Modeling and reasoning with uncertainty

KB systems. Medical example.

We want to build a KB system for the diagnosis of pneumonia.

Problem description:
• Disease: pneumonia
• Patient symptoms (findings, lab tests):
  – Fever, Cough, Paleness, WBC (white blood cells) count,
    Chest pain, etc.

Representation of a patient case:
• Statements that hold (are true) for the patient.
  E.g: Fever = True
       Cough = False
       WBC count = High

Diagnostic task: we want to decide whether the patient suffers
from the pneumonia or not given the symptoms
Uncertainty

To make diagnostic inference possible we need to represent knowledge (axioms) that relate symptoms and diagnosis

Problem: disease/symptoms relations are not deterministic
- They are uncertain (or stochastic) and vary from patient to patient

Two types of uncertainty:
- **Disease** → **Symptoms uncertainty**
  - A patient suffering from pneumonia may not have fever all the times, may or may not have a cough, white blood cell test can be in a normal range.

- **Symptoms** → **Disease uncertainty**
  - High fever is typical for many diseases (e.g. bacterial diseases) and does not point specifically to pneumonia
  - Fever, cough, paleness, high WBC count combined do not always point to pneumonia
Uncertainty

Why are relations uncertain?

- **Observability**
  - It is impossible to observe all relevant components of the world
  - Observable components behave stochastically even if the underlying world is deterministic

- **Efficiency, capacity limits**
  - It is often impossible to enumerate and model all components of the world and their relations
  - Abstractions can make the relations stochastic

*Humans can reason with uncertainty !!!*

- Can computer systems do the same?

---

Modeling the uncertainty.

**Key challenges:**

- How to represent the relations in the presence of uncertainty?
- How to manipulate such knowledge to make inferences?
  - *Humans can reason with uncertainty.*
Methods for representing uncertainty

Extensions of the propositional and first-order logic
– Use, uncertain, imprecise statements (relations)

Example: Propositional logic with certainty factors
Very popular in 70-80s in knowledge-based systems (MYCIN)

• Facts (propositional statements) are assigned a certainty value reflecting the belief in that the statement is satisfied:
  \[ CF(Pneumonia = True) = 0.7 \]
• Knowledge: typically in terms of modular rules

| If | 1. The patient has cough, and |
|    | 2. The patient has a high WBC count, and |
|    | 3. The patient has fever |
| Then with certainty 0.7 | the patient has pneumonia |

Certainty factors

Problem 1:
• Chaining of multiple inference rules (propagation of uncertainty)
  Solution:
  • Rules incorporate tests on the certainty values
    \[ (A \text{ in } [0.5,1]) \land (B \text{ in } [0.7,1]) \rightarrow C \text{ with } CF = 0.8 \]

Problem 2:
• Combinations of rules with the same conclusion
  \[ (A \text{ in } [0.5,1]) \land (B \text{ in } [0.7,1]) \rightarrow C \text{ with } CF = 0.8 \]
  \[ (E \text{ in } [0.8,1]) \land (D \text{ in } [0.9,1]) \rightarrow C \text{ with } CF = 0.9 \]

• What is the resulting \( CF(C) \)?
**Certainty factors**

- **Combination of multiple rules**
  
  \[(A \text{ in } [0.5,1]) \land (B \text{ in } [0.7,1]) \rightarrow C \text{ with } CF = 0.8\]
  
  \[(E \text{ in } [0.8,1]) \land (D \text{ in } [0.9,1]) \rightarrow C \text{ with } CF = 0.9\]

- **Three possible solutions**
  
  \[CF(C) = \max[0.9;0.8] = 0.9\]
  
  \[CF(C) = 0.9 \times 0.8 = 0.72\]
  
  \[CF(C) = 0.9 + 0.8 - 0.9 \times 0.8 = 0.98\]

**Problems:**

- Which solution to choose?
- All three methods break down after a sequence of inference rules

---

**Methods for representing uncertainty**

**Probability theory**

- A well defined theory for modeling and reasoning in the presence of uncertainty
- A natural choice to replace certainty factors

**Facts (propositional statements)**

- Are represented via **random variables** with two or more values
  
  **Example:** Pneumonia is a random variable
  
  **values:** True and False
  
- Each value can be achieved with some probability:
  
  \[P(\text{Pneumonia} = \text{True}) = 0.001\]
  
  \[P(\text{WBC count} = \text{high}) = 0.005\]
Probability theory

- Well-defined theory for representing and manipulating statements with uncertainty
- **Axioms of probability:**
  For any two propositions $A$, $B$.
  1. $0 \leq P(A) \leq 1$
  2. $P(True) = 1$ and $P(False) = 0$
  3. $P(A \lor B) = P(A) + P(B) - P(A \land B)$

