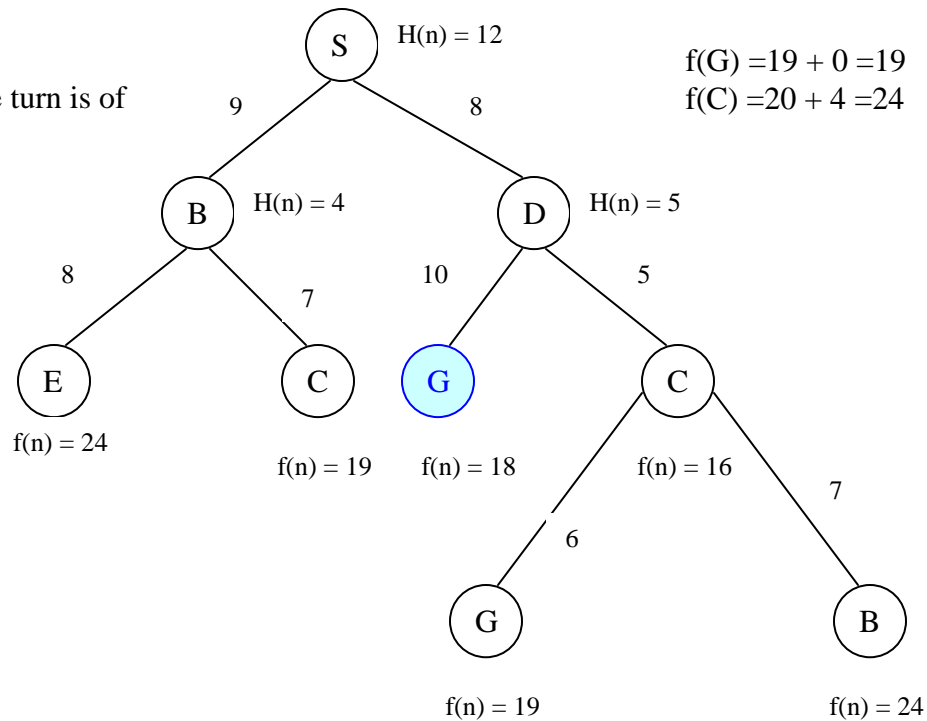
G is goal node, Hence we are done

Evaluating A* Search:
Completeness:
  – Yes A* search always gives us solution
Optimality:
  – A* search gives optimal solution when the heuristic function is admissible heuristic.
Time complexity:
  – Exponential with path length i.e. $O(b^d)$ where d is length of the goal node from start node.
Space complexity:
  – It keeps all generated nodes in memory. Hence space is the major problem not time

<u>Admissible heuristic</u>

A heuristic function is said to be admissible heuristic if iot never overestimates the cost to reach to the goal.
i.e
$h(n) \leq h^*(n)$
Where
 $h(n)$ = Estimated cost to reach to the goal node from node n
$h^*(n)$ = Actual cost to reach to the goal node from node n

Formulating admissible heuristics:

- *n* is a node
- *h* is a heuristic
- *h(n)* is cost indicated by *h* to reach a goal from *n*
- *C(n)* is the actual cost to reach a goal from n
- *h* is admissible if

$$\forall n, h(n) \leq C(n)$$

For Example: 8-puzzle

Figure shows 8-puzzle start state and goal state. The solution is 26 steps long.

Start State   **Goal State**
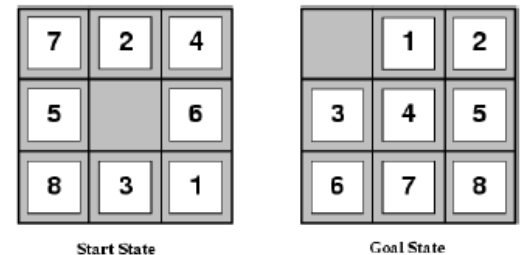
$h_1(n)$ = number of misplaced tiles
$h_2(n)$ = sum of the distance of the tiles from their goal position (notdiagonal).
$h_1(S) = ?$ 8
$h_2(S) = ?$ 3+1+2+2+2+3+3+2 = 18
$h_n(S) = \max\{h1(S), h2(S)\} = 18$

Consistency (Monotonicity)

A heuristic is said to be consistent if for any node N and any successor N' of N , estimated cost to reach to the goal from node N to goal node is less than the sum of step cost from N to N' and estimated cost from node N' to goal node.
i.e
$h(n) \leq d(n, n') + h(n')$
Where
h(n) = Estimated cost to reach to the goal node from node n
d(n, n') = actual cost from n to n'

#Prove that A* search gives optimal solution when the heuristic is admissible.

Suppose suboptimal goal *G2* in the queue.
Let *n* be an unexpanded node on a shortest to optimal goal *G* and C* be the cost of optimal goal node.
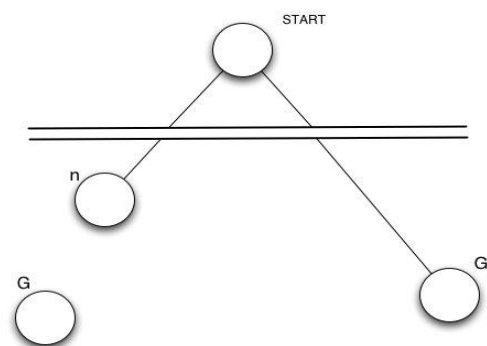
f(G2) = h (G2) + g(G2)
f(G2) = g(G2)   since h(G2)=0
f(G2) > C*   …………..(1)

Again
Since h(n) is admissibl, It doest not overstimatews the cost of completing
the solution path.
=>   f(n) = g(n) + h(n) ≤C* ……………(2)

Now from (1) and (2)
f(n) ≤ C* < f(G2)

Since f(G2) > f(n), A* will never select G2 for expansion. Thus A* gives
us optimal solution when heuristic function is admissible.

Prove that:If h(n) is consistent , then the values of f(n) along the path are
nondecreasing.
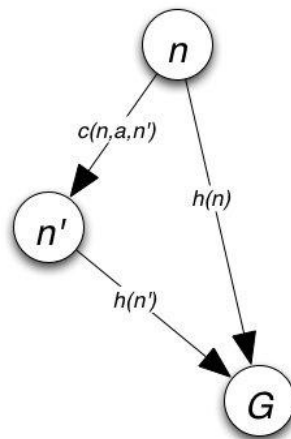
Suppose n' is successor of n, then
g(n') = g(n) + C(n,a,n')
we know that,
f(n') = g(n') + h(n')
f(n') = g(n) + C(n,a,n') + h(n') …………….(1)
A heuristic is consistent if
h(n) ≤ C(n,a,n') + h(n')……………………(2)
Now from (1) and (2)
 f(n') = g(n) + C(n,a,n') + h(n') ≥ g(n) + h(n) = f(n)
    ⇨   f(n') ≥ f(n)
    ⇨   f(n) is nondecreasing along any path.

Hill climbing

Hill climbing is depth-first search with a heuristic measurement that orders choices as nodes are expanded. It always selects the most promising successor of the node last expanded.

For instance, consider that the most promising successor of a node is the one that has the shortest straight-line distance to the goal node G. In figure below, the straight line distances between each city and G is indicated in square brackets.
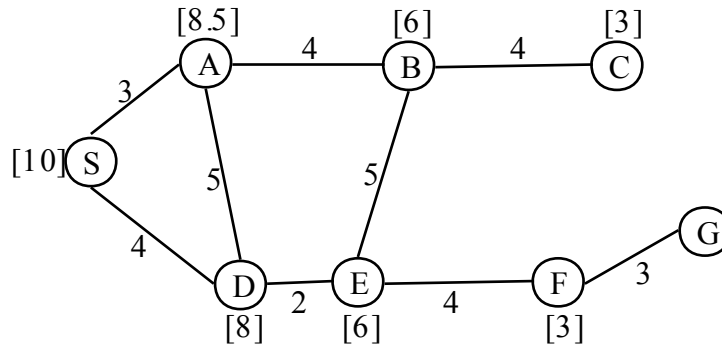

Figure: The map indicating also the straight line distances between each city and G.

The hill climbing search from S to G proceeds as follows:
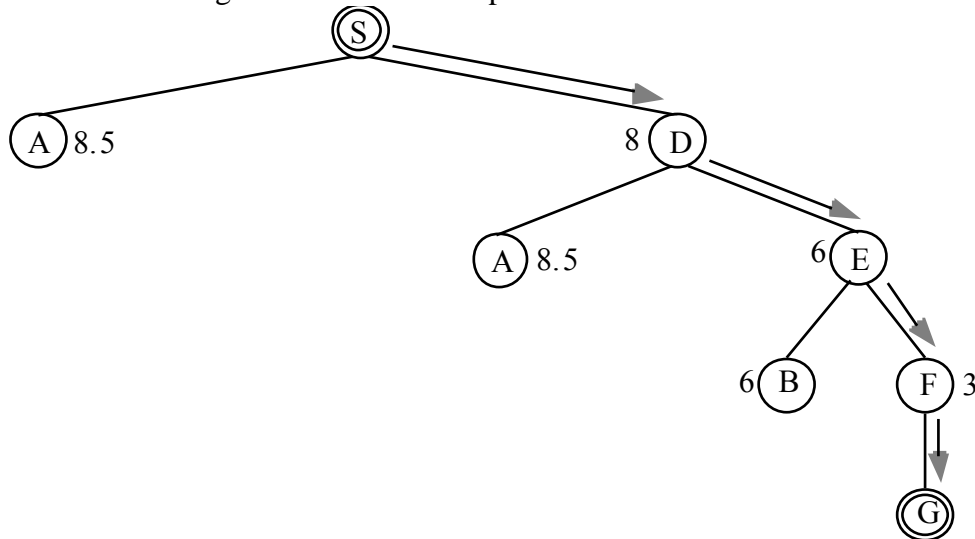
Figure : Hill climbing search of a path from node S to node G.

Exercise

Apply the hill climbing algorithm to find a path from S to G, considering that the most promising successor of a node is its closest neighbor.
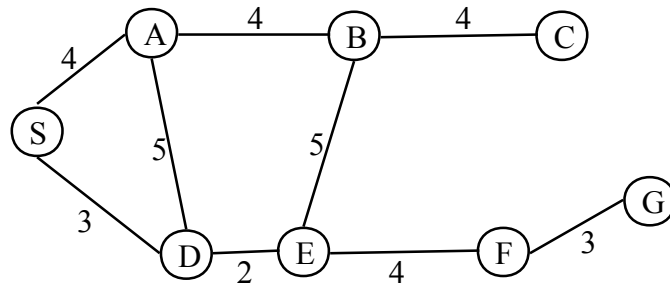
Note:

The difference between the hill climbing search method and the best first search method is the following one:

- the best first search method selects for expansion the most promising leaf node of the current search tree;
- the hill climbing search method selects for expansion the most promising successor of the node last expanded.

Problems with Hill Climbing

:

- Gets stuck at **local minima** when we reach a position where there are no better neighbors, it is not a guarantee that we have found the best solution. **Ridge** is a sequence of local maxima.
- Another type of problem we may find with hill climbing searches is finding a *plateau.* This is an area where the search space is flat so that all neighbors return the same evaluation

Simulated Annealing

It is motivated by the physical annealing process in which material is heated and slowly cooled into a uniform structure. Compared to hill climbing the main difference is that SA allows downwards steps. Simulated annealing also differs from hill climbing in that a move is selected at random and then decides whether to accept it. If the move is better than its current position then simulated annealing will always take it. If the move is worse (i.e. lesser quality) then it will be accepted based on some probability. The probability of accepting a worse state is given by the equation

$$P = \exp(-c\,/t) > r$$

Where
c	=	the change in the evaluation function
t	=	the current value
r	=	a random number between 0 and 1

The probability of accepting a worse state is a function of both the current value and the change in the cost function. The most common way of implementing an SA algorithm is to implement hill climbing with an accept function and modify it for SA

Game Search

Games are a form of *multi-agent environment*
- − What do other agents do and how do they affect our success?
- − Cooperative vs. competitive multi-agent environments.
- − Competitive multi-agent environments give rise to adversarial search often known as *games*

Games – adversary
- − Solution is strategy (strategy specifies move for every possible opponent reply).
- − Time limits force an *approximate* solution
- − Evaluation function: evaluate "goodness" of game position
- − Examples: chess, checkers, Othello, backgammon

Difference between the search space of a game and the search space of a problem: In the first case it represents the moves of two (or more) players, whereas in the latter case it represents the "moves" of a single problem-solving agent.

7.1 An exemplary game: Tic-tac-toe

There are two players denoted by X and O. They are alternatively writing their letter in one of the 9 cells of a 3 by 3 board. The winner is the one who succeeds in writing three letters in line.

The game begins with an empty board. It ends in a win for one player and a loss for the other, or possibly in a draw.

A complete tree is a representation of all the possible plays of the game. The root node is the initial state, in which it is the first player's turn to move (the player X). The successors of the initial state are the states the player can reach in one move, their successors are the states resulting from the other player's possible replies, and so on.

Terminal states are those representing a win for X, loss for X, or a draw.

Each path from the root node to a terminal node gives a different complete play of the game. Figure given below shows the initial search space of Tic-Tac-Toe.
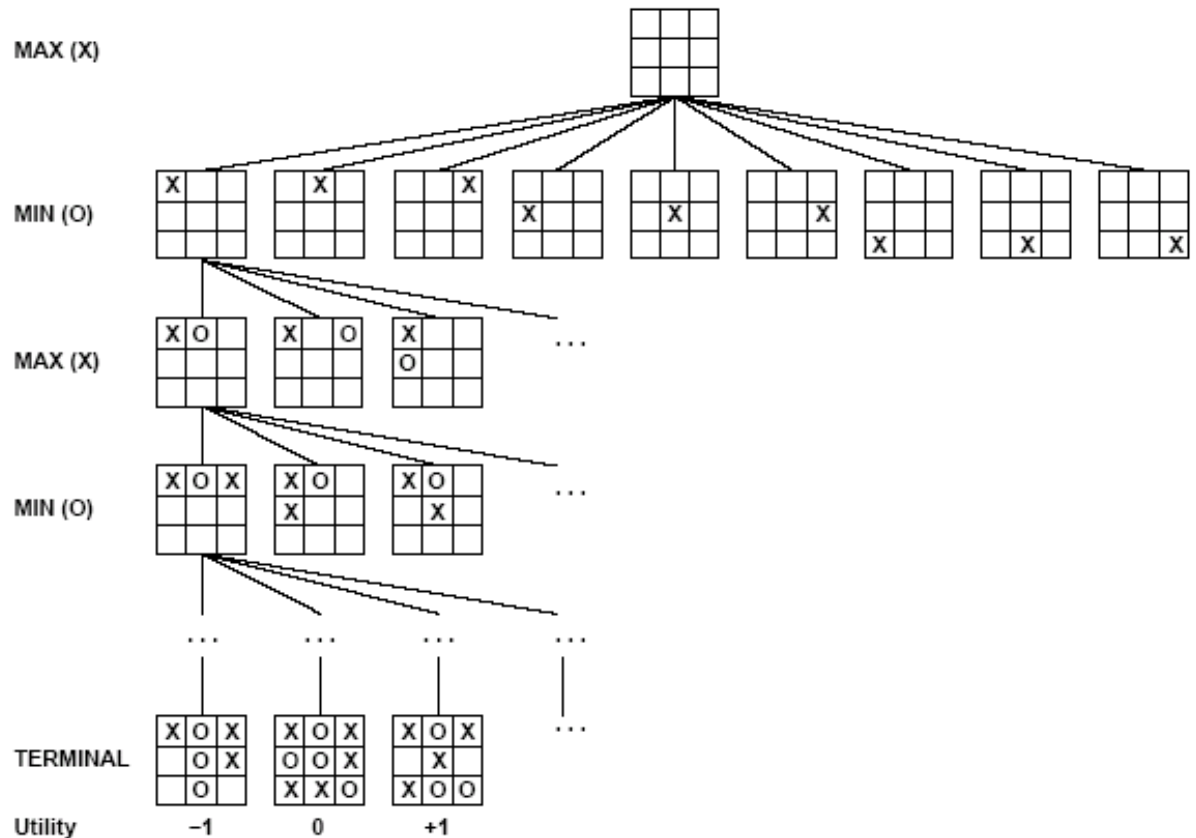
Figure 3.14: Partial game tree for Tic-Tac-Toe

A game can be formally defined as a kind of search problem as below:
- Initial state: It includes the board position and identifies the playesr to move.
- Successor function: It gives a list of (move, state) pairs each indicating a legal move and resulting state.
- Terminal test: Which determines when the game is over. States where the game is ended are called terminal states.
- Utility function: It gives numerical value of terminal states. E.g. win (+1), loose (-1) and draw (0). Some games have a wider variety of possible outcomes eg. ranging from +92 to -192.

7.2 The minimax Algorithm

Let us assign the following values for the game: 1 for win by X, 0 for draw, -1 for loss by X.

Given the values of the terminal nodes (win for X (1), loss for X (-1), or draw (0)), the values of the non-terminal nodes are computed as follows:

- the value of a node where it is the turn of player X to move is the maximum of the values of its successors (because X tries to maximize its outcome);

- the value of a node where it is the turn of player O to move is the minimum of the values of its successors (because O tries to minimize the outcome of X).

Figure below shows how the values of the nodes of the search tree are computed from the values of the leaves of the tree.
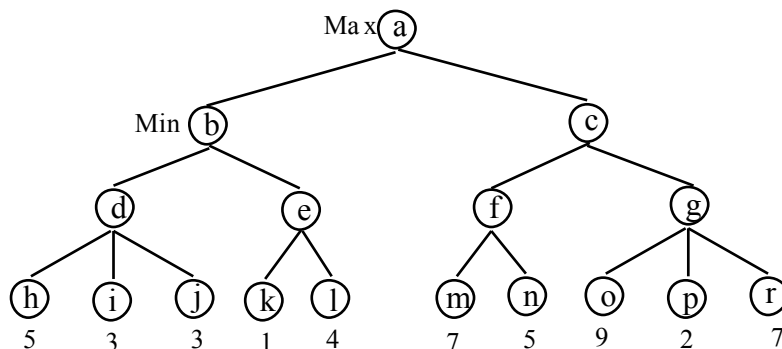The values of the leaves of the tree are given by the rules of the game:
- 1 if there are three X in a row, column or diagonal;
- -1 if there are three O in a row, column or diagonal;
- 0 otherwise
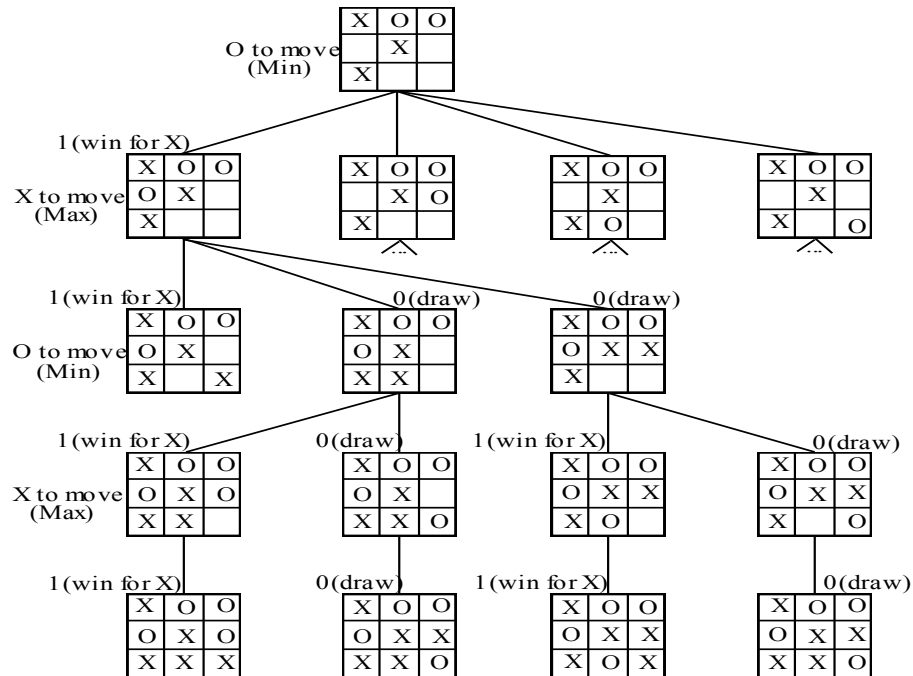
Example

Consider the following game tree (drawn from the point of view of the Maximizing player):

Show what moves should be chosen by the two players, assuming that both are using the mini-max procedure.

Solution:
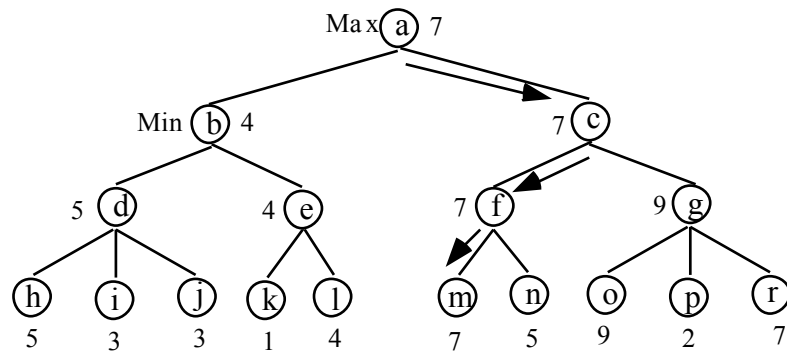
Figure 3.16: The mini-max path for the game tree

Alpha-beta pruning

Alpha-beta pruning is a technique for evaluating nodes of a game tree that eliminates unnecessary evaluations. It uses two parameters, alpha and beta.

An alpha cutoff

To apply this technique, one uses a parameter called alpha that represents a lower bound for the achievement of the Max player at a given node.

Let us consider that the current board situation corresponds to the node A in the following figure.
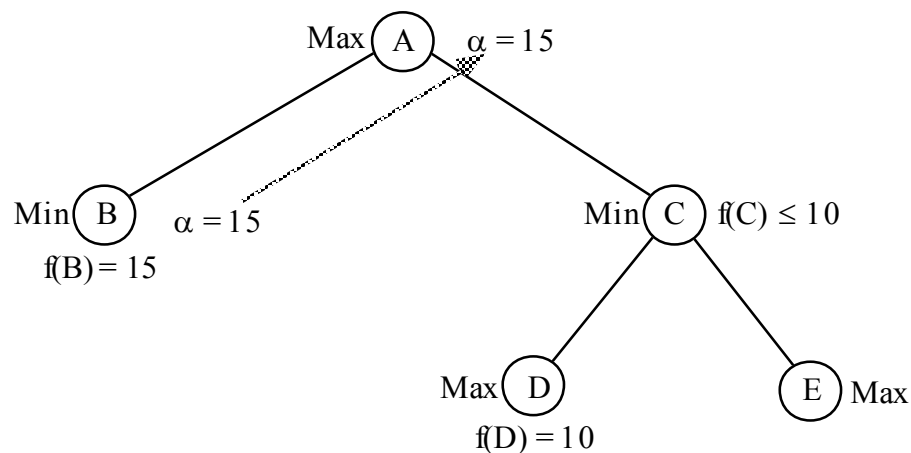
The minimax method uses a depth-first search strategy in evaluating the descendants of a node. It will therefore estimate first the value of the node B. Let us suppose that this value has been evaluated to 15, either by using a static evaluation function, or by backing up from descendants omitted in the figure. If Max will move to B then it is guaranteed to achieve 15. Therefore 15 is a lower bound for the achievement of the Max player (it may still be possible to achieve more, depending on the values of the other descendants of A). Therefore, the value of α at node B is 15. This value is transmitted upward to the node A and will be used for evaluating the other possible moves from A.

To evaluate the node C, its left-most child D has to be evaluated first. Let us assume that the value of D is 10 (this value has been obtained either by applying a static evaluation function directly to D, or by backing up values from descendants omitted in the figure). Because this value is less than the value of α, the best move for Max is to node B, independent of the value of node E that need not be evaluated. Indeed, if the value of E is greater than 10, Min will move to D which has the value 10 for Max. Otherwise, if the value of E is less than 10, Min will move to E which has a value less than 10. So, if Max moves to C, the best it can get is 10, which is less than the value $\alpha = 15$ that would be gotten if Max would move to B. Therefore, the best move for Max is to B, independent of the value of E. The elimination of the node E is an alpha cutoff.

One should notice that E may itself have a huge subtree. Therefore, the elimination of E means, in fact, the elimination of this subtree.

A beta cutoff

To apply this technique, one uses a parameter called beta that represents an upper bound for the achievement of the Max player at a given node.

In the above tree, the Max player moved to the node B. Now it is the turn of the Min player to decide where to move:
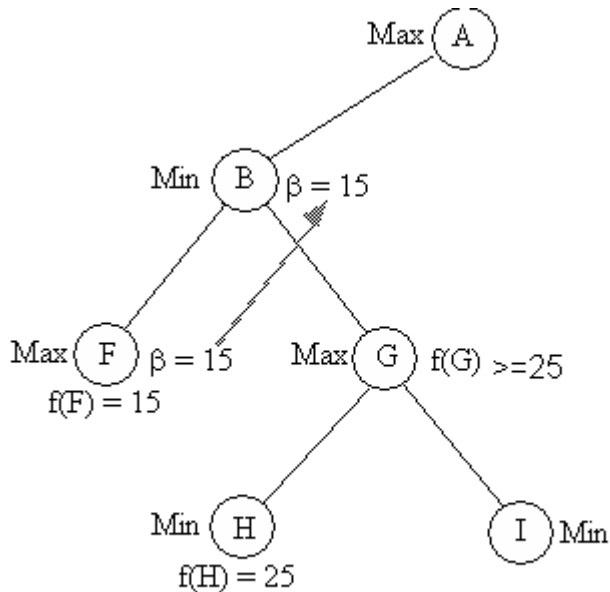
Figure 3.18: Illustration of the beta cut-off.

The Min player also evaluates its descendants in a depth-first order.

