Modeling uncertainty using probabilities

Milos Hauskrecht
milos@cs.pitt.edu
5329 Sennott Square

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
        WBCcount = 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 become 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:
  \[ CF(\text{Pneumonia} = \text{True}) = 0.7 \]
• Knowledge: typically in terms of modular rules

<table>
<thead>
<tr>
<th>If</th>
<th>Then (with certainty 0.7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The patient has cough, and</td>
<td></td>
</tr>
<tr>
<td>2. The patient has a high WBC count, and</td>
<td></td>
</tr>
<tr>
<td>3. The patient has fever</td>
<td></td>
</tr>
<tr>
<td>the patient has pneumonia</td>
<td></td>
</tr>
</tbody>
</table>

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:
  \[ P(Pneumonia = True) = 0.001 \]
  \[ P(WBCcount = high) = 0.005 \]
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. $P(A \lor B) = P(A) + P(B) - P(A \land B)$

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 $<0; 250>$
Probabilities

Unconditional probabilities (prior probabilities)

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

Probability distribution

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

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

<table>
<thead>
<tr>
<th>Pneumonia</th>
<th>(P(\text{Pneumonia}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>0.001</td>
</tr>
<tr>
<td>False</td>
<td>0.999</td>
</tr>
</tbody>
</table>

Probabilities sum to 1 !!!

Example 1:

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

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

Example 2:

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

<table>
<thead>
<tr>
<th>WBCcount</th>
<th>(P(\text{WBCcount}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>0.005</td>
</tr>
<tr>
<td>normal</td>
<td>0.993</td>
</tr>
<tr>
<td>low</td>
<td>0.002</td>
</tr>
</tbody>
</table>
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

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

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

\( P(\text{pneumonia}, \text{WBC count}) \) \(2 \times 3\) 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>

\( P(\text{Pneumonia}) \)

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

**Marginalization** (here summing of columns or rows)
**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:** Pneumonia, Fever, Paleness, WBCcount, Cough

Full joint defines the probability for all possible assignments of values to Pneumonia, Fever, Paleness, WBCcount, Cough

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

---

**Conditional probabilities**

**Conditional probability distribution**

- Defines probabilities for all possible assignments, given a fixed assignment to some other variable values

\[ P(\text{Pneumonia} = \text{true} | \text{WBCcount} = \text{high}) \]

\[ P(\text{Pneumonia} | \text{WBCcount}) \]

3 element vector of 2 elements

\[
\begin{array}{c|ccc}
\text{Pneumonia} & \text{high} & \text{normal} & \text{low} \\
\hline
\text{True} & 0.08 & 0.0001 & 0.0001 \\
\text{False} & 0.92 & 0.9999 & 0.9999 \\
\hline
1.0 & 1.0 & 1.0
\end{array}
\]

\[ P(\text{Pneumonia} = \text{true} | \text{WBCcount} = \text{high}) \]
\[ + P(\text{Pneumonia} = \text{false} | \text{WBCcount} = \text{high}) \]
Conditional probabilities

Conditional probability
- Is defined in terms of the joint probability:

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

- Example:

\[
P(\text{pneumonia} = \text{true} \mid \text{WBC count} = \text{high}) = \frac{P(\text{pneumonia} = \text{true}, \text{WBC count} = \text{high})}{P(\text{WBC count} = \text{high})}
\]

\[
P(\text{pneumonia} = \text{false} \mid \text{WBC count} = \text{high}) = \frac{P(\text{pneumonia} = \text{false}, \text{WBC count} = \text{high})}{P(\text{WBC count} = \text{high})}
\]

\[ \]

Conditional probabilities

- Conditional probability distribution.

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

- 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

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

\[ = P(X_n \mid X_1, \ldots, X_{n-1})P(X_{n-1} \mid X_1, \ldots, X_{n-2})P(X_1, \ldots, X_{n-2}) \]

\[ = \prod_{i=1}^{n} P(X_i \mid X_1, \ldots, X_{i-1}) \]
Bayes rule

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

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

When is it useful?
- When we are interested in computing the diagnostic query from the causal probability
  \[ P(cause \mid effect) = \frac{P(effect \mid cause)P(cause)}{P(effect)} \]
  - Reason: It is often easier to assess causal probability
    - E.g. Probability of pneumonia causing fever vs. probability of pneumonia given fever

Bayes rule

Assume a variable A with multiple values \( a_1, a_2, \ldots a_k \)

Bayes rule can be rewritten as:
\[ P(A = a_j \mid B = b) = \frac{P(B = b \mid A = a_j)P(A = a_j)}{P(B = b)} \]
\[ = \frac{P(B = b \mid A = a_j)P(A = a_j)}{\sum_{i=1}^{k} P(B = b \mid A = a_i)P(A = a_i)} \]

Used in practice when we want to compute:
\[ P(A \mid B = b) \quad \text{for all values of} \quad a_1, a_2, \ldots a_k \]
Bayes Rule in a simple diagnostic inference.

- **Device** (equipment) operating *normally* or *malfunctioning*.
  - Operation of the device sensed indirectly via a sensor
- **Sensor reading** is either *high* or *low*

<table>
<thead>
<tr>
<th>Device status</th>
<th>normal</th>
<th>malfunctioning</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.9</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P(Sensor reading</th>
<th>Device status)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device\Sensor</td>
<td>high</td>
</tr>
<tr>
<td>normal</td>
<td>0.1</td>
</tr>
<tr>
<td>malfunctioning</td>
<td>0.6</td>
</tr>
</tbody>
</table>

**Diagnostic inference:** compute the probability of device operating normally or malfunctioning given a sensor reading

\[
P(\text{Device status} | \text{Sensor reading} = \text{high}) = ?
\]
\[
= \frac{P(\text{Device status} = \text{normal} | \text{Sensor reading} = \text{high})}{P(\text{Device status} = \text{malfunctioning} | \text{Sensor reading} = \text{high})}
\]

- Note that typically the opposite conditional probabilities are given to us: they are much easier to estimate
- **Solution:** apply **Bayes rule** to reverse the conditioning variables