Probabilistic reasoning with Bayesian networks

Lecturers: Silja Renooij & Linda van der Gaag  
*e-mail:* {s.renooij, l.c.vandergaag}@uu.nl

Prerequisites: probability theory & graph theory

Literature: syllabus & slides & studymanual

Form: lectures & exercises (formative self assessment)  
(tip: discuss exercises on Blackboard forum)

Grading: practical assignments & written exam

Additional info: see course website:  
http://www.cs.uu.nl/docs/vakken/prob/
Chapter 1:

Introduction
Reasoning under uncertainty

In numerous application areas of knowledge-based decision-support systems we have

- uncertainty concerning the general domain knowledge;
- problem-specific information that is often uncertain, incomplete and even contradictory.

A decision-support system should be capable of dealing with these types of knowledge.
Application of probability theory

Consider a discrete joint probability distribution $\Pr$ on a set of random variables $\mathbf{V} = \{V_1, \ldots, V_n\}$. In general we have that:

- the representation of $\Pr$ requires exponential space
  consider e.g. $n = 2$ binary-valued variables, or $n = 40$; what if they have 5 values each? (and how do you get the numbers?)

- calculating the (conditional) probability of a value of a variable by conditioning and marginalisation requires exponential time
  consider e.g. computing $\Pr(V_1 = \text{true})$ from $\Pr(\mathbf{V})$, or $\Pr(V_1 = \text{true} \mid V_2 = \text{true})$

This cannot be improved without additional knowledge about the probability distribution.
Diagnosis problem: pioneering in the 1960s

Let $H = \{h_1, \ldots, h_n\}, n \geq 1$, be a set of hypotheses, and let $E = \{e_1, \ldots, e_m\}, m \geq 1$, be a set of relevant findings (evidence).

Determine the 'best' diagnosis given findings $e \subseteq E$.

The approach: Compute for each $h \subseteq H$ the probability

$$
\Pr(h \mid e) = \frac{\Pr(e \mid h) \Pr(h)}{\Pr(e)}
$$

Drawback: An exponential number of probabilities need to be computed; storage is also exponential.
Pioneering in the 1960s

Determine the diagnosis given findings \( e \subseteq E \).

**The approach:** Assume \( h_i \in H \) mutually exclusive, and collectively exhaustive: \( \bigcup_{i=1}^{n} \{h_i\} = \Omega \).

Then, compute for each \( h_i \in H \):

\[
\Pr(h_i \mid e) = \frac{\Pr(e \mid h_i) \Pr(h_i)}{\Pr(e)} = \frac{\Pr(e \mid h_i) \Pr(h_i)}{\sum_{k=1}^{n} \Pr(e \mid h_k) \Pr(h_k)}
\]

**Drawback:** We compute only \( n - 1 \) probabilities, but computation still requires an exponential number of probabilities.
Determine the diagnosis given findings \( e = \{e_p, \ldots, e_q\} \), \( 1 \leq p, q \leq m \).

The approach: Assume in addition that all findings \( e_1, \ldots, e_m \) are conditionally independent given \( h_i \), \( i = 1, \ldots, n \). Then:

\[
\Pr(h_i \mid e) = \frac{\Pr(e_p, \ldots, e_q \mid h_i) \Pr(h_i)}{\sum_{k=1}^{n} \Pr(e_p, \ldots, e_q \mid h_k) \Pr(h_k)}
\]

\[
= \frac{\Pr(e_p \mid h_i) \cdot \ldots \cdot \Pr(e_q \mid h_i) \Pr(h_i)}{\sum_{k=1}^{n} \Pr(e_p \mid h_k) \cdot \ldots \cdot \Pr(e_q \mid h_k) \Pr(h_k)}
\]

Benefit: Only \( m \cdot n \) conditional probabilities and \( n - 1 \) prior probabilities are required for the computation.
GLADYS

GLADYS (GLASGOW DYSPEPSIA SYSTEM) is a system for diagnosing dyspepsia.

The global structure of the system:

- Interview
- Differential diagnosis
- Therapy selection

Probabilistic component developed with data collected from ± 1200 patients.

**Symptoms and diseases**

**Context:** patients with an Ulcer. **Question:** which type?

<table>
<thead>
<tr>
<th></th>
<th>duodenal ulcer ($n = 248$)</th>
<th>gastric ulcer ($n = 43$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>169</td>
<td>17</td>
</tr>
<tr>
<td>female</td>
<td>79</td>
<td>26</td>
</tr>
<tr>
<td><strong>Age:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 26</td>
<td>43</td>
<td>1</td>
</tr>
<tr>
<td>26 - 40</td>
<td>82</td>
<td>5</td>
</tr>
<tr>
<td>41 - 55</td>
<td>87</td>
<td>19</td>
</tr>
<tr>
<td>&gt; 55</td>
<td>36</td>
<td>18</td>
</tr>
<tr>
<td><strong>Daily pain:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td>21</td>
<td>11</td>
</tr>
<tr>
<td>no</td>
<td>214</td>
<td>27</td>
</tr>
<tr>
<td><strong>Effect food on pain:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>worsens</td>
<td>44</td>
<td>11</td>
</tr>
<tr>
<td>no effect</td>
<td>82</td>
<td>9</td>
</tr>
<tr>
<td>relieves</td>
<td>104</td>
<td>17</td>
</tr>
</tbody>
</table>