Let us assume that the value of F has been evaluated to 15. From the point of view of Min, this is an upper bound for the achievement of Min (it may still be possible to make Min achieve less, depending of the values of the other descendants of B). Therefore the value of β at the node F is 15. This value is transmitted upward to the node B and will be used for evaluating the other possible moves from B.

To evaluate the node G, its left-most child H is evaluated first. Let us assume that the value of H is 25 (this value has been obtained either by applying a static evaluation function directly to H, or by backing up values from descendants omitted in the figure). Because this value is greater than the value of β, the best move for Min is to node F, independent of the value of node I that need not be evaluated. Indeed, if the value of I is $v \geq 25$, then Max (in G) will move to I. Otherwise, if the value of I is less than 25, Max will move to H. So in both cases, the value obtained by Max is at least 25 which is greater than β (the best value obtained by Max if Min moves to F).
Therefore, the best move for Min is at F, independent of the value of I. The elimination of the node I is a beta cutoff.

One should notice that by applying alpha and beta cut-off, one obtains the same results as in the case of mini-max, but (in general) with less effort. This means that, in a given amount of time, one could search deeper in the game tree than in the case of mini-max.
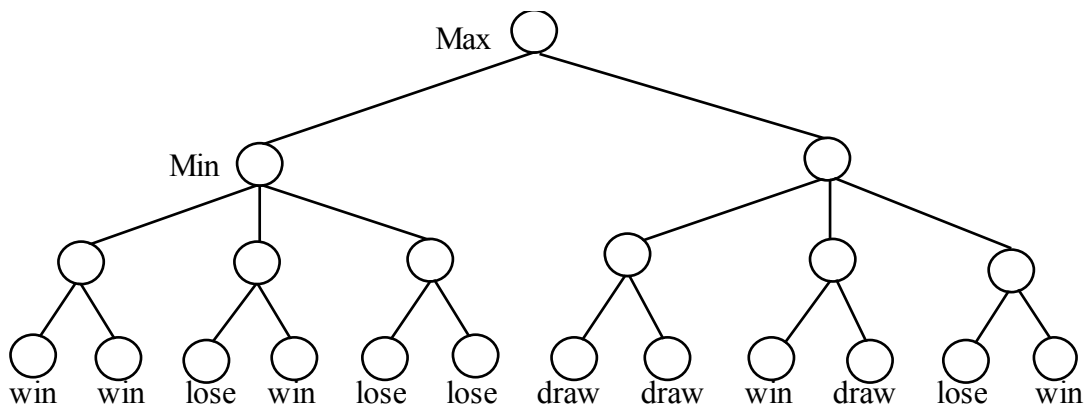
Exercises:

1. Consider the following game tree (drawn from the point of view of the Maximizing player):
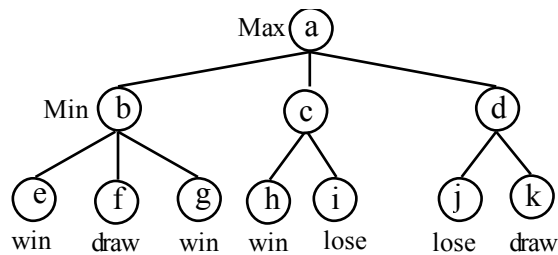

What move should be chosen by the Max player, and what should be the response of the Min player, assuming that both are using the mini-max procedure?

2. Consider the following game tree (drown from the point of view of the Maximizing player):

What move should be chosen by the Max player, and what should be the response of the Min player, assuming that both are using the mini-max procedure?

3. Consider the following two-person game tree, drown from the point of view of the Maximizing player:
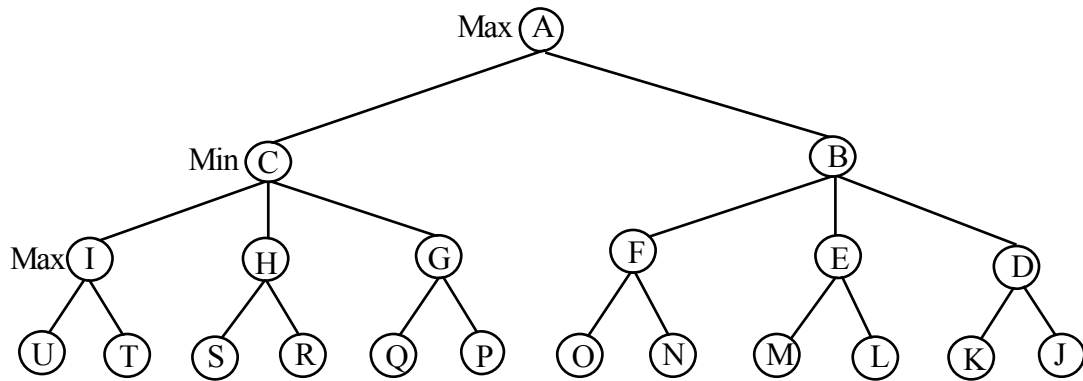
Assume a static evaluation function which produces the following values:

f(J)=win	f(K)=win	f(L)=lose	f(M)=lose	f(N)=win
f(O)=lose	f(P)=lose	f(Q)=win	f(R)=draw	f(S)=draw
f(T)=draw	f(U)=win			

Use the alpha-beta pruning procedure to determine the moves which should be chosen by the two players, assuming a depth-first search (generation) of the tree. Which nodes need not to be examined (generated)?
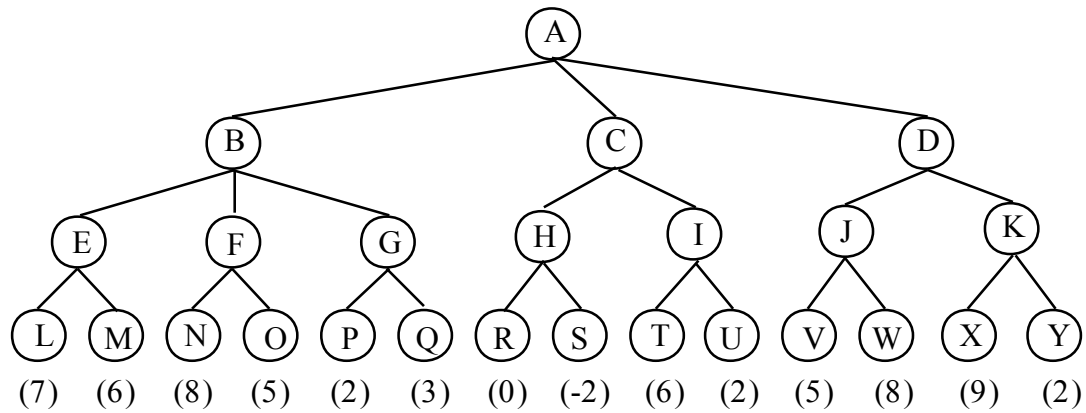
4. Consider the following game tree (drown from the point of view of the Maximizing player):
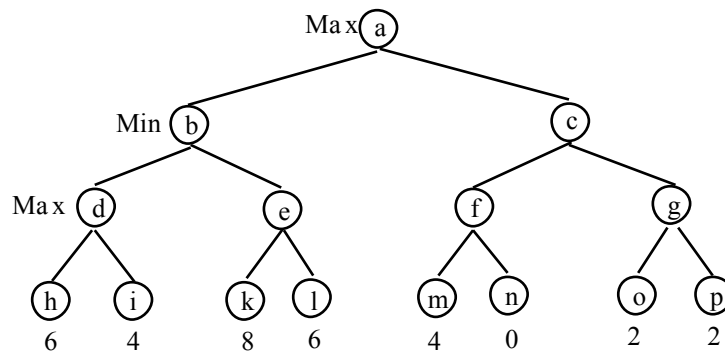
a) Use the mini-max procedure and show what moves should be chosen by the two players.
b) Use the alpha-beta pruning procedure and show what nodes would not need to be examined.
5. Consider the following game tree in which static scores are all from the first player's point of view:

A

B   C   D

E   F   G   H   I   J   K

L   M   N   O   P   Q   R   S   T   U   V   W   X   Y

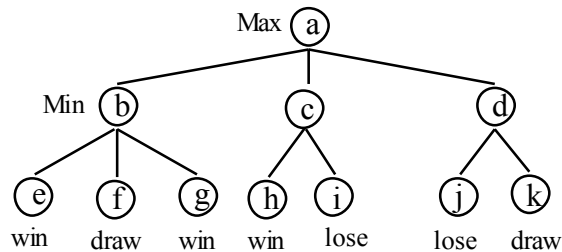(7)   (6)   (8)   (5)   (2)   (3)   (0)   (-2)   (6)   (2)   (5)   (8)   (9)   (2)

Suppose the first player is the maximizing player. What move should be chosen?

what nodes would not need to be examined using the alpha-beta pruning procedure?
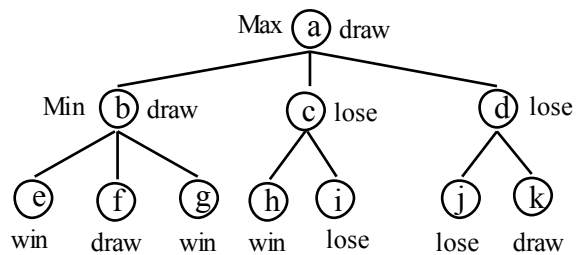
Solutions to exercises

1. Consider the following game tree (drawn from the point of view of the Maximizing player):

What move should be chosen by the Max player, and what should be the response of the Min player, assuming that both are using the mini-max procedure ?

Solution:

Max will move to b and min will respond by moving to f.

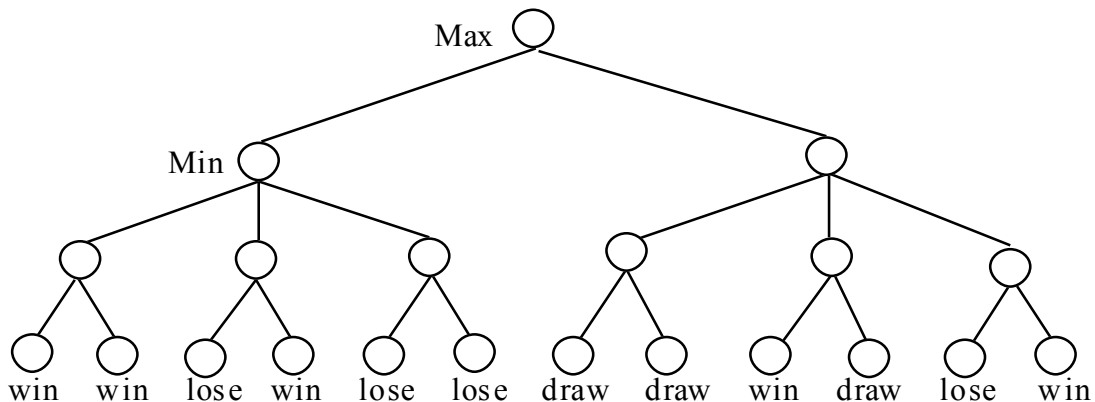2. Consider the following game tree (drown from the point of view of the Maximizing player):

What move should be chosen by the Max player, and what should be the response of the Min player, assuming that both are using the mini-max procedure?
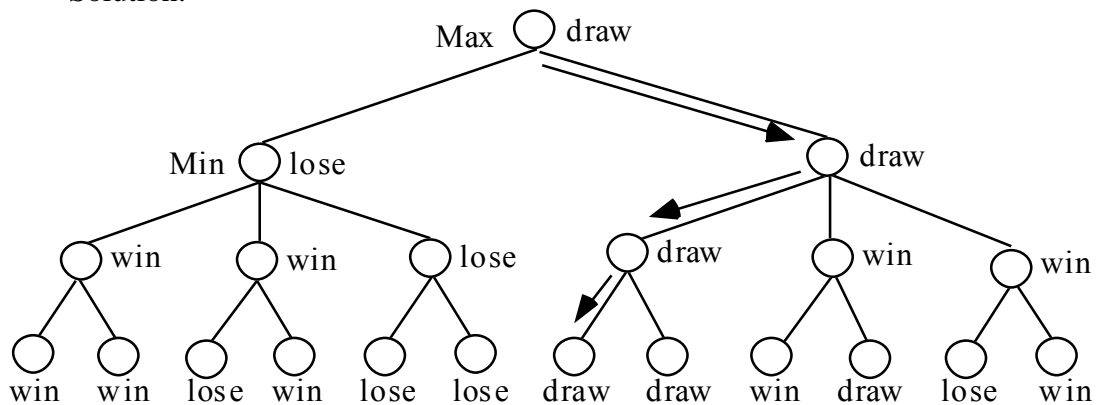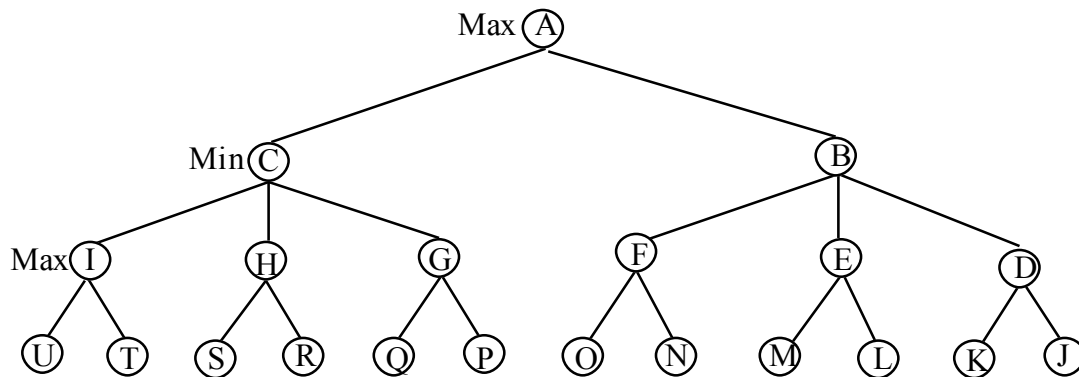
Solution:

3. Consider the following two-person game tree, drown from the point of view of the Maximizing player:

Assume a static evaluation function which produces the following values:

f(J)=win	f(K)=win	f(L)=lose	f(M)=lose	f(N)=win
f(O)=lose	f(P)=lose	f(Q)=win	f(R)=draw	f(S)=draw
f(T)=draw	f(U)=win			

Use the alpha-beta pruning procedure to determine the moves which should be chosen by the two players, assuming a depth-first search (generation) of the tree. Which nodes need not to be examined (generated)?

a) The following figure shows the application of the alpha-beta pruning procedure on:
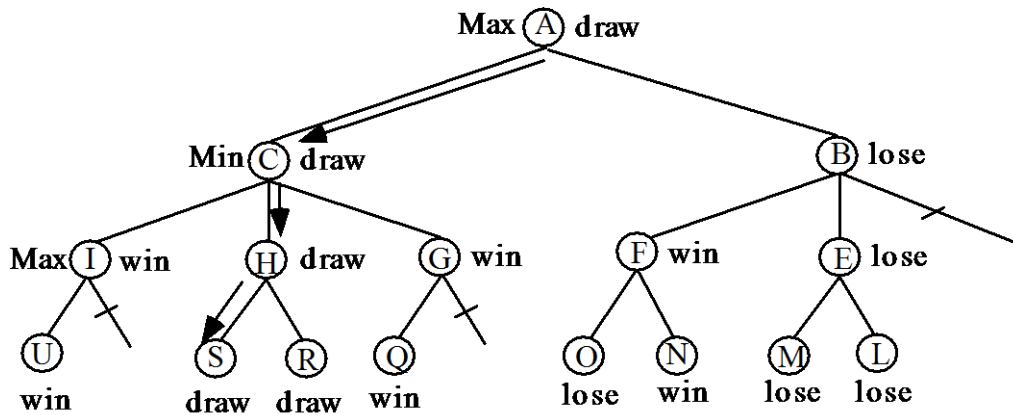
Max moves to C and Min moves to H and Max moves to either S or R.
The game ends in a draw.

4. Consider the following game tree (drown from the point of view of the Maximizing player):

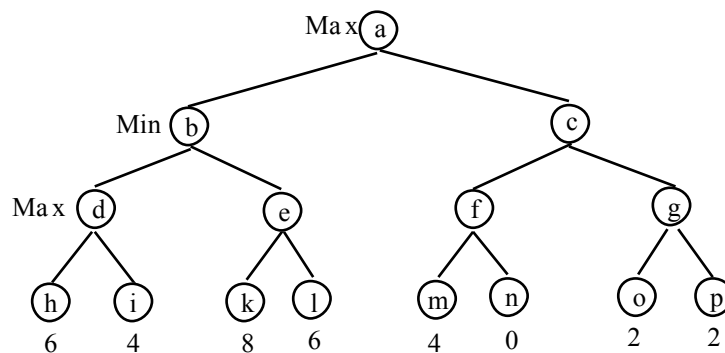a) Use the mini-max procedure and show what moves should be chosen by the two players.

b) Use the alpha-beta pruning procedure and show what nodes would not need to be examined.

c) Consider the mirror image of the above tree and apply again the alpha-beta pruning procedure. What do you notice?

The mirror image is the following tree:
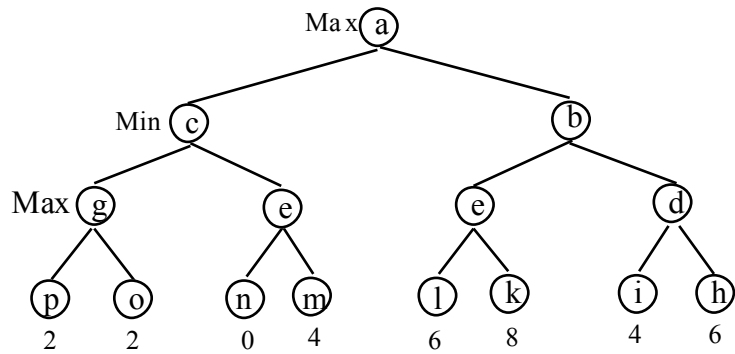
Solution:

a)
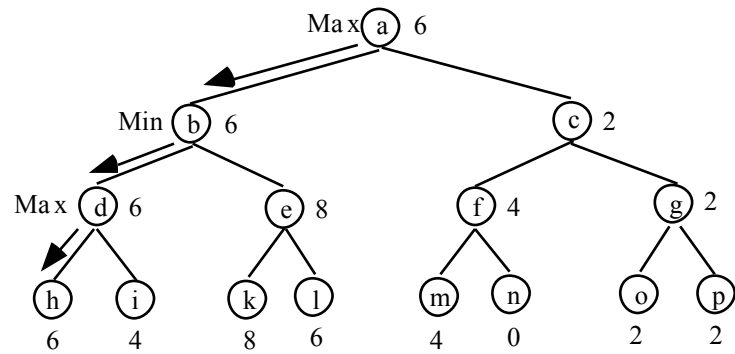
b) The cut branches need not be examined:
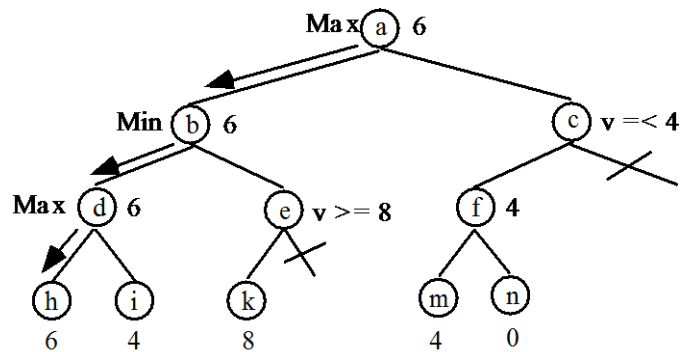
c) Using alpha-beta pruning on the mirror image of the tree does not produce any savings:
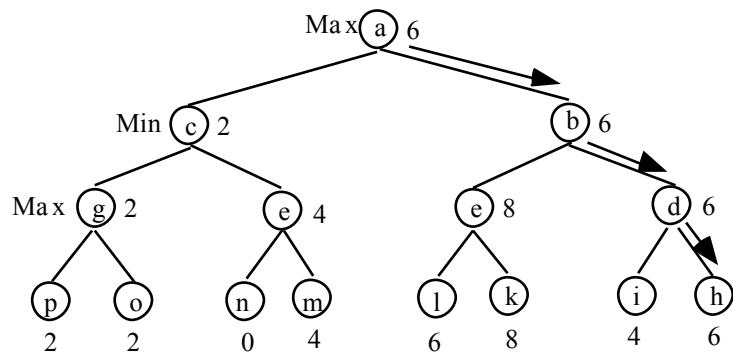
[Chapter 4: Knowledge Representation]
Artificial Intelligence (CSC 355)

Arjun Singh Saud

Central Department of Computer Science & Information Technology
Tribhuvan University

Knowledge Representation and Reasoning: Logic

Logical Agents

General features of a representation

Representational adequacy

The ability to represent all of the kinds of knowledge that are needed in a certain domain.

Inferential adequacy

The ability to represent all of the kinds of inferential procedures (procedures that manipulate the representational structures in such a way as to derive new structures corresponding to new knowledge inferred from old).

Inferential efficiency

The ability to represent efficient inference procedures (for instance, by incorporating into the knowledge structure additional information that can be used to focus the attention of the inference mechanisms in the most promising directions).

Acquisitional efficiency

The ability to acquire new information easily.

Logic

Logic (Knowledge-Based) agents combine general knowledge with current percepts to infer hidden aspects of current state prior to selecting actions
- Crucial in partially observable environments

Knowledge Base

Knowledge Base : set of sentences represented in a knowledge representation language and represents assertions about the world.
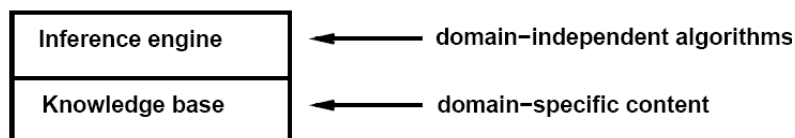
Then it can ask itself what to do-answers should follow from the KB
Agents can be viewed at the knowledge level
       i.e., what they know, regardless of how implemented
Or at the implementation level
       i.e., data structures in KB and algorithms that manipulate them

Abilities KB agent

Agent must be able to:
- Represent states and actions,
- Incorporate new percepts
- Update internal representation of the world
- Deduce hidden properties of the world
- Deduce appropriate actions

Logic

Logics are formal languages for representing information such that conclusions can be drawn: It has syntax and semantics:
- Syntax defines the sentences in the language
- Semantics define the meaning of sentences. that defines truth of a sentence in a world
-

E.g., the language of arithmetic

$x + 2 \geq y$ is a sentence; $x2 + y >$ is not a sentence

$x + 2 \geq y$ is true iff the number $x + 2$ is no less than the number y

$x + 2 \geq y$ is true in a world where x=7; y =1

$x + 2 \geq y$ is false in a world where x=0; y =6

E n t a i l m e n t
Entailment means that one thing follows from another:
      KB |= α

Knowledge base KB entails sentence α if and only if α is true in all worlds where KB is true

E.g., x + y =4 entails 4=x + y

Entailment is a relationship between sentences (i.e., syntax) that is based on semantics

M o d e l s

Logicians typically think in terms of models, which are formally structured worlds with respect to which truth can be evaluated.

m is a model of a sentence α if α is true in m.

$M(\alpha)$ is the set of all models of α

L o g i c a l   i n f e r e n c e

If an inference algorithm i can derive α from KB, we write KB $|_{-i}$ α which is pronounced as " i derives α from KB"

If an algorithm only derives entailed sentences it is called *sound* or *truth preserving*.

–   Otherwise it just makes things up.

i is sound if whenever KB |-i α it is also true that KB|= α

Completeness : the algorithm can derive any sentence that is entailed.

i is complete if whenever KB |= α it is also true that KB|-i α

Propositional Logic:Syntax

Propositional logic is the simplest logic. We use the symbols like P1, P2 to represent sentences. Propositional logic is defined as:

If S is a sentence, ¬S is a sentence (negation)
If S1 and S2 are sentences, S1 ^ S2 is a sentence (conjunction)
If S1 and S2 are sentences, S1 ∨ S2 is a sentence (disjunction)
If S1 and S2 are sentences, S1 ⇒ S2 is a sentence (implication)
If S1 and S2 are sentences, S1 ⇔ S2 is a sentence (biconditional)

Formal grammar for propositional logic can be given as below:

Sentennce   → AutomicSentence | ComplexSedntence
AutomicSentence   → True | False | Symbol
Symbol   → P | Q | R …………
ComplexSentence   → ¬Sentence
                                   | (Sentence ^ Sentence)
                                   | (Sentence ∨ Sentence)
                                   | (Sentence ⇒ Sentence)
                                   | (Sentence ⇔ Sentence)

Propositional Logic: Semantics

Each model specifes true/false for each proposition symbol

Rules for evaluating truth with respect to a model:
¬S is true if, S is false
S1 ^ S2 is true if, S1 is true and S2 is true
S1 ∨ S2 is true if, S1 is true or S2 is true
S1 ⇒ S2 is true if, S1 is false or S2 is true
S1 ⇔ S2 is true if, S1 ⇒ S2 is true and S2 ⇒ S1 is true

Truth Table showing the evaluation of semantics of complex sentences:

P	Q	¬P	P∧Q	P∨Q	P⇒Q	P⇔Q
false	false	true	false	false	true	true
false	true	true	false	true	true	false
true	false	false	false	true	false	false
true	true	false	true	true	true	true

Logical equivalence
Two sentences α and ß are *logically equivalent* (α ≡ ß) iff true they are true inn same set of models
or Two sentences α and ß are *logically equivalent* (α ≡ ß) iff α |= ß and ß |= α.

$$(\alpha \wedge \beta) \equiv (\beta \wedge \alpha) \quad \text{commutativity of } \wedge$$
$$(\alpha \vee \beta) \equiv (\beta \vee \alpha) \quad \text{commutativity of } \vee$$
$$((\alpha \wedge \beta) \wedge \gamma) \equiv (\alpha \wedge (\beta \wedge \gamma)) \quad \text{associativity of } \wedge$$
$$((\alpha \vee \beta) \vee \gamma) \equiv (\alpha \vee (\beta \vee \gamma)) \quad \text{associativity of } \vee$$
$$\neg(\neg\alpha) \equiv \alpha \quad \text{double-negation elimination}$$
$$(\alpha \Rightarrow \beta) \equiv (\neg\beta \Rightarrow \neg\alpha) \quad \text{contraposition}$$
$$(\alpha \Rightarrow \beta) \equiv (\neg\alpha \vee \beta) \quad \text{implication elimination}$$
$$(\alpha \Leftrightarrow \beta) \equiv ((\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha)) \quad \text{biconditional elimination}$$
$$\neg(\alpha \wedge \beta) \equiv (\neg\alpha \vee \neg\beta) \quad \text{de Morgan}$$
$$\neg(\alpha \vee \beta) \equiv (\neg\alpha \wedge \neg\beta) \quad \text{de Morgan}$$
$$(\alpha \wedge (\beta \vee \gamma)) \equiv ((\alpha \wedge \beta) \vee (\alpha \wedge \gamma)) \quad \text{distributivity of } \wedge \text{ over } \vee$$
$$(\alpha \vee (\beta \wedge \gamma)) \equiv ((\alpha \vee \beta) \wedge (\alpha \vee \gamma)) \quad \text{distributivity of } \vee \text{ over } \wedge$$

Validity

A sentence is *valid* if it is true in all models,
        e.g., *True*, A∨¬A, A ⇒ A, (A ∧ (A ⇒ B)) ⇒ B

Valid sentences are also known as tautologies
Every valid sentence is logically equivalent to True

Satisfiability

A sentence is *satisfiable* if it is true in *some* model
- e.g., A ∨ B, C

A sentence is *unsatisfiable* if it is true in *no* models
- e.g., A¬∧A

Validity and satisfiablity are related concepts
- α is valid iff ¬α is unsatisfiable
- α is satisfiable iff ¬α is not valid

Satisfiability is connected to inference via the following:
- $KB \models \alpha$ if and only if ($KB \land \neg \alpha$) is unsatisfiable

Inference rules in PL

Modus Ponens

$$\frac{\alpha \Rightarrow \beta, \alpha}{\beta}$$

And-elimination

$$\frac{\alpha \land \beta}{\alpha}$$

Monotonicity: the set of entailed sentences can only increase as information is added to the knowledge base.