![Venn diagram](image)

Modeling uncertainty with probabilities

**Probabilistic extension of propositional logic**

- **Propositions:**
  - statements about the world
  - Represented by the assignment of values to random variables
- **Random variables:**
  - **Boolean** $Pneumonia$ is either True, False
  - **Multi-valued** $Pain$ is one of \{Nopain, Mild, Moderate, Severe\}
  - **Continuous** $HeartRate$ is a value in $<0;180>$

M. Hauskrecht
Probabilities

Unconditional probabilities (prior probabilities)

\[ P(\text{Pneumonia}) = 0.001 \quad \text{or} \quad P(\text{Pneumonia} = \text{True}) = 0.001 \]
\[ P(\text{Pneumonia} = \text{False}) = 0.999 \]
\[ P(\text{WBC count} = \text{high}) = 0.005 \]

Probability distribution

- Defines probabilities for all possible value assignments to a random variable
- Values are mutually exclusive

\[
\begin{array}{|c|c|}
\hline
\text{Pneumonia} & P(\text{Pneumonia}) \\
\hline
\text{True} & 0.001 \\
\text{False} & 0.999 \\
\hline
\end{array}
\]

Example 1:

\[ P(\text{Pneumonia} = \text{True}) = 0.001 \]
\[ P(\text{Pneumonia} = \text{False}) = 0.999 \]

\[ P(\text{Pneumonia} = \text{True}) + P(\text{Pneumonia} = \text{False}) = 1 \]

Probabilities sum to 1 !!!

Probability distribution

Defines probability for all possible value assignments

Example 1:

\[ P(\text{Pneumonia} = \text{True}) = 0.001 \]
\[ P(\text{Pneumonia} = \text{False}) = 0.999 \]

\[
\begin{array}{|c|c|}
\hline
\text{Pneumonia} & P(\text{Pneumonia}) \\
\hline
\text{True} & 0.001 \\
\text{False} & 0.999 \\
\hline
\end{array}
\]

Example 2:

\[ P(\text{WBC count} = \text{high}) = 0.005 \]
\[ P(\text{WBC count} = \text{normal}) = 0.993 \]
\[ P(\text{WBC count} = \text{high}) = 0.002 \]

\[
\begin{array}{|c|c|}
\hline
\text{WBC count} & P(\text{WBC count}) \\
\hline
\text{high} & 0.005 \\
\text{normal} & 0.993 \\
\text{low} & 0.002 \\
\hline
\end{array}
\]
Joint probability distribution

Joint probability distribution (for a set variables)
• Defines probabilities for all possible assignments of values to variables in the set

Example 2: Assume variables:

- Pneumonia (2 values)
- WBC count (3 values)
- Pain (4 values)

\( P(\text{pneumonia, WBC count, Pain}) \) is represented by \( 2 \times 3 \times 4 \) array

---

Joint probability distribution

Joint probability distribution (for a set variables)
• Defines probabilities for all possible assignments of values to variables in the set

Example: variables Pneumonia and WBC count

\( P(\text{pneumonia, WBC count}) \)

is represented by \( 2 \times 3 \) matrix

<table>
<thead>
<tr>
<th>Pneumonia</th>
<th>high</th>
<th>normal</th>
<th>low</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>0.0008</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>False</td>
<td>0.0042</td>
<td>0.9929</td>
<td>0.0019</td>
</tr>
</tbody>
</table>
Joint probabilities

Marginalization
• reduces the dimension of the joint distribution
• Sums variables out

\[ \mathbf{P}(\text{pneumonia}, \text{WBCcount}) \] 2×3 matrix

\[
\begin{array}{c|ccc}
\text{WBCcount} & \text{high} & \text{normal} & \text{low} \\
\hline
\text{Pneumonia True} & 0.0008 & 0.0001 & 0.0001 \\
\text{Pneumonia False} & 0.0042 & 0.9929 & 0.0019 \\
\end{array}
\]

\[ \mathbf{P}(\text{WBCcount}) \]
Marginalization (here summing of columns or rows)

