Unit-03: Advanced Knowledge Representation and Reasoning & Reasoning under uncertainty

Artificial Intelligence

Dr. Mohammad Fayazur Rahaman
Associate Professor, mfrahaman_ece@mgit.ac.in

Dept. of Electronics and Communications Engineering,
Mahatma Gandhi Institute of Technology, Gandipet, Hyderabad-75

Mar-Jun 2022

Unit-03: Advanced Knowledge Representation and Reasoning & Reasoning under uncertainty [1]

1. Advanced Knowledge Representation and Reasoning

1.1 Knowledge Representation Issues [2]
- Important Attributes
- Relationships among Attributes
- Choosing the Granularity of Representation
- Representing Sets of Objects
- Finding the Right Structures as Needed

1.2 Non-monotonic Reasoning [2]
- Introduction

2. Reasoning Under Uncertainty [1]

2.1 Introduction to Probabilistic Reasoning
- Acting Under Uncertainty
- Basic Probability Notation
- Propositions in probability assertions
- Inference using Full joint distribution

2.2 Bayes Theorem

2.3 Representing Knowledge in an Uncertain Domain

2.4 Bayesian Networks

Where are we ?

Knowledge Representation Issues [2]

Some issues that occur when representing real-world knowledge are

i. Are any attributes of objects so basic that they occur in almost every problem domain
ii. Are there any important relationships that exist among attributes of objects ?
iii. At what level should knowledge be represented ?
iv. How should sets of objects be represented ?
v. Given a large amount of knowledge stored in a database, how can relevant parts be accessed when they are needed ?

Important Attributes

i. There are two attributes that are of very general significance: `instance` and `isa`
   - These attributes are important because they support property inheritance

ii. They represent class membership and class inclusion

Relationships among Attributes

i. The attributes that we use to describe objects are themselves entities that we represent
ii. There are four properties of the attributes
    - Inverses
    - Existence in an `isa` hierarchy
    - Techniques for reasoning about values
    - Single-valued attributes

I. **Inverses:** Entities in the world or related to each other in many different ways
   - But as soon as we decide to describe those relationships as attributes, we commit to a perspective in which we focus on one object and look for binary relationships between it and others
   - Each of these is shown in figures with directed arrow, originating at the object that was being described and terminating at the object representing the value of the specified attribute
   - But we could equally well have focused on the object representing the value
     - If we do that, then there is still a relationship between the two entities, although it is a different one since the original relationship was not symmetric

II. **Existence in an `isa` hierarchy:** Just as there are classes of objects and specialised subset of those classes, there are attributes and specialisation of attributes
   - For example, the attribute height is actually a specialisation of the more general attribute physical-size, which is in turn is specialisation of physical attribute
   - These generalisation-specialisation relationships are important for attributes as they support inheritance
     - In the case of attributes, they support inheriting information such as constraints on the value that the attribute can have and mechanisms for computing those values

III. **Techniques for reasoning about values:** Often the reasoning system must reason about values it has not been given explicitly. Several kinds of information can play a role in this reasoning
   - Information about the type of the value. Ex, the value of `height`
   - Constraints on the value, often stated in terms of related entities. For example, the age of a person cannot be greater than the age of either of parents
   - Rules for computing the value when it is needed, i.e., backward rules
   - Rules that describe action that should be taken if the value ever become known, i.e., forward rules

IV. **Single-valued attributes:** This kind of attribute is guaranteed to take a unique value. For example a baseball player can, at any one time, have only a single height and be a member of only one team. Knowledge representation systems take several different approaches for single-valued attributes:
   - Introduce an explicit notation for temporal interval
   - Assume that the only temporal interval that is of interest is now

Choosing the Granularity of Representation

i. Regardless of the particular representation formalism we choose, it is necessary to answer the question " At what level of detail should the world be representation?"
   - That is "What should be our premitives ?"

ii. A brief example illustrates the problem, Suppose we are interested in the following fact: `John spotted Sue`
   - We could represent this as

```
spotted( agent(John) ),
object(Sue)
```