For any sentence α and β if KB $\models$ α then KB ∧ β $\models$ α.

Resolution

Unit resolution rule:

Unit resolution rule takes a clause – a disjunction of literals – and a literal and produces a new clause. Single literal is also called unit clause.

$$\frac{\ell_1 \lor \cdots \lor \ell_k, \qquad m}{\ell_1 \lor \cdots \lor \ell_{i-1} \lor \ell_{i+1} \lor \cdots \lor \ell_k}$$

Where ℓ_i and m are complementary literals

Generalized resolution rule:
Generalized resolution rule takes two clauses of any length and produces a new clause as below.

$$\frac{\ell_1 \vee \cdots \vee \ell_k, \qquad m_1 \vee \cdots \vee m_n}{\ell_1 \vee \cdots \vee \ell_{i-1} \vee \ell_{i+1} \vee \cdots \vee \ell_k \vee m_1 \vee \cdots \vee m_{j-1} \vee m_{j+1} \vee \cdots \vee m_n}$$

For example

$$\frac{\ell_1 \vee \ell_2, \qquad \neg\ell_2 \vee \ell_3}{\ell_1 \vee \ell_3}$$

Resolution Uses CNF (Conjunctive normal form)
  − Conjunction of disjunctions of literals (clauses)
The resolution rule is sound:
  − Only entailed sentences are derived
Resolution is complete in the sense that it can always be used to either confirm or refute a sentence (it can not be used to enumerate true sentences.)

Conversion to CNF

A sentence that is expressed as a conjunction of disjunctions of literals is said to be in conjunctive normal form (CNF).
A sentence in CNF that contains only k literals per clause is said to be in k-CNF.

 Let's illustrate the conversion to CNF by using an example.

$B \Leftrightarrow (A \vee C)$

- Eliminate $\Leftrightarrow$, replacing $\alpha \Leftrightarrow \beta$ with $(\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha)$.
  − $(B \Rightarrow (A \vee C)) \wedge ((A \vee C) \Rightarrow B)$

- Eliminate $\Rightarrow$, replacing $\alpha \Rightarrow \beta$ with $\neg \alpha \vee \beta$.
  − $(\neg B \vee A \vee C) \wedge (\neg(A \vee C) \vee B)$

- Move $\neg$ inwards using de Morgan's rules and double-negation:
  − $(\neg B \vee A \vee C) \wedge ((\neg A \wedge \neg C) \vee B)$

- Apply distributivity law ($\wedge$ over $\vee$) and flatten:
  − $(\neg B \vee A \vee C) \wedge (\neg A \vee B) \wedge (\neg C \vee B)$

Resolution algorithm

Convert KB into CNF
Add negation of sentence to be entailed into KB i.e. (KB $\wedge \neg\alpha$)
Then apply resolution rule to resulting clauses.
The process continues until:
- There are no new clauses that can be added
  - Hence *KB* **does not** entail α
- Two clauses resolve to entail the empty clause.
  - Hence *KB* **does** entail α

Resolution example

Consider the knowledge base given as below:
$KB = (B \Leftrightarrow (A \vee C)) \wedge \neg B$

Prove that $\neg A$ can be inferred from above KB by using resolution.

Convert KB into CNF

$B \Rightarrow (A \vee C)) \wedge ((A \vee C) \Rightarrow B) \wedge \neg B$

$(\neg B \vee A \vee C) \wedge (\neg(A \vee C) \vee B) \wedge \neg B$

$(\neg B \vee A \vee C) \wedge ((\neg A \wedge \neg C) \vee B) \wedge \neg B$

$(\neg B \vee A \vee C) \wedge (\neg A \vee B) \wedge (\neg C \vee B) \wedge \neg B$

Add negation of sentence to be inferred from KB into KB

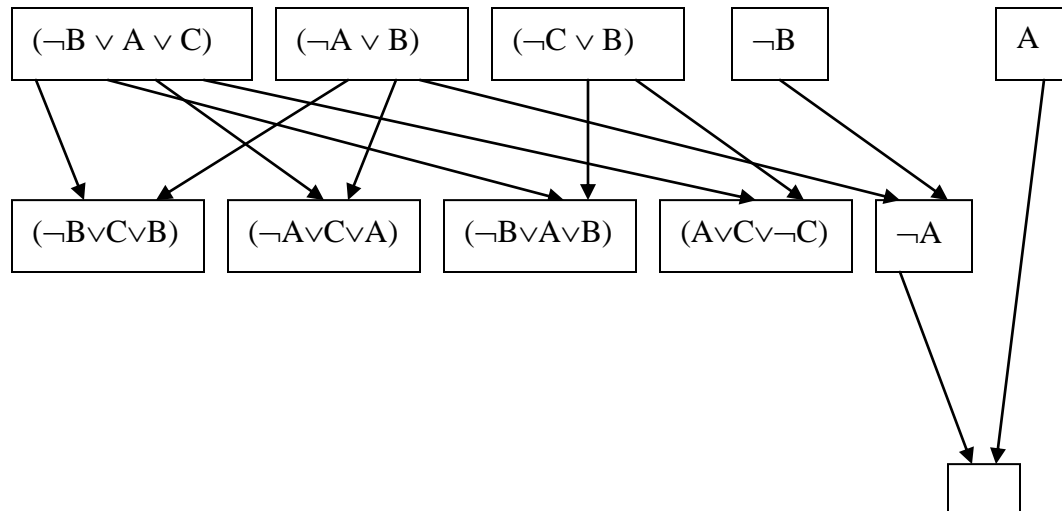Now KB contains following sentences all in CNF
$(\neg B \vee A \vee C)$
$(\neg A \vee B)$
$(\neg C \vee B)$
$\neg B$
A

Now use Resolution algorithm

Forward and backward chaining

The completeness of resolution makes it a very important inference model. But in many practical situations full power of resolution is not needed. Real-world knowledge bases often contain only clauses of restricted kind called **Horn Clause.**
A Horn clauses is disjunction of literals with at most one positive literal
Three important properties of Horn clause are:

- ✓ Can be written as an implication
- ✓ Inference through forward chaining and backward chaining.
- ✓ Deciding entailment can be done in a time linear size of the knowledge base.

Forward chaining

Idea: fire any rule whose premises are satisfied in the *KB*,
  − add its conclusion to the *KB*, until query is found
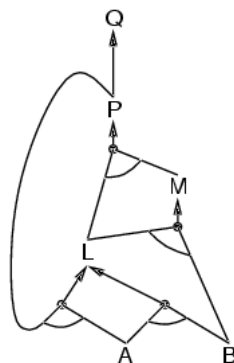
$$P \Rightarrow Q$$
$$L \wedge M \Rightarrow P$$
$$B \wedge L \Rightarrow M$$
$$A \wedge P \Rightarrow L$$
$$A \wedge B \Rightarrow L$$
$$A$$
$$B$$
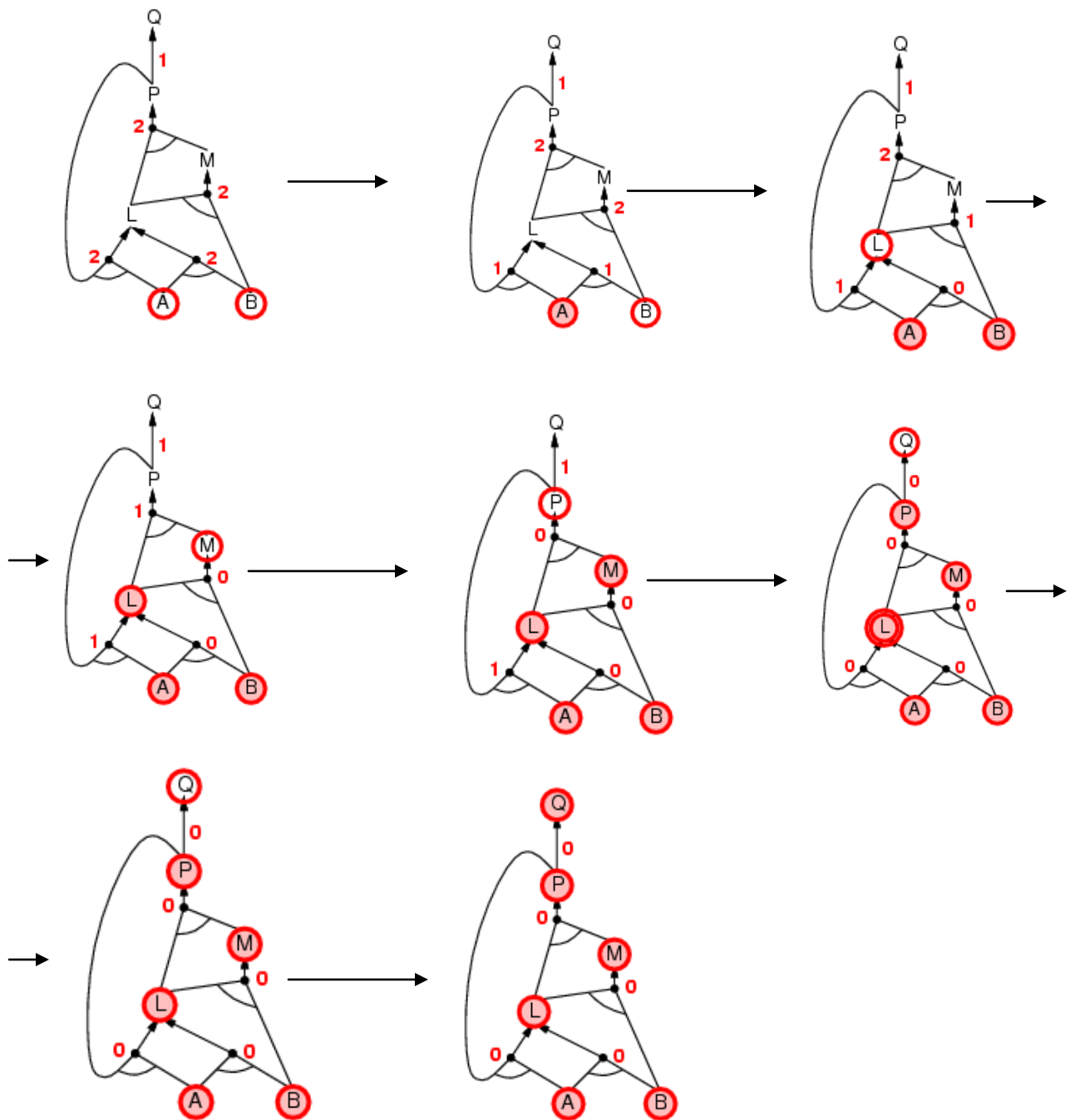
Backward chaining

Idea: work backwards from the query *q*: to prove *q* by BC,
Check if *q* is known already, or
Prove by BC all premises of some rule concluding *q*

Avoid loops: check if new subgoal is already on the goal stack

Avoid repeated work: check if new subgoal
1. has already been proved true, or
2. has already failed

Foe example for above KB

$P \Rightarrow Q$
$L \wedge M \Rightarrow P$
$B \wedge L \Rightarrow M$
$A \wedge P \Rightarrow L$
$A \wedge B \Rightarrow L$
A
B

Prove that Q can be inferred from above KB

We know $P \Rightarrow Q$, try to prove P
$L \wedge M \Rightarrow P$
Try to prove L and M
$B \wedge L \Rightarrow M$
$A \wedge P \Rightarrow L$
Try to prove B, L and A and P
A and B is already known, since $A \wedge B \Rightarrow L$, L is also known
Since, $B \wedge L \Rightarrow M$, M is also known
Since, $L \wedge M \Rightarrow P$, p is known, hence the proof

First-Order Logic

Pros and cons of propositional logic

Propositional logic is declarative

Propositional logic allows partial/disjunctive/negated information
   − (unlike most data structures and databases)

Propositional logic is compositional:
   − meaning of $B \wedge P$ is derived from meaning of B and of P

Meaning in propositional logic is context-independent
   − (unlike natural language, where meaning depends on context)

Propositional logic has very limited expressive power
   − (unlike natural language)
   − E.g., cannot say "pits cause breezes in adjacent squares"
      − except by writing one sentence for each square

First-order logic

Whereas propositional logic assumes the world contains facts,
first-order logic (like natural language) assumes the world contains
   − Objects: people, houses, numbers, colors, baseball games, wars, …
   − Relations: red, round, prime, brother of, bigger than, part of, comes between, …
   − Functions: father of, best friend, one more than, plus, …

Logics in General

The primary difference between PL and FOPL is their ontological commitment:
Ontological Commitment: What exists in the world — TRUTH
   − PL: facts hold or do not hold.
   − FL : objects with relations between them that hold or do not hold
Another difference is:
Epistemological Commitment: What an agent believes about facts — BELIEF

Language	Ontological Commitment	Epistemological Commitment
Propositional logic	facts	true/false/unknown
First-order logic	facts, objects, relations	true/false/unknown
Temporal logic	facts, objects, relations, times	true/false/unknown
Probability theory	facts	degree of belief $\in [0, 1]$
Fuzzy logic	degree of truth $\in [0, 1]$	known interval value

Representing knowledge in first-order logic

The objects from the real world are represented by constant symbols (a,b,c,...).
For instance, the symbol "Tom" may represent a certain individual called Tom.

Properties of objects may be represented by predicates applied to those objects (P(a), ...): e.g "male(Tom)" represents that Tom is a male.

Relationships between objects are represented by predicates with more arguments: "father(Tom, Bob)" represents the fact that Tom is the father of Bob.

The value of a predicate is one of the boolean constants T (i.e. true) or F (i.e. false)."father(Tom, Bob) = T" means that the sentence "Tom is the father of Bob" is true. "father(Tom, Bob) = F" means that the sentence "Tom is the father of Bob" is false.

Besides constants, the arguments of the predicates may be functions (f,g,...) or variables (x,y,...).

Function symbols denote mappings from elements of a domain (or tuples of elements of domains) to elements of a domain. For instance, weight is a function that maps objects to their weight: weight (Tom) = 150.Therefore the predicate greater-than (weight (Bob), 100) means that the weight of Bob is greater than 100. The arguments of a function may themselves be functions.

Variable symbols represent potentially any element of a domain and allow the formulation of general statements about the elements of the domain.

The quantifier's $\forall$ and $\exists$ are used to build new formulas from old ones.
"$\exists x\ P(x)$" expresses that there is at least one element of the domain that makes P(x) true.
"$\exists x\ mother(x, Bob)$" means that there is x such that x is mother of Bob or, otherwise stated, Bob has a mother.
"$\forall x\ P(x)$" expresses that for all elements of the domain P(x) is true.

Quantifiers
Allows us to express properties of collections of objects instead of enumerating objects by name. Two quantifiers are:
Universal: "for all" $\forall$
Existential: "there exists" $\exists$

Universal quantification:

$\forall <Variables> <sentence>$
Everyone at VUB is smart:
$$\forall x\ At(x,VUB) \Rightarrow Smart(x)$$
$\forall x\ P$ is true in a model m iff P is true for all x in the model

Roughly speaking, equivalent to the conjunction of instantiations of *P*

At(KingJohn,VUB) $\Rightarrow$ Smart(KingJohn) $\wedge$ At(Richard,VUB) $\Rightarrow$ Smart(Richard)
        $\wedge$At(VUB,VUB) $\Rightarrow$ Smart(VUB)$\wedge$ ...

Typically, $\Rightarrow$ is the main connective with $\forall$
- A universally quantifier is also equivalent to a set of implications over all objects

Common mistake: using $\wedge$ as the main connective with $\forall$:
$\forall$x At(x, VUB) $\wedge$ Smart(x)
means "Everyone is at VUB and everyone is smart"

Existential quantification

$\exists$<*variables*> <*sentence*>
Someone at VUB is smart:
        $\exists x$ At(x, VUB) $\wedge$ Smart(x)
$\exists x$ *P* is true in a model *m* iff *P* is true for at least one *x* in the model
Roughly speaking, equivalent to the disjunction of instantiations of *P*

At(KingJohn,VUB) $\wedge$ Smart(KingJohn)$\vee$At(Richard,VUB) $\wedge$ Smart(Richard)
$\vee$At(VUB, VUB) $\wedge$ Smart(VUB) $\vee$ ...

Typically, $\wedge$ is the main connective with $\exists$

Common mistake: using $\Rightarrow$ as the main connective with $\exists$:
$\exists x$ At(x, VUB) $\Rightarrow$ Smart(x)
        is true even if there is anyone who is not at VUB!

Properties of quantifiers
$\forall$x $\forall$y is the same as $\forall$y $\forall$x
$\exists$x $\exists$y is the same as $\exists$y $\exists$x

$\exists$x $\forall$y is not the same as $\forall$y $\exists$x
$\exists$x $\forall$y Loves(x,y)
- "There is a person who loves everyone in the world"
$\forall$y $\exists$x Loves(x,y)
- "Everyone in the world is loved by at least one person"

Quantifier duality: each can be expressed using the other
        $\forall$x Likes(x,IceCream) $\equiv$ $\exists\neg$x $\neg$Likes(x,IceCream)
        $\exists$x Likes(x,Broccoli) $\equiv$ $\neg\forall$x $\neg$Likes(x,Broccoli)

Equality

term1 = *term2* is true under a given interpretation if and only if *term1* and *term2* refer to the same object
e.g.
Father (John) = Henary
Here Objects referred by father of john and Henary is same.

Inference in first-order logic

First Order Logic to Predicate Logic

First order inference can be done by converting the knowledge base to PL and using propositional inference.
  − How to convert universal quantifiers?
    − Replace variable by ground term.
  − How to convert existential quantifiers?
    − Skolemization.

Universal instantiation (UI)
Substitute ground term (term without variables) for the variables.
For example consider the following KB
$\forall$ x King (x) $\wedge$ Greedy (x) $\Rightarrow$ Evil(x)
King (John)
Greedy (John)
Brother (Richard, John)

It's UI is:

King (John) $\wedge$ Greedy (John) $\Rightarrow$ Evil(John)
King (Richard) $\wedge$ Greedy (Richard) $\Rightarrow$ Evil(Richard)
King (John)
Greedy (John)
Brother (Richard, John)

Note: Remove universally quantified sentences after universal instantiation.

Existential instantiation (EI)
For any sentence α and variable v in that, introduce a constant that is not in the KB (called skolem constant) and substitute that constant for v.
e.g.
Consider the sentence.
$\exists$ x crown(x) $\wedge$ OnHead(x, John)

After EI

crown(C1) ∧ OnHead(C1, John)   where C1 is Skolem Constant.

EI versus UI

UI can be applied several times to *add* new sentences; the new KB is logically equivalent to the old.

EI can be applied once to replace the existential sentence; the new KB is not equivalent to the old but is satisfiable if the old KB was satisfiable.

Note: After propositionalization, we can use any algorithm fool Pl to derive the conclusion.

Unification and Lifting

Instead of translating the knowledge base to PL, we can redefine the inference rules into FOL.

- − Lifting: they only make those substitutions that are required to allow particular inferences to proceed.
- − E.g. *generalized Modus Ponens*
    - − To introduce substitutions different logical expressions have to be look identical
    - − To identify the substitution we need unification algorithm.

Unification:

We can get the inference immediately if we can find a substitution α such that *King(x)* and *Greedy(x)* match *King(John)* and *Greedy(y)*

α = {x/John,y/John} works

Unify(α ,β) = θ if αθ = θβ

p	q	θ
Knows(John,x)	Knows(John,Jane)	{x/Jane}
Knows(John,x)	Knows(y,OJ)	{x/OJ,y/John}
Knows(John,x)	Knows(y,Mother(y))	{y/John,x/Mother(John)}}
Knows(John,x)	Knows(x,OJ)	{fail}

Last unification is failed due to overlap of variables. x can not take the values of John and OJ at the same time.

We can avoid this problem by renaming to avoid the name clashes (standardizing apart)

e.g.

Unify{Knows(John,x)   Knows(z,OJ) } = {x/OJ, z/John}

Another complication:
To unify *Knows(John,x)* and *Knows(y,z)*,
Unification of Knows*(John,x)* and *Knows(y,z)* gives α = {y/John, x/z } or α={y/John, x/John, z/John}

First unifier gives the result Knows(John,z) and second unifier gives the resultKnows(John, John). Second can be achieved from first by substituting john in place of z. The first unifier is more general than the second.

There is a single most general unifier (MGU) that is unique up to renaming of variables.
MGU = { y/John, x/z }

Generalized Modus Ponens (GMP)

p1', p2', … , pn', (p1 ∧ p2 ∧ … ∧ pn ⇒q)

─────────────────────────────────   where pi'θ = piθ for all i

$\qquad$ qθ

p1' is King(John)   p1 is King(x)
p2' is Greedy(y)   p2 is Greedy(x)
θ is {x/John,y/John}   q is Evil(x)
qθ is Evil(John)

GMP used with KB of definite clauses (exactly one positive literal).

Forward Chaining:

We can use FC foe FOPL also:
Consider the following KB:
The law says that it is a crime for an American to sell weapons to hostile nations. The country Nono, an enemy of America, has some missiles, and all of its missiles were sold to it by Colonel West, who is American.

Prove that Col. West is a criminal

Soln:
it is a crime for an American to sell weapons to hostile nations:
American(x) ∧ Weapon(y) ∧ Sells(x,y,z) ∧ Hostile(z) ⇒ Criminal(x)

Nono … has some missiles, i.e.,
∃x Owns(Nono,x) ∧ Missile(x):
Use Existential Instantiation:
Owns(Nono,M1) and Missile(M1)

… all of its missiles were sold to it by Colonel West
Missile(x) ∧ Owns(Nono,x) ⇒ Sells(West,x,Nono)

Missiles are weapons:
Missile(x) ⇒ Weapon(x)
An enemy of America counts as "hostile":
Enemy(x,America) ⇒ Hostile(x)

West, who is American …
American(West)

The country Nono, an enemy of America …
Enemy(Nono,America)
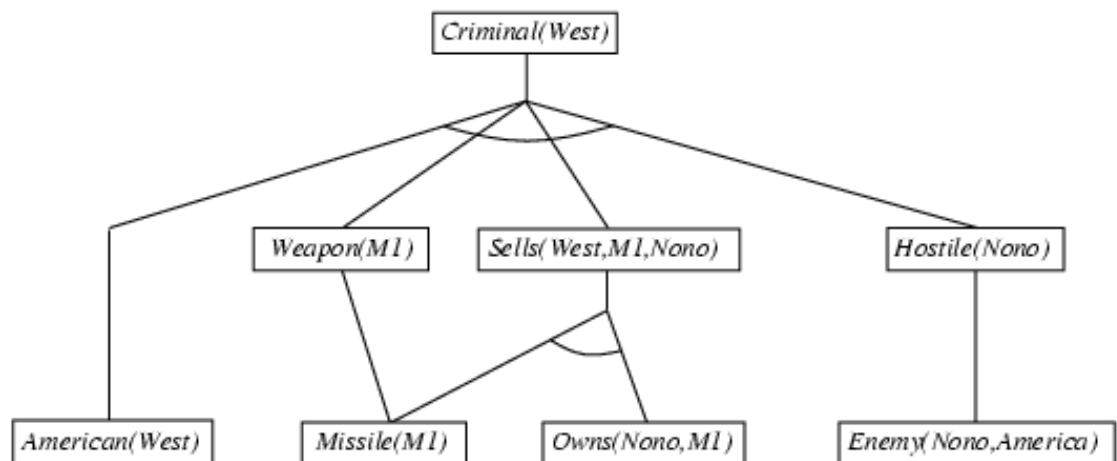
Backward Chaining:

We can also use BC algorithm for FOPL.
Eg. For above KB we can use BC al below:
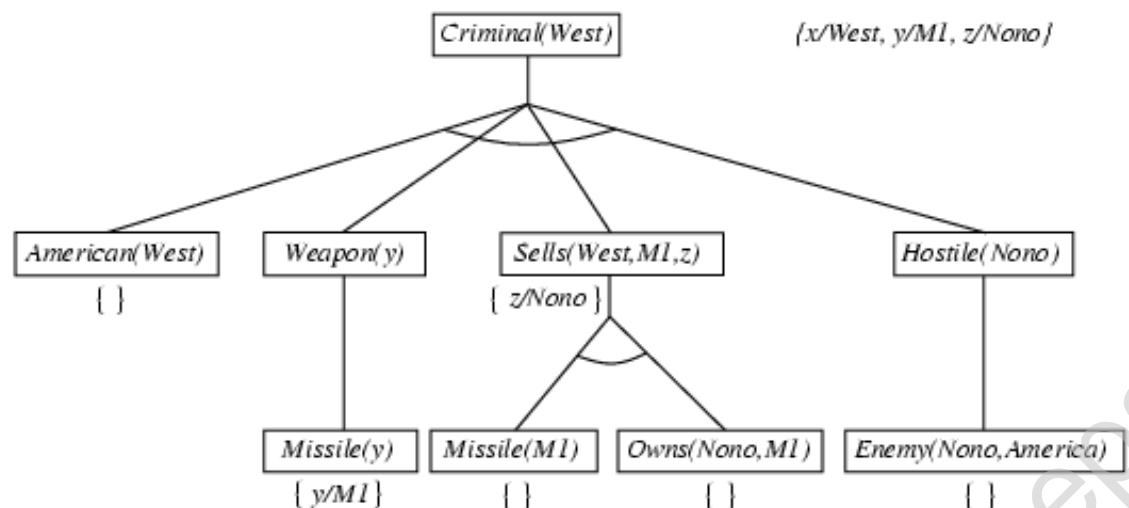
Exercise:

- Robert is an American. All Americans who sells weapons are criminals. All missiles are weapons. Robert sells weapons. Prove that Robert is criminal.
- SRK is a superstar. All superstars are rich. Rich mans have fast cars. Fast cars consume a lot of petrol. Prove that SRK's car consumes a lot of petrol.
- All greedy leaders are autocrat. Autocrats are evils. Prashant is a greedy leader. Ram is a honest leader. Prove that Prashant is evil.
- All students of this class takes a course in BSC IT. BSC IT students are good in AI. Ritu is student of this class. Prove that Ritu is good in AI.

