13. Uncertainty

Logical Agents based on Propositional logic and First order logic consider the propositions (or facts) as true false or unknown.

When an agent knows enough facts about its environment, the logical approach enables its to derive plans that are guaranteed to work.

* Agents seldom have access to the whole truth about the environment. Then, the agents must act under uncertainty.

Eg ① Sensors of Wumpus world report only local information. Most of the world is not immediately observable.
Sometimes the wumpus agent would be unable to discover which of the two squares contains a pit.

② Real-world problems are more complex.
An Agent wants to drive someone to an airport and is considering a plan.

plan A_{90}, leaving home 90 minutes before the flight departs and driving at a reasonable speed.
"No guarantee" that Plan A_{90} will get us to airport on time

A_{90} is a "rational decision"
- it maximises the agent's performance measure given the information it has about the environment

A_{120} might increase agent's belief that it will get to airport on time, but increases the likelihood of a long wait.

Factors governing "rational decision" (the right thing to do)

- relative importance of various goals and
- the likelihood that, and degree to which, they will be achieved.

The remainder of this lecture discusses about the general theories of uncertainty and rational decisions an agent takes in presence of uncertainty

Handling uncertain knowledge

Eg: "Medical Diagnosis" is a task that almost always involves uncertainty.

- Let us try to write down rules for dental diagnosis using first order logic.

$\forall p$ Symptom(p, Toothache) $\Rightarrow$ Disease(p, Cavity)

Note: Not all patients with toothaches have cavities. Some may have a gum disease, an abcess, etc.

A possible rule is
$\forall p$ Symptom (P, Toothache) $\Rightarrow$ Disease(P, cavity) $\vee$
Disease (P, GumDisease) $\vee$
Disease (P, Abcess) ....

But this approach requires us to add unlimited list of possible clauses.

- The rule may be written as a causal rule instead:
$\forall p$ Disease (P, cavity) $\Rightarrow$ Symptom (P, ToothAche)
But not all cavities cause pain.

Trying to ~~using~~ First-order logic to cope with a domain like medical diagnosis fails for 3 main reasons

- Laziness: too much work to list complete set of antecedents or consequents needed. But it is also too hard to use such rules.
- Theoretical ignorance: Medical science has no complete theory for the domain
- Practical ignorance: Even if we know all the rules, we may be still uncertain as a patient didn't undergo all the necessary tests.

Other applications having uncertainty
Law, business, design, automobile repair, gardening, dating etc

An agent's knowledge can at best provide a "degree of belief" in the relevant sentences.

To deal with this, "probability theory" may be used by assigning a numeric degree of belief to each sentence between 0 and 1.

Probability provides a way of summarizing the uncertainty that comes from our laziness and ignorance.

Eg: If there is 80% chance that a patient has a cavity, then he or she has a toothache i.e., probability of 0.8.

The missing 20% summarizes all the other possi causes of toothaches that we were too lazy or ignorant to confirm or deny.

- If there is an unequivocal belief that a statement is false ⇒ probability of the given sentence = 0

- If a statement is believed to be unequivocally true ⇒ probability of the given statement = 1.

* degree of belief vs. degree of truth

A probability of 0.8 does not mean it is "80% true" but rather 80% degree of belief.

Degree of truth may be summarized using Fuzzy Logic

"The probability that a patient has cavity" is 0.8 is about agent's beliefs, not directly about the world.
- These beliefs depend on the percepts the agent receives till date.
- These percepts constitute the evidence on which probabilities are based.

"Before" evidence is obtained, we talk about prior or unconditional probability

"After" evidence is obtained, we talk about posterior or conditional probability.

Uncertainty and rational decisions

The presence of uncertainty radically changes the way an agent makes decisions. A logical agents has a goal and executes a plan that guarantees to achieve the goal.

When uncertainty enters, this is no longer the case.

choices: A_{90} vs. A_{120} vs. A_{1440}

To make such choices, agent must have <u>preferences</u> between different possible <u>outcomes</u> of various plans.

Utility theory to represent and reason with preferences.
[utility - quality of being useful]

Utility theory → every state has a degree of usefulness, or utility, to an agent and the agent will prefer states with higher utility.

