Bayesian networks I

Advanced Artificial Intelligence: Lecture

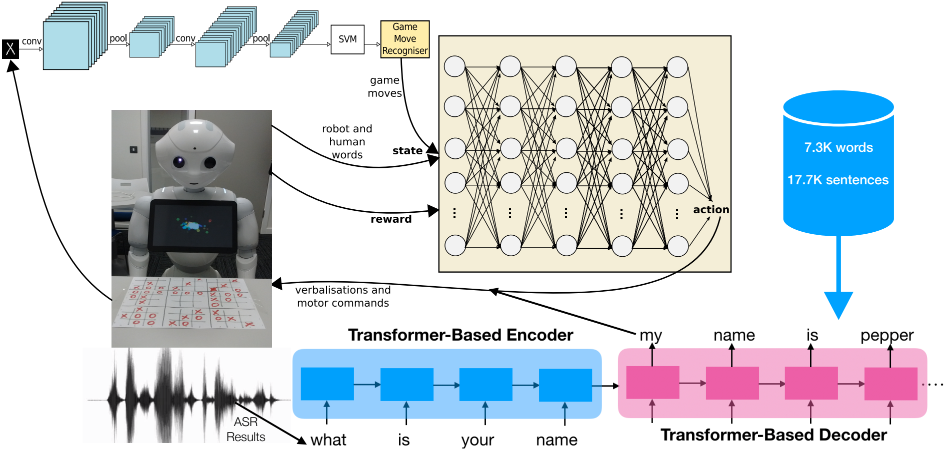
Heriberto Cuayáhuitl
http://staff.lincoln.ac.uk/hcuayahuitl

School of Computer Science, University of Lincoln

9 October 2019

TalkativePepper: A Humanoid Robot with Talkative and Playing Skills

The proposed robot system addresses the topic of natural interaction by running two broad skills in parallel: game playing and chit-chat dialogue.



- Cuayahuitl, H. (2019). A Data-Efficient Deep Learning Approach for Deployable Multimodal Social Robots. In Neurocomputing.
- Cuayahuitl, H. et al. (2019). Ensemble-Based Deep Reinforcement Learning for Chatbots. In Neurocomputing.

Last Week

- Probability Distribution
- Joint Probability
- Conditional Probability
- Prior Probability
- Posterior Probability
- Product Rule
- Sum Rule
- Chain Rule
- Marginalisation
- Normalisation
- Inference
- Bayes Rule
- Probabilistic Model

This Week[1]

- Introduction to Bayesian Networks
- Exact Probabilistic Inference
  - Inference by Enumeration
  - Inference by Variable Elimination

Reading of last week: Russell & Norvig, (2013). Artificial Intelligence: A Modern Approach, **chapter 13**

Reading of this week: Russell & Norvig, (2013). Artificial Intelligence: A Modern Approach, **chapter 14** (til Section 14.4.2)

[1] Slides adapted from the original version of Russell & Norvig

- Bayesian Networks (or Belief Nets) can represent any full joint probability distribution—and they can do so very concisely!

Bayesian Networks

- Bayesian Networks (or Belief Nets) can represent any full joint probability distribution—and they can do so very concisely!

- Syntax:
  - a set of nodes, one per variable
  - a directed, acyclic graph (link $\approx$ "directly influences")
  - a conditional distribution for each node given its parents: $\mathbf{P}(X_i | \mathit{Parents}(X_i))$

Bayesian Networks

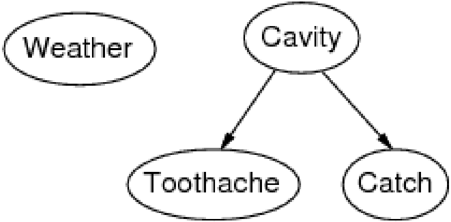
- Bayesian Networks (or Belief Nets) can represent any full joint probability distribution—and they can do so very concisely!

- Syntax:
  - a set of nodes, one per variable
  - a directed, acyclic graph (link $\approx$ "directly influences")
  - a conditional distribution for each node given its parents: $\mathbf{P}(X_i | Parents(X_i))$

- A conditional distribution is represented as a **conditional probability table** (CPT)—a probability distribution over X_i for each combination of parent values.