[Chapter 6: Structured Knowledge Representation]
Artificial Intelligence (CSC 355)

Arjun Singh Saud

Central Department of Computer Science & Information Technology
Tribhuvan University

What is a representation?

A representation consists of two sets and a mapping between them. The elements of each set are objects, relations, classes, laws, actions. The first set is called the represented domain, and the second one is called the representation domain.
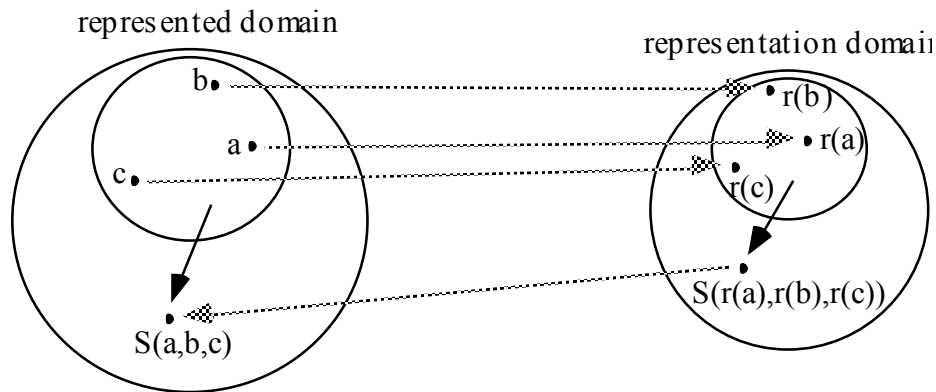
This mapping allows one to reason about the represented domain by performing reasoning processes in the representation domain and transferring the conclusions back into the represented domain (by using reverse mapping between the two sets).

The processes that manipulate the objects and the relationships in the representation domain are also part of the representation.

Knowledge representation is concerned with encoding the human knowledge into the form that is efficiently manipulated by the computer. Following are the kinds of knowledge that is represented in AI systems.

- ⇒ **Objects**: Guitar has Strings
- ⇒ **Events**: Ram played Guitar
- ⇒ **Performance**: Dirt cleaned by vacuum cleaner
- ⇒ **Meta-Knowledge**: Renuka knows that she can read street signs along the way to find where it is
- ⇒ **Facts**: Truths about real worldand what we represent

General features of a representation

- ⇒ **Representational adequacy**: The ability to represent all of the kinds of knowledge that are needed in a certain domain.

- ⇒ **Inferential adequacy**: The ability to represent all of the kinds of inferential procedures (procedures that manipulate the representational structures in such a way as to derive new structures corresponding to new knowledge inferred from old).

$\Rightarrow$ **Inferential efficiency**: The ability to represent efficient inference procedures (for instance, by incorporating into the knowledge structure additional information that can be used to focus the attention of the inference mechanisms in the most promising directions).

$\Rightarrow$ **Acquisitional efficiency**: The ability to acquire new information easily.

Knowledge Representation schemes can be classified as below:
$\Rightarrow$ Logical representation Scheme
$\Rightarrow$ Procedural representation Scheme
$\Rightarrow$ Network representation Scheme
$\Rightarrow$ Structured representation Scheme

Logical representation Scheme
It uses expression in the form of formal notation to represent the knowledge base. Propositional logic and predicate logic are examples of the logical representation Scheme. Give example your self.

Procedural representation Scheme
These representation schemes constitute rule based expert systems.
If type of light is sunlight then light is bright
If type of light is bulb then light is dim
If location is outdoor then light is sunlight
If location is indoor then light is bulb

Network representation Scheme
Network representation schemes represents knowledge in the form of a graph in which the nodes represent objects, situations, or events, and the arcs represent the relationships between them. Semantic networks, conceptual dependencies and conceptual graphs are examples of network representation.(Give example of semantic net yourself)

Structured representation Scheme
This is the extension of network representation schemes where nodes are complex data structures consisting of named slots with attached values. Scripts, frames and objects are examples of structured representation scheme.

Representing knowledge in semantic networks

The underlying idea of semantic networks is to represent knowledge in the form of a graph in which the nodes represent objects, situations, or events, and the arcs represent the relationships between them.

Semantic networks with binary relations

The following is an example of a semantic network with binary relations:
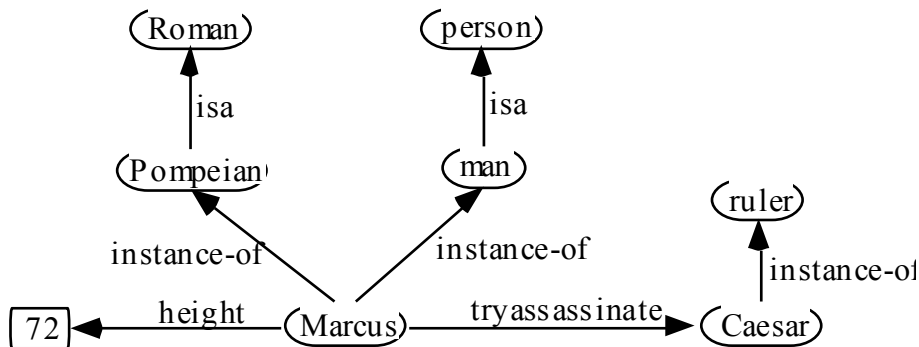
Some nodes represent individual objects (Marcus, Caesar) or constants (72).Other nodes represent classes of individuals (Pompeian, Roman, man, person, ruler). The relationship "instance-of" asserts that an individual is an instance (element) of a class. The relationship "isa" asserts that a class, (e.g. "man"), is a subclass of another class ("person"). "Man isa person" means that any element of man is also an element of person. That is, if someone is a man then he is a person. The relationship "tryassassinate", from Marcus to Caesar, states that Marcus tried to assassinate Caesar. The relationship "height", from Marcus to 72, states that the height of Marcus is 72. Marcus and Caesar are called instances because they represent individual objects.

The following is another example of a semantic network representing knowledge about our solar system. One may notice that this network states that the temperature of the Sun is greater than the temperature of the Earth, without indicating the values of these temperatures:
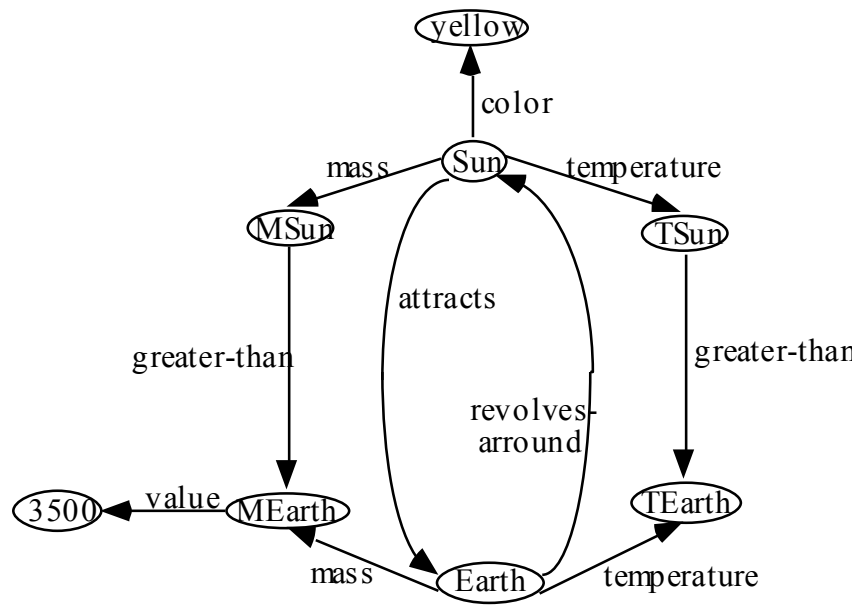
Representing non-binary predicates

Suppose we wish to represent the fact that Clyde, which is a robin, owns a nest. This may be encoded as follows:

.
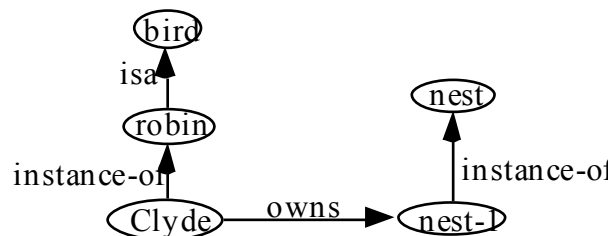
"nest-1" is the nest Clyde owns. It is an instance of "nest" which represents the class of all nests. Suppose one wants to encode the additional information that Clyde owned nest-1 from Spring 90 to Fall 90. This is impossible to do in the above network because the ownership situation is encoded as a link, and links, by their nature, can encode only binary relations.

The solution is to represent this ownership situation as a node characterized by four properties: the owner (Clyde), the object owned (nest-1), the time when the ownership started (Spring 90), and the time when the ownership ended (Fall 90). In this way one may represent the equivalent of a four-place predicate:

                  ownership(owner, ownee, start-time, end-time)

The corresponding semantic network is the following one:
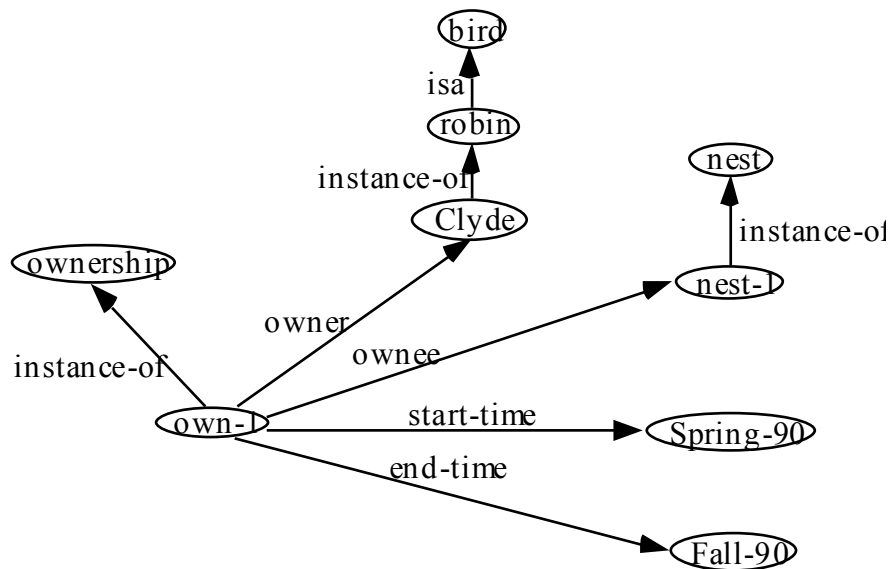
Therefore, the semantic networks are not limited to representing only binary relationships

Network matching

A network fragment is constructed, representing a sought-for object or a query, and then matched against the network database to see if such an object exists. Variable nodes in the fragment are bound in the matching process to the values with which they match perfectly.
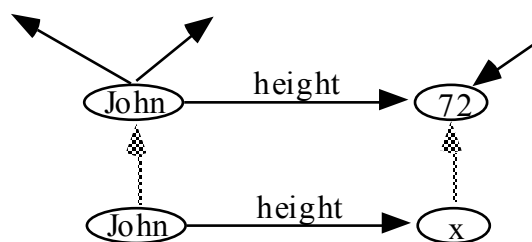
Question:
What is the height of John
Answer:
The height of John is 72.

As another example, let us suppose that we use the network given below as a database and suppose we wish to answer the question "What does Clyde own?".
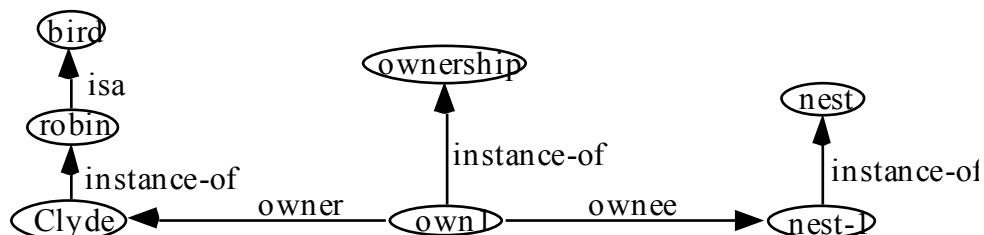
We might construct the network in above figure which represents an instance of ownership in which Clyde is the owner. This fragment is then matched against the network database looking for an own node that has an owner link to Clyde. When it is found, the node that the ownee link points to is bound in the partial match and is the answer to the question.
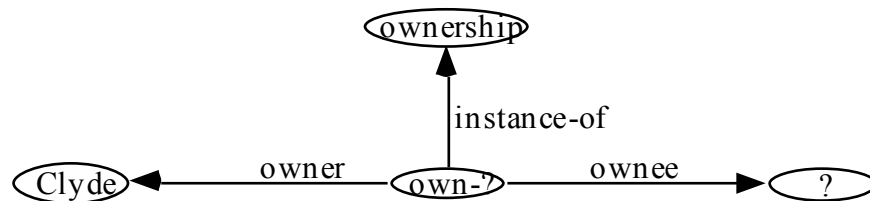
The matcher can make inferences during the matching process to create a network structure that is not explicitly present in the network.

For example, suppose we wish to answer the question
        "Is there a bird who owns a nest?"
We could translate that question into the following network fragment:
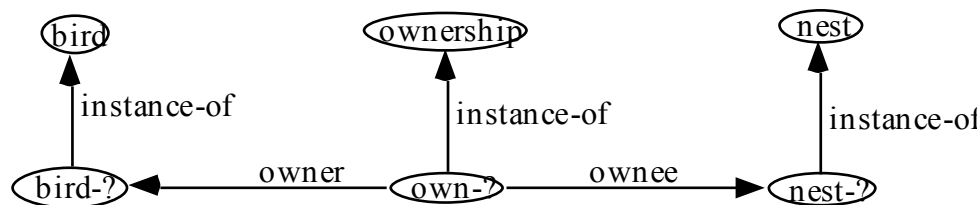
Here, bird-?, nest-?, and own-? nodes represent the yet to be determined bird-owning-nest relation.

Notice that the query network fragment does not match the knowledge base exactly. The deduction procedure would have to construct an "instance-of" link from Clyde to bird to make the match possible. The matcher would bind bird-? to the node Clyde, own-? to own-1, and nest-? to nest-1, and the answer to the question would be "Yes, Clyde".

Exercises

1. Represent the following sentences into a semantic network.
        Birds are animals.
        Birds have feathers, fly and lay eggs.
        Albatros is a bird.
        Donald is a bird.
        Tracy is an albatros.

Solution:
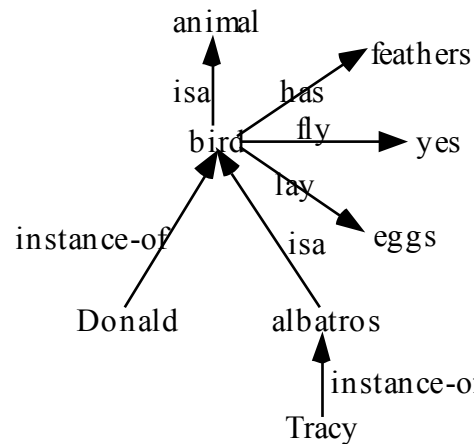
2. Represent the following sentences into a semantic network:

      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.

Solution:
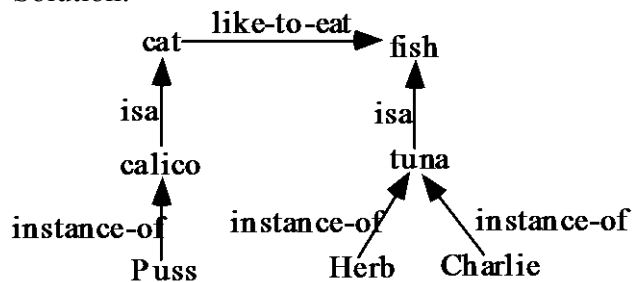
3. Represent the following sentence with a semantic network:
      John lived in Fairfax from May 1985 till February 1990

4. Consider the following network fragment:
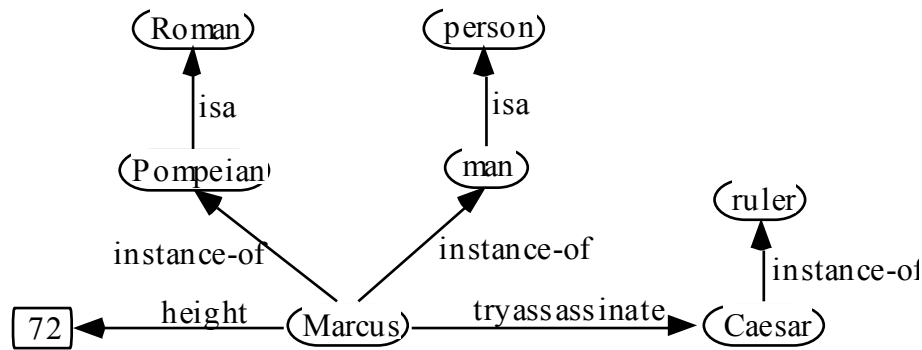
Explain how a semantic network system would answer the questions:

What is the height of Marcus?

Is there a person who tried to assassinate Caesar?

Object-based Approach: Frames

With this approach, knowledge may be represented in a data structure called a **frame**. A *frame* is a data structure containing typical knowledge about a concept or object (Marvin Minsky (mid 1970s)). A frame represents knowledge about real world things (or entities).

Each frame has **a name and slots**. **Slots** are the properties of the entity that has the name, and they have **values** or pointer to other frames (a table like data structure). A particular value may be:
- a default value
- an inherited value from a higher frame
- a procedure, called a daemon, to find a value
- a specific value, which might represent an exception.

When the slots of a frame are all filled, the frame is said to be **instantiated**, it now represents a specific entity of the type defined by the unfilled frame. Empty frames are sometimes called object prototypes

The idea of frame hierarchies is very similar to the idea of class hierarchies found in object-orientated programming. Frames are an application of the object-oriented approach to knowledge-based systems.

Disadvantages

- complex
- reasoning (inferencing) is difficult
- explanation is difficult, expressive limitation

Advantages

- knowledge domain can be naturally structured [a similar motivation as for the O-O approach].
- easy to include the idea of default values, detect missing values, include specialised procedures and to add further slots to the frames

Examples:
(1.)
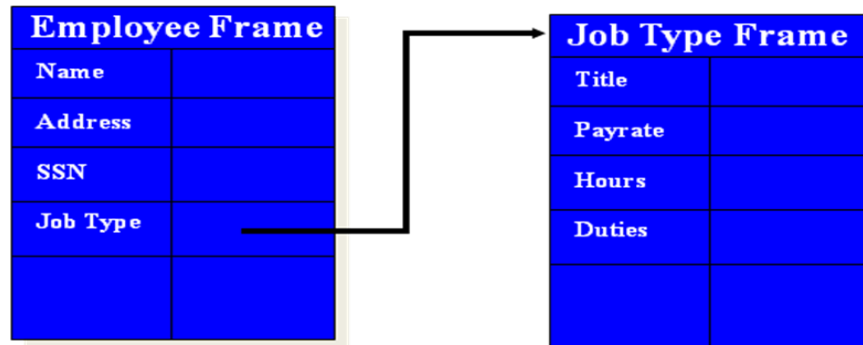
(2.)
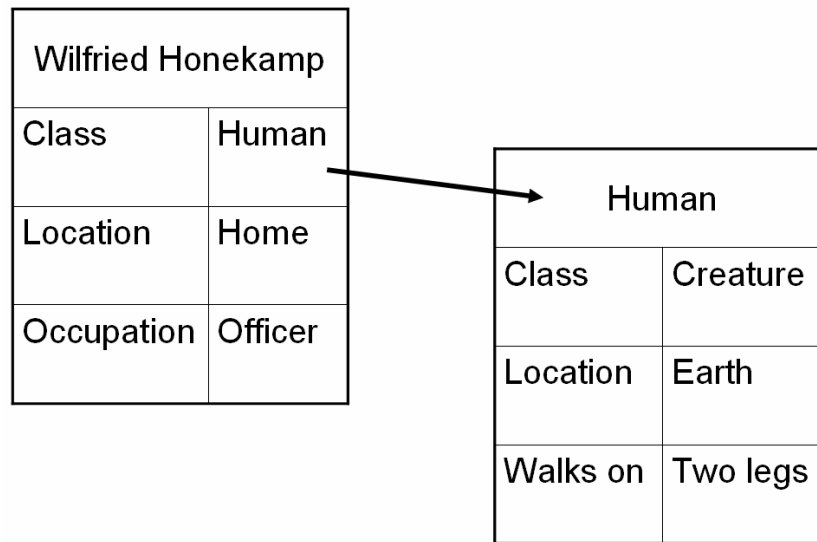
Fig: Two frames describing a human being

Conceptual Dependencies:

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.

- That is to say: *For any 2 (or more) sentences that are identical in meaning there should be only one representation of that meaning.*

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.

dfd

Conceptual Dependency (CD) is a theory of how to represent the kind of knowledge about the events that is typically contained in natural language processing sentences. The goal is to represent the knowledge in a way that

(i)   Facilitates drawing the inferences from the natural language sentences.
(ii) Is independent of the language in which natural language sentences were originally stated.

Because of the two concerns mentioned, the CD representation of a sentence is built not out of primitives corresponding to the words in the sentence, but rather out of conceptual primitives that can be combined to form the semantic meanings of the words in any particular language. The conceptual dependency theory was first developed by Schank in 1973 and was further more developed in 1975 by the same author. It has been implemented in a variety of programs that read and understand natural language text processing. Unlike semantic nets which provide only a structure into which nodes representing information and a specific set of primitives, at a particular level of granularity, out of which representations of particular pieces of information can be constructed.

In CD's the symbols have the following meanings.

Arrows indicate the decision of dependency.

Double arrow indicates two way link between actor and action

P indicates past tense.

ATRANS is one of the primitive symbol acts used by the CD theory. It indicates transfer of possession

O indicates the object case relation

R indicates the recipient case relation.

As a simple example of the way knowledge is represented in conceptual dependency theory, the event represented by the sentence.

'I gave the man a book'

would be represented in conceptual dependency theory, the event represented by the sentence as shown.

In Conceptual dependency, the representations of actions are built from a set of primitive acts. Although there are slight differences in the exact set of primitives actions provided inn the various sources on CD, a typical set is the following, which are taken from the Schanks conceptual dependency theory.

- RANS – transfer of an abstract relationship (give)
- PTRANS – transfer of the physical location of an object (go)
- PROPEL – application of physical force to an object (push)
- MOVE – movement of a body part by ots owner (kick)
- GRASP – grasping of an object by an actor (clutch)
- INGEST – ingestion of an object by an animal (eat)
- EXPEL – Expulsion of something from the body of an animal (cry)
- MTRANS – Transfer of mental information (tell)
- MBUILD – Building new information out of old (decide)
- SPEAK – production of sounds (say)
- ATTEND – Focusing of a sense organ toward a stimulus (listen)

Conceptual Dependency the symbols have the following meanings.

• Single arrows (<—) indicate direction of dependency.
• Double arrows (o) indicates two way link between action and actor.
• P indicates past tenst.
• Primitives (ATRANS, PTRANS, PROPEL) indicates transfer of possession.
• O indicates object relation.
• R indicates recipient relation.

Conceptual Dependency introduced several interesting ideas:

(i) An internal representation that is language free, using primitive ACTs instead.
(ii) A small number of primitive ACTs rather than thousands.
(iii) Different ways of saying the same thing can map to the same internal representation.

There are some problems too:

(i) the expansion of some verbs into primitive ACT structures can be complex
(ii) Graph matching in NP-hard.

Conceptual dependency theory offers a set of primitives conceptualizations from which the world of meaning is build.
These are equal and independent, They are :

ACTs : Actions
PPs : Object (picture procedure)
AAs : Modifiers of actions (action aiders)
PAs : Modifiers of objects i.e., PPS (picture aiders)
AA   : Action aiders are properties or attributes of primitive actions.
PP   : Picture produces are actors or physical objects that perform different acts or produces

Scripts

A *script* is a structure that prescribes a set of circumstances which could be expected to follow on from one another. Scripts are used for representing knowledge about common sequences of events.