<u>Chess</u>

Eg: Utility of a state in which White Won (for an agent playing White) → high
Utility of a state in which White Won (for an agent playing Black) → low

* Preferences expressed as utilities, are combined with probabilities in the general theory of rational decisions
Decision theory = probability theory + utility theory

<u>Fundamental idea of decision theory</u>

An agent is rational iff it chooses the action that yields the highest expected utility, averaged over all all possible outcomes of the action

* <u>Maximum Expected Utility</u> (MEU)

<u>Design for a decision-theoretic agent</u>

- Agent that uses decision theory to select actions.

decision - theoretic agent vs. logical agent
↳ agent's knowledge of current state is uncertain

The agent's <u>belief state</u> is a representation of probabilities of all possible actual states of the world.

As time passes, the agent accumulates more evidence and its belief state changes.

Given the belief state, the agent can make probabilistic predictions of action outcomes and thereby select actions with highest expected utility.

<u>function</u> DT-AGENT(percept) returns an action
    <u>static</u>: belief-state, a probabilistic beliefs about the current state of the world
        action, the agent's action

    update belief-state based on action and percept
    calculate outcome probabilities for actions,
        given action descriptions and current belief-state
    select action with highest expected utility
        given probabilities of outcomes and utility info
    return action

A decision-theoretic agent that selects rational actions

13.2 Basic Probability Notation

Degrees of belief applied to Propositions

Random variable - a part of the world whose "status" is initially unknown.

Eg: cavity

Each random variable has a domain of values.

Eg: domain of Cavity is {true, false}

Random Variables — Boolean Random Variables
- Discrete " "
- Continuous " "

Boolean Random Variables (Spl case of Discrete Random Variable)

Eg: Domain of cavity might be <true, false>

Cavity = true abbreviated as cavity

Cavity = false " " ¬cavity

Discrete Random Variables

Take on values from a countable domain

Eg: Weather = < sunny, rainy, cloudy, snow >

Values must be mutually exclusive and exhaustive

Continuous Random Variables

take on values from real number $X = 4.02$

Complex propositions can be formed by combining elementary propositions using logical connectives.

Eg: Cavity = true ∧ Toothache = false

Atomic Events

An atomic event is complete specification of the state of the world about which the agent is uncertain.

Eg: if world consists of only two boolean variables Cavity and Toothache, then there are four distinct atomic events.

Properties of atomic events:

- They are mutually exclusive
- The set of all possible atomic events is exhaustive (The disjunction of all atomic events is logically equivalent to true)
- Any particular atomic event entails the truth or falsehood of every proposition, whether simple or complex.

Eg: the atomic event cavity ∧ ¬toothache entails truth of cavity and falsehood of cavity ⇒ toothache

- Any proposition is logically equivalent to the disjunction of all atomic events that entail the truth of the proposit

Eg: the proposition cavity is equivalent to the disjunction of atomic events $cavity \wedge toothache$ and $cavity \wedge \neg toothache$

Prior Probability

Unconditional / Prior Probability
Let a be a proposition
P(a) is the degree of belief associated to it in the absence of any other information.
Eg: P(cavity = true) = 0.1 or P(cavity) = 0.1

* P(a) is used when there is no other information
As soon as new information is known, we must reason with "conditional probability" of "a" given that new information.

If Weather = <sunny, rain, cloudy, snow> (a vector of values)
then P(Weather) is

P(Weather = = sunny) = 0.7
P(Weather = = rain) = 0.2
P(Weather = = cloudy) = 0.08
P(Weather = = snow) = 0.02

or simply, P(Weather) = <0.7, 0.2, 0.08, 0.02>
This is "Prior Probability Distribution" of the random variable "Weather"

P(Weather, cavity) is represented as a 4 x 2 table of probabilities ⇒ Joint Probability distribution of Weather and cavity.

For continuous valued variables, it is not possible to list the entire distribution as a table because there are infinite values.
Probabilities for continuous variables are defined by Probability Density Functions

Eg: Temperature distribution

$P(X = 20.5) = U[18, 26](20.5) = 0.125/C$

This does not mean that there is 12.5% chance that maximum temperature will be exactly 20.5° tomorrow.

The probability that the temperature is in a small region around 20.5° is equal

Conditional Probability

Once the agent obtains some information about previously unknown variables, prior probabilities are no longer applicable.

