Lecture 10: Uncertainty - 1

Victor Lesser

CMPSCI 683
Fall 2004

Today's lecture

- The sources of uncertainty in intelligent systems.
- Representing and reasoning with uncertain information.
- Bayesian reasoning.
- Chapter 13.1 - 13.6

"In complex environment, agent almost never have access to the whole truth about their environment (Russell&Norvig)"

Sources of uncertainty

- Imprecise model of the environment
  – medical diagnosis, weather forecast
- Stochastic environment
  – random processes, moving obstacles
- Limited computational resources
  – chess, planning with partial information -- Practical Ignorance
- Limited communication resources
  – distributed systems, MAS without global view -- Practical Ignorance

Sources of uncertainty cont.

- Noisy sensory data
  – object identification and tracking
- Imprecise model of the system
  – Medical science -- Theoretical Ignorance
- Exceptions to our knowledge can never be fully enumerated
  – All birds fly -- Laziness

Probability provides a way of numerically summarizing this uncertainty

Reasoning About Uncertainty

- **Making decisions without knowing everything relevant but using the best of what we do know**
- **Exploiting background and commonsense knowledge, which is knowledge about what is generally true**
- ***Difficult to easily represent in classical logic***
  - *Introduce requirements for vagueness, uncertainty, incomplete and contradictory information*
- **Very different approaches based on type of reasoning required and assumptions about independence of evidence**
- **Crucial to the architecture of an agent that is interacting with the "real" world**
  - The challenge is how to acquire the necessary qualitative and quantitative relationships and devising efficient methods for computing useful answers from uncertain knowledge

Acting Under Uncertainty

- **Because uncertainty is a fact of life in most domains, agents must be able to act in spite of uncertainty.**
- **How should agents behave—What is the "right" thing to do?**
- **The** rational agent **model: agents should do what is expected to maximize their performance measure, given the information they have -- Decision Theory.**
- **Thus, a** rational decision **involves knowing:**
  - The relative *likelihood* of achieving different states/goals -- **Probability Theory**.
  - The relative *importance* (pay-off) for various states/goals -- **Utility Theory**.

Uncertainty in First-Order Logic (FOL)

- **First-Order Logic (FOL) makes the epistemological commitment that facts are either true, false, or unknown.**
  - –Contrast with Probability Theory: Degree of Belief in Proposition, same epistemological commitment as FOL
  - –Contract with Fuzzy Logic: Degree of Truth in Proposition
- **Deductive inference can be done only with categorical facts (definitely true statements).**

  Thus, FOL (and logical agents) cannot deal with *uncertainty*.

  This is a major limitation since virtually all real-world domains involve uncertainty.
- **Eliminating uncertainty would require that:**
  - –the world be accessible, static, and deterministic;
  - –the agent has complete and correct knowledge;
  - –it is practical to do complete, sound inference.

Belief and Evidence

- **Probability is about the agent's belief not directly about the world**
  - Analogous to saying whether a given logical statement is entailed by the knowledge base
- **Beliefs depend on the percepts that the agent has received to date**
- **Percepts constitute the evidence on which probability assertions are based**
- **Probabilities can change when more evidence is acquired**

Uncertainty, *cont'd*

- **Most real domains are inaccessible, dynamic, and non-deterministic (at least from the agent's perspective).**
- **In these domains, it is *impossible* for an agent to know the exact state of its environment.**
  - Also, agents can rarely be assumed to have complete, correct knowledge of a domain.
- **The** qualification problem: **many rules about a domain will be incomplete/incorrect because there are too many conditions to explicitly enumerate them all.**
  - E.g., birds fly (unless they are dead, non-flying types, have broken a wing, are caged, etc.).
- **Finally, even where exact reasoning may be possible, it will typically be impractical computationally.**

Nonmonotonicity

- **FOL assumes that knowledge is *complete* and *consistent*.**
- **Leads to the property of** monotonicity**: once a fact is true/believed, it must remain so.**
  - Thus, adding new knowledge always increases the size of the knowledge base.
- Non-monotonicity**: the addition of new knowledge may require the retraction/removal of previously derived conclusions.**
- **Using incomplete and/or uncertain knowledge leads to non-monotonicity:**
  - Will require that **assumptions** be made;
  - Inferences may not be *deductively valid*.
- **Probability exhibits non-monotonicity:**
  - $P(A \mid E_1, E_2)$ not determined by $P(A \mid E_1)$.