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 specialised 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.

The components of a script include:

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.

The classic example is the restaurant script:

Scene: A restaurant with an entrance and tables.
Actors: The diners, servers, chef.
Props: The table setting, menu, table, chair.
Acts: Entry, Seating, Ordering a meal, Serving a meal, Eating the meal, requesting the check, paying, leaving.

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.

Uncertain Knowledge

Let action A_t = leave for airport t minutes before flight

Will A_t get me there on time?

Problems:

1. Partial observability (road state, other drivers' plans, etc.)
2. Noisy sensors (radio traffic reports)
3. Uncertainty in action outcomes (flat tyre, etc.)
4. Complexity of modeling and predicting traffic

Hence a purely logical approach either

1. Risks falsehood: "A25 will get me there on time" or

---

2. Leads to conclusions that are too weak for decision making: "A25 will get me there on time if there's no accident on the bridge and it doesn't rain and my tires remain intact etc etc."

A1440 might reasonably be said to get me there on time but I'd have to stay overnight in the airport…

Handling Uncertainty

Instead of providing all condition it can express with degree of beliefs in the relevant sentences.

Example:

Say we have a rule

if toothache then problem is cavity

But not all patients have toothaches because of cavities (although perhaps most do)

So we could set up rules like

if toothache and not(gum disease) and not(filling) and ......then problem is cavity

This gets very complicated! a better method would be to say

if toothache then problem is cavity with probability 0.8

Given the available evidence,

A25 will get me there on time with probability 0.04

A most important tool for dealing with degree of beliefs is probability theory, which assigns to each sentence a numerical degree of belief between 0 & 1.

Probability Basics:

Probability of getting 1 when a die is rolled

$P(1) = 1/6$

Probability of getting a number < 4 when a die is rolled

$P(< 4) = P(1) + P(2) + P(3) = 1/6 + 1/6 + 1/6 = 1/2$

Proposition = Set of atomic events in which it is true.

$(a \lor b) = (\neg a \land b) \lor (a \land \neg b) \lor (a \land b)$

$P(a \lor b) = P(\neg a \land b) \lor P(a \land \neg b) \lor P(a \land b)$

$(a \land b) = a \lor b - ($

The **prior or unconditional** probability associated with a proposition is the degree of belief accorded to it in the absence of any other information.

Example:

P(Weather = sunny) = 0.72

P(Cavity = true) = 0.1 or P(cavity) = 0.1

Probability distribution gives values for all possible assignments:

P(Weather) = (0.72, 0.1, 0.08, 0.1)

The conditional probability "P(a|b)" is the probability of "a" given that all we know is "b".

e.g., *P(cavity|toothache) = 0.8* means if a

Patients have toothache and no other information is yet available, then the probability of patient's having the cavity is 0.8

$P(a|b) = P(a \wedge b)/ P(b)$

$P(a \wedge b) = P(a|b) P(b)$

We consider the following domain consisting of three Boolean variables: Toothache, Cavity, and Catch (the dentist's nasty steel probe catches in my tooth).

The full joint distribution is the following 2x2x2 table:

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
Cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

P(cavity v toothache) = P(cavity, toothache, catch) + P(cavity, toothache, ¬catch) +
P(cavity, ¬toothache, catch) + P(cavity, ¬toothache, ¬catch) +
P(¬cavity, toothache, catch) + P(¬cavity, toothache, ¬catch)
= 0.108+0.012+0.072+0.008+0.016+0.064=0.28

P(cavity) = 0.108 +0.012 +0.072 + 0.008 = 0.2

P(¬Toothache) = 0.072 + 0.008 + 0.144 + 0.576 = 0.8

P(Cavity, ¬Toothache) = 0.072 + 0.008 = 0.08

P(¬cavity | Toothache) = P(¬cavity ^ Toothache)/ P(Toothache)

$\qquad$ = (0.016 + 0.064)/(0.108 + 0.012 + 0.016 + 0.064)

$\qquad$ = 0.4

Bayes Rule

P(b|a) = P(a|b) * P(b) / P(a)

[Unit 5: Machine Learning]
Artificial Intelligence (CSC 355)

Bal Krishna Subedi

Central Department of Computer Science & Information Technology
Tribhuvan University

What is Learning?

"Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the same task (or tasks drawn from the same population) more effectively the next time." --Herbert Simon

"Learning is constructing or modifying representations of what is being experienced." --Ryszard Michalski

"Learning is making useful changes in our minds." --Marvin Minsky

Types of Learning:

The strategies for learning can be classified according to the amount of inference the system has to perform on its training data. In increasing order we have

1. **Rote learning** – the new knowledge is implanted directly with no inference at all, e.g. simple memorisation of past events, or a knowledge engineer's direct programming of rules elicited from a human expert into an expert system.

2. **Supervised learning** – the system is supplied with a set of training examples consisting of inputs and corresponding outputs, and is required to discover the relation or mapping between then, e.g. as a series of rules, or a neural network.

3. **Unsupervised learning** – the system is supplied with a set of training examples consisting only of inputs and is required to discover for itself what appropriate outputs should be, e.g. a *Kohonen Network* or *Self Organizing Map*.

Early expert systems relied on rote learning, but for modern AI systems we are generally interested in the supervised learning of various levels of rules.

The need for Learning:

As with many other types of AI system, it is much more efficient to give the system enough knowledge to get it started, and then leave it to learn the rest for itself. We may even end up with a system that learns to be better than a human expert.

The *general learning approach* is to generate potential improvements, test them, and discard those which do not work. Naturally, there are many ways we might generate the potential improvements, and many ways we can test their usefulness. At one extreme, there are model driven (top-down) generators of potential improvements, guided by an understanding of how the problem domain works. At the other, there are data driven (bottom-up) generators, guided by patterns in some set of training data.

Bal Krishna Subedi

Machine Learning:

As regards machines, we might say, very broadly, that a machine learns whenever it changes its structure, program, or data (based on its inputs or in response to external information) in such a manner that its expected future performance improves. Some of these changes, such as the addition of a record to a data base, fall comfortably within the province of other disciplines and are not necessarily better understood for being called learning. But, for example, when the performance of a speech-recognition machine improves after hearing several samples of a person's speech, we feel quite justified in that case saying that the machine has learned.

Machine learning usually refers to the changes in systems that perform tasks associated with artificial intelligence (AI). Such tasks involve recognition, diagnosis, planning, robot control, prediction, etc. The changes might be either enhancements to already performing systems or synthesis of new systems.

Learning through Examples: (A type of Concept learning)

Concept learning also refers to a learning task in which a human or machine learner is trained to classify objects by being shown a set of example objects along with their class labels. The learner will simplify what has been observed in an example. This simplified version of what has been learned will then be applied to future examples. Concept learning ranges in simplicity and complexity because learning takes place over many areas. When a concept is more difficult, it will be less likely that the learner will be able to simplify, and therefore they will be less likely to learn. This learning by example consists of the idea of **version space**.

A **version space** is a hierarchical representation of knowledge that enables you to keep track of all the useful information supplied by a sequence of learning examples without remembering any of the examples.

The **version space method** is a concept learning process accomplished by managing multiple models within a version space.

Version Space Characteristics

In settings where there is a generality-ordering on hypotheses, it is possible to represent the version space by two sets of hypotheses: (1) the **most specific** consistent hypotheses and (2) the **most general** consistent hypotheses, where "consistent" indicates agreement with observed data.

The most specific hypotheses (i.e., the specific boundary **SB**) are the hypotheses that cover the observed positive training examples, and as little of the remaining feature space as possible. These are hypotheses which if reduced any further would *exclude* a *positive* training example, and hence become inconsistent. These minimal hypotheses essentially constitute a (pessimistic) claim that the true concept is defined just by the *positive* data

already observed: Thus, if a novel (never-before-seen) data point is observed, it should be assumed to be negative. (I.e., if data has not previously been ruled in, then it's ruled out.)

The most general hypotheses (i.e., the general boundary **GB**) are those which cover the observed positive training examples, but also cover as much of the remaining feature space without including any negative training examples. These are hypotheses which if enlarged any further would *include* a *negative* training example, and hence become inconsistent.

Tentative heuristics are represented using version spaces. A version space represents all the alternative plausible **descriptions** of a heuristic. A plausible description is one that is applicable to all known positive examples and no known negative example.

A version space description consists of two complementary trees:

1. One that contains nodes connected to overly **general** models, and
2. One that contains nodes connected to overly **specific** models.

Node values/attributes are **discrete**.

Fundamental Assumptions

1. The data is correct; there are no erroneous instances.
2. A correct description is a conjunction of some of the attributes with values.

Diagrammatical Guidelines

There is a **generalization** tree and a **specialization** tree.

Each **node** is connected to a **model**.

Nodes in the generalization tree are connected to a model that matches everything in its subtree.

Nodes in the specialization tree are connected to a model that matches only one thing in its subtree.

Links between nodes and their models denote

- generalization relations in a generalization tree, and
- specialization relations in a specialization tree.

Diagram of a Version Space

In the diagram below, the specialization tree is colored **red**, and the generalization tree is colored **green**.

The most general model matches everything

Negative instances specialize general descriptions

Positive instances prune the general descriptions

Eventually, positive and negative samples may force the general and specific models to converge on a solution

Negative instances prune the specific descriptions

Positive instances generalize specific descriptions

The most specific model matches only one thing

Generalization and Specialization Leads to Version Space Convergence

The key idea in version space learning is that specialization of the general models and generalization of the specific models may ultimately lead to just one correct model that matches all observed positive examples and does not match any negative examples.

That is, each time a negative example is used to specialilize the general models, those specific models that match the negative example are eliminated and each time a positive example is used to generalize the specific models, those general models that fail to match the positive example are eliminated. Eventually, the positive and negative examples may be such that only one general model and one identical specific model survive.

Bal Krishna Subedi

Candidate Elimination Algorithm:

The version space method handles positive and negative examples symmetrically.

Given:

- A representation language.
- A set of positive and negative examples expressed in that language.

Compute: a concept description that is consistent with all the positive examples and none of the negative examples.

Method:

- Initialize G, the set of maximally general hypotheses, to contain one element: the null description (all features are variables).
- Initialize S, the set of maximally specific hypotheses, to contain one element: the first positive example.
- Accept a new training example.
    - If the example is **positive**:
        1. Generalize all the specific models to match the positive example, but ensure the following:
            - The new specific models involve minimal changes.
            - Each new specific model is a specialization of some general model.
            - No new specific model is a generalization of some other specific model.
        2. Prune away all the general models that fail to match the positive example.
    - If the example is **negative**:
        1. Specialize all general models to prevent match with the negative example, but ensure the following:
            - The new general models involve minimal changes.
            - Each new general model is a generalization of some specific model.
            - No new general model is a specialization of some other general model.
        2. Prune away all the specific models that match the negative example.
    - If S and G are both singleton sets, then:
        - if they are identical, output their value and halt.
        - if they are different, the training cases were inconsistent. Output this result and halt.
        - else continue accepting new training examples.

The algorithm stops when:

1. It runs out of data.
2. The number of hypotheses remaining is:
   - o   0 - no consistent description for the data in the language.
   - o   1 - answer (version space converges).
   - o 2^+ - all descriptions in the language are implicitly included.

Problem 1:

Learning the concept of "Japanese Economy Car"

Features: (Country of Origin, Manufacturer, Color, Decade, Type)

Origin	Manufacturer	Color	Decade	Type	Example Type
Japan	Honda	Blue	1980	Economy	Positive
Japan	Toyota	Green	1970	Sports	Negative
Japan	Toyota	Blue	1990	Economy	Positive
USA	Chrysler	Red	1980	Economy	Negative
Japan	Honda	White	1980	Economy	Positive

Solution:

1. **Positive Example**: (Japan, Honda, Blue, 1980, Economy)

Initialize G to a singleton set that includes everything. G = { (?, ?, ?, ?, ?) }
Initialize S to a singleton set that includes the first positive example. S = { (Japan, Honda, Blue, 1980, Economy) }
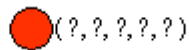
These models represent the most general and the most specific heuristics one might learn. The actual heuristic to be learned, "Japanese Economy Car", probably lies between them somewhere within the version space.

2. **Negative Example**: (Japan, Toyota, Green, 1970, Sports)

Specialize G to exclude the negative example.

Bal Krishna Subedi

$$G = \begin{cases} (?, Honda, ?, ?, ?), \\ (?, ?, Blue, ?, ?), \\ (?, ?, ?, 1980, ?), \\ (?, ?, ?, ?, Economy) \end{cases}$$

S = { (Japan, Honda, Blue, 1980, Economy) }

Refinement occurs by generalizing S or specializing G, until the heuristic hopefully converges to one that works well.
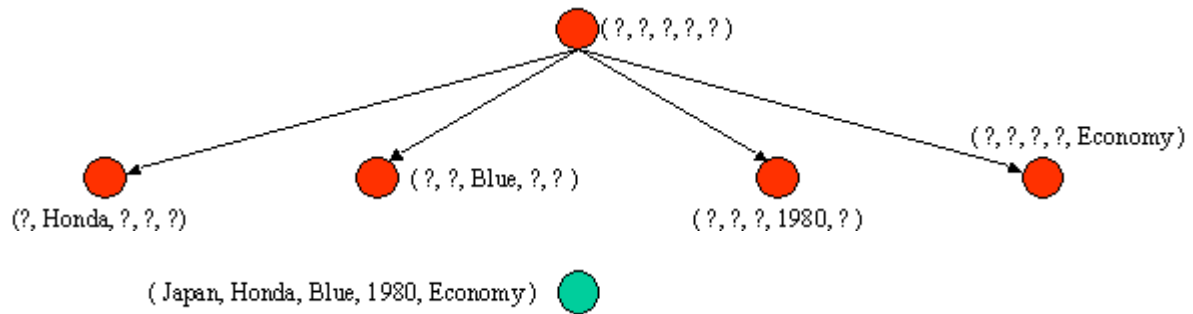
3. **Positive Example**: (Japan, Toyota, Blue, 1990, Economy)

Prune G to exclude descriptions inconsistent with the positive example. (Prune = ✂)
Generalize S to include the positive example.

$$G = \begin{cases} (?, ?, Blue, ?, ?), \\ (?, ?, ?, ?, Economy) \end{cases}$$

S = { (Japan, ?, Blue, ?, Economy) }

4. **Negative Example**: (USA, Chrysler, Red, 1980, Economy)

Specialize G to exclude the negative example (but stay consistent with S)

Bal Krishna Subedi

G = { (?, ?, Blue, ?, ?),
      (Japan, ?, ?, ?, Economy) }
S = { (Japan, ?, Blue, ?, Economy) }

5. **Positive Example**: (Japan, Honda, White, 1980, Economy)

Prune G to exclude descriptions inconsistent with positive example.
Generalize S to include positive example.

G = { (Japan, ?, ?, ?, Economy) }
S = { (Japan, ?, ?, ?, Economy) }
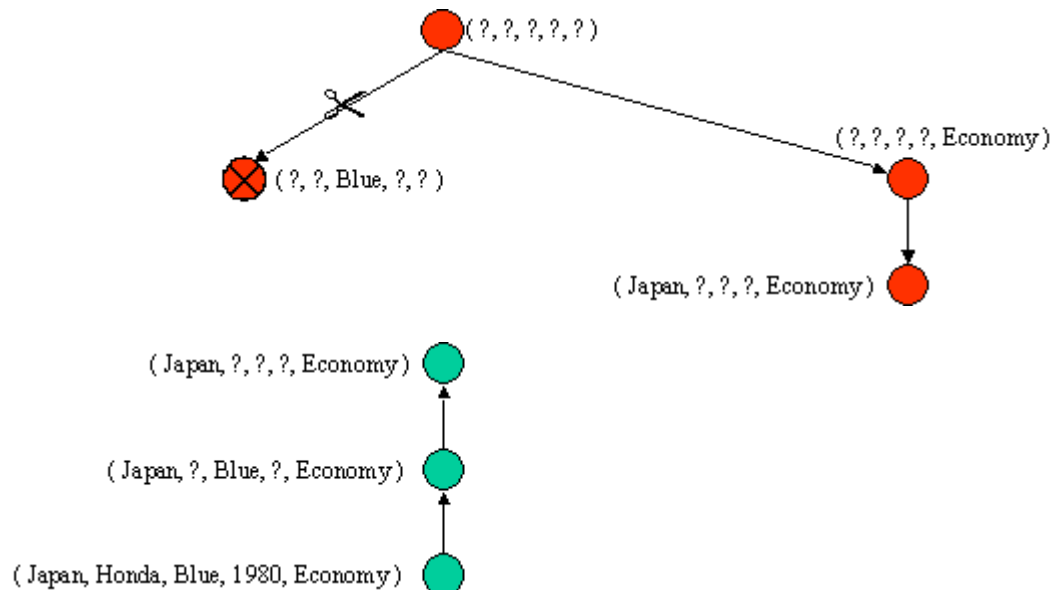
G and S are singleton sets and S = G.
Converged.
No more data, so algorithm stops.
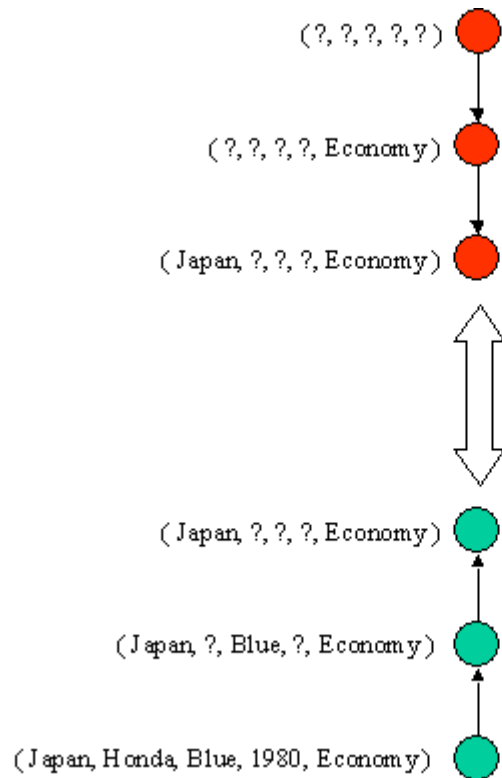
Explanation Based Machine Learning:

Explanation-based learning (**EBL)** is a form of machine learning that exploits a very strong, or even perfect, domain theory to make generaliza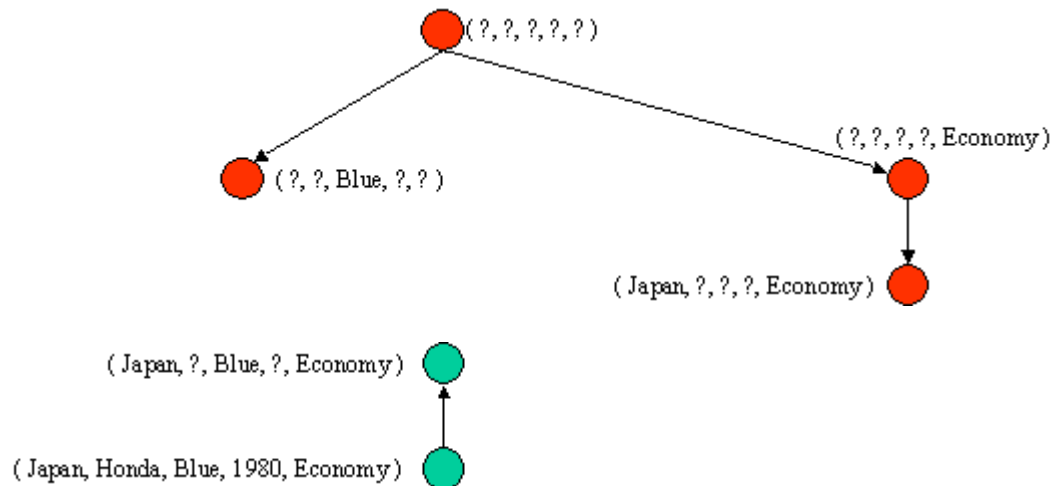tions or form concepts from training examples. This is a type of *analytic* learning. The advantage of explanation-based learning is that, as a deductive mechanism, it requires only a single training example (inductive learning methods often require many training examples)

An Explanation-based Learning (**EBL**) system accepts an example (i.e. a training example) and explains what it learns from the example. The **EBL** system takes only the relevant aspects of the training.

EBL accepts four inputs:

A training example : what the learning *sees* in the world. (specific facts that rule out some possible hypotheses)

A goal concept : a high level description of what the program is supposed to learn. (the set of all possible conclusions)

A operational criterion : a description of which concepts are usable. (criteria for determining which features in the domain are efficiently recognizable, e.g. which features are directly detectable using sensors)

Bal Krishna Subedi

A domain theory : a set of rules that describe relationships between objects and actions in a domain. (axioms about a domain of interest)

From this EBL computes a generalization of the training example that is sufficient not only to describe the goal concept but also satisfies the operational criterion.

This has two steps:

Explanation: the domain theory is used to prune away all unimportant aspects of the training example with respect to the goal concept.

Generalisation: the explanation is generalized as far possible while still describing the goal concept
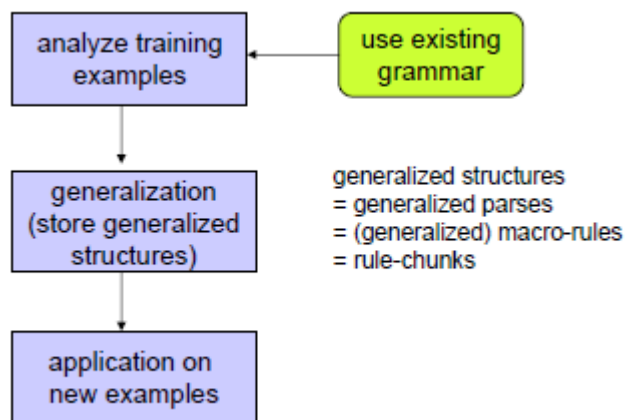
An example of EBL using a perfect domain theory is a program that learns to play chess by being shown examples. A specific chess position that contains an important feature, say, "Forced loss of black queen in two moves," includes many irrelevant features, such as the specific scattering of pawns on the board. EBL can take a single training example and determine what the relevant features are in order to form a generalization.

Learning by Analogy:

Reasoning by analogy generally involves abstracting details from a a particular set of problems and resolving structural similarities between previously distinct problems. Analogical reasoning refers to this process of recognition and then applying the solution from the known problem to the new problem. Such a technique is often identified as *case-based reasoning*. Analogical learning generally involves developing a set of mappings between features of two instances.
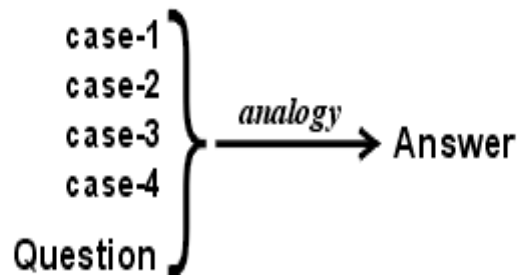
The question in above figure represents some known aspects of a new case, which has unknown aspects to be determined. In deduction, the known aspects are compared (by a version of structure mapping called *unification*) with the premises of some implication. Then the unknown aspects, which answer the question, are derived from the conclusion of the implication. In analogy, the known aspects of the new case are compared with the corresponding aspects of the older cases. The case that gives the best match may be assumed as the best source of evidence for estimating the unknown aspects of the new case. The other cases show alternative possibilities for those unknown aspects; the closer the agreement among the alternatives, the stronger the evidence for the conclusion.

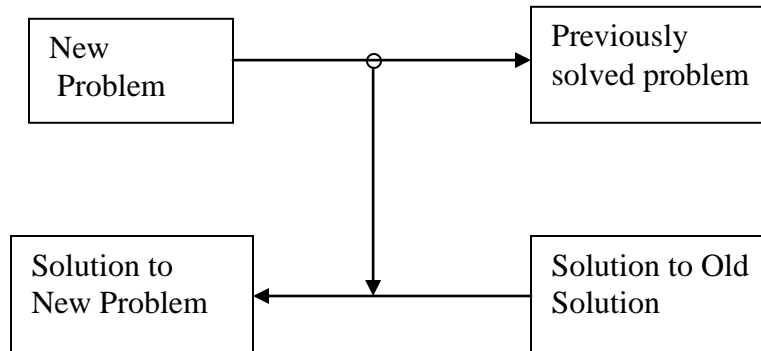1. **Retrieve:** Given a target problem, retrieve cases from memory that are relevant to solving it. A case consists of a problem, its solution, and, typically, annotations about how the solution was derived. For example, suppose Fred wants to prepare blueberry pancakes. Being a novice cook, the most relevant experience he can recall is one in which he successfully made plain pancakes. The procedure he followed for making the plain pancakes, together with justifications for decisions made along the way, constitutes Fred's retrieved case.

2. **Reuse:** Map the solution from the previous case to the target problem. This may involve adapting the solution as needed to fit the new situation. In the pancake example, Fred must adapt his retrieved solution to include the addition of blueberries.

3. **Revise:** Having mapped the previous solution to the target situation, test the new solution in the real world (or a simulation) and, if necessary, revise. Suppose Fred adapted his pancake solution by adding blueberries to the batter. After mixing, he discovers that the batter has turned blue – an undesired effect. This suggests the following revision: delay the addition of blueberries until after the batter has been ladled into the pan.

4. **Retain:** After the solution has been successfully adapted to the target problem, store the resulting experience as a new case in memory. Fred, accordingly, records his newfound procedure for making blueberry pancakes, thereby enriching his set of stored experiences, and better preparing him for future pancake-making demands.

