## UNIT III

Reasoning under uncertainty: Logics of non-monotonic reasoning-Implementation- Basic probability notation - Bayes rule – Certainty factors and rule based systems-Bayesian networks – Dempster - Shafer Theory - Fuzzy Logic

## REASONING UNDER UNCERTAINTY UNCERTAINTY

Let action At = leave for airport t minutes before flight Will At get me there on time?

## Problems:

* 1. partial observability (road state, other drivers' plans, etc.)
	2. noisy sensors (KCBS tra\_c reports)
	3. uncertainty in action outcomes (at tire, etc.)
	4. immense complexity of modelling and predicting tra\_c 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."

## Methods for handling uncertainty

Default or nonmonotonic logic:

Assume my car does not have a at tire

Assume A25 works unless contradicted by evidence

Issues: What assumptions are reasonable? How to handle contradiction?

**Rules with fudge factors**:



Issues: Problems with combination, e.g., Sprinkler causes Rain??

## NON-MONOTONIC REASONING .

A **non-monotonic logic** is a [formal logic](http://en.wikipedia.org/wiki/Formal_logic) whose [consequence](http://en.wikipedia.org/wiki/Logical_consequence) [relation](http://en.wikipedia.org/wiki/Relation_%28mathematics%29) is not [monotonic](http://en.wikipedia.org/wiki/Monotonicity_of_entailment). Most studied formal logics have a monotonic consequence relation, meaning that adding a formula to a theory never produces a reduction of its set of consequences. Intuitively, monotonicity indicates that learning a new piece of knowledge cannot reduce the set of what is known.

A monotonic logic cannot handle various reasoning tasks such as [reasoning by default](http://en.wikipedia.org/wiki/Default_logic) (consequences may be derived only because of lack of evidence of the contrary), [abductive](http://en.wikipedia.org/wiki/Abductive_reasoning) [reasoning](http://en.wikipedia.org/wiki/Abductive_reasoning) (consequences are only deduced as most likely explanations), some important approaches to reasoning about knowledge

## Default reasoning

An example of a default assumption is that the typical bird flies. As a result, if a given animal is known to be a bird, and nothing else is known, it can be assumed to be able to fly. The default assumption must however be retracted if it is later learned that the considered animal is a penguin. This example shows that a logic that models default reasoning should not be monotonic.

Logics formalizing default reasoning can be roughly divided in two categories: logics able to deal with arbitrary default assumptions ([default logic](http://en.wikipedia.org/wiki/Default_logic), [defeasible logic](http://en.wikipedia.org/wiki/Defeasible_logic)/[defeasible](http://en.wikipedia.org/wiki/Defeasible_reasoning) [reasoning](http://en.wikipedia.org/wiki/Defeasible_reasoning)/[argument (logic)](http://en.wikipedia.org/wiki/Argument_%28logic%29), and [answer set programming](http://en.wikipedia.org/wiki/Answer_set_programming)) and logics that formalize the specific default assumption that facts that are not known to be true can be assumed false by default ([closed world assumption](http://en.wikipedia.org/wiki/Closed_World_Assumption) and [circumscription](http://en.wikipedia.org/wiki/Circumscription_%28logic%29)).

## Abductive reasoning

[Abductive reasoning](http://en.wikipedia.org/wiki/Abductive_reasoning) is the process of deriving the most likely explanations of the known facts. An abductive logic should not be monotonic because the most likely explanations are not necessarily correct.

For example, the most likely explanation for seeing wet grass is that it rained; however, this explanation has to be retracted when learning that the real cause of the grass being wet was a sprinkler. Since the old explanation (it rained) is retracted because of the addition of a piece of knowledge (a sprinkler was active), any logic that models explanations is non-monotonic.

## Reasoning about knowledge

If a logic includes formulae that mean that something is not known, this logic should not be monotonic. Indeed, learning something that was previously not known leads to the removal of

the formula specifying that this piece of knowledge is not known. This second change (a removal caused by an addition) violates the condition of monotonicity. A logic for reasoning about knowledge is the [autoepistemic logic.](http://en.wikipedia.org/wiki/Autoepistemic_logic)

## Belief revision

[Belief revision](http://en.wikipedia.org/wiki/Belief_revision) is the process of changing beliefs to accommodate a new belief that might be inconsistent with the old ones. In the assumption that the new belief is correct, some of the old ones have to be retracted in order to maintain consistency. This retraction in response to an addition of a new belief makes any logic for belief revision to be non-monotonic. The belief revision approach is alternative to [paraconsistent logics,](http://en.wikipedia.org/wiki/Paraconsistent_logics) which tolerate inconsistency rather than attempting to remove it.

