UIT2504 Artificial Intelligence

Uncertainty

C. Aravindan
<AravindanC@ssn.edu.in>

Professor of Information Technology
SSN College of Engineering

October 24, 2024

- If you plan to catch a train at 8PM at Chennai Central what time will you leave from the SSN campus?

Limitations of logical reasoning

- If you plan to catch a train at 8PM at Chennai Central what time will you leave from the SSN campus?
- There are several uncertainties!

Limitations of logical reasoning

- If you plan to catch a train at 8PM at Chennai Central what time will you leave from the SSN campus?
- There are several uncertainties!
- In logic, we need to explicitly specify all the exceptions and that leads to *qualification problem*

Limitations of logical reasoning

- If you plan to catch a train at 8PM at Chennai Central what time will you leave from the SSN campus?
- There are several uncertainties!
- In logic, we need to explicitly specify all the exceptions and that leads to *qualification problem*
- $Toothache \Rightarrow Cavity$ is not exactly correct!

Limitations of logical reasoning

- If you plan to catch a train at 8PM at Chennai Central what time will you leave from the SSN campus?
- There are several uncertainties!
- In logic, we need to explicitly specify all the exceptions and that leads to *qualification problem*
- $Toothache \Rightarrow Cavity$ is not exactly correct!
- Similarly, $Cavity \Rightarrow Toothache$ is also not always correct!

Limitations of logical reasoning

- If you plan to catch a train at 8PM at Chennai Central what time will you leave from the SSN campus?
- There are several uncertainties!
- In logic, we need to explicitly specify all the exceptions and that leads to *qualification problem*
- *Toothache* $\Rightarrow$ *Cavity* is not exactly correct!
- Similarly, *Cavity* $\Rightarrow$ *Toothache* is also not always correct!
- We need to consider the probability theory — there is 80% chance that a person who has a toothache has a cavity

- Consider different plans to catch a train at Chennai Central

Rational Decisions

- Consider different plans to catch a train at Chennai Central
- How do we understand the outcome of each plan? — need for *utility theory* — quantify the multiple objectives

Rational Decisions

- Consider different plans to catch a train at Chennai Central
- How do we understand the outcome of each plan? — need for *utility theory* — quantify the multiple objectives
- Agents may choose an outcome with the highest utility

Rational Decisions

- Consider different plans to catch a train at Chennai Central
- How do we understand the outcome of each plan? — need for *utility theory* — quantify the multiple objectives
- Agents may choose an outcome with the highest utility
- But the outcomes are not certain!

- Consider different plans to catch a train at Chennai Central
- How do we understand the outcome of each plan? — need for *utility theory* — quantify the multiple objectives
- Agents may choose an outcome with the highest utility
- But the outcomes are not certain!
- *Decision Theory* $=$ *Probability Theory* $+$ *Utility Theory*

Rational Decisions

- Consider different plans to catch a train at Chennai Central
- How do we understand the outcome of each plan? — need for *utility theory* — quantify the multiple objectives
- Agents may choose an outcome with the highest utility
- But the outcomes are not certain!
- *Decision Theory = Probability Theory + Utility Theory*
- 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 — principle of *maximum expected utility (MEU)*

Decision-theoretic agent

function DT-AGENT(*percept*) **returns** an *action*
persistent: *belief_state*, 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 information
return *action*

Figure: Decision-theoretic agent

Basic Probability Notation

- Uncertainties lead to different possible worlds — *sample space* represented by Ω

Basic Probability Notation

- Uncertainties lead to different possible worlds — *sample space* represented by Ω
- Consider rolling two dice. How many possible worlds are there?

