INFERENCE

Inferencing is often linked to reasoning. It is the process of deriving answers of problems from knowledge that is maintained in a KBS. It may generally be seen as reaching a decision through reasoning.

Automated Reasoning

Reasoning is the process of drawing inferences or conclusions; moving from what is known (fact) to what is unknown (inference)

The program for inferencing is typically called the inference engine or control program. In rule based systems it is also called a rule interpreter.

Purpose: Direct the search through the knowledge base. The engine determines the following:

- Which rules to investigate,
- Which to eliminate, and
- Which will attribute to a match.

Types of Reasoning:

i) Reasoning from signs:

Assuming that a sign or symbol represents or indicates something specific.

Example: making a phone call and receiving a busy signal. The assumption is that someone is already on the line.

ii) Cause and effect:

Inferring that one event causes another

"I was late because my alarm didn't go off." Or reasoning with the result in mind e.g. "If I don't study, I will fail the test."

iii) Reasoning by analogy:

Drawing a parallel between two similar events, people, or traits. Based on comparisons.

Example: If all mothers are women, and Faith is a mother, then Faith must be a woman.

Care should be taken to always check analogies for validity, because these don't always work backwards. E.g. Because Faith is a woman doesn't mean she's a mother.

iv) Reasoning by example:

Basing conclusions on facts, illustrations, or cases.

Example: Conclusions drawn from the results of polls or surveys. When drawing conclusions, always ask, "Do I have enough facts to support this?"

v) Other Types of Reasoning

- ***Formal reasoning:*** use rules, facts; predicate calculus; mathematical logic.
- ***Generalization and abstraction***: use sets; induction and deduction.
- ***Procedural reasoning***: use formulae and algorithms to solve problems.

Reasoning Methods

Reasoning and inference allows generation of new and useful knowledge, conclusions and recommendations. Inference/reasoning methods fall into the general categories.

i). Deductive Reasoning.
ii). Inductive Reasoning.

DEDUCTIVE REASONING

- Conclusions must follow from their premises
- The method used in propositional and predicate calculus (logical reasoning).
- The inference engine combines the rule (or predicates) to arrive at the final answer.
- Starts with a set of axioms, or commonly accepted premises that one cannot derive from a system itself.
- Using the axioms and proven formulae, deductive reasoning can deduce new conclusions.
- You thus build an **axiomatic system**, which can be used to infer new facts or prove that a certain fact is true

Process of Deductive Reasoning

The inference process consists of two parts.

- **Single inference**
  - This is the process of applying inference rules to combine two pieces of knowledge to derive new premises. This may be an intermediate step or final recommendation.
- **Multiple Inference**
  - The sequence or order of applying the single inference process to the entire KB in order to derive final conclusions.

  NB: the knowledge processing capability of an intelligent system is composed by the two parts above.

i) Single Inference in Deductive Reasoning.

Uses the inference rules stated earlier – these mostly derive new statements by combining two previously accepted ones. These include modus ponens, modus tollens etc. These must be a tautology to be true.

e.g A=>B, A is then conclude B
(A=>B) ∧A |-B or ((A=>B) ∧ A)=>B — Modus ponens

e.g A=>B, B=>C, conclude A=>C
i.e. (A=>B) ∧ (B=>C)=>A=>C) — Resolution rule /Hypothetical syllogism

e.g (A=>B∧¬B) conclude ¬A
((A=>B)∧ ¬B) =>¬A — Modus tollens.

- Modus tollens, modus ponens and hypothetical syllogism are used with Propositional Logic.
- FOL uses rules in PL in addition to others – these include And-Elimination, And-introduction, Or-introduction, double – negative elimination, universal elimination, unit resolution, resolution e.t.c

NB: All inference rules can be replaced by a single one called **resolution method**. But you must modify the system of logics to **clausal form**

Deriving information from FOL using Resolution Method and Unification

1) Resolution Method

- An algorithms for proving facts true or false by virtue of contradiction e.g. to prove X is true, we show that the negation of X is not true e.g.
  Fact:
  i) ¬ Feathers (X) V Bird (X) meaning " either X does not have feathers or else X is a bird "
  ii) Feathers (X) meaning "X has feathers"
- To prove X is a bird, we first add an assumption that is the negative of that predicate. The facts then become:
  i) Same as above.
  ii) Same as above
  iii) ¬Bird (X)
- In sentences 1 & 2, ¬Feathers (X) and Feathers (X) cancel each other out. Resolving 1 & 2 produces the resolved, sentence and which is added.
  i) Same as above.
  ii) Same as above
  iii) Same as above
  iv) Bird(X)
- It is clear 3 & 4 cannot both be true: either X is a bird or not-thus a contradiction. Thus we have just proved our first assumption 3 is false 4 must be true.