Reasoning under uncertainty

More realistic approach to many applications, but reasoning under uncertainty is more difficult:

- Non-monotonic: need to examine previously made conclusions based on new or modified evidence
- Non-modular: all the available evidence must be considered.
- Uncertainty measures characterize invisible facts: how do the exceptions to $A \Rightarrow B$ interact with the exceptions to $B \Rightarrow C$ to yield the exceptions to $A \Rightarrow C$?.
  - **Probability as a way of summarizing the uncertainty that comes from our 'laziness and ignorance"**

Symbolic vs Numeric approaches

- **Symbolic approaches represent the different possibilities without measuring their likelihood.**
  - Can potentially combine?
- **There are several different numeric approaches: probabilities, certainty factors, belief functions (**Dempster-Shafer**), fuzzy logic.**
- **We will focus on probabilistic reasoning. (faced lots of early objections)**

Cons (probabilities)

- McCarthy and Hayes claimed that probabilities are "epistemologically inadequate," leading AI researchers to stay away from it. ["Some philosophical problems from the standpoint of artificial intelligence," *Machine Intelligence*, 4:463-502, 1969.]
- Arguments against a probabilistic approach
  - Use of probability requires a massive amount of data
  - Use of probability requires the enumeration of all possibilities
  - *Hides details of character of uncertainty*
  - People are bad probability estimators
  - We do not have those numbers
  - We find their use inconvenient

Pros (probabilities)

"The only satisfactory description of uncertainty is probability. By this it is meant that every uncertainty statement must be in the form of a probability; that several uncertainties must be combined using the rules of probability, and that the calculation of probabilities is adequate to handle all situations involving uncertainty. In particular, alternative descriptions of uncertainty are unnecessary."
-- D.V. Lindey, *Statistical Science* 2:17-24, 1987.

"Probability theory is really about the structure of reasoning."
-- Glen Shafer

Probability versus Causality

"When I began writing Probabilistic Reasoning in Intelligent Systems (1988), I was working within the empiricist tradition. In this tradition, probabilistic relationships constitute the foundations of human knowledge, whereas causality simply provides useful ways of abbreviating and organizing intricate patterns of probabilistic relationships. Today, my view is quite different. I now take causal relationships to be the fundamental building blocks both of physical reality and of human understanding of that reality, and I regard probabilistic relationships as but the surface phenomena of the causal machinery that underlies and propels our understanding of the world."

-- Judea Pearl. CAUSALITY: Models, Reasoning, and Inference**. Cambridge University Press, January 2000.**

MYCIN's Certainty Factors

A rule-based expert system developed in the mid 1970's for automated diagnosis of infectious diseases.

Used certainty factors to represent likelihood.

Example:

if: the stain of the organism is gram-positive, and
the morphology of the organism is coccus, and
the growth conformation of the organism is clumps
then: (.7) the identity of the organism is staphyloccus.

Early and Simplified Approach to Dealing with Uncertainty and Incompleteness of Knowledge and Evidence (data)

Certainty Factors

- Certainty factors are real numbers between –1 and +1 attached to facts and rules.
- Positive and negative values indicate increase and decrease in the degree of belief.
- Certainty factors are relative measures (do not translate to absolute level of belief).

Certainty Factors cont.

- The user provides uncertain observations with certainty factors attached to them.
- Ex. 0.9 organism is gram-positive.
  0.4 morphology of the organism is coccus.
  0.7 the organism grows in clumps.
- Belief in (a conjunction of) premises is calculated by: max[0,min(0.9,0.4,0.7)] = 0.4
- Belief in conclusion = CF x belief in premises = 0.7 x 0.4 = 0.28

Combining Support of Different Rules

$B_{12} = B_1 + B_2 (1 - |B_1|)$ when both positive or negative
when both positive/negative

$B_{12} = (B_1 + B_2)/(1 - \min(|B_1|,|B_2|))$ with opposite signs
opposite signs

- Ex. Combining 0.4 with 0.6 gives: 0.4 + 0.6 (1 - 0.4) = 0.76
- More (positive) evidence will always increase the certainty factor.
- Evidence combination rule is commutative and associative hence order is unimportant.