Basic Probability Notation

- Uncertainties lead to different possible worlds — *sample space* represented by Ω
- Consider rolling two dice. How many possible worlds are there?
- Each possible world in a sample space is usually represented by ω

Basic Probability Notation

- Uncertainties lead to different possible worlds — *sample space* represented by Ω
- Consider rolling two dice. How many possible worlds are there?
- Each possible world in a sample space is usually represented by ω
- Fully specified probability model assigns a numerical probability $P(\omega)$ with each possible world

$$0 \leq P(\omega) \leq 1 \text{ for every } \omega \text{ and } \sum_{\omega \in \Omega} P(\omega) = 1$$

Basic Probability Notation

- Uncertainties lead to different possible worlds — *sample space* represented by Ω
- Consider rolling two dice. How many possible worlds are there?
- Each possible world in a sample space is usually represented by ω
- Fully specified probability model assigns a numerical probability $P(\omega)$ with each possible world

$$0 \leq P(\omega) \leq 1 \text{ for every } \omega \text{ and } \sum_{\omega \in \Omega} P(\omega) = 1$$

- When we roll two dice, each possible world has probability $\frac{1}{36}$

Basic Probability Notation

- In logic, a proposition stands for a set of possible worlds — referred to as *events*

$$\text{For any proposition } \phi, P(\phi) = \sum_{\omega \in \phi} P(\omega)$$

Basic Probability Notation

- In logic, a proposition stands for a set of possible worlds — referred to as *events*

$$\text{For any proposition } \phi, P(\phi) = \sum_{\omega \in \phi} P(\omega)$$

- For example, when we roll two dice, consider the proposition $Total = 11$. What is $P(Total = 11)$?

Conditional Probability

- What we have reviewed so far is *unconditional* or *prior* probabilities

Conditional Probability

- What we have reviewed so far is *unconditional* or *prior* probabilities
- Given an *evidence*, the probability may change! — *conditional* or *posterior* probability

Conditional Probability

- What we have reviewed so far is *unconditional* or *prior* probabilities
- Given an *evidence*, the probability may change! — *conditional* or *posterior* probability
- For example, $P(cavity) = 0.2$, and $P(cavity|toothache) = 0.6$

Conditional Probability

- What we have reviewed so far is *unconditional* or *prior* probabilities
- Given an *evidence*, the probability may change! — *conditional* or *posterior* probability
- For example, $P(cavity) = 0.2$, and $P(cavity|toothache) = 0.6$
- When more evidence is gathered the probability may change further — $P(cavity|toothache \wedge \neg cavity) = 0$

Conditional Probability

- What we have reviewed so far is *unconditional* or *prior* probabilities
- Given an *evidence*, the probability may change! — *conditional* or *posterior* probability
- For example, $P(cavity) = 0.2$, and $P(cavity|toothache) = 0.6$
- When more evidence is gathered the probability may change further — $P(cavity|toothache \wedge \neg cavity) = 0$

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

Conditional Probability

- What we have reviewed so far is *unconditional* or *prior* probabilities
- Given an *evidence*, the probability may change! — *conditional* or *posterior* probability
- For example, $P(cavity) = 0.2$, and $P(cavity|toothache) = 0.6$
- When more evidence is gathered the probability may change further — $P(cavity|toothache \wedge \neg cavity) = 0$

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

- This is usually expressed as *product rule* $P(a \wedge b) = P(a|b)P(b)$

Other probability axioms

- $P(\neg a) = 1 - P(a)$

Other probability axioms

- $P(\neg a) = 1 - P(a)$
- The *inclusion-exclusion principle:* $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$

Random Variables

- *Random variable* is a variable in probability theory that maps each possible world ω to a range (set of possible values)

Random Variables

- *Random variable* is a variable in probability theory that maps each possible world ω to a range (set of possible values)
- For example, when we roll two dice, *Total* is a variable that maps a possible world to $\{2, \ldots, 12\}$

Random Variables

- *Random variable* is a variable in probability theory that maps each possible world ω to a range (set of possible values)
- For example, when we roll two dice, *Total* is a variable that maps a possible world to $\{2, \ldots, 12\}$
- Range of *Weather* might be $\{sun, rain, cloud, snow\}$

