Uncertainty Reasoning

"Nothing is certain but death and taxes" Benjamin Franklin

Non-monotonic Logic

Traditional logic is monotonic .

The set of legal conclusions grows monotonically with the set of facts appearing in our initial database.

When humans reason, we use a defeasible logic.
Almost every conclusion we draw is subject to reversal.
If we find contradicting information later, we'll want to retract earlier inferences.

Nonmonotonic logic , or Defeasible reasoning , allows a statement to be retracted.

Solution: Truth Maintenance

Keep explicit information about which facts/inferences support other inferences.
If the foundation disappears, so must the conclusion.

Uncertainty

On the other hand, the problem might not be in the fact the T/F values can change over time, but rather that we are not *certain* of the T/F value.

Agents almost never have access to the whole truth about their environment

Agents must therefore act in the presence of *uncertainty*

- Some information ascertained from facts
- Some information inferred from facts and knowledge about environment
- Some information is based on assumptions made from experience

Uncertainty Arises Because of Several Factors

Incompleteness

Many rules are incomplete because too many conditions to be explicitly enumerated
Many rules incomplete because some conditions are unknown

Incorrectness

Acting Under Uncertainty

Agents must still act even if world not certainty

If not sure which of two squares have a pit and must enter one of them to reach the gold, the agent will take a chance

If can only act with certainty, most of the time will not act. Consider example that agent wants to drive someone to the airport to catch a flight, and is considering plan A90 that involves leaving home 90 minutes before the flight departs and driving at a reasonable speed. Even though DFW is only 15 miles away, the agent will not be able to reach a definite conclusion such as "Plan A90 will get us to the airport in time, as long as my car doesn't break down or run out of gas, and I don't get into an accident, and there are no accidents on 360, and the plane doesn't leave early, and there's no thunderstorms in the area, ..."

We may stick use this plan if it will maximize gain, given known information

The performance measure here includes getting to the airport in time, not wasting time at the airport, and/or not getting a speeding ticket.

Limitation of Deterministic Logic

Pure logic fails for three main reasons:

Laziness: Too much work to list complete set of antecedents or consequents needed to ensure exceptionless rule, too hard to use the enormous rules that result

Theoretical ignorance: Science has no complete theory for the domain

Practical ignorance: Even if we know all the rules, we may be uncertain about a particular patient because all the necessary tests have not or cannot be run

---

Probability

Probabilities are numeric values between 0 and 1 (inclusive) that represent ideal certainties (not beliefs) of statements, given assumptions about the circumstances in which the statements apply.

These values can be verified by testing, unlike certainty values. They apply in highly controlled situations.

$$Probability(event) = P(event) = \frac{\#\text{ instances of the event}}{\text{total }\#\text{ instances}}$$

Example

For example, if we roll two dice, each showing one of six possible numbers, the number of total unique rolls is 6*6 = 36. We distinguish the dice in some way (a first and second or left and right die). Here is a listing of the joint possibilities for the dice:

```
(1,1)  (1,2)  (1,3)  (1,4)  (1,5)  (1,6)
(2,1)  (2,2)  (2,3)  (2,4)  (2,5)  (2,6)
(3,1)  (3,2)  (3,3)  (3,4)  (3,5)  (3,6)
(4,1)  (4,2)  (4,3)  (4,4)  (4,5)  (4,6)
(5,1)  (5,2)  (5,3)  (5,4)  (5,5)  (5,6)
(6,1)  (6,2)  (6,3)  (6,4)  (6,5)  (6,6)
```

The number of rolls which add up to 4 is 3 ((1,2), (2,2), (3,1)), so the probability of rolling a total of 4 is 3/36 = 1/12.

This does not mean 8.3% true, but 8.3% chance of it being true.

Probability Explanation

P(event) is the probability in the absence of any additional information

Probability depends on evidence.

Before looking at dice: P(sum of 4) = 1/12

After looking at dice: P(sum of 4) = 0 or 1, depending on what we see

All probability statements must indicate the evidence with respect to which the probability is being assessed.