**Probability**

<table>
<thead>
<tr>
<th></th>
<th>duodenal ulcer ($n = 248$)</th>
<th>gastric ulcer ($n = 43$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>probability</td>
<td>0.85</td>
<td>0.15</td>
</tr>
</tbody>
</table>
The idea

Let $Pr$ be a joint distribution on the diagnosis search space including hypothesis $h$ and observed findings $e$.

The prior odds for $h$, and posterior odds for $h$ given $e$, are defined by

$$O(h) = \frac{Pr(h)}{1 - Pr(h)} = \frac{Pr(h)}{Pr(\neg h)} \quad \text{and} \quad O(h \mid e) = \frac{Pr(h \mid e)}{Pr(\neg h \mid e)}$$

Assume that all findings $e_i \in e$ are conditionally independent given $h$, then

$$O(h \mid e) = \frac{Pr(e \mid h) \cdot Pr(h)}{Pr(e \mid \neg h) \cdot Pr(\neg h)} = \prod_i \frac{Pr(e_i \mid h)}{Pr(e_i \mid \neg h)} \cdot O(h)$$

Now consider the following transformation: $10 \cdot \ln O(h \mid e)$...
The idea (cntd)

Applying the transformation $10 \cdot \ln$ to

$$O(h \mid e) = \prod_i \lambda_i \cdot O(h), \text{ where } \lambda_i = \frac{\Pr(e_i \mid h)}{\Pr(e_i \mid \neg h)}$$

results in a score $s$:

$$s = 10 \cdot \ln O(h \mid e) = 10 \cdot \ln O(h) + \sum_i 10 \cdot \ln \lambda_i = w_0 + \sum_i w_i$$

where $w_i$ is a weight for finding $e_i$.

The probability $\Pr(h \mid e)$ is now computed from

$$\Pr(h \mid e) = \frac{O(h \mid e)}{1 + O(h \mid e)} = \frac{e^{\frac{s}{10}}}{1 + e^{\frac{s}{10}}} = \frac{1}{1 + e^{-\frac{s}{10}}}$$
**A scoring system**

<table>
<thead>
<tr>
<th></th>
<th>$h$: duodenal ulcer (du)</th>
<th>$\neg h$: gastric ulcer (gu)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>($n = 248$)</td>
<td>($n = 43$)</td>
</tr>
<tr>
<td>male (m)</td>
<td>169</td>
<td>17</td>
</tr>
<tr>
<td>female (f)</td>
<td>79</td>
<td>26</td>
</tr>
</tbody>
</table>

**Calculation of probabilities, likelihood ratios and weights:**

\[
\Pr(m \mid du) = \frac{169}{248} \sim 0.68, \quad \Pr(m \mid gu) \sim 0.40 \quad \Rightarrow \\
\lambda_m = \frac{\Pr(m \mid du)}{\Pr(m \mid gu)} = \frac{0.68}{0.40} \sim 1.7 \quad \Rightarrow \quad w_m = 10 \cdot \ln \lambda_m \sim 5
\]

\[
\Pr(f \mid du) = \frac{79}{248} \sim 0.32, \quad \Pr(f \mid gu) \sim 0.60 \quad \Rightarrow \\
\lambda_f = \frac{\Pr(f \mid du)}{\Pr(f \mid gu)} = \frac{0.32}{0.60} \sim 0.53 \quad \Rightarrow \quad w_f = 10 \cdot \ln \lambda_f \sim -6
\]
## Symptoms and their weights

<table>
<thead>
<tr>
<th></th>
<th>duodenal ulcer ((n = 248))</th>
<th>gastric ulcer ((n = 43))</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>169</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>female</td>
<td>79</td>
<td>26</td>
<td>−6</td>
</tr>
<tr>
<td><strong>Age:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 26</td>
<td>43</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>26 - 40</td>
<td>82</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>41 - 55</td>
<td>87</td>
<td>19</td>
<td>−2</td>
</tr>
<tr>
<td>&gt; 55</td>
<td>36</td>
<td>18</td>
<td>−10</td>
</tr>
<tr>
<td><strong>Daily pain:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td>21</td>
<td>11</td>
<td>−12</td>
</tr>
<tr>
<td>no</td>
<td>214</td>
<td>27</td>
<td>3</td>
</tr>
<tr>
<td><strong>Effect food on pain:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>worsens</td>
<td>44</td>
<td>11</td>
<td>−4</td>
</tr>
<tr>
<td>no effect</td>
<td>82</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>relieves</td>
<td>104</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td><strong>prior</strong></td>
<td>0.85</td>
<td>0.15</td>
<td>17</td>
</tr>
</tbody>
</table>
An example diagnosis

A 30 year old woman reports to the clinic. She has pain in the abdominal area, but not on a daily basis; the pain worsens as soon as she eats.