Random Variables

- *Random variable* is a variable in probability theory that maps each possible world ω to a range (set of possible values)
- For example, when we roll two dice, *Total* is a variable that maps a possible world to $\{2, \ldots, 12\}$
- Range of *Weather* might be $\{sun, rain, cloud, snow\}$
- Assignment of probability to each possible value of a random variable is called as *probability distribution* — if we order the range of *Weather* as $\langle sun, rain, cloud, snow \rangle$, probability distribution of this random variable may be $\mathbf{P}(Weather) = \langle 0.6, 0.1, 0.29, 0.01 \rangle$

Random Variables

- *Random variable* is a variable in probability theory that maps each possible world ω to a range (set of possible values)
- For example, when we roll two dice, *Total* is a variable that maps a possible world to $\{2, \ldots, 12\}$
- Range of *Weather* might be $\{sun, rain, cloud, snow\}$
- Assignment of probability to each possible value of a random variable is called as *probability distribution* — if we order the range of *Weather* as $\langle sun, rain, cloud, snow \rangle$, probability distribution of this random variable may be $\mathbf{P}(Weather) = \langle 0.6, 0.1, 0.29, 0.01 \rangle$
- For continuous variables, we define *probability density functions*

Random Variables

- *Random variable* is a variable in probability theory that maps each possible world ω to a range (set of possible values)
- For example, when we roll two dice, *Total* is a variable that maps a possible world to $\{2, \ldots, 12\}$
- Range of *Weather* might be $\{sun, rain, cloud, snow\}$
- Assignment of probability to each possible value of a random variable is called as *probability distribution* — if we order the range of *Weather* as $\langle sun, rain, cloud, snow \rangle$, probability distribution of this random variable may be $\mathbf{P}(Weather) = \langle 0.6, 0.1, 0.29, 0.01 \rangle$
- For continuous variables, we define *probability density functions*
- *Joint probability distribution* of *Weather* and *Cavity* stands for a 4×2 table
- $\mathbf{P}(Weather, Cavity) = \mathbf{P}(Weather|Cavity)\mathbf{P}(Cavity)$ stands for $4 \times 2 = 8$ equations

Full joint distributions

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

Figure: Full joint distribution of three variables

Inferences using joint distributions

- Inferences can be carried our using full joint distributions

Inferences using joint distributions

- Inferences can be carried our using full joint distributions
- $P(cavity) =$

Inferences using joint distributions

- Inferences can be carried our using full joint distributions
- $P(cavity) = \ 0.108 + 0.012 + 0.072 + 0.008 = 0.2$

Inferences using joint distributions

- Inferences can be carried our using full joint distributions
- $P(cavity) = \ 0.108 + 0.012 + 0.072 + 0.008 = 0.2$
- This process is called marginalization or summing out

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

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

Inferences using joint distributions

- Inferences can be carried our using full joint distributions
- $P(cavity) = \; 0.108 + 0.012 + 0.072 + 0.008 = 0.2$
- This process is called marginalization or summing out

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

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

- $P(cavity \vee toothache) =$

Inferences using joint distributions

- Inferences can be carried our using full joint distributions
- $P(cavity) = \ 0.108 + 0.012 + 0.072 + 0.008 = 0.2$
- This process is called marginalization or summing out

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

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

- $P(cavity \vee toothache) =$
  $0.018 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28$

Inferences using full joint distributions

- Conditional probabilities can also be computed:

$$
\begin{aligned}
P(cavity|toothache) &= \frac{P(cavity \wedge toothache)}{P(toothache)} \\
&= \frac{0.108 + 0.012}{0.018 + 0.012 + 0.016 + 0.064} \\
&= 0.6
\end{aligned}
$$

Inferences using full joint distributions

- Conditional probabilities can also be computed:

$$\begin{aligned} P(cavity|toothache) &= \frac{P(cavity \land toothache)}{P(toothache)} \\ &= \frac{0.108 + 0.012}{0.018 + 0.012 + 0.016 + 0.064} \\ &= 0.6 \end{aligned}$$

- Normalization constant

$$\begin{aligned} \mathbf{P}(Cavity|toothache) &= \alpha \mathbf{P}(Cavity, toothache) \\ &= \alpha \langle 0.12, 0.08 \rangle \\ &= \langle 0.6, 0.4 \rangle \end{aligned}$$