Example Bayesian Net

The topology of a network[2] encodes conditional independence assertions:



- *Weather* is independent of the other variables
- *Toothache* and *Catch* are conditionally independent given *Cavity*

[2] Cooper, G. and Herskovits, E. A Bayesian method for the induction of probabilistic networks from data. Machine Learning, 9, 1992. `http://www.springerlink.com/content/85c6f40ef659d8b2/fulltext.pdf`.

Example Scenario: Burglary (1/2)

I am at work, my neighbour John calls to say my alarm is ringing, and my neighbour Mary doesn't call. Sometimes the alarm is set off by minor earthquakes. Is there a burglar?

Example Scenario: Burglary (1/2)

I am at work, my neighbour John calls to say my alarm is ringing, and my neighbour Mary doesn't call. Sometimes the alarm is set off by minor earthquakes. Is there a burglar?

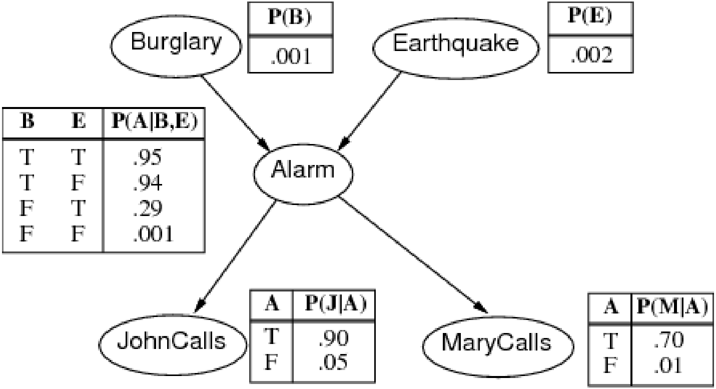
- Random Variables:
  - B: *Burglar*,
  - E: *Earthquake*,
  - A: *Alarm*,
  - J: *JohnCalls*,
  - M: *MaryCalls*

Example Scenario: Burglary (1/2)

I am at work, my neighbour John calls to say my alarm is ringing, and my neighbour Mary doesn't call. Sometimes the alarm is set off by minor earthquakes. Is there a burglar?

- Random Variables:
  - B: *Burglar*,
  - E: *Earthquake*,
  - A: *Alarm*,
  - J: *JohnCalls*,
  - M: *MaryCalls*
- The network topology reflects "causal" knowledge:
  - A burglar can set the alarm on
  - An earthquake can set the alarm on
  - The alarm can cause Mary to call
  - The alarm can cause John to call

Example Scenario: Burglary (2/2)



P(B)
.001

P(E)
.002

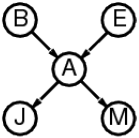
B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

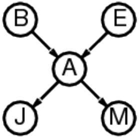
Compactness

- A CPT for Boolean variable X_i with k Boolean parents has 2^k rows for the combinations of parent values



Compactness

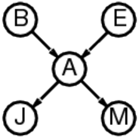
- A CPT for Boolean variable X_i with k Boolean parents has 2^k rows for the combinations of parent values



- Each row requires one number p for $X_i = true$, and one number for $X_i = false$ (i.e. $\neg p = 1 - p$). For the burglary net, $1 + 1 + 4 + 2 + 2 = 10$ numbers (instead of 2^5)

Compactness

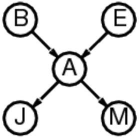
- A CPT for Boolean variable X_i with k Boolean parents has 2^k rows for the combinations of parent values



- Each row requires one number p for $X_i = true$, and one number for $X_i = false$ (i.e. $\neg p = 1 - p$). For the burglary net, $1 + 1 + 4 + 2 + 2 = 10$ numbers (instead of 2^5)
- If each variable has no more than k parents, the complete network requires $n \cdot 2^k$ numbers

Compactness

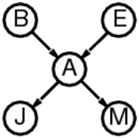
- A CPT for Boolean variable X_i with k Boolean parents has 2^k rows for the combinations of parent values



- Each row requires one number p for $X_i = \mathit{true}$, and one number for $X_i = \mathit{false}$ (i.e. $\neg p = 1 - p$). For the burglary net, $1 + 1 + 4 + 2 + 2 = 10$ numbers (instead of 2^5)
- If each variable has no more than k parents, the complete network requires $n \cdot 2^k$ numbers
- What is the number of probabilities in a Bayesian Network with 30 random variables, each with 5 parents? ... What's the number of probabilities in the full joint distribution?