As new evidence is collected, probability calculates are updated.

Before specific evidence is obtained, we refer to the prior or unconditional probability of the event with respect to the evidence. After the evidence is obtained, we refer to the posterior or conditional probability.

Example

I have three identical boxes labeled H1, H2, and H3
I place 1 black bead and 3 white beads into H1
I place 2 black beads and 2 white beads into H2
I place 4 black beads and no white beads into H3
I draw a box a random, and randomly remove a bead from that box. Given the color of the bead, what can I deduce as to which box I drew?

If I replace the bead, then redraw another bead at random from the same box, how well can I predict its color before drawing it?

These two questions are the foundation of uncertainty reasoning and machine learning .

Example

- Boxes H1, H2, and H3 were my prior models of the world
- The fact that P(H1) = 1/3, P(H2) = 1/3, and P(H3) = 1/3 (uniformly distributed) was my prior distribution
- The color of the bead was a piece of evidence about the true model of the world
- The use of Bayes' rule was a piece of probabilistic inference, giving me a posterior distribution on possible worlds
- Learning is prior + evidence → posterior
- A piece of evidence decreases my ignorance about the world
- Distributions are good ways of describing your state of knowledge. Knowledge that includes an uncertainty measure can mean much better decision making.

Probability Distributions

If we want to know the probability of a variable that can take on multiple values, we may define a probability distribution , or a set of probabilities for each possible variable value.

TemperatureToday = (Below50, 50s, 60s, 70s, 80s, 90sAndAbove)

P(TemperatureToday) = 0.1, 0.1, 0.5, 0.2, 0.05, 0.05

Note that the sum of the probabilities for possible values of a variable must always sum to 1.

Because events are rarely isolated from other events, we may want to define a joint probability distribution, or $P(X_1, X_2, \ldots, X_n)$. Each X_i is a vector of probabilities for values of variable X_i.

The joint probability distribution is an n-dimensional array of combinations of probabilities.

```
          Wet      -Wet
        +-------+-------+
   Rain | 0.6   | 0.4   |  Sum is always 1
        +-------+-------+
  -Rain | 0.4   | 0.6   |  Sum for a variable is unconditional Probability
        +-------+-------+
```

Axioms of Probability

- $0 \leq P(Event) \leq 1$
- Disjunction, a ∨ b – P(a ∨ b) = P(a) + P(b) - P(a ∧ b)

- Negation, P(¬a) = 1 - P(a)
- Conditional probability

  Once evidence is obtained, the agent can use conditional probabilities, P(a|b)
  P(a|b) = probability of a being true given that we known b is true
  The equation $P(a \mid b) = \frac{P(a \wedge b)}{P(b)}$ holds whenever P(B) > 0

An agent who bets according to probabilities that violate these axioms can be forced to bet so as to lose money regardless of outcome [deFinetti, 1931]

Axioms of Probability

- Conjunction

  Product rule:

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

  The only way a and b can both be true is if a is true and we know b is true given a is true (thus b is also true).

  If a and b are independent events (the truth of a has no effect on the truth of b), then P(a∧b) = P(a) * P(b).

  "Wet" and "Raining" are not independent events.
  "Wet" and "Joe made a joke" are pretty close to independent events.

  The chain rule is derived by successive application of the product rule:

$$P(X_1, \ldots, X_n) = P(X_1, \ldots, X_{n-1}) P(X_n \mid X_1, \ldots, X_{n-1}) \tag{1}$$
$$= P((X_1, \ldots, X_{n-2}) P(X_{n-1} \mid X_1, \ldots, X_{n-2}) P(X_n \mid X_1, \ldots, X_{n-1}) \tag{2}$$
$$= \ldots \tag{3}$$
$$= \prod_{i=1}^{n} P(X_i \mid X_1, \ldots, X_{i-1}) \tag{4}$$