Inferences using full joint distributions

- Let X be a query variable, E be the list of evidence variables, e be the list of observed values, and Y be the remaining unobserved variables

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

Independence

- Consider the full joint distribution $\mathbf{P}(Toothache, Catch, Cavity)$ with $2 \times 2 \times 2 = 8$ entries

Independence

- Consider the full joint distribution $\mathbf{P}(Toothache, Catch, Cavity)$ with $2 \times 2 \times 2 = 8$ entries
- Suppose we add a fourth variable *Weather* with 4 possible values — the joint distribution has $8 \times 4 = 32$ entries

Independence

- Consider the full joint distribution $\mathbf{P}(Toothache, Catch, Cavity)$ with $2 \times 2 \times 2 = 8$ entries
- Suppose we add a fourth variable *Weather* with 4 possible values — the joint distribution has $8 \times 4 = 32$ entries
- However, *Weather* is independent of toothache related variables

$$
\begin{aligned}
&P(toothache, catch, cavity, cloud) \\
&\quad = P(cloud | toothache, catch, cavity) P(toothache, catch, cavity)
\end{aligned}
$$

Independence

- Consider the full joint distribution $\mathbf{P}(Toothache, Catch, Cavity)$ with $2 \times 2 \times 2 = 8$ entries
- Suppose we add a fourth variable *Weather* with 4 possible values — the joint distribution has $8 \times 4 = 32$ entries
- However, *Weather* is independent of toothache related variables

$$\begin{aligned} & P(toothache, catch, cavity, cloud) \\ & \quad = P(cloud | toothache, catch, cavity) P(toothache, catch, cavity) \end{aligned}$$

- Because of independence, we have

$$P(cloud | toothache, catch, cavity) = P(cloud)$$

Independence

- Consider the full joint distribution $\mathbf{P}(Toothache, Catch, Cavity)$ with $2 \times 2 \times 2 = 8$ entries
- Suppose we add a fourth variable *Weather* with 4 possible values — the joint distribution has $8 \times 4 = 32$ entries
- However, *Weather* is independent of toothache related variables

$$
\begin{aligned}
& P(toothache, catch, cavity, cloud) \\
& \quad = P(cloud|toothache, catch, cavity)P(toothache, catch, cavity)
\end{aligned}
$$

- Because of independence, we have

$$P(cloud|toothache, catch, cavity) = P(cloud)$$

- So, $P(toothache, catch, cavity, cloud) = P(cloud)P(toothache, catch, cavity)$

Independence

- Consider the full joint distribution $\mathbf{P}(Toothache, Catch, Cavity)$ with $2 \times 2 \times 2 = 8$ entries
- Suppose we add a fourth variable *Weather* with 4 possible values — the joint distribution has $8 \times 4 = 32$ entries
- However, *Weather* is independent of toothache related variables

$$
\begin{aligned}
&P(toothache, catch, cavity, cloud) \\
&\quad = P(cloud|toothache, catch, cavity)P(toothache, catch, cavity)
\end{aligned}
$$

- Because of independence, we have

$$P(cloud|toothache, catch, cavity) = P(cloud)$$

- So, $P(toothache, catch, cavity, cloud) = P(cloud)P(toothache, catch, cavity)$
- So, it is enough to maintain one table with 8 values and another with 4 values