Transformational Analogy:

Suppose you are asked to prove a theorem in plane geometry. You might look for a previous theorem that is very similar and copy its proof, making substitutions when necessary. The idea is to transform a solution to a previous problem in to solution for the current problem. The following figure shows this process,

```
┌──────────┐                    ┌──────────────┐
│ New      │──────────○────────▶│ Previously   │
│ Problem  │          │         │ solved problem│
└──────────┘          │         └──────────────┘
                      │
                      │
┌──────────┐          │         ┌──────────────┐
│ Solution to│◀───────┴─────────│ Solution to Old│
│ New Problem│                   │ Solution      │
└──────────┘                    └──────────────┘
```

Fig: Transformational Analogy

Derivational Analogy:

Notice that transformational analogy does not look at how the old problem was solved, it only looks at the final solution. Often the twists and turns involved in solving an old problem are relevant to solving a new problem. The detailed history of problem solving episode is called derivation, Analogical reasoning that takes these histories into account is called derivational analogy.

```
            ┌──────────┐                    ┌──────────────┐
            │ New      │──────────○────────▶│ Previously   │
            │ Problem  │          │         │ solved problem│
            └──────────┘          │         └──────────────┘
New Derivation  │         ◀───────┴──────────────○   Old derivation
                │                                 │
                ▼                                 ▼
            ┌──────────┐                    ┌──────────────┐
            │ Solution to│                   │ Solution to Old│
            │ New Problem│                   │ Solution      │
            └──────────┘                    └──────────────┘
```
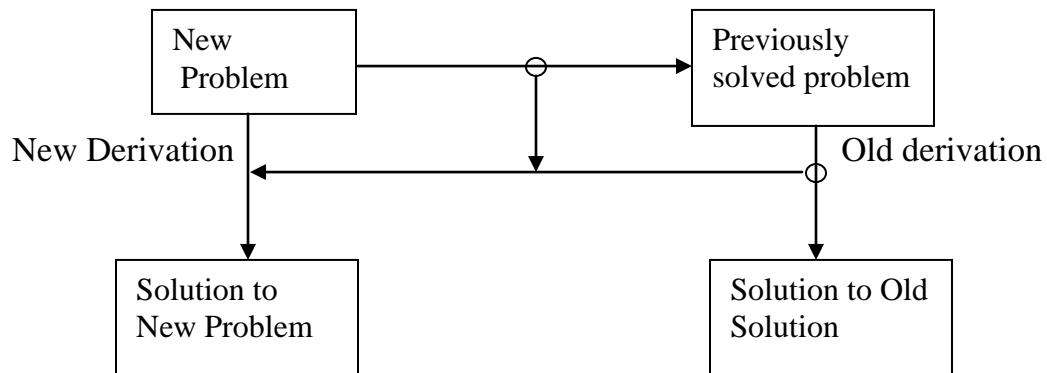
Fig: Derivational Analogy

Refer Book:- E. Rich, K. Knight, S. B. Nair, Tata MacGraw Hill (Pages 371-372)

<u>Learning by Simulating Evolution:</u>

<u>Refer Book:- P. H. Winston, Artificial Intelligence, Addison Wesley. (Around page 220)</u>

<u>Learning by Training Perceptron:</u>

Below is an example of a learning algorithm for a single-layer (no hidden-layer) perceptron. **For multilayer perceptrons, more complicated algorithms such as backpropagation must be used**. Or, methods such as the delta rule can be used if the function is non-linear and differentiable, although the one below will work as well.

The learning algorithm we demonstrate is the same across all the output neurons, therefore everything that follows is applied to a single neuron in isolation. We first define some variables:

- $x(j)$ denotes the j-th item in the n-dimensional input vector
- $w(j)$ denotes the j-th item in the weight vector
- $f(x)$ denotes the output from the neuron when presented with input x
- α is a constant where $0 < \alpha \leq 1$ (learning rate)

Assume for the convenience that the bias term b is zero. An extra dimension $n + 1$ can be added to the input vectors x with $x(n + 1) = 1$, in which case $w(n + 1)$ replaces the bias term.
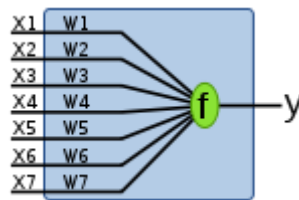
the appropriate weights are applied to the inputs, and the resulting weighted sum passed to a function which produces the output y

Let $D_m = \{(x_1, y_1), \ldots, (x_m, y_m)\}$ be training set of m training examples, where x_i is the input vector to the perceptron and y_i is the desired output value of the perceptron for that input vector.

Bal Krishna Subedi

Learning algorithm steps:

1. Initialize weights and threshold.

- Set $w_i(t)$, $(1 \leq i \leq m)$ to be the weight i at time t, and ø to be the threshold value in the output node.
- Set w(0) to be -ø, the bias, and x(0) to be always 1.
- Set $w_i(1)$ to small random values, thus initialising the weights and threshold.

2. Present input and desired output

- Present input $x_0 = 1$ and $x_1,x_2,...,x_m$ and desired output d(t)

3. Calculate the actual output

- $y(t) = fh[w_0(t) + w_1(t)x_1(t) + w_2(t)x_2(t) + ... + w_m(t)x_m(t)]$

4. Adapts weights

- $w_i(t + 1) = w_i(t) + \alpha[d(t) - y(t)]x_i(t)$, for $0 \leq i \leq m$.

Steps 3 and 4 are repeated until the iteration error is less than a user-specified error threshold or a predetermined number of iterations have been completed.

Bal Krishna Subedi

[Unit 7: Application of AI]
Artificial Intelligence (CSC 355)

Bal Krishna Subedi

Central Department of Computer Science & Information Technology
Tribhuvan University

Expert Systems:

An Expert system is a set of program that manipulates encoded knowledge to solve problem in a specialized domain that normally requires human expertise.

A computer system that simulates the **decision- making process** of a human expert in a specific domain.

An expert system's knowledge is obtained from expert sources and coded in a form suitable for the system to use in its inference or reasoning processes. The expert knowledge must be obtained from specialists or other sources of expertise, such as texts, journals, articles and data bases.

An expert system is an "intelligent" program that solves problems in a narrow problem area by using high-quality, specific knowledge rather than an algorithm.

Block Diagram

There is currently no such thing as "standard" expert system. Because a variety of techniques are used to create expert systems, they differ as widely as the programmers who develop them and the problems they are designed to solve. However, the principal components of most expert systems are **knowledge base, an inference engine, and a user interface,** as illustrated in the figure.
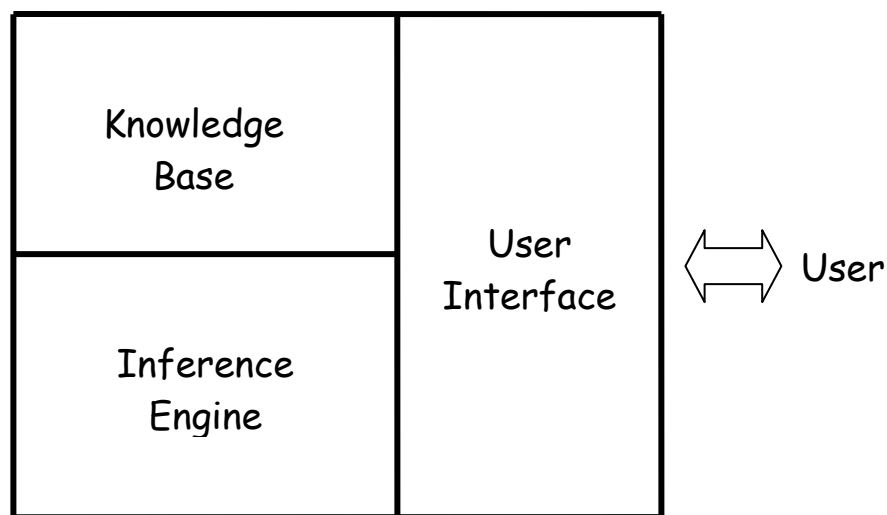

Fig: Block Diagram of expert system

1. Knowledge Base

The component of an expert system that contains the system's knowledge is called its knowledge base. This element of the system is so critical to the way most expert systems are constructed that they are also popularly known as *knowledge-based systems*

A knowledge base contains both declarative knowledge (facts about objects, events and situations) and procedural knowledge (information about courses of action). Depending on the form of knowledge representation chosen, the two types of knowledge may be separate or integrated. Although many knowledge representation techniques have been used in expert systems, the most prevalent form of knowledge representation currently used in expert systems is the *rule-based production* system approach.

To improve the performance of an expert system, we should supply the system with some knowledge about the knowledge it posses, or in other words, meta-knowledge.

2. Inference Engine

Simply having access to a great deal of knowledge does not make you an expert; you also must know **how** and **when** to apply the appropriate knowledge. Similarly, just having a knowledge base does not make an expert system intelligent. The system must have another component that directs the implementation of the knowledge. That element of the system is known variously as the *control structure*, the *rule interpreter*, or the *inference engine*.

The inference engine decides which heuristic search techniques are used to determine how the rules in the knowledge base are to be applied to the problem. In effect, an inference engine "runs" an expert system, determining which rules are to be invoked, accessing the appropriate rules in the knowledge base, executing the rules , and determining when an acceptable solution has been found.

3. User Interface

The component of an expert system that communicates with the user is known as the *user interface*. The communication performed by a user interface is bidirectional. At the simplest level, we must be able to describe our problem to the expert system, and the system must be able to respond with its recommendations. We may want to ask the system to explain its "reasoning", or the system may request additional information about the problem from us.

Beside these three components, there is a Working Memory - a data structure which stores information about a specific run. It holds current facts and knowledge.

Bal Krishna Subedi

Stages of Expert System Development:

Although great strides have been made in expediting the process of developing an expert system, it often remains an extremely time consuming task. It may be possible for one or two people to develop a small expert system in a few months; however the development of a sophisticated system may require a team of several people working together for more than a year.

An expert system typically is developed and refined over a period of several years. We can divide the process of expert system development into five distinct stages. In practice, it may not be possible to break down the expert system development cycle precisely. However, an examination of these five stages may serve to provide us with some insight into the ways in which expert systems are developed.
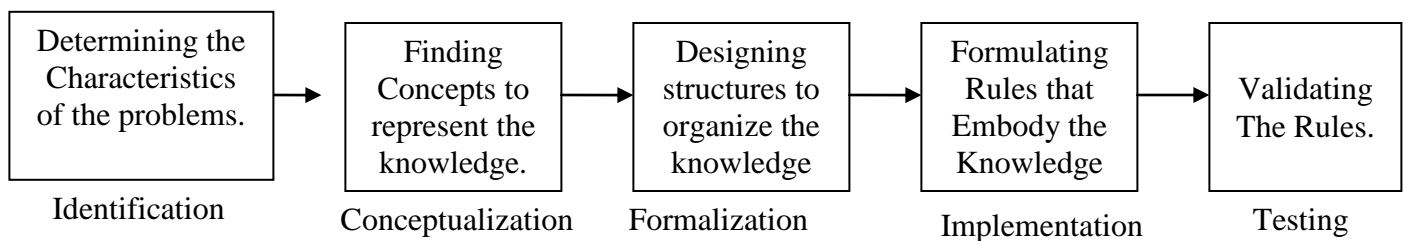
Determining the Characteristics of the problems.	→	Finding Concepts to represent the knowledge.	→	Designing structures to organize the knowledge	→	Formulating Rules that Embody the Knowledge	→	Validating The Rules.
Identification		Conceptualization		Formalization		Implementation		Testing

Fig: Different phases of expert system development

Identification:

Beside we can begin to develop an expert system, it is important that we describe, with as much precision as possible, the problem that the system is intended to solve. It is not enough simply to feel that the system would be helpful in certain situation; we must determine the exact nature of the problem and state the precise goals that indicate exactly how we expect the expert system to contribute to the solution.

Conceptualization:

Once we have formally identified the problem that an expert system is to solve, the next stage involves analyzing the problem further to ensure that its specifics, as well as it generalities, are understood. In the conceptualization stage the **knowledge engineer** frequently creates a diagram of the problem to depict graphically the relationships between the objects and processes in the problem domain. It is often helpful at this stage to divide the problem into a series of sub-problems and to diagram both the relationships among the pieces of each sub-problem and the relationships among the various sub-problems.

Bal Krishna Subedi

Formalization:

In the preceding stages, no effort has been made to relate the domain problem to the artificial intelligence technology that may solve it. During the identification and the conceptualization stages, the focus is entirely on understanding the problem. Now, during the formalization stage, the problem is connected to its proposed solution, an expert system, by analyzing the relationships depicted in the conceptualization stage.

During formalization, it is important that the knowledge engineer be familiar with the following:

- The various techniques of knowledge representation and heuristic search used in expert systems.
- The expert system "tools" that can greatly expedite the development process. And
- Other expert systems that may solve similar problems and thus may be adequate to the problem at hand.

Implementation:

During the implementation stage, the formalized concepts are **programmed** onto the computer that has been chosen for system development, using the predetermined techniques and tools to implement a "first pass" prototype of the expert system.

Theoretically, if the methods of the previous stage have been followed with diligence and care, the implementation of the prototype should be as much an art as it is a science, because following all rules does not guarantee that the system will work the first time it is implemented. Many scientists actually consider the first prototype to be a "throw-away" system, useful for evaluating progress but hardly a usable expert system.

Testing:

Testing provides opportunities to identify the weakness in the structure and implementation of the system and to make the appropriate corrections. Depending on the types of problems encountered, the testing procedure may indicate that the system was

<u>Features of an expert system:</u>

What are the features of a good expert system? Although each expert system has its own particular characteristics, there are several features common to many systems. The following list from Rule-Based Expert Systems suggests seven criteria that are important prerequisites for the acceptance of an expert system .

1. "The program should be **useful**." An expert system should be developed to meet a specific need, one for which it is recognized that assistance is needed.

Bal Krishna Subedi

2. "The program should be **usable**." An expert system should be designed so that even a novice computer user finds it easy to use .

3. "The program should be **educational when appropriate**." An expert system may be used by non-experts, who should be able to increase their own expertise by using the system.

4. "The program should be able to **explain its advice**." An expert system should be able to explain the "reasoning" process that led it to its conclusions, to allow us to decide whether to accept the system's recommendations.

5. "The program should be able to **respond to simple questions**." Because people with different levels of knowledge may use the system , an expert system should be able to answer questions about points that may not be clear to all users.

6. "The program should be able to **learn new knowledge**." Not only should an expert system be able to respond to our questions, it also should be able to ask questions to gain additional information.

7. "The program's knowledge should be **easily modified**." It is important that we should be able to revise the knowledge base of an expert system easily to correct errors or add new information.

Neural Networks:

A neuron is a cell in brain whose principle function is the collection, Processing, and dissemination of electrical signals. Brains Information processing capacity comes from networks of such neurons. Due to this reason some earliest AI work aimed to create such artificial networks. (Other Names are Connectionism; Parallel distributed processing and neural computing).

What is a Neural Network?

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurones) working in unison to solve specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process.

Why use neural networks?

Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyze. Other advantages include:

1. Adaptive learning: An ability to learn how to do tasks based on the data given for training or initial experience.
2. Self-Organisation: An ANN can create its own organisation or representation of the information it receives during learning time.
3. Real Time Operation: ANN computations may be carried out in parallel, and special hardware devices are being designed and manufactured which take advantage of this capability.
4. Fault Tolerance via Redundant Information Coding: Partial destruction of a network leads to the corresponding degradation of performance. However, some network capabilities may be retained even with major network damage

Bal Krishna Subedi

Neural networks versus conventional computers

Neural networks take a different approach to problem solving than that of conventional computers. Conventional computers use an algorithmic approach i.e. the computer follows a set of instructions in order to solve a problem. Unless the specific steps that the computer needs to follow are known the computer cannot solve the problem. That restricts the problem solving capability of conventional computers to problems that we already understand and know how to solve. But computers would be so much more useful if they could do things that we don't exactly know how to do.

Neural networks process information in a similar way the human brain does. The network is composed of a large number of highly interconnected processing elements(neurones) working in parallel to solve a specific problem. Neural networks learn by example. They cannot be programmed to perform a specific task. The examples must be selected carefully otherwise useful time is wasted or even worse the network might be functioning incorrectly. The disadvantage is that because the network finds out how to solve the problem by itself, its operation can be unpredictable.

On the other hand, conventional computers use a cognitive approach to problem solving; the way the problem is to solved must be known and stated in small unambiguous instructions. These instructions are then converted to a high level language program and then into machine code that the computer can understand. These machines are totally predictable; if anything goes wrong is due to a software or hardware fault.

Units of Neural Network:

Nodes(units):
> Nodes represent a cell of neural network.

Links:
> Links are directed arrows that show propagation of information from one node to another node.

Activation:
> Activations are inputs to or outputs from a unit.

Weight:
> Each link has weight associated with it which determines strength and sign of the connection.

Activation function:
> A function which is used to derive output activation from the input activations to a given node is called activation function.
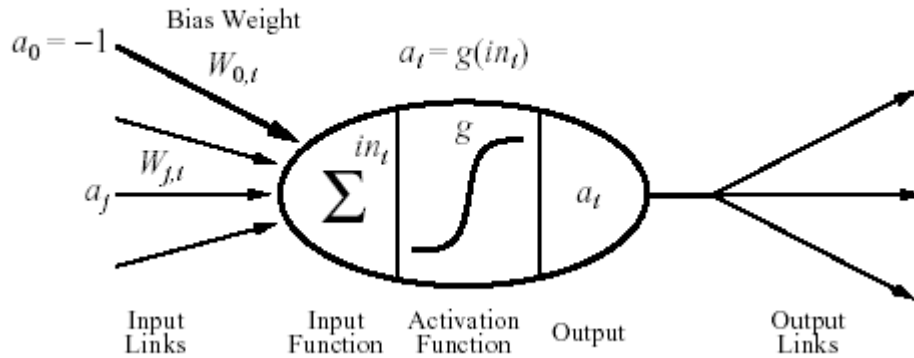
Bias Weight:
> Bias weight is used to set the threshold for a unit. Unit is activated when the weighted sum of real inputs exceeds the bias weight.

Bal Krishna Subedi

Simple Model of Neural Network

A simple mathematical model of neuron is devised by McCulloch and Pit is given in the figure given below:

$$a_i \leftarrow g(in_i) = g\left(\sum_j W_{j,i} a_j\right)$$

It fires when a linear combination of its inputs exceeds some threshold.

A neural network is composed of nodes (units) connected by directed links A link from unit j to i serve to propagate the activation a_j from j to i. Each link has some numeric weight $W_{j,i}$ associated with it, which determines strength and sign of connection.

Each unit first computes a weighted sum of it's inputs:

$$in_i = \sum_{J=0}^{n} W_{j,i}\, a_j$$

Then it applies activation function g to this sum to derive the output:
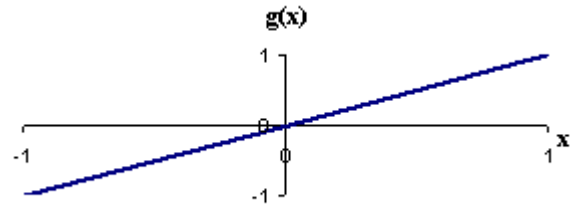
$$a_i = g(in_i) = g\left(\sum_{J=0}^{n} W_{j,i}\, a_j\right)$$

Here, a_j output activation from unit j and $W_{j,i}$ is the weight on the link j to this node. Activation function typically falls into one of three categories:

- Linear
- Threshold (*Heaviside function*)
- Sigmoid
- Sign

For **linear activation functions**, the output activity is proportional to the total weighted output.
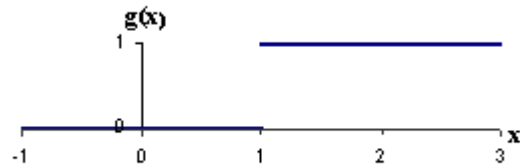
$g(x) = k\,x + c,$ where k and x are constant

For **threshold activation functions**, the output are set at one of two levels, depending on whether the total input is greater than or less than some threshold value.
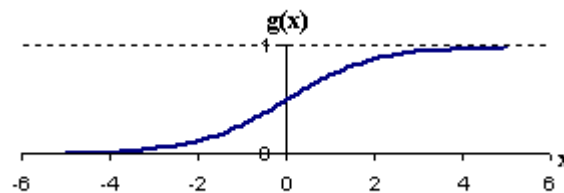
$g(x) = 1$ if x >= k
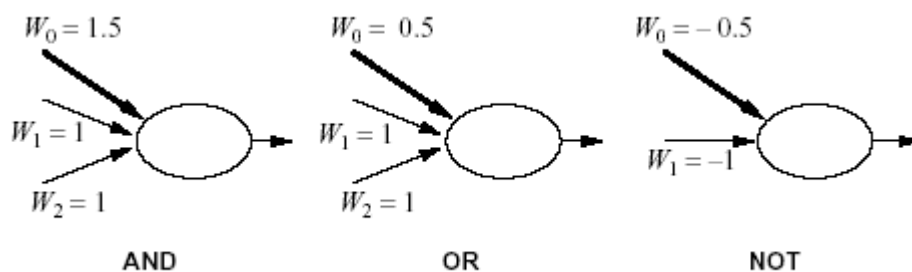
    $= 0$ if x < k

For **sigmoid activation functions**, the output varies continuously but not linearly as the input changes. Sigmoid units bear a greater resemblance to real neurons than do linear or threshold units. It has the advantage of differentiable.

$g(x) = 1/(1 + e^{-x})$

Realizing logic gates by using Neurons:

$W_0 = 1.5$ $W_0 = 0.5$ $W_0 = -0.5$

$W_1 = 1$ $W_1 = 1$ $W_1 = -1$

$W_2 = 1$ $W_2 = 1$

AND   OR   NOT

Bal Krishna Subedi

10

Network structures:

Feed-forward networks:

Feed-forward ANNs allow signals to travel one way only; from input to output. There is no feedback (loops) i.e. the output of any layer does not affect that same layer. Feed-forward ANNs tend to be straight forward networks that associate inputs with outputs. They are extensively used in pattern recognition. This type of organization is also referred to as bottom-up or top-down.
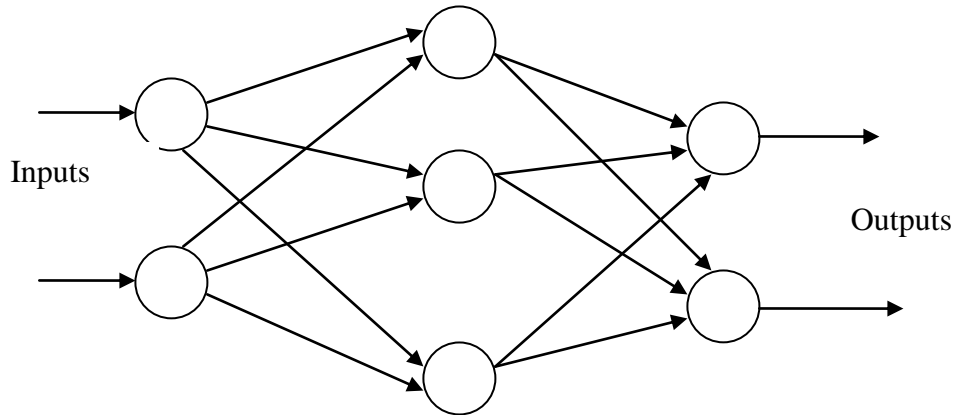
Feedback networks (Recurrent networks:)
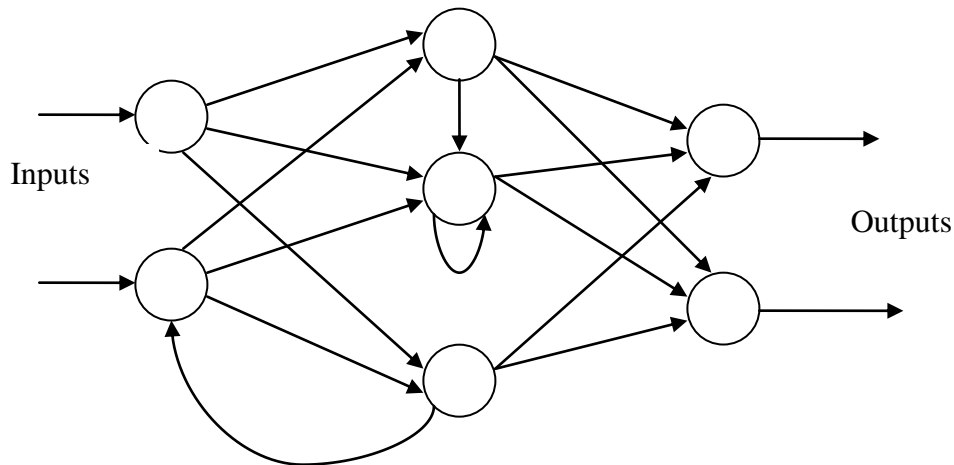
Feedback networks (figure 1) can have signals traveling in both directions by introducing loops in the network. Feedback networks are very powerful and can get extremely complicated. Feedback networks are dynamic; their 'state' is changing continuously until they reach an equilibrium point. They remain at the equilibrium point until the input changes and a new equilibrium needs to be found. Feedback architectures are also referred to as interactive or recurrent.