Calculation of the score:

- the initial score: \(+17\)
- the patient is female: \(-6\)
- her age is 30: \(+10\)
- she is in pain, but not every day: \(+3\)
- food intake worsens the pain: \(-4\)

\[
\begin{align*}
\quad & \quad +10 \\
\quad & \quad +3 \\
\quad & \quad -4 \\
\quad & \quad \hline
\quad & \quad +20
\end{align*}
\]

Given that the patient has one of the two diseases, duodenal ulcer and gastric ulcer, she has with probability

\[
(1 + e^{-\frac{20}{10}})^{-1} \approx 1.14^{-1} \approx 0.88
\]

a duodenal ulcer and a gastric ulcer with probability 0.12.
Reviewing ‘Idiot’s Bayes’

The naive Bayes approach is

- mathematically correct, and
- computationally easy.

However

- underlying assumptions usually unacceptable;
- and, at the time, for larger applications
  - # of hypotheses often large \(\rightarrow\) undoable to compute each \(\Pr(h_i \mid e)\);
  - often not enough information for reliable probability assessments.
The most likely hypothesis given observed findings is determined as follows:

- prune the search space using heuristic rules;
- approximate the missing probabilities required, for example with:
  \[
  \Pr(e_i \land e_j) = \min\{\Pr(e_i), \Pr(e_j)\};
  \]
- select the hypothesis with the highest probability.
Reviewing the quasi-probabilistic models

The quasi-probabilistic models are

• computationally easy, and
• easy to use,

even for larger applications.

However, these models are

• mathematically incorrect, and
• even as an approximation model not convincing.
The rehabilitation of probability theory in the 1980s

Judea Pearl introduces Bayesian belief networks as representational device

- + algorithms for inferring (computing) ’beliefs’ from those represented
- first for trees and polytrees (singly connected graphs)
- then for multiply-connected graphs
- for the latter, the algorithm by Steffen Lauritzen & David Spiegelhalter was the first to find wide-spread use.

Also see “Inference in Bayesian Networks: a Historical Perspective”, by Adnan Darwiche
The Bayesian network framework

A Bayesian network is a very compact representation of a joint probability distribution $Pr$. Such a network comprises

- **qualitative** knowledge of $Pr$: a graphical representation of the independences between the variables involved;
- **quantitative** knowledge of $Pr$: conditional probability distributions that describe $Pr$ ‘locally’ per group of variables.

Associated with a Bayesian network are algorithms for computing probabilities and for processing evidence.
An example: Classical Swine Fever (CSF)

The classical swine fever network is a decision-support system for the early detection of classical swine fever (varkenspest).

- early detection of CSF is important, but hard;
- the network has been developed in cooperation with 2 veterinarians of the Central Veterinary Institute of Wageningen UR;
- part of european EPIZONE project;
- veterinarians all over the country collected data with PDAs.
The Classical swine fever network: initial graphical structure
The Classical swine fever network: probability tables

\[
\Pr(\text{Appetite} \mid \text{BodyTemp} \land \text{Malaise})
\]
Classical swine fever: prior probabilities

Faeces

Prim. Other Infection

Reproduction phase

Respiratory problems
Classical swine fever: diagnostic reasoning
Classical swine fever: prognostic reasoning

Evidence
- Type varken: Zeug
- Viremie KVP: Ja

Posterior probabilities (left click for exact values)

- Evidence probability = 0.000000112713600

Koorts
- 22.7% Nee, 77.3% Ja

Leukopenie
- 84.5% Ja, 15.5% Nee

Graphs showing the posterior probabilities for different symptoms.
A Bayesian network: necessary ingredients

**Definition:**
A Bayesian network is a pair $\mathcal{B} = (G, \Gamma)$ such that

- $G$ is an *acyclic directed graph* with nodes representing a set of *random variables* $\mathbf{V}$;
- $\Gamma = \{ \gamma_{V_i} \mid V_i \in \mathbf{V} \}$ is a set of *assessment functions*.

**Property:**

$$
\Pr(\mathbf{V}) = \prod_{V_i \in \mathbf{V}} \gamma_{V_i}(V_i \mid \rho(V_i))
$$

defines a *joint probability distribution* $\Pr$ on $\mathbf{V}$ such that $G$ is a *directed I-map* for the *independence relation* $I_{\Pr}$ of $\Pr$. 
About this course . . .

The following subjects will be addressed in this course:

- the **syntactics** and **semantics** of a Bayesian network;
- algorithms for **reasoning** with a Bayesian network;
- methods for **constructing** a Bayesian network for a domain of application;
- methods for **evaluating** a Bayesian network’s performance and behaviour;
- algorithms for **controlling** reasoning;
Overview of subjects