Global semantics

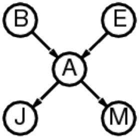
- "Global" semantics defines the full joint distribution as the product of the local conditional distributions:



$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | Parents(X_i))$$

Global semantics

- "Global" semantics defines the full joint distribution as the product of the local conditional distributions:



$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | Parents(X_i))$$

- For example:

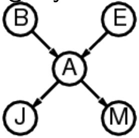
$$\begin{aligned} P(j \wedge m \wedge a \wedge \neg b \wedge \neg e) &= P(j|a)P(m|a)P(a|\neg b, \neg e)P(\neg b)P(\neg e) \\ &= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 \\ &\approx 0.00063 \end{aligned}$$

Inference by Enumeration (1/3)

- It sums out variables from the joint without actually constructing its explicit representation

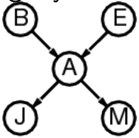
Inference by Enumeration (1/3)

- It sums out variables from the joint without actually constructing its explicit representation
- Simple query on the burglary network:



Inference by Enumeration (1/3)

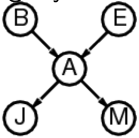
- It sums out variables from the joint without actually constructing its explicit representation
- Simple query on the burglary network:



- $\mathbf{P}(B|j,m) = \frac{\mathbf{P}(B,j,m)}{P(j,m)}$
  $= \alpha \mathbf{P}(B,j,m)$
  $= \alpha \sum_e \sum_a \mathbf{P}(B,e,a,j,m)$

Inference by Enumeration (1/3)

- It sums out variables from the joint without actually constructing its explicit representation
- Simple query on the burglary network:



- $\mathbf{P}(B|j,m) = \frac{\mathbf{P}(B,j,m)}{P(j,m)}$
  $= \alpha \mathbf{P}(B,j,m)$
  $= \alpha \sum_e \sum_a \mathbf{P}(B,e,a,j,m)$
- Rewrite full joint entries using product of CPT entries:
  $\mathbf{P}(B|j,m) = \alpha \sum_e \sum_a \mathbf{P}(B) P(e) \mathbf{P}(a|B,e) P(j|a) P(m|a)$
  $= \alpha \mathbf{P}(B) \sum_e P(e) \sum_a \mathbf{P}(a|B,e) P(j|a) P(m|a)$

Inference by Enumeration (2/3)

- $$\begin{aligned}\mathbf{P}(B|j,m) &= \alpha \ \textstyle\sum_e \ \sum_a \ \mathbf{P}(B)P(e)\mathbf{P}(a|B,e)P(j|a)P(m|a) \\ &= \alpha\mathbf{P}(B) \ \textstyle\sum_e \ P(e) \ \sum_a \ \mathbf{P}(a|B,e)P(j|a)P(m|a) \\ &= \alpha < \mathrm{P}(b|j,m), \mathrm{P}(\neg b|j,m) >\end{aligned}$$

Inference by Enumeration (2/3)

- $$\begin{aligned}\mathbf{P}(B|j,m) &= \alpha \ \sum_e \ \sum_a \ \mathbf{P}(B)P(e)\mathbf{P}(a|B,e)P(j|a)P(m|a) \\ &= \alpha\mathbf{P}(B) \ \sum_e \ P(e) \ \sum_a \ \mathbf{P}(a|B,e)P(j|a)P(m|a) \\ &= \alpha < \mathrm{P}(b|j,m), \mathrm{P}(\neg b|j,m) >\end{aligned}$$

- $\mathrm{P}(b|j,m) = \alpha\mathrm{P}(b) \ \sum_e \ P(e) \ \sum_a \ \mathrm{P}(a|b,e)P(j|a)P(m|a) =$
  $\alpha\mathrm{P}(b) \ \sum_e P(e) \, [\mathrm{P}(a|b,e)P(j|a)P(m|a) + \mathrm{P}(\neg a|b,e)P(j|\neg a)P(m|\neg a$

Inference by Enumeration (2/3)