Performance of MYCIN

- Evaluations of MYCIN show that it is as good or better than most human experts.
- But certainty factors have no operational definition
  - Hard to use in decision making.
- Surprisingly good with appropriate knowledge engineering and limited forms of deduction

Difficulties with Certainty Factors

- **Connecting information derived from different paths**
  - Bi-directional inferences (explaining away)
  - Correlated (non-independent) sources of evidence
- **Retracting conclusions (monotonicity)**

What if the observations are not independent?

Scenario (c):
Events:

S: sprinkler was on last night
W: grass is wet
R: it rained last night

MYCIN-style rules:

```
If:the sprinkler was on last night
then there is suggestive evidence (0.9) that
the grass will be wet this morning

If:the grass is wet this morning
then there is suggestive evidence (0.8) that
it rained last night
```

Combining rules, we get:

MB[W,S]= 0.8 {sprinkler suggest wet}
MB[R,W]= 0.8 · 0.9 = 0.72 {wet suggests rain}

So sprinkler made us believe rain.

Review of Elementary Probabilistic Reasoning

- Let A_t = leave for airport t minutes before the flight. Will A_t get me there on time?
- **Problems**: inaccessible world (road state) noisy sensors (traffic reports) uncertain actions (blow-out).
- Suppose:
  $P(A_{25})=.04$, $P(A_{90})=.6$, $P(A_{120})=.9$, $P(A_{1440})=.9995$
  Which action to choose?
- Depends on preferences (“utilities”)
  Decision theory = probability + utility

Axioms of Probability Theory

Assign a numerical degree of belief to propositions and ground sentences

$0 \le P(A) \le 1$

$P(True) = 1 \quad P(False) = 0$

$P(A \lor B) = P(A) + P(B) - P(A \land B)$

Other properties can be derived:

$1 = P(True) = P(A \lor \neg A) = P(A) + P(\neg A) - P(A \land \neg A) = P(A) + P(\neg A)$

So: $P(\neg A) = 1 - P(A)$

Probability theory

- Random experiments and uncertain outcomes.
- Event Set - refer to possible outcomes of a random experiment.
- Elementary events - the most detailed events of interest.
- The number of distinct events and their definitions are totally subjective and depend on the decision-maker.

Random variables

- Represent the result of random experiments.
- Notation: x, y, z represent particular values of the variables X, Y, Z.
- Sample space - the domain of a random variable (set of all elementary events).
- Ex. Sample space = graduating students.
  Elementary events = {John, Mary, ...}
  Event set = Females graduating in civil engineering

Random Variables

- For random variables we are interested in equalities like $P(X = x_1) = 0.7$
- $\sum_{x_i} P(X = x_i) = 1$ since the values are exhaustive and mutually exclusive.
- Can refer to the probabilities of all values at once as a *vector*: $P(X) = \langle 0.7, 0.1, 0.2 \rangle$.
- E.g., for *Weather*= ⟨*sunny, cloudy, rainy, snowy*⟩ can have P(*W*eather) = ⟨ 0.7,0.2,0.08,0.02⟩.
- Propositions are Boolean random variables.

Probability distributions

- An assignment of probability to each event in the sample space.
- Discrete vs. continuous distributions.
- Ex. P(Weather) = (0.7, 0.2, 0.08, 0.02)
  [sunny,rain,cloudy,snow]
- Q. What are those numbers?
  Where do they come from?

Objective Probability

- **Probabilities are precise properties of the universe.**
- **Value can be obtained by reasoning, for example, if a coin is perfect, use symmetry.**
- **When probability of elementary events are equally likely**
  - Pr[event] = size of event set / size of sample space.
- **Exist only in "artificial" domains.**
- **Require high degree of symmetry.**

Subjective Probability

- **Represent degrees of belief**
- **More realistic approach to representing "expert opinion".**
- **Examples:**
  - The likelihood of a patient recovering from a heart attack.
  - The quality of life in a certain city.