Marginalization
• reduces the dimension of the joint distribution

\[ P(X_1, X_2, \ldots X_{n-1}) = \sum_{\{X_n\}} P(X_1, X_2, \ldots , X_{n-1}, X_n) \]
• We can continue doing this

\[ P(X_2, \ldots X_{n-1}) = \sum_{\{X_1, X_n\}} P(X_1, X_2, \ldots , X_{n-1}, X_n) \]

What is the maximal joint probability distribution?
• Full joint probability
Full joint distribution

The joint distribution for all variables in the problem
– It defines the complete probability model for the problem

Example: pneumonia diagnosis

Variables: \textit{Pneumonia, Fever, Paleness, WBCcount, Cough}

Full joint defines the probability for all possible assignments of values to these variables

\[
P(\text{Pneumonia} = T, \text{WBCcount} = \text{High}, \text{Fever} = T, \text{Cough} = T, \text{Paleness} = T) \\
P(\text{Pneumonia} = T, \text{WBCcount} = \text{High}, \text{Fever} = T, \text{Cough} = T, \text{Paleness} = F) \\
P(\text{Pneumonia} = T, \text{WBCcount} = \text{High}, \text{Fever} = T, \text{Cough} = F, \text{Paleness} = T) \\
\ldots \quad \text{etc}
\]

• How many probabilities are there?

Exponential in the number of variables
Full joint distribution

- Any joint probability for a subset of variables can be obtained via marginalization

\[ P(\text{Pneumonia}, \text{WBC count}, \text{Fever}) = \sum_{c, p \in \{T, F\}} P(\text{Pneumonia}, \text{WBC count}, \text{Fever}, \text{Cough} = c, \text{Paleness} = p) \]

- Is it possible to recover the full joint from the joint probabilities over a subset of variables?

Joint probabilities

- Is it possible to recover the full joint from the joint probabilities over a subset of variables?

\[
\begin{pmatrix}
\text{Pneumonia} & \text{high} & \text{normal} & \text{low} \\
\text{True} & 0.005 & 0.993 & 0.002 \\
\text{False} & 0.005 & 0.993 & 0.002 \\
\end{pmatrix}
\]

\[
P(\text{Pneumonia}) = 0.001 \\
P(\text{WBC count}) = 0.999
\]
Joint probabilities and independence

- Is it possible to recover the full joint from the joint probabilities over a subset of variables?
- Only if the variables are independent!!!

\[ P(\text{pneumonia}, \text{WBC\text{count}}) \quad 2 \times 3 \text{ matrix} \]

<table>
<thead>
<tr>
<th>Pneumonia</th>
<th>high</th>
<th>normal</th>
<th>low</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>False</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

\[ P(\text{WBC\text{count}}) = 0.005 \quad 0.993 \quad 0.002 \]

\[ P(\text{Pneumonia}) = 0.001 \quad 0.999 \]

Conditional probabilities

- **Conditional probability distribution.**
  \[ P(A \mid B) = \frac{P(A, B)}{P(B)} \quad \text{s.t. } P(B) \neq 0 \]

- **Product rule.** Join probability can be expressed in terms of conditional probabilities
  \[ P(A, B) = P(A \mid B)P(B) \]

- **Chain rule.** Any joint probability can be expressed as a product of conditionals
  \[ P(X_1, X_2, \ldots X_n) = P(X_n \mid X_1, \ldots X_{n-1})P(X_1, \ldots X_{n-1}) \]
  \[ = P(X_n \mid X_1, \ldots X_{n-1})P(X_{n-1} \mid X_1, \ldots X_{n-2})P(X_1, \ldots X_{n-2}) \]
  \[ = \prod_{i=1}^{n} P(X_i \mid X_1, \ldots X_{i-1}) \]
Conditional probabilities

Conditional probability
• Is defined in terms of the joint probability:
\[ P(A \mid B) = \frac{P(A, B)}{P(B)} \text{ s.t. } P(B) \neq 0 \]
• Example:
\[
\begin{align*}
P(\text{pneumonia} = \text{true} \mid \text{WBC count} = \text{high}) &= \frac{P(\text{pneumonia} = \text{true}, \text{WBC count} = \text{high})}{P(\text{WBC count} = \text{high})} \\
P(\text{pneumonia} = \text{false} \mid \text{WBC count} = \text{high}) &= \frac{P(\text{pneumonia} = \text{false}, \text{WBC count} = \text{high})}{P(\text{WBC count} = \text{high})}
\end{align*}
\]

