An Introduction to Bayesian Network Inference using Variable Elimination

Jhonatan Oliveira
Department of Computer Science
University of Regina
Outline

• Introduction
• Background
• Bayesian networks
• Variable Elimination
• Repeated Computation
• Conclusions
Bayesian networks are probabilistic graphical models used when reasoning under uncertainty.
Uncertainty

- Conflicting information
- Missing information
Uncertainty

- **Conflicting** information
- Missing information
Uncertainty

• Conflicting information

• **Missing** information
Real World Applications
Real World Applications

TrueSkill™

XBOX LIVE
Real World Applications

Turbo Codes
Real World Applications

Mars Exploration Rover
Probability theory: introducing joint probability distribution, chain rule, and conditional independence
Joint Probability Distribution

• A multivariate function over a finite set of variables

• Assigns a real number between 0 and 1 to each configuration (combination of variable’s values) of the variables

• Summing all assigned real numbers yields 1
## Joint Probability Distribution

<table>
<thead>
<tr>
<th>Lights On</th>
<th>Family Out</th>
<th>Dog Out</th>
<th>Bowel Problem</th>
<th>Hear Bark</th>
<th>( P(L,F,D,B,H) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Joint Probability Distribution

<table>
<thead>
<tr>
<th>Lights On</th>
<th>Family Out</th>
<th>Dog Out</th>
<th>Bowel Problem</th>
<th>Hear Bark</th>
<th>P(L,F,D,B,H)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.08</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.19</td>
</tr>
</tbody>
</table>

1st Query

2nd Query
## Joint Probability Distribution

<table>
<thead>
<tr>
<th>Lights On</th>
<th>Family Out</th>
<th>Dog Out</th>
<th>Bowel Problem</th>
<th>Hear Bark</th>
<th>$P(...)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.08</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.19</td>
</tr>
</tbody>
</table>

The **size** issue = 32 probabilities
Chain Rule

\[ P(\ldots) = P(L) \cdot P(F|L) \cdot P(D|L,F) \cdot P(B|L,F,D) \cdot P(H|L,F,D,B) \]
Chain Rule

The \textbf{size} issue = 62 probabilities
Conditional Independence

Given:

- family out
- dog out
- dog out
- hear bark
Conditional Independence

Given:

Independence $I(\text{family out, dog out, hear bark})$:
Conditional Independence

- Given $I(X,Y,Z)$:
  - $P(X|Y,Z) = P(X|Y)$

- Given $I(L,F,D)$
  - $P(D|L,F) = P(D|F)$

![Conditional Independence Diagram]
Chain Rule & Conditional Independence

\[ P(L, F, D, B, H) \]

\[ P(L) \ P(F|L) \ \boxed{P(D|L,F)} \ P(B|L,F,D) \ P(H|L,F,D,B) \]

\[ I(D,F,L) \]

\[ P(L) \ P(F|L) \ P(D|F) \ \boxed{P(B|L,D,F)} \ P(H|L,F,D,B) \]

\[ I(B, ,F) \]

\[ P(L) \ P(F|L) \ P(D|F) \ \boxed{P(B|L,D)} \ P(H|L,F,D,B) \]

?
A graphical interpretation of probability theory
Directed Acyclic Graph

- Family out
- Bowel problem
- Dog out
- Hear bark
- Lights on
Testing Independences

A set of variables $X$ is d-separated from a set of variables $Y$ in the DAG if all paths from $X$ to $Y$ are blocked.
Testing Independences

Is $F$ d-separated from $H$ given $D$?

Yes, namely, $I(F,D,H)$ holds in $P(L,F,D,B,H)$
Testing Independences

The size issue = 18 probabilities
Bayesian Network

A directed acyclic graph $\mathbf{B}$ and a set of conditional probability tables $P(U) = P(v | \text{Pa}(v))$, where $v$ is in $\mathbf{B}$ and $\text{Pa}(v)$ are the parents of $v$. 
Bayesian Network

\[ P(L,F,D,B,H) = P(L|F) \cdot P(F) \cdot P(B) \cdot P(D|B,F) \cdot P(H|D) \]
Inference

\[ P(L|F) \]

\[ P(F) \]

\[ P(B) \]

\[ P(D|B,F) \]

\[ P(H|D) \]

\[ P(L,F,D,B,H) \]

\[ \text{part} \]

\[ P(L) \]
Inference

P(H|D)

P(F)

P(B)

P(D|B,F)

P(L|F)

P(L,F)

F

P(L)

X

P(L)
Inference

Multiplication

<table>
<thead>
<tr>
<th>L</th>
<th>F</th>
<th>P(L,F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.8</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>F</th>
<th>P(F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.8</td>
</tr>
<tr>
<td>1</td>
<td>0.3</td>
</tr>
</tbody>
</table>

\[ P(L | F) = \frac{P(L,F)}{P(F)} \]

\[ P(F) \times P(L | F) = P(L,F) \]

<table>
<thead>
<tr>
<th>L</th>
<th>F</th>
<th>P(L,F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.64</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0.09</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.16</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.21</td>
</tr>
</tbody>
</table>
**Inference**