Probabilities as Frequencies

- **Probability as frequency of occurrence**
- **Pr[event] = number of time event occurs / number of repeated random experiments**
- **Problem: Need to gather infinite amount of data and assume that the probability does not change over time.**
- **Some experiments cannot be repeated:**
  - o **Success of oil drilling at a particular location**
  - o **Success of marketing a new PC operating system**
  - o **Success of the UMass basketball team in 2005**

Prior Probabilities

•**The appropriate probability to associate with a proposition depends on the knowledge (information) that is available.**

•**P(A) denotes the** prior probability **(**prior**):**
- The probability that *A* is true *in the absence of any (specific) knowledge.*

•**Once an agent has some knowledge (evidence), the prior is no longer applicable.**

Conditional (posterior) probability

- **P(*A* | *E*) denotes the c**onditional probability **(**posterior probability**): the probability that *A* is true *given that all we know is E*..**
  - Probabilistic reasoning is inherently non-monotonic because there are no constraints on how conditional probabilities can vary: e.g., we can have P(*A* | E_1) = 1, but P(A | E_1 + E_2) = 0.
  - Contrast this with FOL in which if KB_1 |= α then KB_1 + *E* |= α.
- **Ex. P(Cavity|Toothache) = 0.8**
- **Similarly, can use P(A|B,C), etc.**

Conditional (posterior) Probability

- **The notation P(X|Y) refers to the two dimensional table: P(X=x_i|Y=y_i)**
- **Conditional probability can be defined in terms of unconditional probabilities:**
- **P(A|B) = P(A,B)/P(B) when P(B) > 0, or**
- **P(A,B) = P(A|B) P(B) (the product rule)**

The Joint Probability Distribution

- **A probabilistic model consists of a set of random variables that can take on particular combinations of values with certain probabilities.**
- **An** atomic event **is an assignment of values to all the variables—e.g.,** $X_1 = x_{1i_1}, \ldots, X_n = x_{ni_n}$.
- **Atomic events are mutually exclusive and collectively exhaustive.**
- **The** joint probability distribution (joint) P$(X_1,\ldots,X_n)$ **assigns probabilities to all possible atomic events.**
- **Thus, it completely specifies the probability assignments for all propositions in the domain:**

  P(A ∧ B) = P(A,B)

  P(A ∨ B) = P(A) + P(B) - P(A,B)

  P(A) = $\sum_i$P(A,B_i) -- marginalization or summing out

  P(A) = $\sum_i$P(A | B_i) P(B_i) -- conditioning

Joint probability distributions

- **Given X_1, ..., X_n, the joint probability distribution P(X_1, ..., X_n) assigns probabilities to each set of possible values of the variables. Example:**

	Toothache	¬Toothache
Cavity	0.04	0.06
¬Cavity	0.01	0.89

- **From the joint distribution we can compute the probability of any complex proposition such as: P(Cavity v Toothache) or P(Cavity | Toothache)**
- **Why not use the joint probability distribution?**

Bayes' Rule

- **From the product rule:**
  P(A,B) = P(A|B) P(B) = P(B|A) P(A)
  Hence: P(B|A) = (P(A|B) P(B))/P(A)
- **Don't really need P(A): Normalization**
  P(B|A) = α P(A|B) P(B); P($\neg$ B|A) = α P(A| $\neg$ B) P($\neg$ B);
  or:
  $$P(y_i \mid x) = \frac{P(x \mid y_i)P(y_i)}{\sum_j P(x \mid y_j)P(y_j)}$$
- **We can condition on background knowledge:**
  P(B|A,E) = (P(A|B,E) P(B|E)) / P(A|E)

Why is Bayes' Rule Useful?

- **P(object | image) proportional to:**
  P(image | object) P(object)
- **P(sentence | audio) proportional to:**
  P(audio | sentence) P(sentence)
- **P(fault | symptoms) ...**
  P(symptoms | fault) P(fault)

Abductive Inference!!

