인공지능

10차시: Uncertainty and Probability

서울대학교 컴퓨터공학부
담당 교수: 장병탁

Seoul National University
Byoung-Tak Zhang
Introduction

- Agents need to handle **uncertainty**, due to partial observability and non-determinism.

- In previous lectures, the logical agents handled uncertainty by
  - **Belief state**—a representation of the set of all possible world states
  - **Contingency plan** that handles every possible eventuality

- **Drawbacks of this deterministic approach**
  - Every logically possible explanation lead to large belief-state representations
  - A correct contingent plan can grow arbitrarily large
  - There might be no plan to guarantee to achieve a goal

- In this lecture, we study methods for **handling uncertainty with degrees of belief**. This requires quantifying uncertainty and using probability theory and probabilistic reasoning methods.
Outline (Lecture 10)

- 10.1 Acting under Uncertainty ........................................ 4
- 10.2 Basic Probability Notation ........................................ 8
- 10.3 Inference Using Full Joint Distribution .............. 18
- 10.4 Independence and Bayes’ Rule .............................. 21
- Summary ........................................................................ 26
- Homework ...................................................................... 27
10.1 Acting under Uncertainty (1/4)

1) Uncertainty

- Let action \( A_t \) = leave for airport \( t \) minutes before flight. Will \( A_t \) get me there on time?

- Problems:
  - Partial observability (road state, other drivers’ plans, etc.)
  - Noisy sensors (TBS traffic reports)
  - Uncertainty in action outcomes (flat tire, etc.)
  - Immense complexity of modelling and predicting traffic
2) Methods for handling uncertainty

- Default or nonmonotonic logic
  - Assume $A_{25}$ works unless contradicted by evidence
- Rules with fudge factors
  - $A_{25} \leftrightarrow_{0.3} \text{AtAirportOnTime}$
  - $\text{Sprinkler} \leftrightarrow_{0.99} \text{WetGrass}$
  - $\text{WetGrass} \leftrightarrow_{0.7} \text{Rain}$
- Probability
  - Given the available evidence, $A_{25}$ will get me there on time with probability 0.04
3) Summarizing uncertainty

Example: medical diagnosis

Toothache $\Rightarrow$ Cavity

Toothache $\Rightarrow$ Cavity $\lor$ GumProblem $\lor$ Abscess …

Cavity $\Rightarrow$ Toothache

Logic fails to deal with this for three main reasons:
- Laziness: hard to list a complete set of rules
- Theoretical ignorance: no complete theory
- Practical ignorance: not all necessary tests

Probability provides a way of summarizing the uncertainty that comes from our laziness and ignorance, thereby solving the qualification problems.
4) Uncertainty and rational decisions

» Utility theory
  - An agent must first have preferences between the different possible outcomes of the various plans. We use utility theory to represent and reason with preferences.
  - utility being “the quality of being useful”

» Decision theory
  - Preferences, as expressed by utilities, are combined with probabilities in the general theory of rational decisions called decision theory:
    \[
    \text{Decision theory} = \text{probability theory} + \text{utility theory}
    \]

» Maximum expected utility (MEU)
  - The fundamental idea of decision theory is that an agent is rational if and only if it chooses the action that yields the highest expected utility, averaged over all the possible outcomes of the action. This is the principle of maximum expected utility (MEU).
10.2 Basic Probability Notation (1/10)

1) Probability

Probabilistic assertions **summarize** effects of

- Laziness: failure to enumerate exceptions, qualifications, etc.
- Ignorance: lack of relevant facts, initial conditions, etc.

Subjective or Bayesian probability:

- Probabilities relate propositions to one's own state of knowledge

\[ P(A_{25}|\text{no reported accidents}) = 0.06 \]

- Probabilities of propositions change with new evidence:

\[ P(A_{25}|\text{no reported accidents, 5 a.m.}) = 0.15 \]
2) Probability basics

Begin with a set $\Omega$ - the sample space

- e.g., 6 possible rolls of a die.
- $\omega \in \Omega$ is a sample point / possible world / atomic event

A probability space or probability model is a sample space with an assignment $P(\omega)$ for every $\omega \in \Omega$ s.t.

$$0 \leq P(\omega) \leq 1, \quad \sum_{\omega} P(\omega) = 1$$

$$P(A) = \sum_{\{\omega \in A\}} P(\omega)$$

An event $A$ is any subset of $\Omega$
3) Random variables

- A **random variable** is a function from sample points to some range, e.g., the reals or Booleans
  
  - e.g., $Odd(1) = true$.  