2) Unification

An operation that tries to find consistent variable bindings (substitutions) for two terms

- a substitution is the simultaneous replacement of variable instances by terms, providing a "binding" for the variable
- without unification, the matching between rules would be restricted to constants

e.g.

(1) hates (X, Y)
(2) hates (Wainaina, Beer)

Unify 2 with 1 by binding Wainaina to X and Beer to Y.

(3) Hates (Z, Beer)

Unify 6 with 1 X with Z, Y with Beer.

Note: Prolog uses a generalized version of unification algorithm called **match** (i.e. substitutions may occur only in one direction).

ii) Multiple Inferences in Deductive reasoning.

Most commonly used methods are:

1) Backward chaining.
2) Forward chaining.

1. Backward Chaining

It is an inference mechanism that begins with the goal (conclusion) and sets out to look for evidence to establish or disprove the goal.

I.e. the inference engine starts from the consequent of a rule and goes backward to the antecedent or **IF** part of the rule.

Backward Chaining Algorithm

i). State a specific goal (question)
ii). Find rules which resolve the goal
iii). At runtime, answer questions to satisfy the antecedents of the rules as required
iv). Obtain a result (goal resolved or not)

Example

Consider a knowledge base that consists of the following rules

i). $A \rightarrow C$
ii). $D \rightarrow E$
iii). $B \wedge C \rightarrow F$
iv). $E \vee F \rightarrow G$

- The goal is to establish whether G is true.
- The inference engine check to see which rule has G as its consequent.
  Rule 4 fires
- To conclude from this that G is true, E or F must be shown to be true.
- E becomes the current goal of the system
- The inference engine checks to see which rule has E as its consequent
  Rule 2 fires
- The current goal becomes D
- The inference engine checks to see which rule has D as its consequent
- There is none.
- It asks the user to provide information about D
- D is True
- The inference engine concludes that E is true and G is also true.
- D is False
- The system fails to establish the truth of E
- In this case, it backtracks and picks up F to get at the goal G.
- B and C become the current goal.
- B is not the consequent of any rule and the system asks the user if it is true.
- B is False
  No matter what C is, the conditional part of rule 3 is false and the inference engine fails to establish whether G is true.
- B is True
  The inference engine pursues C etc

- There is a clear set of statements, which must be confirmed or denied.
- A large number of questions could be asked of the user, but typically only a few are necessary to resolve a situation.
- It is desirable to have interactive dialogue with the user.
- Rule execution depends on data gathering that may be expensive or difficult.

Characteristics of backward chaining:

i). Good for Diagnosis.
ii). Looks from present to past.
iii). Works from consequent to antecedent.
iv). Is goal-driven, top-down reasoning.
v). It facilitates a depth-first search.

2. Forward Chaining

It refers to an inference mechanism which begins with evidence (facts), then sets out to draw a conclusion. The system requires the user to provide facts pertaining to the problem. I.e. the inference engine tries to match each fact with the antecedent or if part of a rule. If the match succeeds, the rule fires and the consequent of that rule is established and is added to the known facts of the case. The process continues until the inference engine has drawn all possible conclusions by matching facts to antecedents of rules in the knowledge base.

Forward Chaining Algorithm.

i). Enter new data
ii). Fire forward chaining rules
iii). Rule actions infer new data values
iv). Go to step 2
v). Repeat until no new data can be inferred
vi). If no solution, rule base is insufficient

Example

Consider a knowledge base that consists of the following rules

i). $A \rightarrow C$
ii). $D \rightarrow E$
iii). $B \wedge C \rightarrow F$
iv). $E \vee F \rightarrow G$

Establish whether G is true given that A and B are the known facts

- Using rule 1 inference engine concludes that C is true.
- Using rule 3 it concludes that F is true.
- Using rule 4 it then concludes that G is true.

Characteristics of forward chaining:

- Good for monitoring, planning, and control
- Looks from present to future.
- Works from antecedent to consequent.
- Is data-driven, bottom-up reasoning.
- Works forward to find what solutions follow from the facts.
- It facilitates a breadth-first search.

Backward chaining Vs Forward chaining.

Backward Chaining: - Useful when number of goals is small. In some systems, the user has the choice of telling the inference engine which goal it must pursue – goal selection by user.

Forward chaining: - Performs well when no of goals is large the user has a given set of facts at the start & wants to find the implications of these facts.