**Marginalization**

<table>
<thead>
<tr>
<th>L</th>
<th>F</th>
<th>P(L,F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.2</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.4</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

\[ P(L) + P(F) = P(L,F) \]

<table>
<thead>
<tr>
<th>L</th>
<th>P(F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
</tr>
</tbody>
</table>
Inference Algorithms

- $P(L|F)$
- $P(F)$
- $P(B)$
- $P(D|B,F)$
- $P(H|D)$

- Shafer-Shennoy
- Lauritzen and Spiegelhalter
- Hugin
- Lazy Propagation
- **Variable Elimination**

$P(L)$
Eliminates all variables that are not in the query
Variable Elimination Algorithm

**Input:** factorization $F$, elimination ordering $L$, query $X$, evidence $Y$

**Output:** $P(X|Y)$

For each variable $v$ in $L$:
- multiply all CPTs in $F$ involving $v$ yielding CPT $P_1$
- marginalize $v$ out of $P_1$
- remove all CPTs from $F$ involving $v$
- append $P_1$ to $F$

Multiply all remaining CPTs in $F$ yielding $P(X,Y)$

**return** $P(X|Y) = P(X,Y) / P(Y)$
Variable Elimination Algorithm

\[ P(H \mid L) = ? \]

\[ P(L,F,D,B,H) = P(L\mid F) \cdot P(F) \cdot P(B) \cdot P(D\mid B,F) \cdot P(H\mid D) \]
Variable Elimination Algorithm

Input

Factorization: \( P(L|F) \ P(F) \ P(B) \ P(D|B,F) \ P(H|D) \)

Query variable: \( H \)

Evidence variable: \( L=1 \)

Elimination ordering: \( B, \ F, \ D \)
Variable Elimination Algorithm

Eliminating B

\[ P(B,D|F) = P(B) P(D|B,F) \]
\[ P(D|F) = \text{marginalize B from } P(B,D|F) \]
Factorization: \( P(L|F) P(F) P(H|D) P(D|F) \)

Eliminating F

\[ P(D,F,L) = P(L|F) P(F) P(D|F) \]
\[ P(D,L) = \text{marginalize F from } P(D,F,L) \]
Factorization: \( P(H|D) P(D,L) \)
Variable Elimination Algorithm

Eliminating D

\[ P(D,H,L) = P(H|D) \ P(D,L) \]
\[ P(H,L) = \text{marginalize D from } P(D,H,L) \]
Factorization: \( P(H,L) \)

Output

\[ P(L) = \text{marginalize H from } P(H,L) \]
\[ P(H|L) = \frac{P(H,L)}{P(L)} \]
Variable Elimination can perform repeated computation
Variable Elimination Algorithm

\[ P(H \mid F) \]

\[ P(L,F,D,B,H) = P(L \mid F) \cdot P(F) \cdot P(B) \cdot P(D \mid B,F) \cdot P(H \mid D) \]
Variable Elimination Algorithm

Input

Factorization: \( P(L|F) \ P(F) \ P(B) \ P(D|B,F) \ P(H|D) \)

Query variable: \( H \)

Evidence variable: \( F=1 \)

Elimination ordering: \( L, B, D \)
Variable Elimination Algorithm

Eliminating L

$1(F) = \text{marginalize } L \text{ from } P(L|F)$

Factorization:

$P(F) \ P(B) \ P(D|B,F) \ P(H|D)$

Eliminating B

$P(B,D|F) = P(B) \ P(D|B,F)$

$P(D|F) = \text{marginalize } B \text{ from } P(B,D|F)$

Factorization:

$P(F) \ P(H|D) \ P(D|F)$
Variable Elimination Algorithm

Eliminating D

\[
P(D, H|F) = P(H|D) \ P(D|F)
\]
\[
P(H|F) = \text{marginalize } D \text{ from } P(D,H|F)
\]
Factorization: \( P(F) \ P(H|F) \)

Multiply all: \( P(F,H) = P(F) \ P(H|F) \)

Output

\[
P(F) = \text{marginalize } H \text{ from } P(F, H)
\]
\[
P(H|F) = \frac{P(F,H)}{P(F)}
\]
Repeated Computation

Eliminating B

\[ P(B, D|F) = P(B) P(D|B,F) \]
\[ P(D|F) = \text{marginalize } B \text{ from } P(B, D|F) \]
Factorization: \( P(L|F) P(F) P(H|D) P(D|F) \)

Eliminating B

\[ P(B, D|F) = P(B) P(D|B,F) \]
\[ P(D|F) = \text{marginalize } B \text{ from } P(B, D|F) \]
Factorization: \( P(F) P(H|D) P(D|F) \)
Repeated Computation

- Store past computation
- Find relevant computation for new query
- Retrieve computation that can be reused
Variable Elimination as a Join Tree

Answering $P(H|L)$
Variable Elimination as a Join Tree

Answering $P(H|F)$
Conclusions

• Bayesian networks are useful probabilistic graphical models

• Inference can be performed by Variable Elimination

• Future work will investigate how to avoid repeated computation during Variable Elimination
References

• Bonaparte Project: http://www.bonaparte-dvi.com/


• Microsoft True Skill: http://research.microsoft.com/en-us/projects/trueskill/