- Such a representation make it easy to answer questions such as: `Who spotted Sue` ?
- But now suppose we want to know: `Did John see Sue`?
  - Given only one the fact we have, we cannot discover the answer
- We could add other facts, such as

$$\texttt{spotted(x,y)} \rightarrow \texttt{saw(x,y)}$$

Representing Sets of Objects

i. It is important to be able to represent sets of objects for several reasons

ii. One is that there are some properties that are true of sets, that are not true of the individual members of a set
   - As examples, consider the assertions "`There are more sheep then people in Australia`" and "`English speakers can be found all over the world` "

- The only way to represent the facts described in the sentences is to attach assertions to the sets representing people, sheep, and English speakers
  - Since no single English speaker can be found all over the world

Finding the Right Structures as Needed

i. Suppose we have a script that describes the typical sequence of events in a restaurant. The script would enable us to take a text such as `John went to Steak and Ale last night. He ordered a large rare steak, paid his bill, and left.` and answer "yes" to the question `Did John eat dinner last night?`
   - Notice that nowhere in the story was John's eating anything mentioned explicitly
   - But the fact that when one goes to a restaurant one eats will be contained in the restaurant script
   - If we know in advance to use the restaurant script, then we can answer the question easily
   - But in order to be able to reason about a variety of things, a system must have many scripts for every thing.
   - How will it select appropriate one each time.
     - For example, nowhere in our story what is the word "restaurant" mentioned

ii. In order to have access to the right structure for describing a particular situation, it is necessary to solve all of the following problems
   - How to perform an initial selection of the most appropriate structure
   - How to fill in appropriate details from the current situation
   - How to find a better structure if the one chosen initially turns out not to be appropriate
   - What to do if none of the available structures is appropriate
   - When to create and remember a new structure

Where are we ?

Introduction

I. In the previous sections we have seen techniques for reasoning with a complete, consistent, and unchanging model of the world
   - However in many problems domains it is not possible to create such models
   - In the section we explore techniques for solving problems with incomplete and uncertain models

II. For example we use ABC murder story which clearly illustrates many of the main issues that the reasoning techniques must deal with

ABC Murder Story

i. Let A, B, and C be the suspects in a murder case
   - A has an alibi, in the register of a respectable hotel in Albany
   - B also has an alibi, for his brother in law testified that B was visiting him in Brooklyn at that time
   - C pleads alibi too, claiming to have been watching a ski meet in Catskills, but we have only his words for that

- So we believe,

That A did not commit the crime (1)

`That B did not commit the crime` (2)

`That A or B or C did` (3)

- Further, C documents his alibi - he had the good luck to have been caught by television in the sidelines at the ski meet. A new belief is thus thrusted upon us

`That C did not` (4)

iii. Our beliefs (1) through (4) are inconsistent, so we must choose one for rejection. Which has the weakest evidence ?
   - The basis for (1) in the hotel register is good, since it is a fine old hotel
   - The basis for (2) is weaker, since B's brother-in-law might be lying
   - The basis for (3) is perhaps twofold: that there is no sign of burglary and that only A, B, and C seem to have stood to gain from the murder apart from burglary
     - This exclusion of burglary seems conclusive, but the other consideration does not: there could be some fourth beneficiary
   - For (4), finally, the basis is conclusive: the evidence from television
   - Thus (2) and (3) are the weak points

iv. To resolve the inconsistency of (1) through (4) we should reject (2) or (3), thus either incriminating B or widening our net for some new suspect

v. This story illustrates some of the problems posed by uncertain, fuzzy, and often changing knowledge

vi. Verity of logical frameworks and computational methods have been proposed for handling such problems

Non-monotonic Reasoning [2]

i. In nonmonotonic reasoning, the axioms or the rules of inference are extended to make it possible to reason with incomplete information
   - These systems preserve, the property that, at any given moment, a statement is either believed to be true, believed to be false, or not believed to be either

ii. Conventional reasoning systems, such as first order predicate logic, are designed to work with information that has three important properties
   I. It is complete with respect to the domain of interest, i.e., all the facts that are necessary to solve a problem are present in the system or can be derived from the conventional rules of first-order logic
   II. It is consistent
   III. The only way you can change is that new facts can be added as they become available.
      - If these new facts are consistent with all the other facts that have already been asserted, then nothing will be retracted. This property is called **monotonicity**

iii. If any of these properties is not satisfied, conventional logic based reasoning systems become inadequate

iv. On the other hand, nonmonotonic reasoning systems are designed to be able to solve problems in which all of these properties may be missing
   - In order to do this, we must address several key issues, including the following:

I. *How can the knowledge base be extended to allow inferences to be made on the basis of lack of knowledge as well as on the presence of it ?*
   i. Specifically, we need to make clear the distinction between:

      `It is known that` $\neg P$

      `It is not known whether` P

   ii. First order logic allows reasoning to be based on the first of these
   iii. We need an extended system that allows reasoning to be based on the second as well
      - We call any inference that depends on the lack of some piece of knowledge as nonmonotonic inference
      - A non-monotonic inference may be defeated (invalid) by the addition of new information that violates assumptions that were made

II. *How can the knowledge base be updated properly when a new fact is added to the system (or when an old one is removed?)*
   i. In nonmonotonic systems, since the addition of a fact can cause previously discovered rules to become invalid

III. *How can knowledge be used to help resolve conflicts when there are several inconsistent nonmonotonic inferences that could be drawn ?*
   i. It turns out that when inferences can be based on the lack of knowledge as well as on its presence, contradictions are much more likely to occur then they were in conventional logical system

	Monotonic Reasoning	Non-monotonic Reasoning
1.	In monotonic reasoning, once the conclusion is taken, then it will remain the same even if we add some other information to existing information in our knowledge base. In monotonic reasoning, adding knowledge does not decrease the set of prepositions that can be derived.	In Non-monotonic reasoning, some conclusions may be invalidated if we add some more information to our knowledge base.
2.	To solve monotonic problems, we can derive the valid conclusion from the available facts only, and it will not be affected by new facts.	Logic will be said as non-monotonic if some conclusions can be invalidated by adding more knowledge into our knowledge base.
3.	Monotonic reasoning is used in conventional reasoning systems, and a logic-based system is monotonic.	Non-monotonic reasoning deals with incomplete and uncertain models. "Human perceptions for various things in daily life, "is a general example of non-monotonic reasoning.
4.	Example: `Earth revolves around the Sun.` It is a true fact, and it cannot be changed even if we add another sentence in knowledge base like, `"The moon revolves around the earth"` Or `"Earth is not round,"` etc.	Example: Let suppose the knowledge base contains the following knowledge: `Birds can fly`, `Penguins cannot fly`, `Pitty is a bird` So from the above sentences, we can conclude that `Pitty can fly.` However, if we add one another sentence into knowledge base `"Pitty is a penguin"`, which concludes `"Pitty cannot fly"`, so it invalidates the above conclusion.

	Monotonic Reasoning	Non-monotonic Reasoning
5.	**Advantages of Monotonic Reasoning:** In monotonic reasoning, each old proof will always remain valid. If we deduce some facts from available facts, then it will remain valid for always.	**Advantages of Non-monotonic reasoning:** For real-world systems such as Robot navigation, we can use non-monotonic reasoning. In Non-monotonic reasoning, we can choose probabilistic facts or can make assumptions.
6.	**Disadvantages of Monotonic Reasoning:** We cannot represent the real world scenarios using Monotonic reasoning. Since we can only derive conclusions from the old proofs, so new knowledge from the real world cannot be added.	**Disadvantages of Non-monotonic Reasoning:** In non-monotonic reasoning, the old facts may be invalidated by adding new sentences. It cannot be used for theorem proving.

Where are we ?

Acting Under Uncertainty

i. Agents in the real world need to handle **uncertainity**
   - An agent may never know for sure what state it is in, or where it will end up after a sequence of action

ii. For example, diagnosing a dental patients to take almost always involves uncertainity
   - Consider the following simple rule:

$$\texttt{Toothache} \Rightarrow \texttt{Cavity}.$$

   - However, **not all** patients with toothaches have cavities, i.e, might have several other problems

$$\texttt{Toothache} \Rightarrow \texttt{Cavity} \vee \texttt{GumProblem} \vee \texttt{Abscess} \cdots .$$