INDUCTIVE REASONING

- **Definition**: Inductive reasoning is the process of going from specific cases to general rules e.g. the use of quantitative methods e.g in statistics.
- Learning is through experience – and this is an example for induction process- where exposure to a number of similar circumstances that lead to a certain consequence leads to generalization and development of rules for dealing with that group of circumstances.

Methods for inductive reasoning

i) **Decision trees: -** Here you must have a set of representative cases for which the goal or decision value is known.
ii) **ID3: -** Tree with minimum number of nodes. Show the tree.
iii) **Case Based approach: -** Reasoning by analogy

Exercise

Consider the following investment decision-making given in the following symbolic form:

- A=100,000 Shillings
- B=Younger than 40 years of age
- C= Education at University level
- D= Annual income of at least 800,000 Shillings
- E=Invest at Treasury bills
- F=Invest at growth stocks
- G=Invest at NSE

The rules are given as follows: -

R1: IF A AND C THEN E
R2: IF D AND C THEN F
R3: IF B AND E THEN F
R4: IF B THEN C
R5: IF F THEN G

An investor has 100,000 Shillings and is younger than 40 years and wants an advice from an ES on whether to invest in NSE. Using backward chaining inference procedures advise the investor.

UNCERTAINTY

It occurs in various situations where the relevant information is deficient in some way. The following are possible sources of uncertainty:

i). Partial information.
ii). Information not reliable.
iii). Representation language unreliable.
iv). Information from multiple sources is conflicting.

Other sources include: Error, lack of confidence, imprecision, unreliability, variability, vagueness, ignorance, ambiguity etc.

Uncertainty is usually present in most tasks that require intelligent behavior e.g.

- Planning
- Reasoning
- Problem solving
- Decision Making
- Classification

Management of uncertainty is central to the development of computer based systems that execute the above tasks.

AI classification of uncertainty

Uncertainty has two distinctions in AI

i). Top level
  - **Unary:** Apply to individual propositions
  - **Set theoretic:** Apply to sets of propositions

ii). Second level
  - **Ignorance:** Lack knowledge
  - **Conflict:** Conflicting knowledge

iii). Tertiary Level
  - Leads to further Distributions

Handling Uncertainty: The following are various method of dealing with uncertain reasoning:

i). Uncertain answers
ii). Confidence factors
iii). Probabilities
iv). Fuzzy Logic

i). **Uncertain Answers**
  - It is a simple method of dealing with uncertainty, which allows Yes / No / Unknown answers
  - 'Unknown' is processed by triggering extra inferencing to determine answer when user cannot

ii). **Confidence/Certainty Factors**
  - Confidence expressed as a number between 0 and 1
  - Allows uncertainty to be expressed in information and in reasoning
  - Not necessarily based on any real evidence!

Example I
- Assume that if A is true, then B is true
- But we are only 80% certain that A is true
- Clearly we can only be 80% certain that B is true!

$$A \overset{0.8}{\Rightarrow} B$$

Example II
- What if we were only 80% certain that A => B, and only 60% certain that B => C?
- How certain can is it that B and C are true?
  - √ Each uncertain step in the reasoning process must make us less certain in the result - this is mimicked by multiplying Confidence Factors together.
  - √ Multiplying fractions reduces the total at each step
  - √ Thus, given we are only 80% certain of A, we can only be 64% certain that B is true i.e. (0.8 * 0.8 = 0.64)
  - √ and only 38% certain that C is true i.e. (0.8*0.8*0.6 = 0.38)
- However, two independent pieces of corroborating evidence must make us more certain of the result. i.e.

$$\text{Given } A \overset{0.8}{\Rightarrow} C \quad \text{AND} \quad B \overset{0.8}{\Rightarrow} C$$

- Given A and B as true how confident are we of C?
- Clearly the answer should be higher than 0.8 but less than 1(calculate)

Advantages of CFs

i). CFs allow people to express varying degrees of confidence
ii). Easy to manipulate
iii). Can work well when trying to rank several possible solutions
iv). Must be careful not to place too much emphasis on the actual numbers generated

Disadvantages of CFs

i). CF's come from the opinions of one or more experts and thus have very little basis in fact
ii). People are unreliable at assigning confidence or certainty values i.e. two people will assign very different numbers and will often themselves be inconsistent from day to day

Bayes' Theorem
It provides a statistical theory of evidence, which is used to update the probabilities attached to certain facts in the light of new evidence. The fundamental is that of conditional probability given as:

P (H/E),

This is read as the probability that a Hypothesis H is true given that we have observed evidence E for the hypothesis. I.e. It represents uncertainty by describing a model of the application domain as a set of possible outcomes known as the Hypothesis.