- $\mathbf{P}(B|j,m) = \alpha\ \sum_e\ \sum_a\ \mathbf{P}(B)P(e)\mathbf{P}(a|B,e)P(j|a)P(m|a)$
  $= \alpha\mathbf{P}(B)\ \sum_e\ P(e)\ \sum_a\ \mathbf{P}(a|B,e)P(j|a)P(m|a)$
  $= \alpha < \mathrm{P}(b|j,m), \mathrm{P}(\neg b|j,m) >$

- $\mathrm{P}(b|j,m) = \alpha\mathrm{P}(b)\ \sum_e\ P(e)\ \sum_a\ \mathrm{P}(a|b,e)P(j|a)P(m|a) =$
  $\alpha\mathrm{P}(b)\ \sum_e P(e)\,[\mathrm{P}(a|b,e)P(j|a)P(m|a) + \mathrm{P}(\neg a|b,e)P(j|\neg a)P(m|\neg a$

- $=$
  $\alpha\mathrm{P}(b)\ P(e)\,[\mathrm{P}(a|b,e)P(j|a)P(m|a) + \mathrm{P}(\neg a|b,e)P(j|\neg a)P(m|\neg a)] +$
  $P(\neg e)\,[\mathrm{P}(a|b,\neg e)P(j|a)P(m|a) + \mathrm{P}(\neg a|b,\neg e)P(j|\neg a)P(m|\neg a)]$

Inference by Enumeration (2/3)

- $$\mathbf{P}(B|j,m) = \alpha \sum_e \sum_a \mathbf{P}(B)P(e)\mathbf{P}(a|B,e)P(j|a)P(m|a)$$
  $$= \alpha\mathbf{P}(B) \sum_e P(e) \sum_a \mathbf{P}(a|B,e)P(j|a)P(m|a)$$
  $$= \alpha < \mathrm{P}(b|j,m), \mathrm{P}(\neg b|j,m) >$$
- $\mathrm{P}(b|j,m) = \alpha\mathrm{P}(b) \sum_e P(e) \sum_a \mathrm{P}(a|b,e)P(j|a)P(m|a) =$
  $\alpha\mathrm{P}(b) \sum_e P(e) [\mathrm{P}(a|b,e)P(j|a)P(m|a) + \mathrm{P}(\neg a|b,e)P(j|\neg a)P(m|\neg a$
- $=$
  $\alpha\mathrm{P}(b)\, P(e) [\mathrm{P}(a|b,e)P(j|a)P(m|a) + \mathrm{P}(\neg a|b,e)P(j|\neg a)P(m|\neg a)] +$
  $P(\neg e) [\mathrm{P}(a|b,\neg e)P(j|a)P(m|a) + \mathrm{P}(\neg a|b,\neg e)P(j|\neg a)P(m|\neg a)]$
- $= \alpha[0.001 \times [0.002 \times [0.095 \times 0.9 \times 0.7 + 0.05 \times 0.05 \times 0.01] +$
  $[0.998 \times [0.94 \times 0.9 \times 0.7 + 0.06 \times 0.05 \times 0.01]]] = \alpha[0.001 \times$
  $[0.002 \times [0.5985 + 0.000025] + 0.998 \times [0.5922 + 0.00003]]]$
  $= \alpha[0.001 \times [0.001197 + 0.591045]] = \alpha 0.000592243$

Inference by Enumeration (3/3)

$\mathrm{P}(\neg b|j,m) = \alpha \mathrm{P}(\neg b)\ \sum_e\ P(e)\ \sum_a\ \mathrm{P}(a|\neg b, e) P(j|a) P(m|a) =$
$\alpha \mathrm{P}(\neg b)\ \sum_e P(e)\,[\mathrm{P}(a|\neg b, e) P(j|a) P(m|a) + \mathrm{P}(\neg a|\neg b, e) P(j|\neg a) P(m|\neg a$
$=$
$\alpha \mathrm{P}(\neg b)\, P(e)\,[\mathrm{P}(a|\neg b, e) P(j|a) P(m|a) + \mathrm{P}(\neg a|\neg b, e) P(j|\neg a) P(m|\neg a)] +$
$P(\neg e)\,[\mathrm{P}(a|\neg b, \neg e) P(j|a) P(m|a) + \mathrm{P}(\neg a|\neg b, \neg e) P(j|\neg a) P(m|\neg a)]$
$= \alpha[0.999 \times [0.002 \times [0.29 \times 0.9 \times 0.7 + 0.71 \times 0.05 \times 0.01] +$
$[0.998 \times [0.001 \times 0.9 \times 0.7 + 0.999 \times 0.05 \times 0.01]]]$
$= \alpha[0.999 \times [0.002 \times [0.1827 + 0.0000355] + 0.998 \times [0.00063 +$
$0.00004995]]]$
$= \alpha[0.999 \times [0.00036611 + 0.001127241]] = \alpha 0.001491858$