## IMPLEMENTATION ISSUES IN NON MONOTONIC REASONING:

The logical frameworks are not enough for implementing non monotonic reasoning in problem solving programs. These are the some weakness for logical systems. The four important problems that arise in real systems are us follows.

* The first is how to derive exactly those monotonic conclusions that are relevant to solving the problem at hand while not wasting not wasting time on those that, while they may be licensed by the logic.
* The second problem is how to update our knowledge incrementally as problem solving progresses.
* The third problem is that in nonmonotonic reasoning systems, it often happens that more than one interpretation of the known facts is licensed by the available inference rules.
* The final problem is that, in general, these theories are not computationally effective.

## Implementation – Depth first search

The implementation of DFS is directed by

* + Dependency directed Backtracking
	+ Justification based truth maintenance systems
	+ Logic based truth maintenance systems.

## Dependency directed Backtracking:

The following scenarios are likely to occur often in non monotonic reasoning (DFS).

* Analyze the fact, F which cannot be derived monotonic from what we already know but which can be derived by making some assumption A which seems plausible. So we make assumption A, derive F, and then derive some additional facts G and H from F.
* We later derive some other facts M and N, but they are completely derive some additional facts G and H from F. We later derive some other facts M and N, but they are completely independent of A and F.
* A little while later, a new fact comes in that invalidates A. We need to rescind our proof of F, and also our proofs of G and H since they depended on F.
* But what about M and N? They didn’t depend on F, so there is no logical need to invalidate them. But if we use a conventional backtracking scheme, we have to back up past conclusions in the order in which they derived them.
* To get around this problem, we need to backup past M and N, thus undoing them in order to get back to F,G,H and A. To get around this problem, we need a slightly different notion of backtracking, one that is based on logical dependencies rather than the chronological order in which decisions were made. We call this new method dependency directed backtracking.

## Justification based truth maintenance systems



* This is a simple TMS in that it does not know anything about the structure of the assertions themselves.
* Each supported belief (assertion) in has a justification.
* Each justification has two parts:
* An IN-List -- which supports beliefs held.
* An OUT-List -- which supports beliefs not held.
* An assertion is connected to its justification by an arrow.
* One assertion can feed another justification thus creating the network.
* Assertions may be labeled with a belief status.
* An assertion is valid if every assertion in the IN-List is believed and none in the OUT- List are believed.
* An assertion is non-monotonic is the OUT-List is not empty or if any assertion in the IN- List is non-monotonic.

The main key reasoning operations that a JTMS performed are as follows:

* Consistent labeling.
* Contradiction.
* Applying rules to derive conclusions.
* Creating justifications for the results of applying rules.
* Choosing among alternative ways of resolving a contradiction.
* Detecting contradictions.

## Logic-Based Truth Maintenance Systems (LTMS)

Similar to JTMS except:

* Nodes (assertions) assume no relationships among them except ones explicitly stated in justifications.
* JTMS can represent P and P simultaneously. An LTMS would throw a contradiction here.
* If this happens network has to be reconstructed.

## BASIC PROBABILITY NOTATION

An important goal for many problem solving systems is to collect evidence as the system goes along to modify its behavior on basis of evidence.

To model this behavior, we need a statistical theory of evidence. Bayesian theory is such a theory.

The fundamental notion of Bayesian statistics is than of conditional probability.

**P(H\E)** = the Probability that hypothesis H is true given evidence E.

**P(E\H)** = the Probability that we will observe evidence E given that hypothesis i is true.

**P(H)** = the priori probability that hypothesis i is true in absence of any specific evidence. These probabilities are called prior probabilities or priors.

**K** = the number of possible hypothesis.

# Baye’s Theorem then states that

## P(H\E)= P(E\H) P(H)

**Σn=1 P (E\Hn) P(Hn)**

## EXAMPLE:

We are interested in examining the geological evidence at a particular location to determine whether that would be a good place to dig to find a desired mineral.

1. If we know the probability of each minerals in prior and its physical characteristics then we can use baye’s formula to compute.
2. From based upon the evidence we collect how likely it is various minerals are present in particular place can be identified.