Example

- A medical expert attempting to diagnose a patient illness.
  - √ The Bayesian approach would treat each patient illness e.g. influenza, bronchitis etc as possible hypothesis in the problem space.
  - √ Each hypothesis has an initial probability of occurring (even if it may be difficult to find).
- The Bayesian method of inference requires an initial estimate of the probability for each of the possible hypotheses in the problem space – **Prior Probability.**
- The prior probability must be known, or estimated in advance.
- When prior probabilities are known Bayesian inference updates hypotheses by using the answers to questions supplied by the user during the running of the system e.g.
  - √ "What is the body temperature of the patient?".
  - √ "Is the patient suffering from dizziness?".
- This is known as **evidence**.
- Each item of evidence updates the probability of each of the hypotheses being considered.

The following is the formula for Bayes theorem:

$$P(H_i/E) = \frac{P(E/H_i)\times P(H_i)}{\left[\sum_{i=1}^{n} P(E/H_i)\times P(H_i)\right]}$$

Where

- P (Hi/E) – Probability that hypothesis Hi is true given evidence E
- P(E/Hi) – Probability that we will observe evidence E given the Hi is true
- P(Hi) – Prior probability that hypothesis i is true in the absence of any evidence:
- n – No of possible hypothesis.

Example

There are 4 coins in a jar. Each coin is either a Sh. 1 or Sh. 5. There is at least one of each present in the jar. One is required to make inferences about the actual composition of coins in the jar by selecting one coin taken from the jar at random.

Solution

Hypotheses

- H1 – Sh. 1 and three Sh. 5
- H2 – two Sh. 1 and two Sh. 5
- H3 – three Sh.1 and one Sh. 5

Note

- The hypothesis are mutually exclusive and exhaustive since none can occur together
- The probability of each hypothesis is the same since they all have an equally likely chance of being selected. Probability of each is:

$$P(H1) = P(H2) = P(H3) = \frac{1}{3}$$

- A coin is selected at random from the jar and it turns out to be a Sh. 5 coin. How does this event change the probabilities of the hypotheses?
- Selecting Sh. 5 – **Evidence** E.

- The conditional probability of E occurring given H1 (Hypothesis 1).

$P(E/H1) = \frac{3}{4}$ $\quad P(E/H2) = \frac{2}{4} = \frac{1}{2}$ $\quad P(E/H3) = \frac{1}{4}$

Applying Bayes Theory: $P(H1/E) = \frac{P(E/H_1) \times P(H_1)}{\left[\sum_{i=1}^{n} P(E/H_i) \times P(H_i)\right]}$

$$P(H1/E) = \frac{\left(\frac{3}{4} \times \frac{1}{3}\right)}{\left(\frac{3}{4} \times \frac{1}{3}\right) + \left(\frac{2}{4} \times \frac{1}{3}\right) + \left(\frac{1}{4} \times \frac{1}{3}\right)} = \frac{\frac{1}{4}}{\frac{6}{12}} = \frac{1}{2}$$

- The probability of the hypothesis H1 has increased from 1/3 to 1/2 as a result of observing 5p coin.
- The probability of hypothesis H2 and H3can be found in the same way

$$P(H_2/E) = \frac{P(E/H_2) \times P(H_2)}{\left[\sum_{i=1}^{n} P(E/H_i) \times P(H_i)\right]} = \frac{\left(\frac{2}{4} \times \frac{1}{3}\right)}{\left(\frac{3}{4} \times \frac{1}{3}\right) + \left(\frac{2}{4} \times \frac{1}{3}\right) + \left(\frac{1}{4} \times \frac{1}{3}\right)} = \frac{\frac{1}{6}}{\frac{6}{12}} = \frac{1}{3}$$

$$P(H_3/E) = \frac{P(E/H_3) \times P(H_3)}{\left[\sum_{i=1}^{n} P(E/H_i) \times P(H_i)\right]} = \frac{\left(\frac{1}{4} \times \frac{1}{3}\right)}{\left(\frac{3}{4} \times \frac{1}{3}\right) + \left(\frac{2}{4} \times \frac{1}{3}\right) + \left(\frac{1}{4} \times \frac{1}{3}\right)} = \frac{\frac{1}{12}}{\frac{6}{12}} = \frac{1}{6}$$

Advantages/Disadvantages

i). Bayes theorem is mathematically sound therefore results using this method can be justified.

ii). However, this method needs data collected from previous results and will only work where this is available.

iii). This is in itself a strong point as the data can be proven where CFs are only expressions of opinion.

Note: Another method of handling uncertainty is the use of fuzzy logic