- $P$ induces a **probability distribution** for any r.v. $X$:

  $$P(X = x_i) = \sum_{\omega : X(\omega) = x_i} P(\omega)$$

  - e.g., $P(Odd = true) = P(1) + P(3) + P(5) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$
4) Propositions

- Given Boolean random variables $A$ and $B$:
  - event $a = \text{set of sample points where } A(\omega) = \text{true}$
  - event $\neg a = \text{set of sample points where } A(\omega) = \text{false}$
  - event $a \land b = \text{points where } A(\omega) = \text{true} \land B(\omega) = \text{true}$

- With Boolean variables, sample point = propositional logic model
  - e.g., $A(\omega) = \text{true}, B = \text{false}$, or $a \land \neg b$.

- Proposition = disjunction of atomic events in which it is true
  - e.g., $(a \lor b) \equiv (\neg a \land b) \lor (a \land \neg b) \lor (a \land b)$
  - $\Rightarrow P(a \lor b) = P(\neg a \land b) + P(a \land \neg b) + P(a \land b)$
5) Syntax for propositions

» **Propositional** or **Boolean** random variables
  - e.g., *Cavity* (do I have a cavity?)
  - *Cavity = true* is a proposition, also written *cavity*

» **Discrete** random variables (finite or infinite)
  - e.g., *Weather* is one of \{sunny, rain, cloudy, snow\}
  - *Weather = rain* is a proposition
  - Values must be exhaustive and mutually exclusive

» **Continuous** random variables (bounded or unbounded)
  - e.g., *Temp = 21.6*; also allow, e.g., *Temp < 22.0*.
6) Prior probability

- Prior or unconditional probabilities of propositions
  - e.g., \( P(Cavity = \text{true}) = 0.1 \) and \( P(Weather=\text{sunny}) = 0.72 \)

- Probability distribution gives values for all possible assignments:
  - e.g., \( P(Weather) = \langle 0.72, 0.1, 0.08, 0.1 \rangle \) (normalized, i.e., sums to 1)

- Joint probability distribution for a set of r.v.s gives the probability of every atomic event on those r.v.s (i.e., every sample point)

<table>
<thead>
<tr>
<th>( Weather = )</th>
<th>sunny</th>
<th>rain</th>
<th>cloudy</th>
<th>snow</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Cavity = \text{true} )</td>
<td>0.144</td>
<td>0.02</td>
<td>0.016</td>
<td>0.02</td>
</tr>
<tr>
<td>( Cavity = \text{false} )</td>
<td>0.576</td>
<td>0.08</td>
<td>0.064</td>
<td>0.08</td>
</tr>
</tbody>
</table>
7) Probability for continuous variables

Express distribution as a parameterized function of value:

\[ P(X = x) = U[18,26](x) = \text{uniform density between } 18 \text{ and } 26 \]

Here \( P \) is a density; integrates to 1.

\[ P(X = 20.5) = 0.125 \text{ really means } \lim_{dx \to 0} \frac{P(20.5 \leq X \leq 20.5 + dx)}{dx} = 0.125 \]
8) Probability for continuous variables, contd.

- Gaussian density

\[ P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]
9) Conditional probability

Conditional or posterior probabilities

- e.g., $P(\text{cavity}|\text{toothache}) = 0.8$
- i.e., given that \text{toothache} is all I know
- NOT “if \text{toothache} then 80% chance of \text{cavity}”
- If we know more, e.g., \text{cavity} is also given, then we have
  $$P(\text{cavity} | \text{toothache, cavity}) = 1$$
10.2 Basic Probability Notation (10/10)

10) Conditional probability, contd.

» Definition

\[ P(a|b) = \frac{P(a \land b)}{P(b)} \text{ if } P(b) \neq 0 \]

» Product rule

\[ P(a \land b) = P(a|b)P(b) = P(b|a)P(a) \]

» Chain rule

\[ P(X_1, \ldots, X_n) = P(X_1, \ldots, X_{n-1})P(X_n|X_1, \ldots, X_{n-1}) \]
\[ = P(X_1, \ldots, X_{n-2})P(X_{n-1}|X_1, \ldots, X_{n-2})P(X_n|X_1, \ldots, X_{n-1}) \]
\[ = \ldots \]
\[ = \prod_{i=1}^{n} P(X_i|X_1, \ldots, X_{i-1}) \]
10.3 Inference Using Full Joint Distributions (1/3)

1) Inference by enumeration

\[
\begin{array}{c|c|c|c|c}
 & \text{toothache} & \neg \text{toothache} \\
\hline
\text{catch} & 0.108 & 0.012 & 0.072 & 0.008 \\
\text{\neg catch} & 0.016 & 0.064 & 0.144 & 0.576 \\
\end{array}
\]

\[P(\text{toothache}) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2\]

\[P(\text{cavity} \lor \text{toothache}) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28\]

\[P(\neg \text{cavity} | \text{toothache}) = \frac{P(\neg \text{cavity} \land \text{toothache})}{P(\text{toothache})} = \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4\]
2) Normalization

Denominator can be viewed as a normalization constant $\alpha$

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

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

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

$$= \alpha (0.12, 0.08) + (0.6, 0.4)$$
3) Inference by enumeration, contd.