## Key to use Baye’s Theorem:

P(A\B)= conditional probability of A given we have only evidence of B.

## Example: Solving Medical Diagnosis problem S: Patient has Spots

**M: Patient has Measles F: Patient has High Fever**

* 1. Presence of spots serves as evidence in favor of Measles. It also serves evidence for fever, measles cause fever.
	2. Either spot or fever alone causes evidence in favor of measles.
	3. If both are present we need to take both into account in determining total weight of evidence.

## P(H\E,e)= P(H\E) P(e/E,H)

**P(e/E)**

## BAYES' RULE



**Probability**

Given the available evidence, A25 will get me there on time with probability 0:04 (Fuzzy logic handles degree of truth NOT uncertainty e.g., WetGrass is true to degree 0:2)

Probabilistic assertions summarize effects of

laziness: failure to enumerate exceptions, qualifications, etc. ignorance: lack of relevant facts, initial conditions, etc.

# CERTAINTY FACTOR AND RULE BASED SYSTEM

* 1. It is one of practical way of compromising pure Bayesian system.
	2. The approach we discuss was pioneered in the MYCIN system, which attempt to recommend appropriate therapies for patient with bacterial infections.
	3. It interacts with the physician to acquire the clinical data it needs.
	4. MYCIN is an example of an expert system since it normally done by a expert system.
	5. MYCIN uses rules to reason clinical data from its goal of finding significant disease causing organisms.
	6. Once it finds the organisms, it then attempt to select a therapy by which the disease (s) may be treated.

## CERTAINTY FACTOR:

A Certainty factor (CF [h,e]) is defined in terms of 2 components

1. MB[h,e] --- a Measure of Belief (between 0 and 1) of belief in hypothesis h given the evidence e. MB measures the extent to which the evidence supports the hypothesis. It is zero if the evidence fails to support the hypothesis.
2. MD[h,e] --- a Measure of DisBelief (between 0 and 1) of Disbelief in hypothesis h given the evidence e. MD measures the extent to which the evidence supports the negation of the hypothesis. It is zero if the evidence supports the hypothesis.

From these two measures we can define the certainty factor

**CF[h,e]= MB[h,e]-MD[h,e]** eq no 1

## Combining Uncertainty rules:

1. MYCIN reflect the experts’ assessments of the strength of evidence in support of hypothesis.
2. MYCIN reasons however there CF’s need to be combined to reflect the operations of multiple pieces of evidence and multiple rules are applied to the problem.

The measure of belief and disbelief is given by S1 and S2

**MB(h, S1∧S2)**{MB(h,S1) + MB(h, S2) [1-MB(h,S1) …………Equation no 2

**MD(h, S1∧S2)**{MD(h,S1) + MD(h, S2) [1-MD(h,S1)……….. Equation no 3

Suppose we make an initial observation that confirms our belief in h with **MB(h, S1)= 0.3**

## MD[h, s1]=0, MB[h,s2]=0.2 and CF[h,s1]=0.3

Substituting these in **equation no 2** weget MB(h, S1**∧S2) =** {0.3 +0.2 (1-0.3)

= 0.3 + 0.2 (0.7)

= 0.44… Equation no 4

MD(h, S1**∧ S2)= 0.0** Equation no 5

Substituting eq4 and eq5 in eq1 we get **CF[h,e]= MB[h,e]-MD[h,e]** = 0.44 -0

## Certainty Factor (h,e)= 0.44 equation no 6

The formula MYCIN uses for the MB of the conjunction and disjunction of 2 hypothesis.

## By using Bayes theorem:

**MB[h,e]=** {**max [p(h/e), p(h)]- p(h)**

## P(h/e) 1-p(h)

From eq 4, 5 and 6 can be written as single rule rather that 3 separate rules.