Bal Krishna Subedi

Feed-forward example
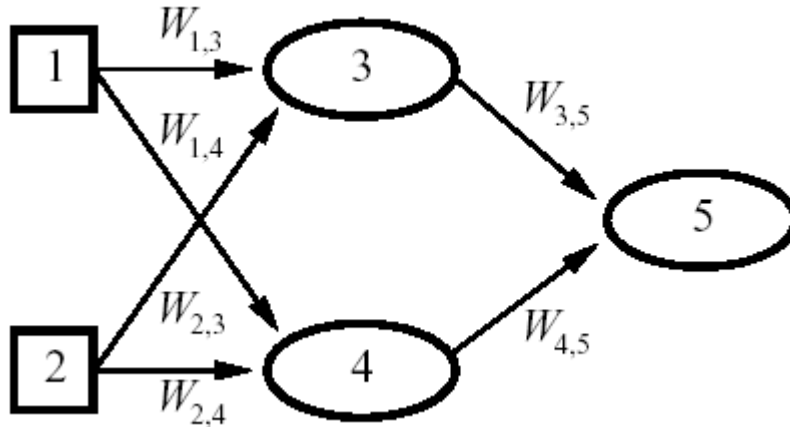
Here;

$a5 = g(W_{3;5} \ a_3 + W_{4;5} \ a_4)$

$\quad = g(W_{3;5} \ g(W_{1;3} \ a_1 + W_{2;3} \ a_2) + W4;5 \ g(W_{1;4} \ a_1 + W_{2;4} \ a_2)$

Types of Feed Forward Neural Network:

Single-layer neural networks (perceptrons)

A neural network in which all the inputs connected directly to the outputs is called a single-layer neural network, or a perceptron network. Since each output unit is independent of the others each weight affects only one of the outputs.
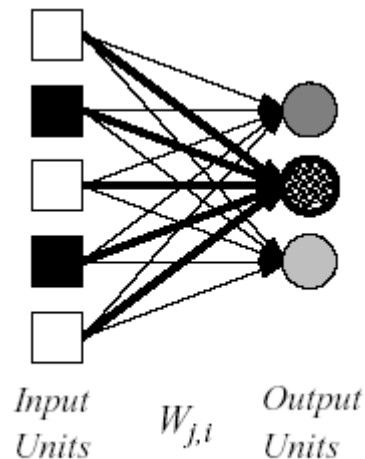
Multilayer neural networks (perceptrons)

The neural network which contains input layers, output layers and some hidden layers also is called multilayer neural network. The advantage of adding hidden layers is that it enlarges the space of hypothesis. Layers of the network are normally fully connected.
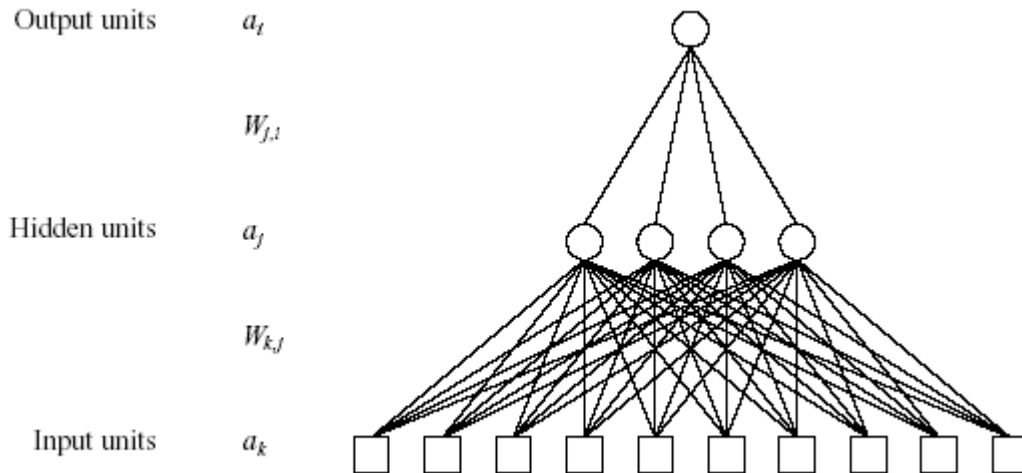
Once the number of layers, and number of units in each layer, has been selected, training is used to set the network's weights and thresholds so as to minimize the prediction error made by the network

Training is the process of adjusting weights and threshold to produce the desired result for different set of data.

Learning in Neural Networks:

Learning: One of the powerful features of neural networks is learning. **Learning in neural networks is carried out by adjusting the connection weights among neurons**. It is similar to a biological nervous system in which learning is carried out by changing synapses connection strengths, among cells.

The operation of a neural network is determined by the values of the interconnection weights. There is no algorithm that determines how the weights should be assigned in order to solve specific problems. Hence, the weights are determined by a learning process

Learning may be classified into two categories:

    (1) Supervised Learning
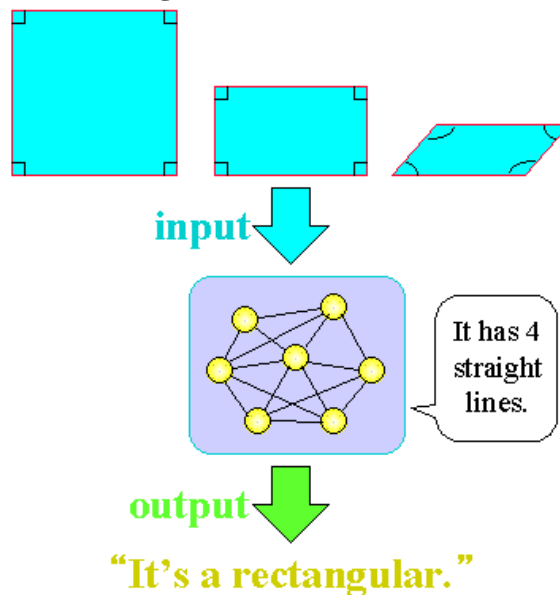   **(2) Unsupervised Learning**

Supervised Learning:

In supervised learning, the network is presented with inputs together with the target (teacher signal) outputs. Then, the neural network tries to produce an output as close as possible to the target signal by adjusting the values of internal weights. The most common supervised learning method is the "error correction method".

Error correction method is used for networks which their neurons have discrete output functions. Neural networks are trained with this method in order to reduce the error (difference between the network's output and the desired output) to zero.
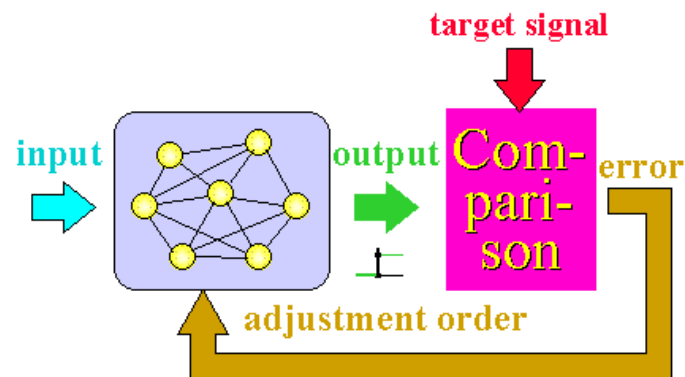
Unsupervised Learning:

In unsupervised learning, there is no teacher (target signal) from outside and the network adjusts its weights in response to only the input patterns. A typical example of unsupervised learning is **Hebbian learning.**

Consider a machine (or living organism) which receives some sequence of inputs x1, x2, x3, . . ., where xt is the sensory input at time t. In supervised learning the machine is given a sequence of input & a sequence of desired outputs y1, y2, . . . , and the goal of the machine is to learn to produce the correct output given a new input. While, in unsupervised learning the machine simply receives inputs x1, x2, . . ., but obtains neither supervised target outputs, nor rewards from its environment. It may seem somewhat mysterious to imagine what the machine could possibly learn given that it doesn't get any feedback from its environment. However, it is possible to develop of formal framework for unsupervised learning based on the notion that the machine's goal is to build representations of the input that can be used for decision making, predicting future inputs, efficiently communicating the inputs to another machine, etc. In a sense, unsupervised learning can be thought of as finding patterns in the data above and beyond what would be considered pure unstructured noise.

Hebbian Learning:

The oldest and most famous of all learning rules is Hebb's postulate of learning:

"When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic changes take place in one or both cells such that A's efficiency as one of the cells firing B is increased"

From the point of view of artificial neurons and artificial neural networks, Hebb's principle can be described as a method of determining how to alter the weights between model neurons. **The weight between two neurons increases if the two neurons activate simultaneously—and reduces if they activate separately.** Nodes that tend to be either both positive or both negative at the same time have strong positive weights, while those that tend to be opposite have strong negative weights.

Hebb's Algorithm:

Step 0: initialize all weights to 0

Step 1: Given a training input, s, with its target output, t, set the activations of the input units: $x_i = s_i$

Step 2: Set the activation of the output unit to the target value: $y = t$

Step 3: Adjust the weights: $w_i(new) = w_i(old) + x_i y$

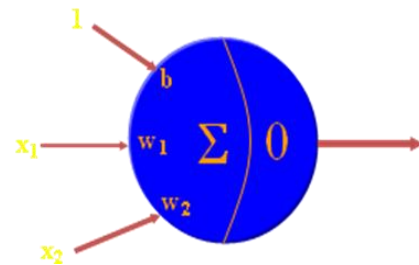**Step 4**: Adjust the bias (just like the weights): $b(new) = b(old) + y$

Example:

PROBLEM: Construct a Hebb Net which performs like an AND function, that is, only when both features are "active" will the data be in the target class.

TRAINING SET (with the bias input always at 1):

```
x1 x2 bias   Target
 1  1   1       1
 1 -1   1      -1
-1  1   1      -1
-1 -1   1      -1
```


Training-First Input:

- Initialize the weights to 0

Present the first input:
(1 1 1) with a target of 1

Update the weights:

$w_1(new) = w_1(old) + x_1t$
$= \quad 0 + 1 = 1$
$w_2(new) = w_2(old) + x_2t$
$= \quad 0 + 1 = 1$
$b(new) \quad = b(old) + t$
$= \quad 0 + 1 = 1$
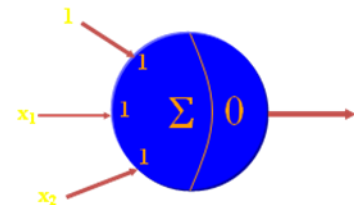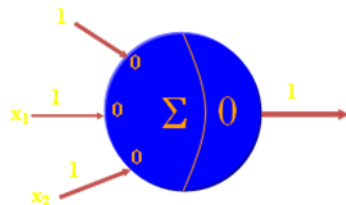
Training- Second Input:

- **Present the second input: (1 -1 1) with a target of -1**
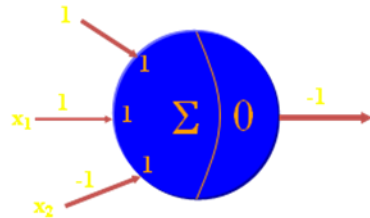
Update the weights:

$w_1(new) = w_1(old) + x_1 t$

$= \quad 1 + 1(-1) = 0$

$w_2(new) = w_2(old) + x_2 t$

$= \quad 1 + (-1)(-1) = 2$

$b(new) \quad = b(old) + t$

$= \quad 1 + (-1) = 0$
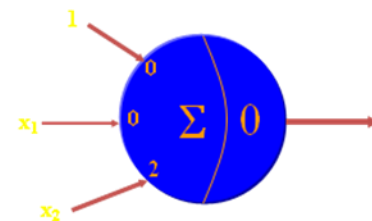
Training- Third Input:

- **Present the third input: (-1 1 1) with a target of -1**
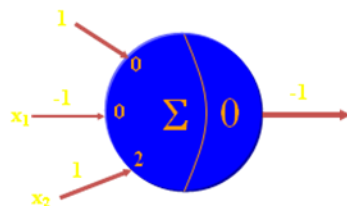
Update the weights:

$w_1(new) = w_1(old) + x_1 t$

$= \quad 0 + (-1)(-1) = 1$

$w_2(new) = w_2(old) + x_2 t$

$= \quad 2 + 1(-1) = 1$

$b(new) \quad = b(old) + t$

$= \quad 0 + (-1) = -1$
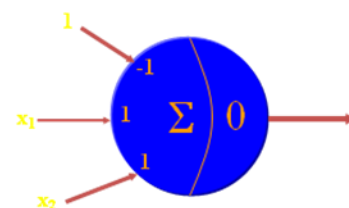
Training- Fourth Input:

- **Present the fourth input:**
  (-1 -1 1) with a target of -1
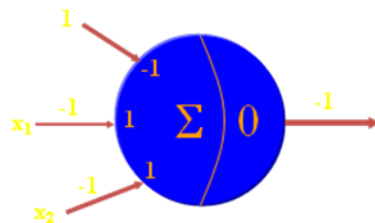
Update the weights:

$w_1(new) = w_1(old) + x_1 t$

$\quad = \quad 1 + (-1)(-1) = 2$

$w_2(new) = w_2(old) + x_2 t$

$\quad = \quad 1 + (-1)(-1) = 1$

$b(new) \quad = b(old) + t$

$\quad = \quad -1 + (-1) = -2$
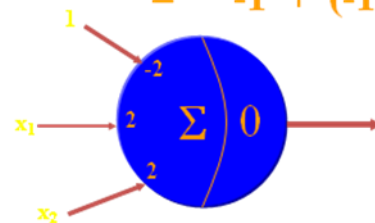
Final Neuron:

- **This neuron works:**

x1	x2	bias	Target
1	1	1	1
1	-1	1	-1
=1	1	1	=1
=1	=1	1	=1
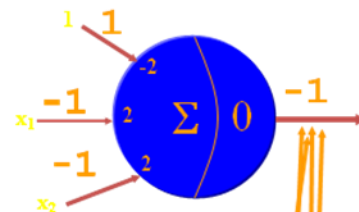
$1*2 + 1*2 + 1*(-2) = 2 > 0$

$(-1)*2 + 1*2 + 1*(-2) = -2 < 0$

$1*2 + (-1)*2 + 1*(-2) = -2 < 0$

$(-1)*2 + (-1)*2 + 1 *(-2) = -6 < 0$

Bal Krishna Subedi

Perceptron Learning Theory:

The term "Perceptrons" was coined by Frank RosenBlatt in 1962 and is used to describe the connection of simple neurons into networks. These networks are simplified versions of the real nervous system where some properties are exagerrated and others are ignored. For the moment we will concentrate on Single Layer Perceptrons.

So how can we achieve learning in our model neuron? We need to train them so they can do things that are useful. To do this we must allow the neuron to learn from its mistakes. There is in fact a learning paradigm that achieves this, it is known as supervised learning and works in the following manner.

   i.   set the weight and thresholds of the neuron to random values.
  ii.   present an input.
 iii.   caclulate the output of the neuron.
 iv.   alter the weights to reinforce correct decisions and discourage wrong decisions, hence reducing the error. So for the network to learn we shall increase the weights on the active inputs when we want the output to be active, and to decrease them when we want the output to be inactive.
  v.   Now present the next input and repeat steps iii. - v.

Perceptron Learning Algorithm:

The algorithm for Perceptron Learning is based on the supervised learning procedure discussed previously.

Algorithm:

   i.   Initialize weights and threshold.

       Set $w_i(t)$, $(0 <= i <= n)$, to be the weight i at time t, and ϕ to be the threshold value in the output node. Set w_0 to be $-\phi$, the bias, and x_0 to be always 1.

       Set $w_i(0)$ to small random values, thus initializing the weights and threshold.

  ii.   Present input and desired output

       Present input x_0, x_1, x_2, ..., x_n and desired output $d(t)$

 iii.   Calculate the actual output

$$y(t) = g\,[w_0(t)x_0(t) + w_1(t)x_1(t) + ... + w_n(t)x_n(t)]$$

 iv.   Adapts weights

Bal Krishna Subedi

$w_i(t+1) = w_i(t) + \alpha[d(t) - y(t)]x_i(t)$, where $0 <= \alpha <= 1$ (learning rate) is a positive gain function that controls the adaption rate.

Steps iii. and iv. are repeated until the iteration error is less than a user-specified error threshold or a predetermined number of iterations have been completed.

Please note that the weights only change if an error is made and hence this is only when learning shall occur.

Delta Rule:

The **delta rule** is a gradient descent learning rule for updating the weights of the artificial neurons in a single-layer perceptron. It is a special case of the more general backpropagation algorithm. For a neuron j with activation function $g(x)$ the delta rule for j's ith weight w_{ji} is given by

$$\Delta w_{ji} = \alpha(t_j - y_j)g'(h_j)x_i,$$

where α is a small constant called *learning rate*, $g(x)$ is the neuron's activation function, t_j is the target output, h_j is the weighted sum of the neuron's inputs, y_j is the actual output, and x_i is the ith input. It holds $h_j = \sum x_i w_{ji}$ and $y_j = g(h_j)$.

The delta rule is commonly stated in simplified form for a perceptron with a linear activation function as

$$\Delta w_{ji} = \alpha(t_j - y_j)x_i$$

Backpropagation

It is a supervised learning method, and is an implementation of the **Delta rule**. It requires a teacher that knows, or can calculate, the desired output for any given input. It is most useful for feed-forward networks (networks that have no feedback, or simply, that have no connections that loop). The term is an abbreviation for "backwards propagation of errors". Backpropagation requires that the activation function used by the artificial neurons (or "nodes") is differentiable.

As the algorithm's name implies, the errors (and therefore the learning) propagate backwards from the output nodes to the inner nodes. So technically speaking, backpropagation is used to calculate the gradient of the error of the network with respect to the network's modifiable weights. This gradient is almost always then used in a simple *stochastic gradient descent algorithm, is a general optimization algorithm, but is typically used to fit the parameters of a machine learning model, to find weights that minimize the error*. Often the term "backpropagation" is used in a more general sense, to refer to the

entire procedure encompassing both the calculation of the gradient and its use in stochastic gradient descent. Backpropagation usually allows quick convergence on satisfactory local minima for error in the kind of networks to which it is suited.

Backpropagation networks are necessarily multilayer perceptrons (usually with one input, one hidden, and one output layer). In order for the hidden layer to serve any useful function, multilayer networks must have non-linear activation functions for the multiple layers: a multilayer network using only linear activation functions is equivalent to some single layer, linear network.

Summary of the backpropagation technique:

1. Present a training sample to the neural network.
2. Compare the network's output to the desired output from that sample. Calculate the error in each output neuron.
3. For each neuron, calculate what the output should have been, and a *scaling factor*, how much lower or higher the output must be adjusted to match the desired output. This is the local error.
4. Adjust the weights of each neuron to lower the local error.
5. Assign "blame" for the local error to neurons at the previous level, giving greater responsibility to neurons connected by stronger weights.
6. Repeat from step 3 on the neurons at the previous level, using each one's "blame" as its error.

Characteristics:

- A multi-layered perceptron has three distinctive characteristics
  - The network contains one or more layers of hidden neurons
  - The network exhibits a high degree of connectivity
  - Each neuron has a smooth (differentiable everywhere) nonlinear activation function, the most common is the sigmoidal nonlinearity:

$$y_j = \frac{1}{1 + e^{s_j}}$$

- The backpropagation algorithm provides a computational efficient method for training multi-layer networks
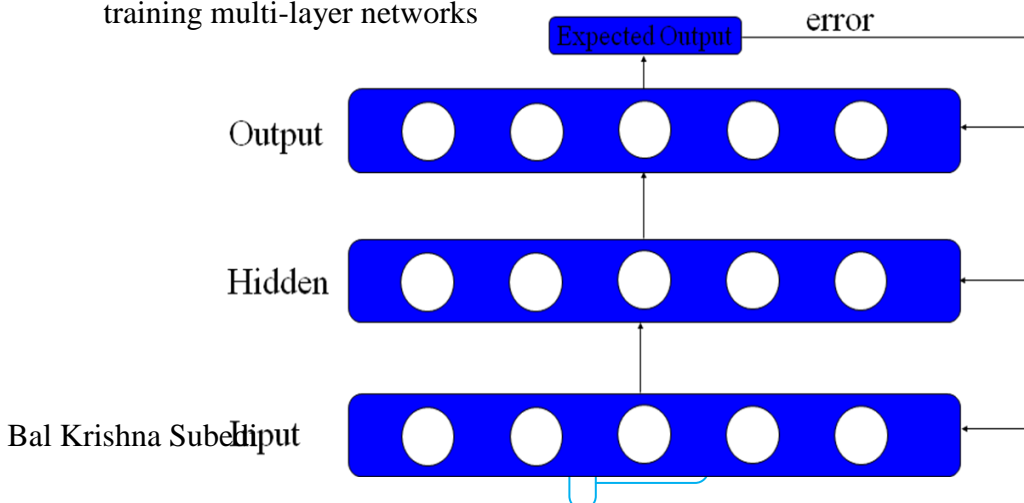
Algorithm:

Step 0: Initialize the weights to small random values

Step 1: Feed the training sample through the network and determine the final output

Step 2: Compute the error for each output unit, for unit k it is:

$$\delta_k = (t_k - y_k)f'(y\_in_k)$$

Actual output — Required output — Derivative of f

Step 3: Calculate the weight correction term for each output unit, for unit k it is:

$$\Delta w_{jk} = \alpha \delta_k z_j$$

Hidden layer signal

A small constant

Step 4: Propagate the delta terms (errors) back through the weights of the hidden units where the delta input for the jth hidden unit is:

$$\delta\_in_j = \sum_{k=1}^{m} \delta_k w_{jk}$$

The delta term for jth hidden unit is: $\delta_j = \delta\_in_j f'(z\_in_j)$

Step 5: Calculate the weight correction term for the hidden units: $\Delta w_{ij} = \alpha \delta_i x_i$

Step 6: Update the weights: $w_{ik}(new) = w_{ik}(old) + \Delta w_{ik}$

Step 7: Test for stopping (maximum cylces, small changes, etc)

Note: There are a number of options in the design of a backprop system;
– Initial weights – best to set the initial weights (and all other free parameters) to random numbers inside a small range of values (say –0.5 to 0.5)
– Number of cycles – tend to be quite large for backprop systems
– Number of neurons in the hidden layer – as few as possible
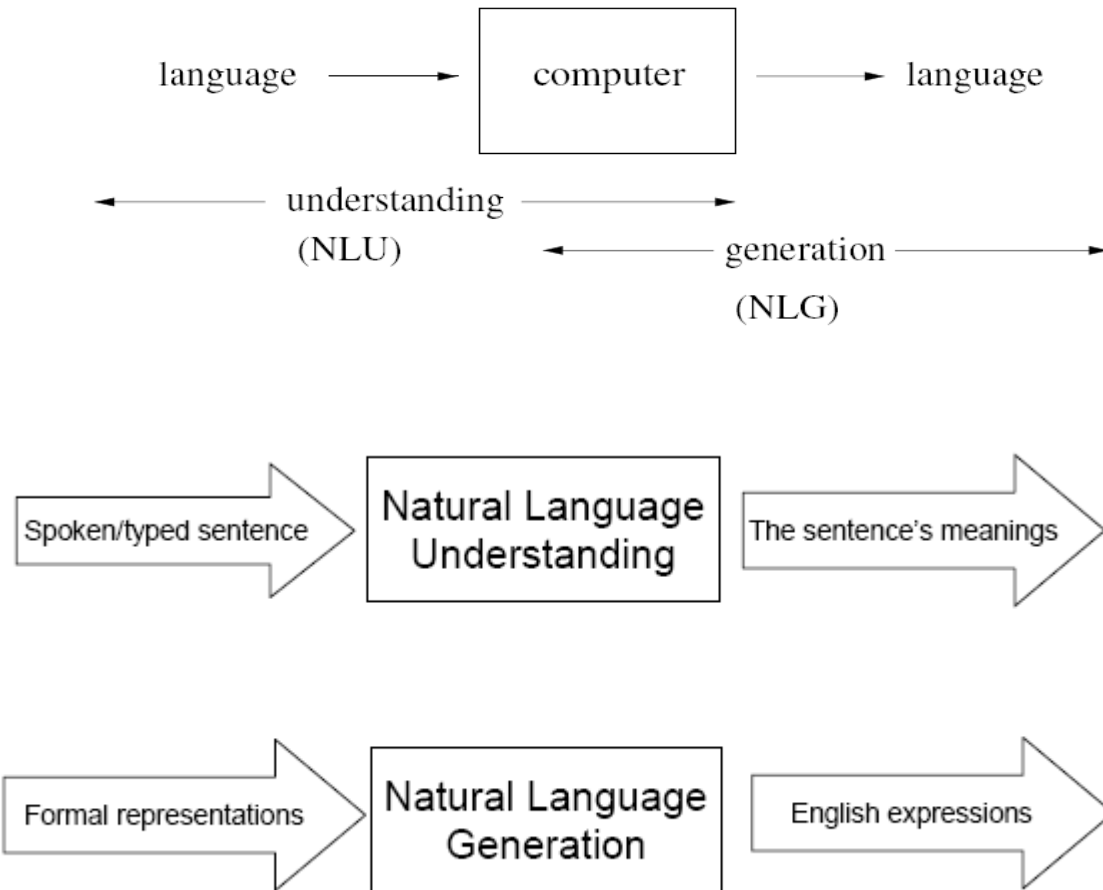
Natural Language Processing:

Perception and communication are essential components of intelligent behavior. They provide the ability to effectively interact with our environment. Humans perceive and communicate through their five basic senses of sight, hearing, touch, smell and taste, and their ability to generate meaningful utterances. Developing programs that understand a natural language is a difficult problem. Natural languages are large. They contain infinity of different sentences. No matter how many sentences a person has heard or seen, new ones can always be produced. Also, there is much ambiguity in a natural language. Many words have several meanings and sentences can have different meanings in different contexts. This makes the creation of programs that understand a natural language, one of the most challenging tasks in AI.

Developing programs to understand natural language is important in AI because a natural form of communication with systems is essential for user acceptance. AI programs must be able to communicate with their human counterparts in a natural way, and natural language is one of the most important mediums for that purpose. So, Natural Language Processing (NLP) is the field that deals with the computer processing of natural languages, mainly evolved by people working in the field of Artificial Intelligence.

Natural Language Processing (NLP), is the attempt to extract the fuller meaning representation from the free text. Natural language processing is a technology which involves converting spoken or written human language into a form which can be processed by computers, and vice versa. Some of the better-known applications of NLP include:

- **Voice recognition software** which translates speech into input for word processors or other applications;
- **Text-to-speech synthesizers** which read text aloud for users such as the hearing-impaired;
- **Grammar and style checkers** which analyze text in an attempt to highlight errors of grammar or usage;
- **Machine translation systems** which automatically render a document such as a web page in another language.

computers using natural language as input and/or output

Natural Language Generation:

"Natural Language Generation (NLG), also referred to as text generation, is a subfield of natural language processing (NLP; which includes computational linguistics)

Natural Language Generation (NLG) is the natural language processing task of generating natural language from a machine representation system such as a knowledge base or a logical form.

In a sense, one can say that an NLG system is like a translator that converts a computer based representation into a natural language representation. However, the methods to produce the final language are very different from those of a compiler due to the inherent expressivity of natural languages.

Bal Krishna Subedi

NLG may be viewed as the opposite of natural language understanding. The difference can be put this way: whereas in natural language understanding the system needs to disambiguate the input sentence to produce the machine representation language, in NLG the system needs to make decisions about how to put a concept into words.