Abduction as the Basis of Interpretation

Abduction: if *A*s can *cause* *B*s and know of a *B* then hypothesize *A* as an explanation for the *B*

Abductive inferences are uncertain/plausible inferences (as opposed to deductive/logical inferences)

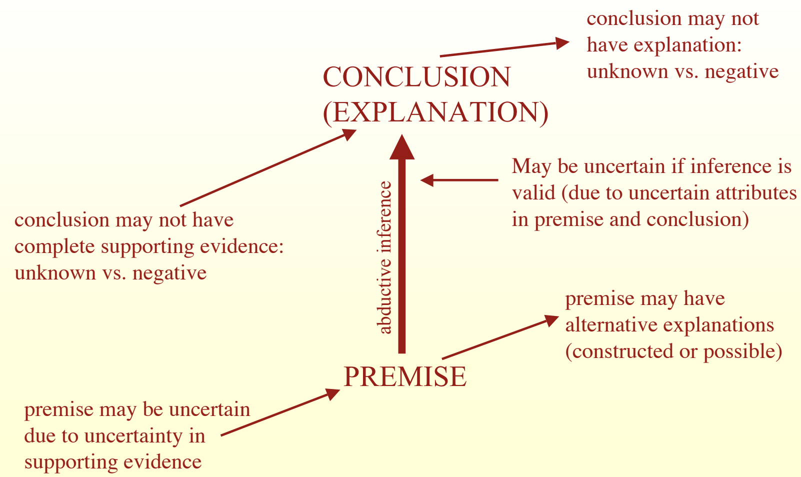
The existence of *B* provides *evidence* for *A*— i.e., a reason to believe *A*

Evidence from abductive inference is uncertain because there may be some other cause/explanation for *B*

Abduction is the basis for medical diagnosis:

If disease *D* can cause symptom *S* then if a patient has symptom *S* hypothesize that she suffers from disease *D*

Model of Abductive Uncertainty



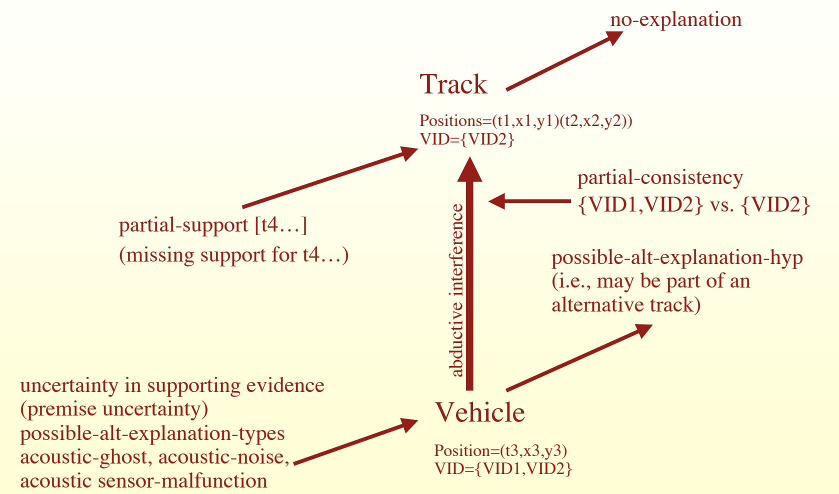
Sources of Uncertainty

Hypothesis B based on the evidence $\{A^j\}$, where the comple evidence is $\{A^i\}$ and $\{A^j\} \subset \{A^i\}$.

Potential sources of uncertainty in hypothesis:

- Partial evidence - i.e., $\{A^j\} \neq \{A^i\}$.
- Uncertain evidence satisfies the inference axiom i.e., uncertain some $A^k \in \{A^j\}$ *is* $\in \{A^i\}$.
- Uncertain premise - i.e., some $A^k \in \{A^j\}$ is uncertain.
- Possible alternative interpretations for evidence - i.e., for some $A^k \in \{A^j\}$ the correct inference is $A^k \Rightarrow$ C.
- Possible altrernative evidence for the hypothesis - i.e., for some $A^k \in \{A^j\}$ the correct evidence is actually $\{A^l\}$.

Instance of Abductive Uncertainty



Example

3 pennies are placed in a box (2-headed, 2-tailed, fair). A coin is selected at random and tossed. What is the probability that the 2H coin was selected given that the outcome is H?