$$
\begin{aligned}
\mathbf{P}(B|j,m) &= \alpha < P(b|j,m), P(\neg b|j,m) > \\
&= \alpha < 0.000592243, 0.001491858 > \\
&= < 0.2842, 0.7158 >
\end{aligned}
$$

Enumeration Algorithm

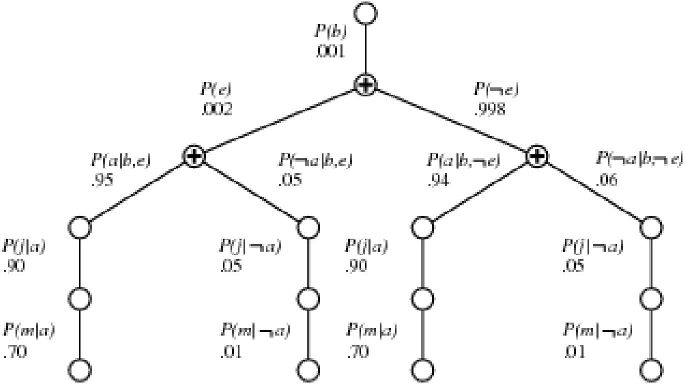
function ENUMERATION-ASK($X, \mathbf{e}, bn$) **returns** a distribution over X
  inputs: X, the query variable
    $\mathbf{e}$, observed values for variables $\mathbf{E}$
    bn, a Bayesian network with variables $\{X\} \cup \mathbf{E} \cup \mathbf{Y}$

  $\mathbf{Q}(X) \leftarrow$ a distribution over X, initially empty
  for each value x_i of X **do**
    extend $\mathbf{e}$ with value x_i for X
    $\mathbf{Q}(x_i) \leftarrow$ ENUMERATE-ALL(VARS[bn], $\mathbf{e}$)
  return NORMALIZE($\mathbf{Q}(X)$)

---

function ENUMERATE-ALL($vars, \mathbf{e}$) **returns** a real number
  if EMPTY?($vars$) **then return** 1.0
  $Y \leftarrow$ FIRST($vars$)
  if Y has value y in $\mathbf{e}$
    then return $P(y \mid Pa(Y)) \times$ ENUMERATE-ALL(REST($vars$), $\mathbf{e}$)
    else return $\Sigma_y \; P(y \mid Pa(Y)) \times$ ENUMERATE-ALL(REST($vars$), $\mathbf{e}_y$)
      where $\mathbf{e}_y$ is $\mathbf{e}$ extended with $Y = y$

Evaluation Tree



Enumeration is inefficient due to repeated computation, e.g. the tree above computes $P(j|a)P(m|a)$ for each value of E ...

Inference by Variable Elimination (1/2)

- Idea: do the calculation once and save the results for later use

Inference by Variable Elimination (1/2)

- Idea: do the calculation once and save the results for later use
- Variable elimination evaluates expressions in right-to-left order, and uses factors $\mathbf{f}_i$ (matrices) as follows:

$$\mathbf{P}(B \mid j, m) = \alpha \underbrace{\mathbf{P}(B)}_{\mathbf{f}_1(B)} \sum_e \underbrace{P(e)}_{\mathbf{f}_2(E)} \sum_a \underbrace{\mathbf{P}(a \mid B, e)}_{\mathbf{f}_3(A,B,E)} \underbrace{P(j \mid a)}_{\mathbf{f}_4(A)} \underbrace{P(m \mid a)}_{\mathbf{f}_5(A)}$$

Inference by Variable Elimination (1/2)

