Lecture 16: Uncertainty 1

Victor R. Lesser
CMPSCI 683
Fall 2010

Announcements

- 3 more Homeworks
  - MDP Module (due Monday Nov 15, to be posted by Friday Nov 5)
  - Reasonging Under Uncertainty Module (due Wed Dec 1, posted around Nov 15)
  - Learning Module (due ? Friday Dec 10, posted around Wed Dec 1)

Today's Lecture

- Review of sources of uncertainty in intelligent systems.
- Bayesian reasoning.

Ubiquity of Uncertainty

- 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/models 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.

Sources of uncertainty

- Imprecise model of the environment
  - weather forecasting -- *Theoretical Ignorance*
- Stochastic environment
  - random processes, moving obstacles -- *Theoretical Ignorance*
- Noisy sensory data
  - object identification and tracking -- *Theoretical Ignorance*
- Imprecise model of the system
  - Medical science -- *Theoretical Ignorance*

Sources of uncertainty cont.

- Limited computational resources
  - chess, planning with partial information -- *Practical Ignorance*
- Limited communication resources
  - distributed systems, MAS without global view -- *Practical 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
  - Crucial to the architecture of an agent that is interacting with the "real" world
- 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

The challenge is how to acquire the necessary qualitative and quantitative relationships and to devise 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
  - Contrast 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.

Cons (probabilities)

- McCarthy and Hayes claimed that probabilities are "epistemologically inadequate," leading AI researchers to stay away from it for awhile!!. ["Some philosophical problems from the standpoint of artificial intelligence," *Machine Intelligence*, 4:463-502, 1969.]
- Arguments against a probabilistic approach *(no longer valid?)*
  - 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.

Review of Key Ideas in Probability Theory as applied to AI reasoning

Axioms of probability theory

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

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

$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

Other properties can be derived:

$1 = P(True)$

$= P(A \vee \neg A) = P(A) + P(\neg A) - P(A \wedge \neg A)$

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

Probability theory

- **Random experiments** and uncertain outcomes.
- **Events** - refer to possible outcomes of a random experiment.
  - Collections of Elementary Events
- **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

- Value of a Random Variable -- represent the result of a random experiment.
- 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).
  - Sample space = graduating students.
  - Elementary events = {John, Mary, ...}
  - Event set = Females graduating in civil engineering

Probability distributions

- An assignment of probability to each event in the sample space.
- Discrete vs. continuous distributions.
  - We will in this module talk about discrete 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?

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

P (Cavity, ¬Toothache)= .06

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 2009

Conditional probability

- Prior probability
  - P(Cavity) = 0.05
- Posterior/Conditional probability
  - posterior probability of a random event or an uncertain proposition is the conditional probability that is assigned after *the relevant evidence is taken into account*
  - P(Cavity|Toothache) = 0.8
- 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*)

Conditionality with 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)
  - *Identify all atomic events where proposition is true and add up their probabilities*
- Can not directly calculate P(Cavity | Toothache)?
  - *Why the need to normalize?*

Examples of Using Joint Probabilities Distribution

	Toothache	¬Toothache
Cavity	0.04	0.06
¬Cavity	0.01	0.89

P(Cavity v Toothache) =

.04+.01+.06=.11

	Toothache	¬Toothache
Cavity	0.04(.04)	0.06
¬Cavity	0.01	0.89

P(Cavity=t |Toothache=t) =

P(Cavity=t , Toothache=t)/ P(Toothache=t)= .04/(.04+.01)=.8

More on Calculating with Joint Probability Distributions

- Completely specifies the probability assignments for all propositions in the domain:

  $P(A \wedge B) = P(A,B)$

  $P(A \vee B) = P(A) + P(B) - P(A,B)$

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

  $P(A) = \sum_i P(A \mid B_i)\, P(B_i)$ -- conditioning
  - based on product rule

Why not use the joint probability distribution?

Bayes' Rule

P(A,B,C,D,..) = P(A|B,C, D,..) P(B,C, D,..) ; product rule
P(A,B) = P(A|B)P(B) = P(B|A)P(A)

Thus, **Bayes' Rule**: $P(B \mid A) = \dfrac{P(A \mid B)P(B)}{P(A)}$

This allows us to compute a *conditional probability from its inverse so as to reflect causality.*

E.g., $P(disease \mid symptom) = \dfrac{P(symptom \mid disease)P(disease)}{P(symptom)}$

Bayes' rule is typically written as: $\mathbf{P}(B \mid A) = \alpha \mathbf{P}(A \mid B)\mathbf{P}(B)$
(α is the **normalization constant** needed to make the $\mathbf{P}(B=b_i \mid A)$ entries sum to 1, it eliminates the need to know $\mathbf{P}(A)$;
if computing all the probabilities values of B=True and False then just add up and normalize; don't need to know constant)

Bayes' Rule continued

- Don't really need P(A): Normalization

  $P(B=T|A) = \alpha\, P(A|B=T)\, P(B=T)$;
  $P(B=F|A) = \alpha\, P(A|\, B=F)\, P(B=F)$;

  or: $P(y_i \mid x) = \dfrac{P(x \mid y_i)P(y_i)}{\sum_j P(x \mid y_j)P(y_j)}$ [marginalization and conditioning of P(x)]

- Condition on background knowledge E:
  $P(B|A,E) = (P(A|B,E)\, P(B|E)) / P(A|E)$

Why is Bayes' Rule Useful?