   - Further trying to turn the rule into a causal rule, that is

$$\texttt{Cavity} \Rightarrow \texttt{Toothache}.$$

   is also **not right**; as not all cavities cause pain

iii. Trying to use logic with a domain with uncertainity like medical diagnosis fails for 3 main reasons
   I. **Laziness**: It is too much work to list the complete set of consequences to make an exceptionless rule
   II. **Theoretical ignorance:** Medical science has no complete theory for the domain
   III. **Practical ignorance:** We might be uncertain about a particular patient because not all the tests have been done

iv. The agents knowledge can at best provide only a **degree of belief** in the relevant sentences
   - The main tool for dealing with degrees of belief is **probability theory**
   - *The theory of probability provides a way of summarising the uncertainty that comes from our laziness and ignorance*
   - For example, we say
     - `"The probability that the patient has a cavity, given that she has a toothache, is 0.8"`
     - `" The probability that the patient has a cavity, given that she has toothache and history of gum disease, is 0.4 "`
     - Note that the statements do not **contradict** each other; it is a separate assertion about a different knowledge state

Basic Probability Notation

i. **Sample space**: In probability theory, the set of all possible worlds is called a **sample space**

- The possible worlds are **mutually exclusive** and **exhaustive**
- For example, if we are about to roll two dice, there are 36 possible worlds to consider: i.e., sample space Ω is

$$\Omega = \{(1,1),(1,2),\cdots,(6,6)\}$$

and ω refers to elements of the space

ii. **Probability model**: A fully specified probability model associates a numerical probability $P(\omega)$ with each possible world. Where

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

iii. **Event**: Probabilistic assertions and queries or not usually about particular possible worlds, but about sets of them i.e., called **events** in probability theory, and **proposition** in logic language

- For example the probability that the two dice add up to 11, or the probability that doubles are rolled, and so on
- The probability associated with a proposition is defined to be the sum of probabilities of the worlds in which it holds: i.e., for any proposition ϕ

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

  - When rolling fair dice, we have

$$\begin{aligned} P(\texttt{Total} = 11) &= P((5,6)) + P((6,5)) \\ &= 1/36 + 1/36 = 1/18 \\ P(\texttt{doubles}) &= 1/4 \end{aligned}$$

iv. **Unconditional probability or Prior probability**: Probabilities such as $P(\texttt{Total} = 11)$, $P(\texttt{doubles})$ are called **unconditional** or **prior** probabilities; they refer to degrees of belief in propositions in the **absence of any other information**

v. **Evidence**: Most of the time, however we have some information, usually called **evidence**

- For example, the first die may already be showing a 5 and we are waiting for the other one to stop spinning

vi. **Conditional probability or Posterior probability**: When some evidence is available, we are interested in the **conditional** or **posterior** probability of rolling `doubles` given that the first die is a 5, i.e.,

$$P(\texttt{doubles} \mid \texttt{Die}_1 = 5)$$

- Further the assertion that $P(\texttt{cavity}|\texttt{toothache}) = 0.6$,
  - Does not mean “Whenever `toothache` is true, conclude that `cavity` is true with probability 0.6 ", rather
  - it mean “Whenever `toothache` is true and we have *no further information*, conclude that `cavity` is true with probability 0.6"
- Mathematically, conditional probability is or defined in terms of unconditional probabilities as,

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

that is

$$P(\texttt{doubles} \mid \texttt{Die}_1 = 5) = \frac{P(\texttt{doubles} \wedge \texttt{Die}_1 = 5)}{P(\texttt{Die}_1 = 5)}$$

vii. **Product rule**: The definition of conditional probability, can be written in a different form called the **product rule**:

$$P(a \wedge b) = P(a \mid b)\, P(b)$$

It comes from the fact that for a and b to be true, we need b to be true, and we also need a to be true given b

Propositions in probability assertions

i. **Random variable**: Variables in probability theory are called as **random variables**. Example: `Total`, $\texttt{Die}_1$
   - Every random variable is a function that maps from the domain of possible worlds Ω to some range

ii. **Range**: The **range** is the set of possible values that a random variable can take on
   - The range of `Total` for two dice is the set $\{2, \cdots, 12\}$
   - and the range for $\texttt{Die}_1$ is $\{1, \cdots, 6\}$.
   - A boolean random variable has the range $\{true, false\}$
     - For example, the proposition that doubles are rolled can be written as `Doubles` = true
     - Propositions are abbreviated simply as, `Doubles` or $\neg$`Doubles`
   - Further, ranges can be set of arbitrary tokens, for example
     - the range of `Age` can be the set $\{\texttt{infant}, \texttt{teen}, \texttt{adult}\}$, and
     - the range of `Weather` might be $\{\texttt{sun, rain, cloud, snow}\}$
   - We can **combine** elementary prepositions by using the connectives of proportional logic
     - For example can express " `The probability that the patient has a cavity, given that she is a teenager with no tooth one is 0.1` " as follows