MYCIN independence assumption can make moment 3 separate rules **CF each was 0.6**

MB [h, S1**∧S2] = 0.6+ 0.6 (1-0.6)**

**= 0.6 + 0.6(0.4)**

**= 0.84**

MB [h, (S1**∧S2)∧S3] = 0.84+ 0.6 (1-0.84)**

**= 0.84 + 0.6(0.16)**

**= 0.936**

Let us consider a concrete example S: Sprinkler was on last night

W: grass is wet

R: it rained last night

We can write MYCIN rules that describe predictive relation among 3 events

1. If sprinkle was on last night then grass is wet in morning = evidence is 09%
2. If the grass is wet in morning then it is rained last night = 0.9-0.1=0.8% MB(W,S) =0.8

MD(R,W)= 0.8\* 0.9

= 0.72

## BAYESIAN NETWORKS

A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions

## Syntax:

a set of nodes, one per variable

a directed, acyclic graph (link \_ \directly inuences")

a conditional distribution for each node given its parents:



In the simplest case, conditional distribution represented as a conditional probability table (CPT) giving the distribution over Xi for each combination of parent values

Example

Topology of network encodes conditional independence assertions:



Weather is independent of the other variables

Toothache and Catch are conditionally independent given Cavity

***Example***

I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set of by minor earthquakes. Is there a burglar?

Variables: Burglar, Earthquake, Alarm, JohnCalls, MaryCalls Network topology reacts \causal" knowledge:

* A burglar can set the alarm off
* An earthquake can set the alarm off
* The alarm can cause Mary to call
* The alarm can cause John to call



## Compactness



**Constructing Bayesian networks**

Need a method such that a series of locally testable assertions of conditional independence guarantees the required global semantics

## Example











Deciding conditional independence is hard in noncausal directions (Causal models and conditional independence seem hardwired for humans!)

Assessing conditional probabilities is hard in noncausal directions Network is less compact: 1 + 2 + 4 + 2 + 4=13 numbers needed

Example: Car diagnosis

Initial evidence: car won't start

Testable variables (green), \broken, so \_x it" variables (orange) Hidden variables (gray) ensure sparse structure, reduce parameters

## Example: Car insurance



**Inference in Bayesian networks Inference tasks**



##  DEMPSTER - SHAFER

* The **Dempster–Shafer theory (DST)** is a mathematical theory of [evidence](http://en.wikipedia.org/wiki/Evidence). It allows one to combine evidence from different sources and arrive at a degree of belief (represented by a belief function) that takes into account all the available evidence. The theory was first developed by [Arthur P. Dempster](http://en.wikipedia.org/wiki/Arthur_P._Dempster) and [Glenn Shafer.](http://en.wikipedia.org/wiki/Glenn_Shafer)
* In a narrow sense, the term **Dempster–Shafer theory** refers to the original conception of the theory by Dempster and Shafer. However, it is more common to use the term in the wider sense of the same general approach, as adapted to specific kinds of situations. In particular, many authors have proposed different rules for combining evidence, often with a view to handling conflicts in evidence better.
* Dempster–Shafer theory is a generalization of the [Bayesian theory of subjective probability](http://en.wikipedia.org/wiki/Bayesian_probability); whereas the latter requires probabilities for each question of interest, belief functions base degrees of belief (or confidence, or trust) for one question on the probabilities for a related question.

These degrees of belief may or may not have the mathematical properties of probabilities; how much they differ depends on how closely the two questions are related.[4] Put another way, it is a way of representing [epistemic](http://en.wikipedia.org/wiki/Epistemology) plausibilities but it can yield answers that contradict those arrived at using [probability theory.](http://en.wikipedia.org/wiki/Probability_theory)

## Dempster–Shafer theory is based on two ideas:

* 1. Obtaining degrees of belief for one question from subjective probabilities for a related question.
	2. Dempster's rule for combining such degrees of belief when they are based on independent items of evidence.

In essence, the degree of belief in a proposition depends primarily upon the number of answers (to the related questions) containing the proposition, and the subjective probability of each answer. Also contributing are the rules of combination that reflect general assumptions about the data.

In this formalism a **degree of belief** (also referred to as a **mass**) is represented as a **belief function** rather than a [Bayesian](http://en.wikipedia.org/wiki/Bayesianism) [probability distribution](http://en.wikipedia.org/wiki/Probability_distribution). Probability values are assigned to *sets* of possibilities rather than single events: their appeal rests on the fact they naturally encode evidence in favor of propositions.

Dempster–Shafer theory assigns its masses to all of the non-empty subsets of the entities that compose a system.[*clarification needed*]

## Belief and plausibility:

Shafer's framework allows for belief about propositions to be represented as intervals, bounded by two values, *belief* (or *support*) and *plausibility*:

*belief* ≤ *plausibility*.