We use conditional / posterior probabilities

$P(a|b)$ read as "the probability of a, give that we all know is b"

$P(\text{cavity} | \text{toothache}) = 0.8$

probability that the patient has cavity if the patient is observed to have toothache and no other information is available, is 0.8.

conditional probabilities is defined in terms of unconditional probabilities

$$P(a|b) = \frac{P(a \wedge b)}{P(b)}$$ which holds whenever $P(b) > 0$

$\Rightarrow P(a \wedge b) = P(a|b) \cdot P(b)$ (Product Rule)

|||ly, $P(a \wedge b) = P(b|a) \cdot P(a)$

Notation for conditional distributions

$P(X|Y)$ gives the values of $P(X = x_i | Y = y_i)$ for each possible

$P(X = x_1 \wedge Y = y_1) = P(X = x_1 | Y = y_1) \cdot P(Y = y_1)$

$\vdots$

combining into a single equation

$P(X, Y) = P(X|Y) \cdot P(Y)$

The Axioms of Probability (Kolgomorov's Axioms)

① All probabilities are between 0 and 1.

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

② Necessarily true (valid) propositions have probability 1
Necessarily false (unsatisfiable) " " " 0

$P(true) = 1 \qquad P(false) = 0$

③ Probability of a disjunction

$$P(a \vee b) = P(a) + P(b) - P(a \wedge b)$$

using the axioms of probability

$$P(a \vee \neg a) = P(a) + P(\neg a) - P(a \wedge \neg a)$$
$$P(true) = P(a) + P(\neg a) - P(false)$$
$$1 = P(a) + P(\neg a) - 0$$
$$\Rightarrow P(\neg a) = 1 - P(a)$$

Let D be a discrete Variable

$Dom(D) = \langle d_1, \ldots, d_n \rangle$

then $\sum_{i=1}^{n} P(D = d_i) = 1$

The probability of a proposition is equal to the sum o
the probabilities of the atomic events in which it holds

$$P(a) = \sum_{e_i \in e(a)} P(e_i)$$

Inference using Full Joint Distributions

simple method for probabilistic inference

Eg: domain containing only 3 boolean variables:
Toothache, Cavity, and Catch
↓ Full Joint distribution table

	toothache		¬ toothache	
	catch	¬ catch	catch	¬ catch
cavity	0.108	0.012	0.072	0.008
¬ cavity	0.016	0.064	0.144	0.576

* all probabilities in joint distribution sum upto 1.

$$P(cavity \vee toothache) = \underbrace{0.108 + 0.012}_{P(toothache)} + 0.072 + 0.008$$

$$+ 0.016 + 0.064 = 0.28$$

Unconditional or marginal probability of cavity

$$P(cavity) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2$$

Marginalization or summing out

$$P(Y) = \sum_{z} P(Y, z)$$

conditioning

$$P(Y) = \sum_{z} P(Y|z) \cdot P(z)$$

Eg:

$$P(cavity | toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)}$$

$$= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

$P(\text{cavity} \mid \text{toothache})$ (through normalization)

$= \alpha\, P(\text{cavity}, \text{toothache})$

$= \alpha[P(\text{cavity}, \text{toothache}, \text{catch}) + P(\text{Cavity}, \text{toothache}, \neg\text{catch})]$

$= \alpha[(0.108, 0.016) + (0.012, 0.064)]$

$= \alpha\langle 0.12, 0.08\rangle = \langle 0.6, 0.4\rangle$

$P(X \mid e) = \alpha\, P(X, e) = \alpha \sum_{y} P(X, e, y)$

function for answering probabilistic queries for discrete variables.

function ENUMERATE-JOIN-ASK(X, e, P) returns a distribution over X

inputs: X, the query variable

e, observed values for variables E

P, a joint distribution ~~of~~ on variables

$\{X\} \cup E \cup Y$ /* Y = hidden variables */

Q(x) ← a distribution over X, initially empty

for each value x_i of X do

Q(x_i) ← ENUMERATE-JOINT(x_i, e, Y, [ ], P)

return NORMALIZE(Q(x))

function ENUMERATE-JOINT(x, e, vars, values, P) returns a real number

if EMPTY?(vars) then return P(x, e, values)

Y ← FIRST(vars)

return $\sum_y$ ENUMERATE-JOINT(x, e, REST(vars), [y|values], P)

For a domains of 'n' Boolean variables, we require an input table of size $O(2^n)$ and $O(2^n)$ time

∴ full joint distribution in tabular form not a practical tool for building reasoning systems.