$$P(\texttt{cavity} \mid \neg\texttt{toothache} \wedge \texttt{teen}) = 0.1$$

$$P(\texttt{cavity} \mid \neg\texttt{toothache}, \texttt{teen}) = 0.1$$

iii. **Probability distribution**: The **probability distribution** $\mathbf{P}$, of a random variable is an assignment of a probability for each possible value of that random variable
   - For Example:

$$\mathbf{P}(\texttt{Weather}) = \langle 0.6, 0.1, 0.29, 0.01 \rangle$$

iv. **Joint probability distribution**: A **Joint probability distribution** is a distribution on multiple variables
   - For example, $\mathbf{P}(\texttt{Weather, Cavity})$ denotes the probabilities of all combinations of the values of `Weather` and `Cavity`
   - This is a 4×2 table of probabilities

v. **Full joint probability distribution**: A probability model is completely determined by the joint distribution for all of the random variables - the so-called **full joint probability distribution**
   - For example, joint distribution $\mathbf{P}(\texttt{Cavity, Toothache, Weather})$ can be represented as a $2 \times 2 \times 4$ table with 16 entries
   - The fully joint distribution **suffices** to calculate the probability of every proposition by using the product rule

vi. **Probability of negation**: We can derive the relationship between the probability of a proposition and the probability of its **negation** as

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

vii. **Inclusion-exclusion principle**: The probability of a **disjunction** is also sometimes called the inclusion-exclusion principle and is given by

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

Inference using Full joint distribution

i. **Probabilistic inference Query**: The method of computation of **posterior probabilities** for **query** propositions given **evidence** is called as **probabilistic inference**

- The full joint distribution is used as the **knowledge base** to derive answers to all the questions
- The $2 \times 2 \times 2$ table below lists a full joint distribution for a **simple example** consisting of just three boolean variables `Toothache`, `Cavity`, and `Catch` (the dentist's steel probe that catches in the tooth)

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

Figure 12.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

- Note that to calculate the probability of **any proposition**, simply **identify** those possible worlds in which the **proposition** is true and **add up** their probabilities
  - For example there are six possible worlds in which `cavity` $\vee$ `toothache` holds:

$$P(\texttt{cavity} \vee \texttt{toothache}) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28$$

ii. **Marginal probability**: Adding the entries in the first row gives the **unconditional** or **marginal probability** of `cavity`

$$P(\texttt{cavity}) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2$$

- This process is called **marginalization**, or summing out - Because we sum up the probabilities for each possible value of the other variables, thereby taking them out of the equation

iii. **Marginalization**: The general marginalization rule for any set of variables $\mathbf{Y}$ and $\mathbf{Z}$ is given by

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

- For example, we can obtain the distribution as $\mathbf{P}(\texttt{Cavity})$, as follows

$$\begin{aligned}\mathbf{P}(\texttt{Cavity}) = {} & \mathbf{P}(\texttt{Cavity, toothache, catch}) + \\ & \mathbf{P}(\texttt{Cavity, toothache, } \neg \texttt{ catch}) + \\ & \mathbf{P}(\texttt{Cavity, } \neg \texttt{ toothache, catch}) + \\ & \mathbf{P}(\texttt{Cavity, } \neg \texttt{ toothache, } \neg \texttt{ catch}) \\ = {} & \langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle + \\ & \langle 0.072, 0.144 \rangle + \langle 0.008, 0.576 \rangle \\ = {} & \langle 0.2, 0.8 \rangle\end{aligned}$$

iv. **Conditioning**: Using the product rule, we can replace $\mathbf{P}(\mathbf{Y}, \mathbf{z})$ in the above equation with $\mathbf{P}(\mathbf{Y} \mid \mathbf{z})\mathbf{P}(\mathbf{z})$, and obtain a rule called **conditioning** :

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

- In most cases, we are interested in computing **conditional probabilities** of some variables, given **evidence** about others
- For example, we can compute the conditional probability $P(\texttt{cavity}|\texttt{toothache})$

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

v. Given the **full joint distribution**, the answers to all the probabilistic queries for discrete variables can be **derived**
- However for a domain described by n boolean variables, it requires an input **table of size** $\mathcal{O}(2^n)$ and takes $\mathcal{O}(2^n)$ **time to process** the table
- In a realistic problem we could easily have $n = 100$, requiring a table with $2^{100} \approx 10^{30}$ entries, which is **impractical**

vi. **Independence:** Two propositions a and b are called **independent** (or marginal independent) if the following properties holds:

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

Where are we ?