Factoring Distributions

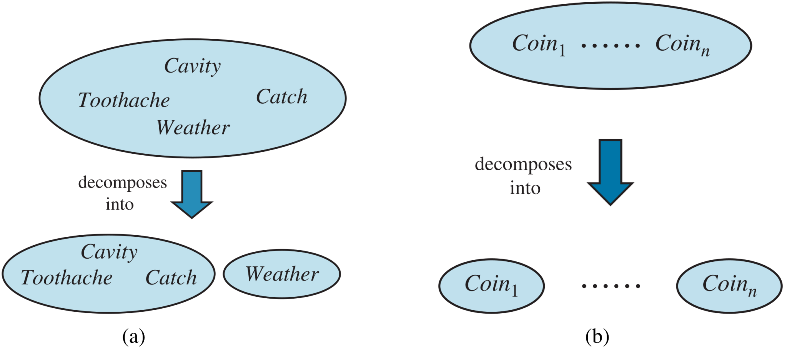

Figure: Factoring Distributions

Bayes's Rule

- From the product rule of probability, we have

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

Bayes's Rule

- From the product rule of probability, we have

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

- From these, we have

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

Bayes's Rule

- From the product rule of probability, we have

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

- From these, we have

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

- This Bayes's rule forms the basis for probabilistic reasoning in AI

Diagnostic inferences

- Bayes rule is very useful for diagnostic inferences

$$P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(effect)}$$

Diagnostic inferences

- Bayes rule is very useful for diagnostic inferences

$$P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(effect)}$$

- With normalization, this can be rewritten as

$$\mathbf{P}(Y|X) = \alpha \mathbf{P}(X|Y)\mathbf{P}(Y)$$

Diagnostic inferences

- Bayes rule is very useful for diagnostic inferences

$$P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(effect)}$$

- With normalization, this can be rewritten as

$$\mathbf{P}(Y|X) = \alpha \mathbf{P}(X|Y)\mathbf{P}(Y)$$

- How can we use it for cases where we have more evidences?

Diagnostic inferences

- Bayes rule is very useful for diagnostic inferences

$$P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(effect)}$$

- With normalization, this can be rewritten as

$$\mathbf{P}(Y|X) = \alpha \mathbf{P}(X|Y)\mathbf{P}(Y)$$

- How can we use it for cases where we have more evidences?

$$\mathbf{P}(Cavity|toothache \wedge catch) = \\ \alpha \mathbf{P}(toothache \wedge catch|Cavity)\mathbf{P}(Cavity)$$

Diagnostic inferences

- Bayes rule is very useful for diagnostic inferences

$$P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(effect)}$$

- With normalization, this can be rewritten as

$$\mathbf{P}(Y|X) = \alpha \mathbf{P}(X|Y)\mathbf{P}(Y)$$

- How can we use it for cases where we have more evidences?

$$\mathbf{P}(Cavity|toothache \wedge catch) = \alpha \mathbf{P}(toothache \wedge catch|Cavity)\mathbf{P}(Cavity)$$

- We get into complexity issues again!

Bayes' Rule with conditional independence

- *Conditional independence* comes to the rescue

$$\mathbf{P}(toothache \wedge catch | Cavity) = \mathbf{P}(toothache | Cavity)\mathbf{P}(catch | Cavity)$$

Bayes' Rule with conditional independence

- *Conditional independence* comes to the rescue

$$\mathbf{P}(toothache \wedge catch | Cavity) = \mathbf{P}(toothache | Cavity)\mathbf{P}(catch | Cavity)$$

- So, the diagnostic inference can be generalized as

$$\mathbf{P}(Cause | \mathbf{e}) = \alpha \mathbf{P}(Cause) \prod_j \mathbf{P}(e_j | Cause)$$

Bayes' Rule with conditional independence

- *Conditional independence* comes to the rescue

$$\mathbf{P}(toothache \wedge catch | Cavity) = \mathbf{P}(toothache | Cavity)\mathbf{P}(catch | Cavity)$$

- So, the diagnostic inference can be generalized as

$$\mathbf{P}(Cause|\mathbf{e}) = \alpha \mathbf{P}(Cause) \prod_j \mathbf{P}(e_j|Cause)$$

- Such a probability distribution is called as a *naive Bayes* model when the criteria of conditional independence is always assumed

Summary

- We have discussed how agents can deal with uncertainty
- Logic is not convenient for dealing with uncertainty
- Decision theory = utility theory + probability theory
- We have reviewed the basics of probability theory
- Full joint distributions may be used for inferences, but the complexity is unmanageable — independence among the random variables may help
- Bayes's rule provides a convenient way to deal with diagnostic inferences — however, may not be scalable with multiple evidences
- Naive Bayes Models, that assume conditional independence, are useful for probabilistic inferences in practice

What Next?

- Read chapter 12 of the text book!