Independence

Eg: Full joint distribution → P(Toothache, Catch, Cavity, Weather)

which has 32 entries
(Toothache - 2, catch - 2, Cavity - 2, Weather - 4)

How is P(toothache, catch, cavity, Weather = cloudy) and P(toothache, catch, cavity) related?

P(toothache, catch, cavity, Weather = cloudy)
= P(Weather = cloudy | toothache, catch, cavity) P(toothache, catch, cavity)

$P(\text{Weather} = \text{cloudy} \mid \underbrace{\text{toothache, catch, cavity}}_{\text{unrelated events}}) = P(\text{Weather} = \text{cloudy})$

For all possible values of Weather, we get

~~P(toothache, catch,~~
P(Toothache, Catch, cavity, Weather) = P(Weather) P(toothache, catch, Cavity)

⇒ independence / marginal independence / absolute independence

Independence between propositions a and b can be written as
$P(a|b) = P(a)$ or $P(b|a) = P(b)$ or $P(a \wedge b) = P(a)\,P(b)$

Cavity
Toothache Catch
Weather

$\Downarrow$ decomposes into

Cavity
Toothache Catch

Weather

Eg: joint distribution of outcome of 'n' coin flips $P(C_1, \ldots, C_n)$ can be represented as product of 'n' single-variable distributions $P(C_i)$

$Coin_1 \ldots\ldots Coin_n$

$\Downarrow$ decomposes into

$coin_1$ .... $coin_n$

Bayes' Rule and its Use

$$P(a \wedge b) = P(a|b)\,P(b)$$
$$P(a \wedge b) = P(b|a)\,P(a)$$
$$\Rightarrow P(b|a)\,P(a) = P(a|b)\,P(b)$$
$$\Rightarrow P(b|a) = \frac{P(a|b)\,P(b)}{P(a)}$$

Bayes' Rule underlies all modern AI systems for probabilistic inference.

$$P(Y|X) = \frac{P(X|Y) \cdot P(Y)}{P(X)}$$

If the above general version is conditioned on some background evidence 'e':

$$P(Y|X, e) = \frac{P(X|Y, e)\, P(Y|e)}{P(X|e)}$$

<u>case study</u>

A doctor knows meningitis causes stiff neck, 50% of the time $\Rightarrow$ $P(s|m) = 0.5$

The doctor also knows some unconditional facts:
probability that a patient has meningitis is 1/50,000
$\Rightarrow P(m) = 1/50000$
Prior probability that any patient has stiff neck is 1/20
$\Rightarrow P(s) = 1/20$

$$P(m|s) = \frac{P(s|m) \cdot P(m)}{P(s)} = \frac{0.5 \times 1/50000}{1/20} = 0.0002$$

The general form of Bayes' Rule with normalisation is
$$P(Y|X) = \alpha\, P(X|Y) \cdot P(Y)$$

<u>Using Bayes' rule: Combining evidence</u>

Using full joint distribution for large number of variables is infeasible conditional independence

$P(toothache \wedge catch \mid cavity) = P(toothache \mid cavity) P(catch \mid cavity)$

Note: Toothache & catch variables are not independent, because if a probe catches in the tooth, then there is probably a cavity.
But these variables are independent given the presence or absence of a cavity.

$$P(X, Y \mid Z) = P(X \mid Z) \cdot P(Y \mid Z)$$

$P(Toothache, catch, cavity)$
$= P(Toothache, catch \mid cavity) P(cavity)$ (Product rule)
$= P(Toothache \mid cavity) P(catch \mid cavity) P(cavity)$ (conditional independence)

Size of representation grows to $O(n)$ instead of $O(2^n)$

Naive Bayes Model

$$P(cause, Effect_1, \ldots, Effect_n) = P(cause) \prod_i P(Effect_i \mid cause)$$

Naive Bayesian classifier based on this.