Bayes Theorem or Bayes' Rule

i. The **product rule** can actually be written in two forms :

$$P(a \wedge b) = P(a|b)\, P(b) \quad \text{and}$$
$$= P(b|a)\, P(a)$$

- Equating the two right hand sides, we get an equation called **Bayes' Rule** (also Bayes' law or Bayes' theorem)

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

- The more general case of Bayes' rules for multi-valued variables can be written in the $\mathbf{P}$ notation as

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

ii. It allows us to **compute** the single time $P(b|a)$ in terms of three terms: $P(a|b)$, $P(b)$ and $P(a)$

- There are many cases where **we do have good** probability estimates for these three numbers and **need to compute** the fourth

iii. Often, we perceive as **evidence** the `effect` of some unknown `cause` and we would like to determine the `cause`. In that case, Bayes rule becomes

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

iv. **Causal** and **Diagnostic**:

- The conditional probability $P(\texttt{effect}|\texttt{cause})$ quantifies the relationship in the causal direction
- Whereas $P(\texttt{cause}|\texttt{effect})$ describes the **diagnostic** direction.
- For example, the doctor knows $P(\texttt{symptoms}|\texttt{disease})$ and wants to derive a diagnosis, $P(\texttt{disease}|\texttt{symptoms})$

Example

v. Consider a doctor knows that the disease meningitis causes a patient to have a stiff neck say 70% of the time. The doctor also knows some **unconditional** facts, that
   - the **prior probability** that any patient has meningitis $P(m)$ is $1/50,000$ and
   - The **prior probability** that any patient has a stiff neck $P(s)$ is 1%

vi. Then we have

$$\begin{aligned}
\text{Given:} \quad & P(s|m) = 0.7 \\
\text{Given:} \quad & P(m) \quad = 1/50000 \\
\text{Given:} \quad & P(s) \quad = 0.01 \\
\textbf{Compute}: \quad & P(m|s) = \frac{P(s|m)\, P(m)}{P(s)} = \frac{0.7 \times 1/50000}{0.01} \\
& \qquad\quad = 0.0014
\end{aligned}$$

   - That is, we expect only 0.14% of patients with stiff neck to have meningitis

Where are we ?

Representing Knowledge in an Uncertain Domain

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

Figure 12.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

i. we saw that the full joint probability distribution can answer any question about the domain, but can become intractably large as the number of variables grows
ii. We also saw that independence and conditional independence relationships among variables can greatly reduce the number of probabilities that need to be specified in order to define the full joint distribution

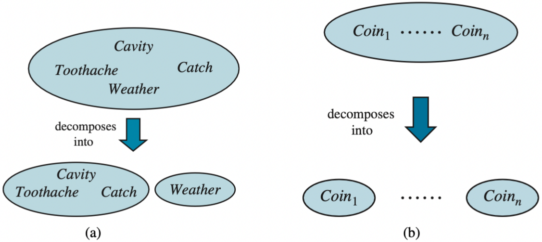


Figure 12.4 Two examples of factoring a large joint distribution into smaller distributions, using absolute independence. (a) Weather and dental problems are independent. (b) Coin flips are independent.

iii. This section introduces a data structure called a Bayesian network to represent the dependencies among variables.
iv. Bayesian networks can represent essentially any full joint probability distribution and in many cases can do so very concisely.
v. A Bayesian network is a directed graph in which each node is annotated with quantitative probability information. The full specification is as follows:
    I. Each node corresponds to a random variable, which may be discrete or continuous.
    II. Directed links or arrows connect pairs of nodes. If there is an arrow from node X to node Y, X is said to be a parent of Y. The graph has no directed cycles and hence is a directed acyclic graph, or DAG.
    III. Each node X_i has associated probability information $\theta(X_i|\text{Parents}(X_i))$ that quantifies the effect of the parents on the node using a finite number of parameters.
vi. The intuitive meaning of an arrow is typically that X has a direct influence on Y, which suggests that causes should be parents of effects
vii. Once the topology of the Bayes net is laid out, we need only specify the local probability information for each variable, in the form of a conditional distribution given its parents
viii. The full joint distribution for all the variables is defined by the topology and the local probability information