- Let $\mathbf{X}$ be all the variables. Typically, we want
  - the posterior joint distribution of the query variables $\mathbf{Y}$
  - given specific values $\mathbf{e}$ for the evidence variables $\mathbf{E}$

- Let the hidden variables be $\mathbf{H} = \mathbf{X} - \mathbf{Y} - \mathbf{E}$

- Then the required summation of joint entries is done by summing out the hidden variables:
  \[
P(\mathbf{Y}| \mathbf{e}) = \alpha P(\mathbf{Y}, \mathbf{e}) = \alpha \sum_{\mathbf{h}} P(\mathbf{Y}, \mathbf{e}, \mathbf{h})
  \]

- The terms in the summation are joint entries because $\mathbf{Y}$, $\mathbf{E}$, and $\mathbf{H}$ together exhaust the set of random variables
10.4 Independence and Bayes’ Rule (1/5)

1) Independence

A and B are independent if

\[ P(A|B) = P(A) \text{ or } P(B|A) = P(B) \text{ or } P(A, B) = P(A)P(B) \]

\[ P(\text{Toothache, Catch, Cavity, Weather}) = P(\text{Toothache, Catch, Cavity})P(\text{Weather}) \]
2) Conditional independence

- $P(\text{Toothache}, \text{Cavity}, \text{Catch})$ has $2^3 - 1 = 7$ independent entries.
- If I have a cavity, the probability that the probe catches in it doesn’t depend on whether I have a toothache:
  \begin{equation}
  (1) \quad P(\text{catch}|\text{toothache}, \text{cavity}) = P(\text{catch}|\text{cavity})
  \end{equation}
- The same independence holds if I haven’t got a cavity:
  \begin{equation}
  (2) \quad P(\text{catch}|\text{toothache}, \neg \text{cavity}) = P(\text{catch}|\neg \text{cavity})
  \end{equation}
- Catch is **conditionally independent** of Toothache given Cavity:
  \begin{equation}
  P(\text{Catch}|\text{Toothache}, \text{Cavity}) = P(\text{Catch}|\text{Cavity})
  \end{equation}
3) Conditional independence, contd.

Write out full joint distribution using chain rule:

\[
P(\text{Toothache}, \text{Catch}, \text{Cavity})
\]
\[
= P(\text{Toothache}|\text{Catch}, \text{Cavity})P(\text{Catch}, \text{Cavity})
\]
\[
= P(\text{Toothache}|\text{Catch}, \text{Cavity})P(\text{Catch}|\text{Cavity})P(\text{Cavity})
\]
\[
= P(\text{Toothache}|\text{Cavity})P(\text{Catch}|\text{Cavity})P(\text{Cavity})
\]

i.e., \(2 + 2 + 1 = 5\) independent numbers (equations 1 and 2 remove 2)

In most cases, the use of conditional independence reduces the size of the representation of the joint distribution from exponential in \(n\) to linear in \(n\).

Conditional independence is our most basic and robust form of knowledge about uncertain environments.
4) Bayes’ rule

- **Product rule**

  \[ P(a \land b) = P(a|b)P(b) = P(b|a)P(a) \]

- **Bayes’ rule**

  \[ P(a|b) = \frac{P(b|a)P(a)}{P(b)} \text{ or in distribution form } P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} = \alpha P(X|Y)P(Y) \]

- **Useful for assessing** diagnostic **probability from** causal **probability**

  \[ P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(effect)} \]
5) Bayes’ rule and conditional independence

\[ P(\text{Cavity}|\text{toothache} \land \text{catch}) = \alpha P(\text{toothache} \land \text{catch}|\text{Cavity}) P(\text{Cavity}) \]

\[ = \alpha P(\text{toothache}|\text{Cavity}) P(\text{catch}|\text{Cavity}) P(\text{Cavity}) \]

An example of a \textbf{naive Bayes model}:

\[ P(\text{Cause}, E_{1}, \ldots, E_{n}) = P(\text{Cause}) \prod_{i} P(\text{Effect}_{i}|\text{Cause}) \]
Uncertainty arises because of both laziness and ignorance. It is inescapable in nondeterministic, or partially observable environments.

Probabilities express the agent’s inability to reach a definite decision regarding the truth of a sentence. Probabilities summarize the agent’s beliefs relative to the evidence.

Basic probability statements include prior probabilities and conditional probabilities over simple and complex propositions.

The full joint probability distribution specifies the probability of each complete assignment of values to random variables.

Bayes’ rule allows unknown probabilities to be computed from known conditional probabilities, usually in the causal direction.

Conditional independence brought about by direct causal relationships in the domain might allow the full joint distribution to be factored into smaller, conditional distributions.

The naive Bayes model assumes the conditional independence of all effect variables, given a single cause variable, and grows linearly with the number of effects.

A wumpus-world agent can calculate probabilities for unobserved aspects of the world.
Exercises

13.3
13.8
13.16
출처

사진