The different types of generation techniques can be classified into four main categories:

- Canned text systems constitute the simplest approach for single-sentence and multi-sentence text generation. They are trivial to create, but very inflexible.
- Template systems, the next level of sophistication, rely on the application of pre-defined templates or schemas and are able to support flexible alterations. The template approach is used mainly for multi-sentence generation, particularly in applications whose texts are fairly regular in structure.
- Phrase-based systems employ what can be seen as generalized templates. In such systems, a phrasal pattern is first selected to match the top level of the input, and then each part of the pattern is recursively expanded into a more specific phrasal pattern that matches some subportion of the input. At the sentence level, the phrases resemble phrase structure grammar rules and at the discourse level they play the role of text plans.
- Feature-based systems, which are as yet restricted to single-sentence generation, represent each possible minimal alternative of expression by a single feature. Accordingly, each sentence is specified by a unique set of features. In this framework, generation consists in the incremental collection of features appropriate for each portion of the input. Feature collection itself can either be based on unification or on the traversal of a feature selection network. The expressive power of the approach is very high since any distinction in language can be added to the system as a feature. Sophisticated feature-based generators, however, require very complex input and make it difficult to maintain feature interrelationships and control feature selection.

Many natural language generation systems follow a hybrid approach by combining components that utilize different techniques.

Natural Language Understanding:

Developing programs that understand a natural language is a difficult problem. Natural languages are large. They contain infinity of different sentences. No matter how many sentences a person has heard or seen, new ones can always be produced. Also, there is much ambiguity in a natural language. Many words have several meaning such as can, bear, fly, bank etc, and sentences have different meanings in different contexts.

Example :-   A *can* of juice.   I *can* do it.

This makes the creation of programs that understand a natural language, one of the most challenging tasks in AI. Understanding the language is not only the transmission of words. It also requires inference about the speakers' goal, knowledge as well as the context of the

Bal Krishna Subedi

interaction. We say a program understand natural language if it behaves by taking the correct or acceptable action in response to the input. A word functions in a sentence as a part of speech. Parts of the speech for the English language are nouns, pronouns, verbs, adjectives, adverbs, prepositions, conjunctions and interjections. Three major issues involved in understanding language.

- A large amount of human knowledge is assumed.
- Language is pattern based, phonemes are components of the words and words make phrases and sentences. Phonemes, words and sentences order are not random.
- Language acts are the product of agents (human or machine).

## Levels of knowledge used in Language Understanding

A language understanding knowledge must have considerable knowledge about the structures of the language including what the words are and how they combine into phrases and sentences. It must also know the meanings of the words and how they contribute to the meanings of the sentence and to the context within which they are being used. The component forms of knowledge needed for an understanding of natural languages are sometimes classified according to the following levels.

- **Phonological**
  - Relates sound to the words we recognize. A phoneme is the smallest unit of the sound. Phones are aggregated to the words.
- **Morphological**
  - This is lexical knowledge which relates to the word construction from basic units called morphemes. A morpheme is the smallest unit of meaning. Eg:- *friend* + ly = friendly
- **Syntactic**
  - This knowledge relates to how words are put together or structure red together to form grammatically correct sentences in the language.
- **Semantic**
  - This knowledge is concerned with the meanings of words and phrases and how they combine to form sentence meaning.
- **Pragmatic**

- This is high – level knowledge which relates to the use of sentences in different contexts and how the context affects the meaning of the sentence.

- **World**

  - Includes the knowledge of the physical world, the world of human social interaction, and the roles of goals and intentions in communication.

Basic Parsing Techniques

Before the meaning of a sentence can be determined, the meanings of its constituent parts must be established. This requires knowledge of the structure of the sentence, the meaning of the individual words and how the words modify each other. The process of determining the syntactical structure of a sentence is known as parsing. Parsing is the process of analyzing a sentence by taking it apart word – by – word and determining its structure from its constituent parts and sub parts. The structure of a sentence can be represented with a syntactic tree. When given an input string, the lexical parts or terms (root words), must first be identified by type and then the role they play in a sentence must be determined. These parts can be combined successively into larger units until a complete tree has been computed.

N = NOUN

V = VERB

DT =
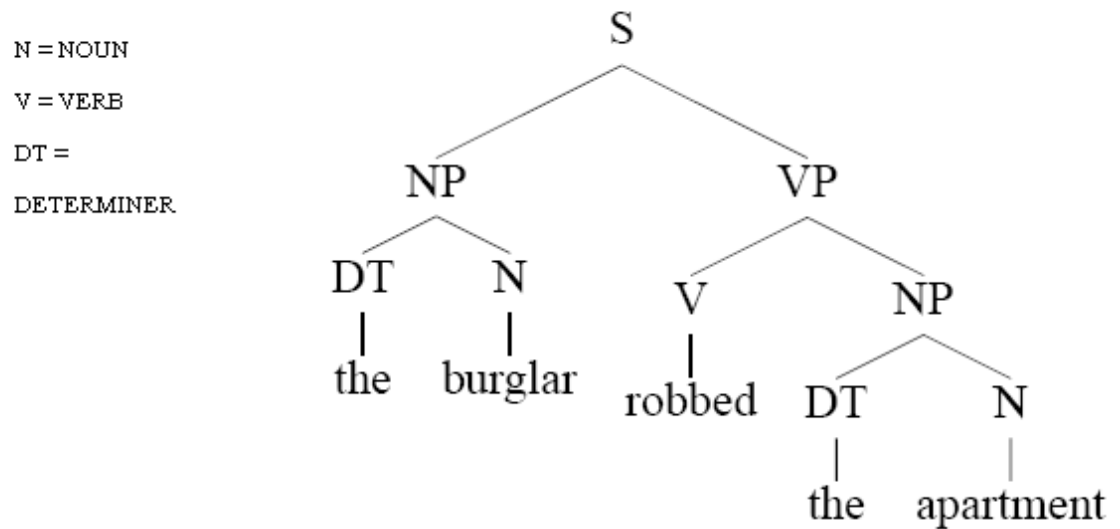
DETERMINER

Noun Phrases (NP): "the burglar", "the apartment"

Verb Phrases (VP): "robbed the apartment"

Sentences (S):   "the burglar robbed the apartment"

To determine the meaning of a word, a parser must have access to a lexicon. When the parser selects the word from the input stream, it locates the world in the lexicon and obtains the word's possible functions and features, including the semantic information.
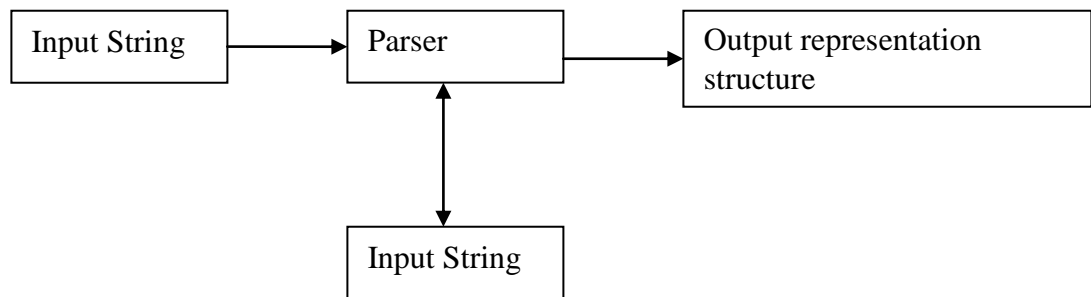

Figure :- Parsing an input to create an output structure

Lexeme (Lexicon) & word forms:

The distinction between these two senses of "word" is arguably the most important one in morphology. The first sense of "word", the one in which *dog* and *dogs* are "the same

Bal Krishna Subedi

word", is called a lexeme. The second sense is called *word form*. We thus say that *dog* and *dogs* are different forms of the same lexeme. *Dog* and *dog catcher*, on the other hand, are different lexemes, as they refer to two different kinds of entities. The form of a word that is chosen conventionally to represent the canonical form of a word is called a lemma, or citation form.

A lexicon defines the words of a language that a system knows about. This is includes common words and words that are specific to the domain of the application. Entries include meanings for each word and its syntactic and morphological behavior.

Morphology:

Morphology is the identification, analysis and description of the structure of words (words as units in the lexicon are the subject matter of lexicology). While words are generally accepted as being (with clitics) the smallest units of syntax, it is clear that in most (if not all) languages, words can be related to other words by rules. For example, English speakers recognize that the words *dog*, *dogs*, and *dog catcher* are closely related. English speakers recognize these relations from their tacit knowledge of the rules of word formation in English. They infer intuitively that *dog* is to *dogs* as *cat* is to *cats*; similarly, *dog* is to *dog catcher* as *dish* is to *dishwasher* (in one sense). The rules understood by the speaker reflect specific patterns (or regularities) in the way words are formed from smaller units and how those smaller units interact in speech. In this way, morphology is the branch of linguistics that studies patterns of word formation within and across languages, and attempts to formulate rules that model the knowledge of the speakers of those languages.

Morphological analysis is the process of recognizing the suffixes and prefixes that have been attached to a word.

We do this by having a table of affixes and trying to match the input as:
   prefixes +root + suffixes.
   - For example: adjective + ly -> adverb. E.g.: [Friend + ly]=friendly
   - We may not get a unique result.
   - "-s, -es" can be either a plural noun or a 3ps verb
   - "-d, -ed" can be either a past tense or a perfect participle

Morphological Information:

- Transform part of speech
  - *green, greenness (adjective, noun)*
  - *walk, walker (verb, noun)*

- Change features of nouns
  - *boat, boats (singular, plural)*

- Bill slept ‚ Bill's bed

- – (subjective case, possessive case)

- Change features of verbs
  - – Aspect
    - *I walk. I am walking.* (present, progressive)
  - – Tense
    - *I walked. I will walk. I had been walking. (past, future, past progressive)*
  - – Number and person
    - *I walk. They walk. (first person singular, third person plural)*

Syntactic Analysis:

Syntactic analysis takes an input sentence and produces a representation of its grammatical structure. A grammar describes the valid parts of speech of a language and how to combine them into phrases. The grammar of English is nearly context free.

A computer grammar specifies which sentences are in a language and their parse trees. A parse tree is a hierarchical structure that shows how the grammar applies to the input. Each level of the tree corresponds to the application of one grammar rule.

It is the starting point for working out the meaning of the whole sentence. Consider the following two sentences.

1. "The dog ate the bone."
2. "The bone was eaten by the dog."

Understanding the structure (via the syntax rules) of the sentences help us work out that it's the bone that gets eaten and not the dog. Syntactic analysis determines possible grouping of words in a sentence. In other cases there may be many possible groupings of words. Consider the sentence "John saw Mary with a telescope". Two different readings based on the groupings.

1. John saw (Mary with a telescope).
2. John (saw Mary with a telescope).

A sentence is syntactically ambiguous if there are two or more possible groupings. Syntactic analysis helps determining the meaning of a sentence by working out possible word structure. Rules of syntax are specified by writing a *grammar* for the language. A parser will check if a sentence is correct according to the grammar. It returns a representation (parse tree) of the sentence's structure. A grammar specifies allowable sentence structures in terms of basic categories such as noun and verbs. A given grammar, however, is unlikely to cover all possible grammatical sentences. Parsing sentences is to help determining their meanings, not just to check that they are correct. Suppose we want a grammar that recognizes sentences like the following.

John ate the biscuit.
The lion ate the zebra.
The lion kissed John

But reject incorrect sentences such as
Ate John biscuit the.
Zebra the lion the ate.
Biscuit lion kissed.

A simple grammar that deals with this is given below
sentence --> noun_phase, verb phrase.
noun_phrase --> proper_noun.
noun_phrase --> determiner, noun.
verb_phrase --> verb, noun_phrase.
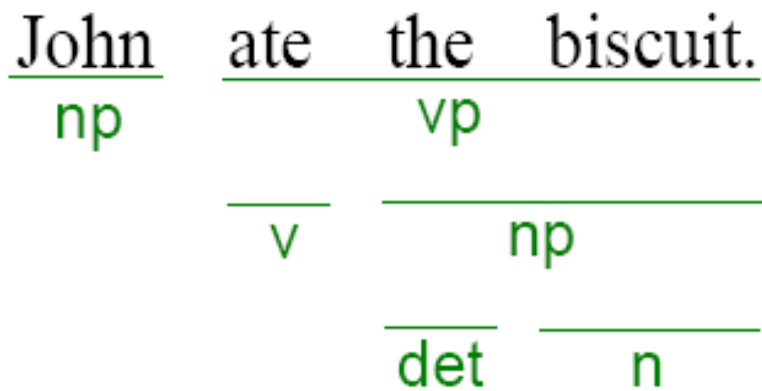proper_noun --> [mary].
proper_noun --> [john].
noun --> [zebra].
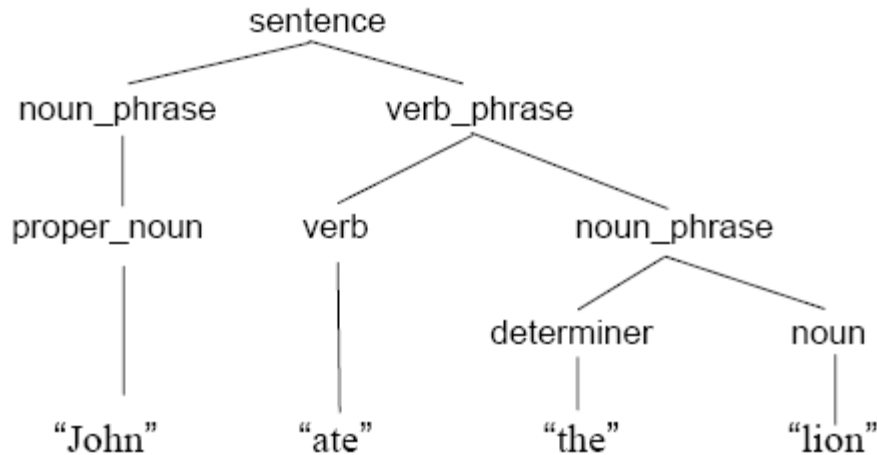noun --> [biscuit].
verb --> [ate].
verb --> [kissed].
determiner --> [the].

Incorrect sentences like "biscuit lion kissed" will be excluded by the grammar.

- A *parse trees* illustrates the syntactic structure of the sentence.

Semantic Analysis:

Semantic analysis is a process of converting the syntactic representations into a meaning representation.

This involves the following tasks:
   – Word sense determination
   – Sentence level analysis
   – Knowledge representation

- *Word sense*

   Words have different meanings in different contexts.

   Mary had a bat in her office.

   - bat = `a baseball thing'

   - bat = `a flying mammal'

- *Sentence level analysis*

   Once the words are understood, the sentence must be assigned some meaning

   I saw an astronomer with a telescope.

- *Knowledge Representation*

   Understanding language requires lots of knowledge**.**

Bal Krishna Subedi

- Using predicate logic, for example, one can represent sentences like

"John likes Mary"	`likes(john,mary)`
"The man likes Mary"	`man(m1)∧likes(m1,mary)`
"A man likes Mary"	`∃X(man(X)∧likes(X,mary)`

"A tall bearded man likes Mary"

`∃X(man(X)∧tall(X)∧bearded(X)∧likes(X,mary)`

- Using semantic net, one can represent sentence "Tarzan kissed Jane" as
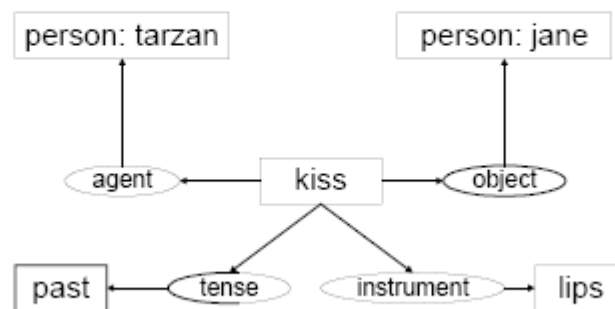
Parameters in Natural Language Processing:

- Auditory Inputs
- Segmentation
- Syntax Structure
- Semantic Structure
- Pragmatic Analysis

- Auditory Inputs:

Three of our five senses – sight, hearing and touch – are used as *major inputs*. These are usually referred to as the visual, auditory and tactile inputs respectively. They are sometimes called input channels; however, as previously mentioned, the term "channel" is used in various ways, so I will avoid it.

Bal Krishna Subedi

In the fashion of video devices, audio devices are used to either capture or create sound. In some cases, an audio output device can be used as an input device, in order to capture produced sound.

- Microphone
- MIDI keyboard or other digital musical instrument

- Segmentation:

Text segmentation is the process of dividing written text into meaningful units, such as words, sentences, or topics. The term applies both to mental processes used by humans when reading text, and to artificial processes implemented in computers, which are the subject of natural language processing. The problem is non-trivial, because while some written languages have explicit word boundary markers, such as the word spaces of written English and the distinctive initial, medial and final letter shapes of Arabic, such signals are sometimes ambiguous and not present in all written languages.

Word segmentation is the problem of dividing a string of written language into its component words. In English and many other languages using some form of the Latin alphabet, the space is a good approximation of a word delimiter. (Some examples where the space character alone may not be sufficient include contractions like *can't* for *can not*.)

However the equivalent to this character is not found in all written scripts, and without it word segmentation is a difficult problem. Languages which do not have a trivial word segmentation process include Chinese, Japanese, where sentences but not words are delimited, and Thai, where phrases and sentences but not words are delimited.

In some writing systems however, such as the Ge'ez script used for Amharic and Tigrinya among other languages, words are explicitly delimited (at least historically) with a non-whitespace character.

Word splitting is the process of parsing concatenated text (i.e. text that contains no spaces or other word separators) to infer where word breaks exist.

Sentence segmentation **is the problem of dividing a string of written language into its component sentences. In English and some other languages, using punctuation, particularly the full stop character is a reasonable approximation. However, even in English this problem is not trivial due to the use of the full stop character for abbreviations, which may or may not also terminate a sentence. For example *Mr.* is not its own sentence in "*Mr. Smith went to the shops in Jones Street.*" When processing plain text, tables of abbreviations that contain periods can help prevent incorrect assignment of sentence boundaries. As with word segmentation, not all written languages contain punctuation characters which are useful for approximating sentence boundaries.**

Bal Krishna Subedi

Other segmentation problems: Processes may be required to segment text into segments besides words, including morphemes (a task usually called morphological analysis), paragraphs, topics or discourse turns.

A document may contain multiple topics, and the task of computerized text segmentation may be to discover these topics automatically and segment the text accordingly. The topic boundaries may be apparent from section titles and paragraphs. In other cases one needs to use techniques similar to those used in document classification. Many different approaches have been tried.

- Syntax Structure:

Same concept as in the syntactic analysis above

- Semantic Structure:

Same concept as in the semantic analysis above

- Pragmatic Analysis:

This is high level knowledge which relates to the use of sentences in different contexts and how the context affects the meaning of the sentences. It is the study of the ways in which language is used and its effect on the listener. Pragmatic comprises aspects of meaning that depend upon the context or upon facts about real world.

Pragmatics – Handling Pronouns

Handling pronouns such as "he", "she" and "it" is not always straight forward. Let us see the following paragraph.

"John buys a new telescope. He sees Mary in the distance. He gets out his telescope. He looks at her through it".

Here, "*her*" refers to *Mary* who was not mentioned at all in the previous sentences. John's telescope was referred to as "*a new telescope*", "*his telescope*" and "*it*".

Let us see one more example

"When is the next flight to Sydney?"
"Does it have any seat left?"

Here, "*it*", refers to a particular flight to Sydney, not Sydney itself.

Pragmatics – Ambiguity in Language

A sentence may have more than one structure such as

Bal Krishna Subedi

"I saw an astronomer with a telescope."

This English sentence has a prepositional phrase "*with a telescope*" which may be attached with either with verb to make phrase "*saw something with telescope*" or to object noun phrase to make phrase "*a astronomer with a telescope*". If we do first, then it can be interpreted as "*I saw an astronomer who is having a telescope*", and if we do second, it can be interpreted as "*Using a telescope I saw an astronomer*".

Now, to remove such ambiguity, one possible idea is that we have to consider the context. If the knowledge base (KB) can prove that whether the telescope is with astronomer or not, then the problem is solved.

Next approach is that; let us consider the real scenario where the human beings communicate. If A says the same sentence "*I saw an astronomer with a telescope.*" To B, then in practical, it is more probable that, B (listener) realizes that "*A has seen astronomer who is having a telescope*". It is because, normally, the word "*telescope*" belongs to "*astronomer*", so it is obvious that B realizes so.

If A has says that "*I saw a lady with a telescope.*" In this case, B realizes that "*A has seen the lady using a telescope*", because the word "*telescope*" has not any practical relationship with "*lady*" like "*astronomer*".

So, we may be able to remove such ambiguity, by defining a data structure, which can efficiently handle such scenario. This idea may not 100% correct but seemed more probable.

Machine Vision:

Machine vision (MV) is the application of computer vision to industry and manufacturing. Whereas computer vision is the general discipline of making computers see (understand what is perceived visually), machine vision, being an engineering discipline, is interested in digital input/output devices and computer networks to control other manufacturing equipment such as robotic arms and equipment to eject defective products.

Machine vision is the ability of a computer to "see." A machine-vision system employs one or more video cameras, analog-to-digital conversion (ADC), and digital signal processing (DSP). The resulting data goes to a computer or robot controller. Machine vision is similar in complexity to voice recognition . The machine vision systems use video cameras, robots or other devices, and computers to visually analyze an operation or activity. Typical uses include automated inspection, optical character recognition and other non-contact applications.

Two important specifications in any vision system are the sensitivity and the resolution. Sensitivity is the ability of a machine to see in dim light, or to detect weak impulses at invisible wavelengths. Resolution is the extent to which a machine can differentiate between objects. In general, the better the resolution, the more confined the field of vision. Sensitivity and resolution are interdependent. All other factors held constant, increasing the sensitivity reduces the resolution, and improving the resolution reduces the sensitivity.

One of the most common applications of Machine Vision is the inspection of manufactured goods such as semiconductor chips, automobiles, food and pharmaceuticals. Just as human inspectors working on assembly lines visually inspect parts to judge the quality of workmanship, so machine vision systems use digital cameras, smart cameras and image processing software to perform similar inspections.

Machine vision systems have two primary hardware elements: the camera, which serves as the eyes of the system, and a computer video analyser. The recent rapid acceleration in the development of machine vision for industrial applications can be attributed to research in the areas of computer technologies. The first step in vision analysis is the conversion of analog pixel intensity data into digital format for processing. Next, an appropriate computer algorithm is employed to understand the image data and provide appropriate analysis or action.

Machine vision encompasses computer science, optics, mechanical engineering, and industrial automation. Unlike computer vision which is mainly focused on machine-based image processing, machine vision integrates image capture systems with digital input/output devices and computer networks to control manufacturing equipment such as robotic arms. Manufacturers favour machine vision systems for visual inspections that require high-speed, high-magnification, 24-hour operation, and/or repeatability of measurements.

A typical machine vision system will consist of most of the following components:

- One or more digital or analogue cameras (black-and-white or colour) with suitable optics for acquiring images, such as lenses to focus the desired field of view onto the image sensor and suitable, often very specialized, light sources
- Input/Output hardware (e.g. digital I/O) or communication links (e.g. network connection or RS-232) to report results
- A synchronizing sensor for part detection (often an optical or magnetic sensor) to trigger image acquisition and processing and some form of actuators to sort, route or reject defective parts
- A program to process images and detect relevant features.

The aim of a machine vision inspection system is typically to check the compliance of a test piece with certain requirements, such as prescribed dimensions, serial numbers, presence of components, etc. The complete task can frequently be subdivided into independent stages, each checking a specific criterion. These individual checks typically run according to the following model:

1. Image Capture
2. Image Preprocessing
3. Definition of one or more (manual) regions of interest
4. Segmentation of the objects
5. Computation of object features
6. Decision as to the correctness of the segmented objects

Naturally, capturing an image, possible several for moving processes, is a pre-requisite for analysing a scene. In many cases these images are not suited for immediate examination and require pre-processing to change certain sizing specific structures etc. In most cases it is at least approximately known which image areas have to be analysed, i.e. the location of a mark to be read or a component to be verified. These are called Regions of Interest (ROIs) (sometimes Area of Interest or AOIs). Of course, such a region can also comprise the entire image if required.

Machine vision is used in various industrial and medical applications. Examples include:

- Electronic component analysis
- Signature identification
- Optical character recognition
- Handwriting recognition
- Object recognition
- Pattern recognition
- Materials inspection
- Currency inspection
- Medical image analysis

Computer Vision:

Computer vision is the science and technology of machines that see, where *see* in this case means that the machine is able to extract information from an image that is necessary to solve some task. As a scientific discipline, computer vision is concerned with the theory behind artificial systems that extract information from images. The image data can take many forms, such as video sequences, views from multiple cameras, or multi-dimensional data from a medical scanner.

As a technological discipline, computer vision seeks to apply its theories and models to the construction of computer vision systems. Examples of applications of computer vision include systems for:

- Controlling processes (e.g., an industrial robot or an autonomous vehicle).
- Detecting events (e.g., for visual surveillance or people counting).
- Organizing information (e.g., for indexing databases of images and image sequences).
- Modeling objects or environments (e.g., industrial inspection, medical image analysis or topographical modeling).
- Interaction (e.g., as the input to a device for computer-human interaction).

Computer vision is closely related to the study of biological vision. The field of biological vision studies and models the physiological processes behind visual perception in humans and other animals. Computer vision, on the other hand, studies and describes the processes implemented in software and hardware behind artificial vision systems. Interdisciplinary exchange between biological and computer vision has proven fruitful for both fields.

Computer vision is, in some ways, the inverse of computer graphics. While computer graphics produces image data from 3D models, computer vision often produces 3D models from image data. There is also a trend towards a combination of the two disciplines, e.g., as explored in augmented reality.

Sub-domains of computer vision include scene reconstruction, event detection, video tracking, object recognition, learning, indexing, motion estimation, and image restoration.