- Idea: do the calculation once and save the results for later use
- Variable elimination evaluates expressions in right-to-left order, and uses factors $\mathbf{f}_i$ (matrices) as follows:

$$\mathbf{P}(B \mid j, m) = \alpha \underbrace{\mathbf{P}(B)}_{\mathbf{f}_1(B)} \sum_e \underbrace{P(e)}_{\mathbf{f}_2(E)} \sum_a \underbrace{\mathbf{P}(a \mid B, e)}_{\mathbf{f}_3(A,B,E)} \underbrace{P(j \mid a)}_{\mathbf{f}_4(A)} \underbrace{P(m \mid a)}_{\mathbf{f}_5(A)}$$

- $\mathbf{f}_4(A) =< P(j|a), P(j|\neg a) >=< 0.90, 0.05 >$
  $\mathbf{f}_5(A) =< P(m|a), P(m|\neg a) >=< 0.70, 0.01 >$

Inference by Variable Elimination (1/2)

- Idea: do the calculation once and save the results for later use
- Variable elimination evaluates expressions in right-to-left order, and uses factors $\mathbf{f}_i$ (matrices) as follows:

$$\mathbf{P}(B \mid j, m) = \alpha \underbrace{\mathbf{P}(B)}_{\mathbf{f}_1(B)} \sum_e \underbrace{P(e)}_{\mathbf{f}_2(E)} \sum_a \underbrace{\mathbf{P}(a \mid B, e)}_{\mathbf{f}_3(A,B,E)} \underbrace{P(j \mid a)}_{\mathbf{f}_4(A)} \underbrace{P(m \mid a)}_{\mathbf{f}_5(A)}$$

- $\mathbf{f}_4(A) =< P(j|a), P(j|\neg a) >=< 0.90, 0.05 >$
  $\mathbf{f}_5(A) =< P(m|a), P(m|\neg a) >=< 0.70, 0.01 >$
- Rewriting the query equation above we get:
  $\mathbf{P}(B|j,m) = \alpha \mathbf{f}_1(B) \times \sum_e \mathbf{f}_2(E) \sum_a \mathbf{f}_3(A,B,E) \times \mathbf{f}_4(A) \times \mathbf{f}_5(A)$,
  where "$\times$" denotes a pointwise product operation

Inference by Variable Elimination (2/2)

- Summing out A from the product of $\mathbf{f}_3$, $\mathbf{f}_4$, and $\mathbf{f}_5$ we get:

$$\begin{aligned}\mathbf{f}_6(B,E) &= \textstyle\sum_a \mathbf{f}_3(A,B,E) \times \mathbf{f}_4(A) \times \mathbf{f}_5(A) \\ &= (\mathbf{f}_3(a,B,E) \times \mathbf{f}_4(a) \times \mathbf{f}_5(a)) + \\ &\quad\ (\mathbf{f}_3(\neg a,B,E) \times \mathbf{f}_4(\neg a) \times \mathbf{f}_5(\neg a))\end{aligned}$$

Inference by Variable Elimination (2/2)

- Summing out A from the product of $\mathbf{f}_3$, $\mathbf{f}_4$, and $\mathbf{f}_5$ we get:

$$\begin{aligned}\mathbf{f}_6(B,E) &= \textstyle\sum_a \mathbf{f}_3(A,B,E) \times \mathbf{f}_4(A) \times \mathbf{f}_5(A) \\ &= (\mathbf{f}_3(a,B,E) \times \mathbf{f}_4(a) \times \mathbf{f}_5(a)) + \\ &\quad\ (\mathbf{f}_3(\neg a,B,E) \times \mathbf{f}_4(\neg a) \times \mathbf{f}_5(\neg a))\end{aligned}$$

- Therefore $\mathbf{P}(B|j,m) = \alpha \mathbf{f}_1(B) \times \sum_e \mathbf{f}_2(E) \times \mathbf{f}_6(B,E)$

Inference by Variable Elimination (2/2)

- Summing out A from the product of $\mathbf{f}_3$, $\mathbf{f}_4$, and $\mathbf{f}_5$ we get:

$$\begin{aligned}\mathbf{f}_6(B,E) &= \textstyle\sum_a \mathbf{f}_3(A,B,E) \times \mathbf{f}_4(A) \times \mathbf{f}_5(A)\\ &= (\mathbf{f}_3(a,B,E) \times \mathbf{f}_4(a) \times \mathbf{f}_5(a)) + \\ &\quad\ (\mathbf{f}_3(\neg a,B,E) \times \mathbf{f}_4(\neg a) \times \mathbf{f}_5(\neg a))\end{aligned}$$