Lunar Lander Example

```
A lunar lander crashes somewhere in your town (one of the cells at
random in the grid).  The crash point is uniformly random (the probability
is uniformly distributed, meaning each location has an equal probability of
being the crash point).

            +--+--+--+--+--+--+--+--+--+
            |  |  |  |  |  |  |  |  |  |
```

```
+--+--+--+--+--+--+--+--+--+
|  |  |  |  |  |  |  |  |  |
+--+--+--+--+--+--+--+--+--+
|  |  |  |  |  |D |D |D |  |
+--+--+--+--+--+--+--+--+--+
| R| R| R| R| R|DR|DR|DR| R|
+--+--+--+--+--+--+--+--+--+
| R| R| R| R| R|DR|DR|DR| R|
+--+--+--+--+--+--+--+--+--+
|  |  |  |  |  |D |D |D |  |
+--+--+--+--+--+--+--+--+--+
D is the event that it crashes downtown.
R is the event that it crashes in the river.

* What is P(R)?

18/54
* What is P(D)?

12/54

* What is P(D&R)?

6/54

* What is P(D|R)?

* What is P(R|D)?

* What is P(R&D)/P(D)?
```

Axioms of Probability

- Bayes' Rule

  Given a hypothesis (H) and evidence (E), and given that $P(E) \neq 0$, what is $P(H|E)$?

  Many times rules and information are uncertain, yet we still want to say something about the consequent; namely, the degree to which it can be believed. A British cleric and mathematician, Thomas Bayes, suggested an approach.

  Recall the two forms of the product rule:

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

If we equate the two right-hand sides and divide by P(a), we get

$$P(b \mid a) = \frac{P(a \mid b)P(b)}{P(a)}$$

Answer

Observation: I draw a white bead.
P(H1—W) = P(H1)P(W—H1) / P(W)
= (1/3 * 3/4) / 5/12 = 3/12 * 12/5 = 36/60 = 3/5
P(H2—W) = P(H2)P(W—H2) / P(W)
= (1/3 * 1/2) / 5/12 = 1/6 * 12/5 = 12/30 = 2/5
P(H3—W) = P(H3)P(W—H3) / P(W)
= (1/3 * 0) / 5/12 = 0 * 12/5 = 0

Example

Bayes' rule is useful when we have three of the four parts of the equation. In this example, a doctor knows that meningitis causes a stiff neck in 50% of such cases. The prior probability of having meningitis is 1/50,000 and the prior probability of any patient having a stiff neck is 1/20. What is the probability that a patient has meningitis if they have a stiff neck?

```
        H = "Patient has meningitis"
E = "Patient has stiff neck"

                 P(E|H)*P(H)
P(H|E) = -----------
                   P(E)

        P(H|E) = (0.5*.00002)/.05 = .0002
```

Example

We wish to know probability that John has malaria, given that he has a slightly unusual symptom: a high fever.

We have 4 kinds of information: a) probability that a person has malaria regardless of symptoms, and b) probability that a person has the symptom of fever given that he has malaria, and c) probability that a person has symptom of fever, given that he does NOT have malaria. Finally, d) John has high fever.

```
   H = "John has malaria"
   E = "John has a high fever"

We use P(H) probability that a person has malaria,
P(E|H), and P(E|-H).


                    P(E|H)*P(H)
    P(H|E) = -----------
                       P(E)

where P(E) = P(E|H)P(H) + P(E|-H)P(-H)   (we know that P(H&-H) = 0).
```

The probability that John has malaria given that he has a high fever
is equal to
the ratio of the probability that he has both the fever and malaria, to
the probability that he has a fever regardless of whether or not he has malaria.

The probability of having a high fever is computed as the sum of the conditional probabilities of having the fever given malaria or given no malaria, weighted by the probability of having malaria or not.

Suppose P(H) = 0.0001, P(E|H) = 0.75, P(E|-H) = 0.14

Then P(E) = 0.75 * 0.0001 + 0.14 * 0.9999 = 0.14006
and P(H—E) = (0.75 * 0.0001) / 0.14006 = 0.0005354

On the other hand, if John did not have a fever, his probability of having malaria would be

```
                     P(-E|H)P(H)    (1-0.75)(0.0001)
    P(H|-E) = -----------  = ----------------
                       P(-E)          (1-0.14006)
```

= 0.000029
which is much smaller.