*Belief* in a hypothesis is constituted by the sum of the masses of all sets enclosed by it (*i.e.* the sum of the masses of all subsets of the hypothesis).[*clarification needed*]

It is the amount of belief that directly supports a given hypothesis at least in part, forming a lower bound. Belief (usually denoted *Bel*) measures the strength of the evidence in favor of a set of propositions. It ranges from 0 (indicating no evidence) to 1 (denoting certainty).

*Plausibility* is 1 minus the sum of the masses of all sets whose intersection with the hypothesis is empty. It is an upper bound on the possibility that the hypothesis could be true, *i.e.* it “could possibly be the true state of the system” up to that value, because there is only so much evidence that contradicts that hypothesis.

Plausibility (denoted by *Pl*) is defined to be

### Pl(s)=1-Bel(~s) Equation 1

It also ranges from 0 to 1 and measures the extent to which evidence in favor of *~s* leaves room for belief in *s*.

## For example: Consider a simplified Diagnosis problem to cause FEVER All : allergy

**Flu: flu Cold: cold**

## Pneu: pneumonia

Θ might consist of the set {All, Flu, Cold, Pneu}. Each contains 0.2 % evidence to cause fever. (i.e) **All = 0.2, Flu= 0.2, Cold= 0.2, Pneu=0.2** Our goal is to attach some measure of belief to elements of Θ.

Let us see how m works for our diagnosis problem. Assume that we have no information about how to choose among four hypothses when we start the diagnosis task. Then we define m as:

## { Θ} (1.0) (i.e 100 percent evidence to cause fever)………Eq 2 Fever might be such a piece of evidence. We update m as follows:

*{ Flu, Cold, Pneu} (0.2+0.2+0.2=0.6) Eq 3*

*Where Flu, Cold, Pneu serves 60 % evidence to cause fever and remaining 40% is assigned to value* **Θ.**

### { Θ} (0.4) eq 4

At this point we assigned to the set {flu, cold, Pneu} the appropriate belief. The remainder of the belief still resides in the larger set Θ. Thus Bel(p) is our overall belief that the correct answer lies somewhere in the set p.

We are given two Belief functions m1 and m2. suppose m1 corresponds to our belief after observing fever. From equation 3 and 4.

## { Flu, Cold, Pneu} (0.6)

**{ Θ} (0.4)**

Suppose m2 corresponds to our belief after observing allergy in addition weget

## Allergy =0.2 therefore (0.6 + 0.2 =0.8) {All, Flu, Cold} (0.8)… eq no 5

**The we can compute the combinationm3 using the following table (in which we further abbreviations disease names**

### M1 (x) \* M2(x) eq no 6

|  |  |
| --- | --- |
|  | ***M2(x)*** |
| ***M1 (x)*** | **{A,FC}** | **(0.8)** | **Θ** | **(0.2)** |
| **{F,C,P}** | **(0.6)** | *{F,C}* | *(0.48)* | *{F,C,P}* | *(0.12)* |
| **Θ** | **(0.4)** | *{A,F,C}* | *(0.32)* | Θ | *(0.08)* |

As a result of applying m1 and m2, we produced a new piece of evidence over fever.

{Flu, Cold} (0.48) eq no 7

{All, Flu, Cold} (0.32) eq no 8

{Flu, Cold, Pneu} (0.12) eq no 9

Θ (0.08) eq no 10

Now let m3 corresponds to our belief given just the evidence for allergy From eq no 5 allergy= 0.8+ 0.1 =0.9 remaining 0.1% serves evidence for Θ.

{All} (0.9) eq no 11

Θ (0.1) eq no 12

We can apply the numerator of the combination rule to produce (where \* denotes the empty set)

|  |  |
| --- | --- |
|  | ***M2(x)*** |
| ***M1 (x)*** | **{A }** | **(0.9)** | **Θ** | **(0.1)** |
| **{F,C}** | **(0.48)** | *φ* | *(0.432)* | *{F,C}* | *(0.048)* |
| **{A,F,C}** | **(0.32)** | *{A,F,C}* | *(0.288)* | *{A,F,C}* | *(0.032)* |
| **{F,C,P}** | **(0.12)** | *φ* | *(0.108)* | *(F,C,P)* | *(0.012)* |
| **Θ** | **(0.08)** | *{A}* | *(0.072)* | **Θ** | *(0.008)* |

***Pl(s)=1-Bel(~s) {i.e}*** *φ= 0.54* **= 1- 0.54 = 0.46**

In this example the percentage value of m5 that was initially assigned to the empty set was large (over half)