- Therefore $\mathbf{P}(B|j,m) = \alpha \mathbf{f}_1(B) \times \sum_e \mathbf{f}_2(E) \times \mathbf{f}_6(B,E)$

- Summing out E from the product of $\mathbf{f}_2$ and $\mathbf{f}_6$ we get:

$$\begin{aligned}\mathbf{f}_7(B) &= \textstyle\sum_e \mathbf{f}_2(E) \times \mathbf{f}_6(B,E)\\ &= (\mathbf{f}_2(e) \times \mathbf{f}_6(B,e)) + (\mathbf{f}_2(\neg e) \times \mathbf{f}_6(B,\neg e))\end{aligned}$$

$$\mathbf{P}(B|j,m) = \alpha \mathbf{f}_1(B) \times \mathbf{f}_7(B)$$

Inference by Variable Elimination (2/2)

- Summing out A from the product of $\mathbf{f}_3$, $\mathbf{f}_4$, and $\mathbf{f}_5$ we get:

$$\begin{aligned}\mathbf{f}_6(B,E) &= \textstyle\sum_a \mathbf{f}_3(A,B,E)\times\mathbf{f}_4(A)\times\mathbf{f}_5(A)\\ &= (\mathbf{f}_3(a,B,E)\times\mathbf{f}_4(a)\times\mathbf{f}_5(a)) +\\ &\quad\ (\mathbf{f}_3(\neg a,B,E)\times\mathbf{f}_4(\neg a)\times\mathbf{f}_5(\neg a))\end{aligned}$$

- Therefore $\mathbf{P}(B|j,m) = \alpha\mathbf{f}_1(B)\times\sum_e \mathbf{f}_2(E)\times\mathbf{f}_6(B,E)$

- Summing out E from the product of $\mathbf{f}_2$ and $\mathbf{f}_6$ we get:

$$\begin{aligned}\mathbf{f}_7(B) &= \textstyle\sum_e \mathbf{f}_2(E)\times\mathbf{f}_6(B,E)\\ &= (\mathbf{f}_2(e)\times\mathbf{f}_6(B,e)) + (\mathbf{f}_2(\neg e)\times\mathbf{f}_6(B,\neg e))\end{aligned}$$

$$\mathbf{P}(B|j,m) = \alpha\mathbf{f}_1(B)\times\mathbf{f}_7(B)$$

- We only need to know how to do operations on factors!

Operations on Factors

- Pointwise product:

A	B	$\mathbf{f}_1(A,B)$	B	C	$\mathbf{f}_2(B,C)$	A	B	C	$\mathbf{f}_3(A,B,C)$
T	T	.3	T	T	.2	T	T	T	$.3 \times .2 = .06$
T	F	.7	T	F	.8	T	T	F	$.3 \times .8 = .24$
F	T	.9	F	T	.6	T	F	T	$.7 \times .6 = .42$
F	F	.1	F	F	.4	T	F	F	$.7 \times .4 = .28$
						F	T	T	$.9 \times .2 = .18$
						F	T	F	$.9 \times .8 = .72$
						F	F	T	$.1 \times .6 = .06$
						F	F	F	$.1 \times .4 = .04$

Figure 14.10 Illustrating pointwise multiplication: $\mathbf{f}_1(A,B) \times \mathbf{f}_2(B,C) = \mathbf{f}_3(A,B,C)$.

Operations on Factors

- Pointwise product:

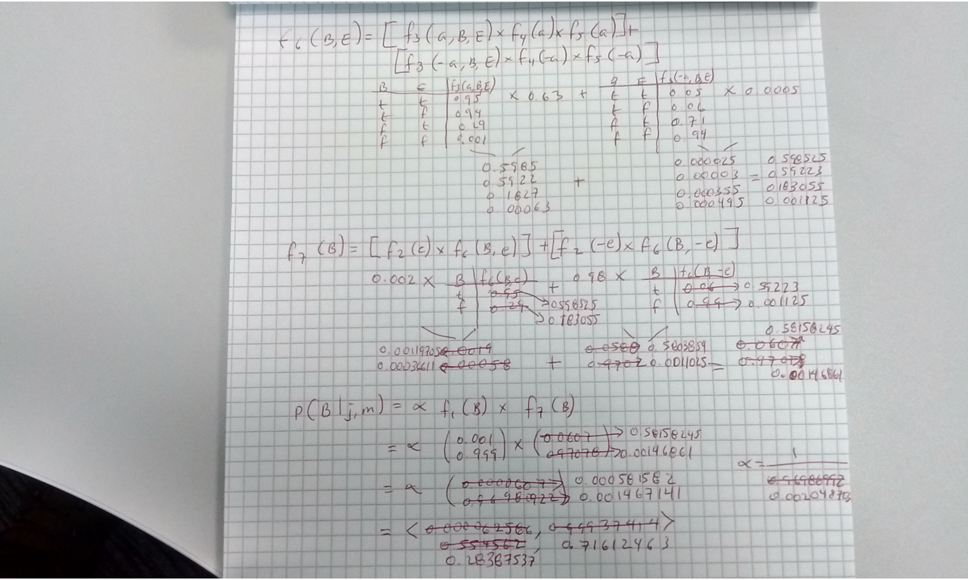
A	B	$\mathbf{f}_1(A,B)$	B	C	$\mathbf{f}_2(B,C)$	A	B	C	$\mathbf{f}_3(A,B,C)$
T	T	.3	T	T	.2	T	T	T	$.3 \times .2 = .06$
T	F	.7	T	F	.8	T	T	F	$.3 \times .8 = .24$
F	T	.9	F	T	.6	T	F	T	$.7 \times .6 = .42$
F	F	.1	F	F	.4	T	F	F	$.7 \times .4 = .28$
						F	T	T	$.9 \times .2 = .18$
						F	T	F	$.9 \times .8 = .72$
						F	F	T	$.1 \times .6 = .06$
						F	F	F	$.1 \times .4 = .04$

Figure 14.10 Illustrating pointwise multiplication: $\mathbf{f}_1(A,B) \times \mathbf{f}_2(B,C) = \mathbf{f}_3(A,B,C)$.

- Summing out a variable:

$$\begin{aligned}\mathbf{f}(B,C) &= \sum_a \mathbf{f}_3(A,B,C) = \mathbf{f}_3(a,B,C) + \mathbf{f}_3(\neg a,B,C) \\ &= \begin{pmatrix} .06 & .24 \\ .42 & .28 \end{pmatrix} + \begin{pmatrix} .18 & .72 \\ .06 & .04 \end{pmatrix} = \begin{pmatrix} .24 & .96 \\ .48 & .32 \end{pmatrix} .\end{aligned}$$

Inference by Variable Elimination: Example



The Variable Elimination Algorithm

function ELIMINATION-ASK(X, **e**, bn) **returns** a distribution over X
  inputs: X, the query variable
    e, observed values for variables **E**
    bn, a Bayesian network specifying joint distribution $\mathbf{P}(X_1, \ldots, X_n)$

  $factors \leftarrow [\,]$
  for each var **in** ORDER(bn.VARS) **do**
    $factors \leftarrow$ [MAKE-FACTOR(var, **e**)|$factors$]
    if var is a hidden variable **then** $factors \leftarrow$ SUM-OUT(var, $factors$)
  return NORMALIZE(POINTWISE-PRODUCT($factors$))

Keywords

- **Bayesian Network**
- **Directed Acyclic Graph (DAG)**
- **Network Topology (Network Structure)**
- **Conditional Probability Table (CPT)**
- **Independence and Conditional Independence**
- **Query/Evidence/Hidden Random Variables**
- **Inference by Enumeration**
- **Inference by Variable Elimination**
- **Factors**
- **Pointwise Product**
- **Summing Out (Marginalisation)**
- **Normalisation**

Next

Workshop:

- Excercises about Bayesian networks, Q&A
- Wednesday from 10:30-12hrs, room=INB1301

Reading of this week: Russell & Norvig, (2013). Artificial Intelligence: A Modern Approach, **chapter 14** (til Section 14.4.2)

Reading of next week: Russell & Norvig, (2013). Artificial Intelligence: A Modern Approach, **chapter 14** (continuation)