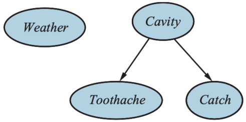


Figure 13.1 A simple Bayesian network in which *Weather* is independent of the other three variables and *Toothache* and *Catch* are conditionally independent, given *Cavity*.

ix. Intuitively, the network represents the fact that Cavity is a direct cause of Toothache and Catch, whereas no direct causal relationship exists between Toothache and Catch.

Where are we ?

Bayesian Networks

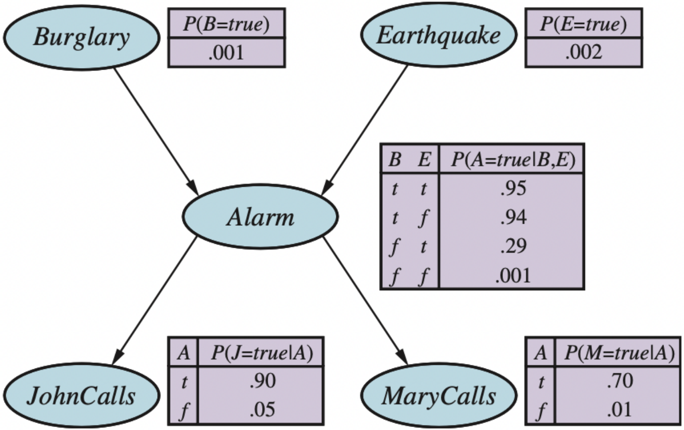


Figure 13.2 A typical Bayesian network, showing both the topology and the conditional probability tables (CPTs). In the CPTs, the letters *B*, *E*, *A*, *J*, and *M* stand for *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, and *MaryCalls*, respectively.

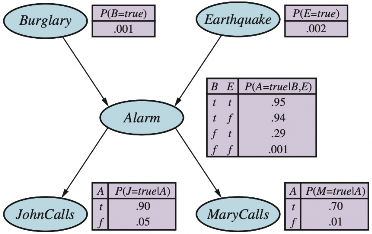
i. Now consider the following example. You have a new burglar alarm installed at home.
   - It is fairly reliable at detecting a burglary, but is occasionally set off by minor earthquakes

ii. You also have two neighbors, John and Mary, who have promised to call you at work when they hear the alarm
   - John nearly always calls when he hears the alarm, but sometimes confuses the telephone ringing with the alarm and calls then, too
   - Mary, on the other hand, likes rather loud music and often misses the alarm altogether

iii. Given the evidence of who has or has not called, we would like to estimate the probability of a burglary

iv. The network structure shows that burglary and earthquakes directly affect the probability of the alarm's going off, but whether John and Mary call depends only on the alarm

v. The local probability information attached to each node in Figure takes the form of a conditional probability table (CPT)



P(B=true)
.001

P(E=true)
.002

B	E	P(A=true|B,E)
t	t	.95
t	f	.94
f	t	.29
f	f	.001

A	P(J=true|A)
t	.90
f	.05

A	P(M=true|A)
t	.70
f	.01

Figure 13.2 A typical Bayesian network, showing both the topology and the conditional probability tables (CPTs). In the CPTs, the letters *B*, *E*, *A*, *J*, and *M* stand for *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, and *MaryCalls*, respectively.

vi. Each row in a CPT contains the conditional probability of each node value for a conditioning case
   - A conditioning case is just a possible combination of values for the parent nodes—a miniature possible world
   - Each row must sum to 1, because the entries represent an exhaustive set of cases for the variable

vii. In general, a table for a Boolean variable with k Boolean parents contains 2^k independently specifiable probabilities
   - A node with no parents has only one row, representing the prior probabilities of each possible value of the variable.

Text Books

[1] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*.
Third Edition, Prentice-Hall, 2009.

[2] S. N. Elaine Rich, Kevin Knight, *Artificial Intelligence*.
Third Edition, The McGraw Hill Publications, 2009.

[3] G. F. Luger, *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*.
Sixth Edition, Pearson Education, 2009.

Thank you