Conditional probabilities

Conditional probability distribution
• Defines probabilities for all possible assignments, given a fixed assignment to some other variable values

\[ P(\text{Pneumonia} = \text{true} \mid \text{WBC count} = \text{high}) \]

\[
P(\text{Pneumonia} \mid \text{WBC count}) \text{ 3 element vector of 2 elements}
\]

<table>
<thead>
<tr>
<th>Pneumonia</th>
<th>high</th>
<th>normal</th>
<th>low</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>0.08</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>False</td>
<td>0.92</td>
<td>0.9999</td>
<td>0.9999</td>
</tr>
</tbody>
</table>

\[ P(\text{Pneumonia} = \text{true} \mid \text{WBC count} = \text{high}) + P(\text{Pneumonia} = \text{false} \mid \text{WBC count} = \text{high}) \]
Bayes rule

Conditional probability.
\[ P(A|B) = \frac{P(A,B)}{P(B)} \]
\[ P(A,B) = P(B|A)P(A) \]

Bayes rule:
\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)} \]

When is it useful?

• When we are interested in computing the diagnostic query from the causal probability
  \[ P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(effect)} \]

• **Reason:** It is often easier to assess causal probability
  – E.g. Probability of pneumonia causing fever vs. probability of pneumonia given fever

Probabilistic inference

Various inference tasks:

• **Diagnostic task.** (from effect to cause)
  \[ P(Pneumonia|Fever=T) \]

• **Prediction task.** (from cause to effect)
  \[ P(Fever|Pneumonia=T) \]

• **Other probabilistic queries** (queries on joint distributions).
  \[ P(Fever) \]
  \[ P(Fever, ChestPain) \]
Inference

Any query can be computed from the full joint distribution !!!

• **Joint over a subset of variables** is obtained through marginalization

\[ P(A = a, C = c) = \sum_i \sum_j P(A = a, B = b_j, C = c, D = d_j) \]

• **Conditional probability over set of variables**, given other variables’ values is obtained through marginalization and definition of conditionals

\[ P(D = d \mid A = a, C = c) = \frac{P(A = a, C = c, D = d)}{P(A = a, C = c)} \]

\[ = \frac{\sum_i P(A = a, B = b_i, C = c, D = d)}{\sum_i \sum_j P(A = a, B = b_i, C = c, D = d_j)} \]

---

Modeling uncertainty with probabilities

• Defining the full joint distribution makes it possible to represent and reason with uncertainty in a uniform way

• We are able to handle an arbitrary inference problem

**Problems:**

– **Space complexity.** To store a full joint distribution we need to remember \( O(d^n) \) numbers.

\( n \) – number of random variables, \( d \) – number of values

– **Inference (time) complexity.** To compute some queries requires \( O(d^n) \) steps.

– **Acquisition problem.** Who is going to define all of the probability entries?
Medical diagnosis example

- **Space complexity.**
  - Pneumonia (2 values: T,F), Fever (2: T,F), Cough (2: T,F), WBC count (3: high, normal, low), paleness (2: T,F)
  - Number of assignments: $2^2 \times 2^2 \times 3 \times 2 = 48$
  - We need to define at least 47 probabilities.

- **Time complexity.**
  - Assume we need to compute the marginal of Pneumonia=T from the full joint

\[
P(P\text{neumonia} = T) = \\
= \sum_{i \in T,F} \sum_{j \in T,F} \sum_{k=h,n,l} \sum_{u \in T,F} P(F\text{ever} = i, C\text{ough} = j, W\text{BC count} = k, \text{Pale} = u)
\]

  - Sum over: $2 \times 2 \times 3 \times 2 = 24$ combinations

---

Modeling uncertainty with probabilities

- **Knowledge based system era (70s – early 80’s)**
  - **Extensional non-probabilistic models**
  - Solve the space, time and acquisition bottlenecks in probability-based models
  - Froze the development and advancement of KB systems and contributed to the slow-down of AI in 80s in general

- **Breakthrough (late 80s, beginning of 90s)**
  - **Bayesian belief networks**
    - Give solutions to the space, acquisition bottlenecks
    - Partial solutions for time complexities
  - Bayesian belief network
Bayesian belief networks (BBNs)

**Bayesian belief networks.**
- Represent the full joint distribution over the variables more compactly with a smaller number of parameters.
- Take advantage of conditional and marginal independences among random variables

- **A and B are independent**
  \[ P(A, B) = P(A)P(B) \]

- **A and B are conditionally independent given C**
  \[ P(A, B | C) = P(A | C)P(B | C) \]
  \[ P(A | C, B) = P(A | C) \]