Monty Hall Problem

Utility Theory

Utility Theory reasons based on costs.

If I believe that the probabilities of getting me to my destination on time are
P(A25) = 0.04, P(A90) = 0.70, P(A120) = 0.95, P(1440) = 0.9999, which action should I choose?
Depends on my **preferences** for missing a flight vs. airport cuisine, etc.
Utility theory is used to represent and infer preferences

Probability Theory + Utility Theory = Decision Theory

Game Theory

Decision problems in which utility of an action depends on environment AND on actions of other agents

Assume agents make decisions simultaneously without knowledge of decisions of other agents

---

Prisoner's Dilemma

Problem drawn from political science and game theory

Two players, each with a choice of cooperating with the other or defecting
Each receives payoff according to payoff matrix for their decision
When both cooperate, both rewarded equal, intermediate payoff (reward, R)
When one player defects, he/she receives highest payoff (temptation, T)
and other gets poor payoff (sucker, S)
When both player defect they receive intermediate penalty P

Make problem more interesting by repeating with same players, use history to guide future decisions (iterated prisoner's dilemma)

Some strategies:

Tit For Tat: Cooperate on first move then do whatever opponent did on previous move, performed best in tournament

Golden Rule: Always cooperate

Iron Rule: Always defect

Examples

In the first example, the other player chooses randomly

Where Do Probabilities Come From?

- Frequency
- Subjective judgment

Consider the probability that the sun will still exist tomorrow. There are several ways to compute this.

Choice of experiment is know as the reference class problem.

Belief Networks

A belief network (Bayes net) represents the dependence between variables.

Components of a belief network graph:

Nodes

These represent variables

Links

X points to Y if X has a direct influence on Y

Conditional probability tables

Each node has a CPT that quantifies the effects the parents have on the node

The graph has no directed cycles

Example

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

Variables: $Burglar$, $Earthquake$, $Alarm$, $JohnCalls$, $MaryCalls$

Network topology reflects "causal" knowledge:

Example

Suppose you are going home, and you want to know the probability that the lights are on given the dog is barking and the dog does not have a bowel problem. If the family is out, often the lights are on. The dog is usually in the yard when the family is out and when it has bowel troubles. If the dog is in the yard, it probably barks.

Use the variables:

f = family out

l = light on

b = bowel problem

d = dog out

h = hear bark

There should be a graph with five nodes.

Example

We know

- l is directly influenced by f and is independent of b,d,h given f
  Add link from f to l
- d is directly influenced by f and b, independent of l and h
  Add link from f to d and b to d
- h is directly influenced by d, independent of f, l, b, and d
  Add link from d to h

```
f     b    Once we specify the topology (or learn it from data),
|\    |    we need to specify the conditional probability table for each node
| \   |
|  \  |
v   v v
l    d
     |          p(f) = 0.15, 0.85           p(b) = 0.01, 0.99
     |          p(l|f) = 0.60, 0.40         p(l|-f) = 0.05, 0.95
     |          p(d|f,b) = 0.99, 0.01       p(d|f,-b) = 0.90, 0.10
     v          p(d|-f,b) = 0.97, 0.03      p(d|-f,-b) = 0.30, 0.70
     h          p(h|d) = 0.70, 0.30         p(h|-d) = 0.01, 0.99
```

Example

Smart Home Example

The Bad (and Challenging) News

General querying of Bayes nets is NP-Complete
The best known algorithm is exponential in the number of variables
Pathfinder system

Heckerman, 1991

Diagnostic system for lymph-node diseases

60 diseases, 100 symptoms and test rulests

14,000 probabilities

8 hours to determine variables, 35 hours for topology, 40 hours for CPTs

Outperforms world experts in diagnosis

Being extended to several dozen other medical domains

Netica

- Nature nodes, decision nodes, utility nodes
- Links
- Learn values from observations
- Probabilities (percentages) must sum to 100.0
- Compile
- Make observation
- Calculate posterior probabilities

Netica Smart Home example

Dempster-Shafer Theory

- Measure certainty
- Belief(X) = -1...1
- Belief(X) = 1 means you are certain X is true
- Belief(X) = -1 means you are certain X is not true
- Belief(X) = 0 means you do not know whether X is true or not
- Facts and rules have beliefs, propagate belief values

Fuzzy Logic

Fuzzy Logic is a multivalued logic that allows intermediate values to be defined between conventional evaluations like yes/no, true/false, black/white, etc.