Appropriate View of Causality

- 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)

Knowledge easier to obtain

Abductive Inference!!

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

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

Causal vs. Diagnostic Knowledge

S = patient has a stiff neck
M = patient has meningitis

P(S|M) = .5
P(M) = 1/50,000
P(S) = 1/20

$$P(M|S) = \frac{P(S|M)P(M)}{P(S)} = \frac{.5 \times 1/50{,}000}{1/20} = .0002$$

Suppose given only P(M|S) based on actual observation of data… what happens if there is a sudden outbreak of meningitis:

⇒ P(M) goes up significantly and P(S/M) not affected

"*Diagnostic knowledge is often more tenuous (makes assumptions about the environment) than Causal knowledge.*"

Combining evidence

- Consider a diagnosis problem with multiple symptoms:
  - $P(d|s_i,s_j) = P(d)P(s_i,s_j|d)/P(s_i,s_j)$; Bayes' Rule
- 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.
- If we can make *independence assumptions*:
  - $P(s_i|s_j) = P(s_i) \rightarrow P(s_i,s_j) = P(s_i)P(s_j)$;
  - *conditional independence assumptions*:
  - $P(s_i|s_j,d) = P(s_i|d)$ $P(s_i,s_j|d) = P(s_i|d)\,P(s_j|d)$
- With conditional independence, Bayes' rule becomes:
  - $P(Z|X,Y) = \alpha\, P(Z)\, P(X|Z)\, P(Y|Z)$

Relate to Markov Assumption

Example

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

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

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

Assuming conditional independence: $P(s_i,s_j|d) = P(s_i|d)\,P(s_j|d)$)

- P(Toothache,Catch|Cavity) =P(Catch|Cavity) P(Toothache|Cavity)

Bayes' Rule: Incremental Evidence Accumulation

- Probabilistic inference involves computing probabilities that are not explicitly stored by the reasoning system.
- *P*(*hypothesis* | *evidence*) is a common value we want, and we want to compute *this incrementally as evidence accumulates*.
- Possible with conditional independence

$$\mathbf{P}(H \mid E_1,E_2) = \alpha\mathbf{P}(E_2 \mid H)\underline{\mathbf{P}(E_1 \mid H)\mathbf{P}(H)}$$

[$\mathbf{P}(E_1 \mid H)\mathbf{P}(H)$ is just the based on E_1 and is *$P(H| E_1)$ which would be the result after receiving only E_1*]

Abduction as the Basis of Interpretation

Abduction: if *A*s can *cause B*s, *P(B|A) >0,* and know of a *B* then hypothesize *A* as an explanation for the *B, P(A|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*

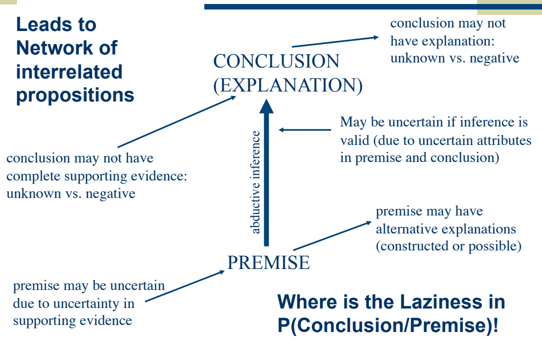
Evidence from abductive inference is uncertain because there may be some other cause/explanation for *B*

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

Leads to Network of interrelated propositions



Where is the Laziness in P(Conclusion/Premise)!

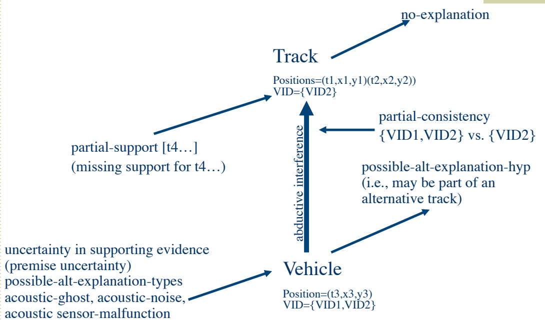
Sources of Uncertainty

Hypothesis B based on the evidence, $\{A^j\}$, where the complete 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



Next Lecture

- Bayes Nets