P(2H|H) =

$$\frac{P(H|2H)\,P(2H)}{P(H|2H)P(2H) + P(H|2T)P(2T) + P(H|F)P(F)}$$

= 1 * 1/3 / [1 * 1/3 + 0 * 1/3 + 1/2 * 1/3] = 2/3

$$P(y_i \mid x) = \frac{P(x \mid y_i)P(y_i)}{\sum_j P(x \mid y_j)P(y_j)}$$

Combining evidence

- Consider a diagnosis problem with multiple symptoms:
  $P(d|s_i,s_j) = P(d)P(s_i,s_j|d)/P(s_i,s_j)$
- For each pair of symptoms, we need to know $P(s_i,s_j|d)$ and $P(s_i,s_j)$. Large amount of data is needed.
- Need to make **independence assumptions**:
  $P(s_i|s_j) = P(s_i)$
- Or **conditional independence assumptions**:
  $P(s_i|s_j,d) = P(s_i|d)$ $\quad$ $P(s_i,s_j|d) = P(s_i|d)\,P(s_j|d)$
  implicitly d causes s_i and s_j
- With conditional independence, Bayes' rule becomes:
  $\mathbf{P}(Z|X,Y) = \alpha\, \mathbf{P}(Z)\, \mathbf{P}(X|Z)\, \mathbf{P}(Y|Z)$

Example

Given: P(Cavity|Toothache) = 0.8
P(Cavity|Catch) = 0.95

Compute: P(Cavity|Toothache,Catch)
= P(Toothache,Catch|Cavity) P(Cavity) / P(...)

– Need to know P(Toothache,Catch|Cavity)??

Assuming conditional independence:
P(X|Y,Z) = P(X|Z) and Bayes' rule becomes:
P(Z|X,Y) = α P(Z) P(X|Z) P(Y|Z)

– P(Catch|Cavity) P(Toothache|Cavity)

The three prisoner's paradox

Three prisoners, A, B, and C, have been tried for murder, and their verdicts will be read and their sentences executed tomorrow morning. They know that only one of them will be declared guilty and will be hanged while the other two will be set free; the identity of the condemned prisoner is revealed to a very reliable prison guard, but not to the prisoners themselves.

The three prisoner's paradox

In the middle of the night, Prisoner A calls the guard and makes the following request: "Please give this letter to one of my friends - to one who is to be released. You and I know that at least one of them will be freed." The guard takes the letter and promises to do as told. An hour later prisoner A calls the guard again and asks, "Can you tell me which of my friends you gave the letter to? It should give me no clue regarding my own status because each of my friends has an equal chance of receiving my letter."

The three prisoner's paradox

The guard answers, "I gave the letter to prisoner B; he will be released tomorrow." Prisoner A returns to his bed and thinks, "Before I talked to the guard, my chances of being executed were one in three. Now that I was told that B will be released, only C and I remain, and my chances of dying have gone from 33.3% to 50%. What did I do wrong? I made certain not to ask any information relevant to my own fate..."

Problem: Did the guard reveal any information to prisoner A regarding his fate**?**

The three prisoner's paradox

Let IX stand for "prisoner X will be declared innocent"
Let GX stand for "prisoner X will be declared guilty"

Then:

$$P(GA|IB) = \frac{P(IB|GA)*P(GA)}{P(IB)} = \frac{P(GA)}{P(IB)} = \frac{1/3}{2/3} = \frac{1}{2}$$

The three prisoner's paradox

But, IB = "B will be declared innocent"
was inferred from a more direct observation,
I'B = "Guard said that B received the letter"

if we compute P(GA|I'B) we get the correct answer:

$$P(GA|I'B) = \frac{P(I'B|GA)*P(GA)}{P(I'B)} = \frac{1/2*1/3}{1/2} = \frac{1}{3}$$

More on Conditional Probability

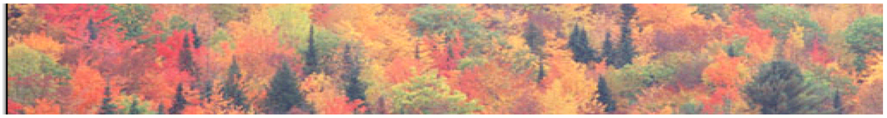
- Consider the following results of a test of two types of drugs on a group of people:

	Drug A		Drug B	
	lived	died	lived	died
Men	550	500	20	10
Women	100	200	280	390

Survival for men: DRUG A = 0.52, DRUG B = 0.67
Survival for women: DRUG A = 0.33, DRUG B = 0.42
For the total population: DRUG A = 0.48, DRUG B = 0.43
Is this a paradox?

Next lecture

- Probabilistic reasoning with belief networks.



Extra Slides