Fuzzy Logic was initiated in 1965 by Lotfi A. Zadeh, professor of computer science at the University of California in Berkeley.

The concept of fuzzy sets is associated with the term "graded membership".

This has been used as a model for inexact, vague statements about the elements of an ordinary set.

Fuzzy logic prevalent in products:

- Washing machines
- Video cameras
- Razors
- Subway systems

Fuzzy Sets

In a fuzzy set the elements have a DEGREE of existence.

Some typically fuzzy sets are "large numbers", "tall men", "young children", "approximately equal to 10", "mountains", etc.

Ordinary Sets

```
              +- 1   if x in A
      f_A(x) =|
              +- 0   if x not in A
```

P(X) is the "power set" of X (all subsets of X)
2^X represents all functions from X into {0,1}
$f_{A \cup B} = \max(f_A, f_B)$

Fuzzy Sets

f_A(x) = i, where $0 \leq i \leq 1$
if $f_A(x) > f_A(y)$, then x is "more in" the set than y
if $f_A(x) = 1$, then $x \in A$ if $f_A(x) = 0$, then $x \notin A$ if $f_A(x) = \lambda$, where $0 < \lambda < 1$, then $x \in_\lambda A$

Degree of membership sometimes determined as a function (degree of tall calculated as a function of height)

```
   tall(x) = {0                if height(x) < 5'
                height(x)-5'/2'  if 5' <= height(x) <= 7'
                1                if height(x) > 7'

A graph of this looks like:

1.0 +                   +--------------------
    |                  /
    |                 /
0.5 +                /
    |               /
    |              /
0.0 +-------------+-----+--------------------
                  |     |
                 5.0   7.0
                height, ft. ->
```

Fuzzy Set Relations

One set A is a "subset" of set B if for every x, f_A(x) $\leq$ f_B(x)
Sets A and B are equal if for every element x, f_A(x) = f_B(x).

OR / Union: A $\cup$ B is the smallest fuzzy subset of X containing both A and B, and is defined by $f_{A \cup B}(x) = max(f_A(x), f_B(x))$
AND / Intersection: The intersection A $\cap$ B is the largest fuzzy subset of X contained in both A and B, and is defined by $f_{A \cap B}(x) = min(f_A(x), f_B(x))$
NOT: truth($\neg$ x) = 1.0 - truth(x)
IMPLICATION: A $\rightarrow$ B = $\neg$A $\vee$ B, so truth(A $\rightarrow$ B) = $max(1.0 - f_A(x), f_B(x))$

Examples

Robot Exercise

What seems easy, now that we have reasoning tools, is not also so easy to implement.

Simulate a robot using four students.

Student one: Robot vision system

Student two: Robot brain

Student three: Robot left arm

Student four: Robot right arm

Task: Stack one box on top of another

Robot Exercise

Ground rules

The brain can talk to any or all of the other three “robot parts”. The arms follow the directions of the brain. They should pretend not to hear the vision system.

The arms can answer yes/no questions from the brain, and can tell whether or not they are touching something. They cannot distinguish the table from the boxes.

The vision system cannot volunteer information, but it can answer (as literally as possible) any question posed to it by the brain. I point out that this vision system is also very smart, in that it can easily distinguish the table from the boxes, can estimate distances pretty well, etc.

The vision system can move around to get a good view. Stereo vision may be allowed using two student volunteers with fixed vision systems.

Robot Exercise

What were the most difficult parts of this exercise?

World knowledge is important
Box-centered vs. arm-centered vision system
Vision system usually is stationary
Difficult to construct a plan without knowing preconditions