Probabilistic Graphical Models for Diagnosis and Decision Making

EURANDOM, Eindhoven, August 2009

Section 0

Background

References

I will not give references of credit. Below some references for study.

- Judea Pearl, *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann Publishers, 1988.
- Steffen L. Lauritzen, *Graphical Models*, Oxford University Press, 1996.
- Enrique Castillo, José M. Gutiérrez, and Ali S. Hadi, *Expert Systems and Probabilistic Network Models*, Springer-Verlag, 1997.
- Robert G. Cowell, A. Philip Dawid, and Steffen L. Lauritzen, *Probabilistic Networks and Expert Systems*, Springer-Verlag, 1999.
- Kevin B. Korb and Ann E. Nicholson, *Bayesian Artificial Intelligence*, Chapman & Hall 2004.
- Richard E. Neapolitan, *Learning Bayesian Networks*, Pearson Prentice Hall, 2004.
- Finn V. Jensen and Thomas D. Nielsen *Bayesian Networks and Decision Graphs*, 2nd edition, Springer-Verlag, 2007.

Belief updating

A very basic task in artificial intelligence is *belief updating*. In the framework of probability theory it says:

- Hypothesis variable A with prior distribution $P(A)$.
- Case specific knowledge: $e = \{B = \beta, \ldots, C = \gamma\}$
- What is $P(A|e)$?

In AI, the approach to probabilities is that of subjective probabilities.

Conditional probabilities

Every probability is conditioned on a context. For example, if we throw a dice:

"$P(\{\text{six}\}) = \frac{1}{6}$" = "$P(\text{six}|\text{symmetric dice}) = \frac{1}{6}$".

In general, if $\mathcal{A}$ and $\mathcal{B}$ are events and $P(\mathcal{A}|\mathcal{B}) = x$, then:

"In the context of $\mathcal{B}$ we have that $P(\mathcal{A}) = x$"

Note: It is not "whenever $\mathcal{B}$ we have $P(\mathcal{A}) = x$", but rather: if $\mathcal{B}$, and everything else known is irrelevant to $\mathcal{A}$, then $P(\mathcal{A}) = x$.

Basic probability calculus

Let $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ be events.

The fundamental rule: $P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A}|\mathcal{B})P(\mathcal{B})$.

The conditioned fundamental rule: $P(\mathcal{A} \cap \mathcal{B}|\mathcal{C}) = P(\mathcal{A}|\mathcal{B} \cap \mathcal{C})P(\mathcal{B}|\mathcal{C})$.

Basic probability calculus

Let $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ be events.

The fundamental rule: $P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A}|\mathcal{B})P(\mathcal{B})$.

The conditioned fundamental rule: $P(\mathcal{A} \cap \mathcal{B}|\mathcal{C}) = P(\mathcal{A}|\mathcal{B} \cap \mathcal{C})P(\mathcal{B}|\mathcal{C})$.

Conditional independence: $P(\mathcal{A}|\mathcal{B} \cap \mathcal{C}) = P(\mathcal{A}|\mathcal{C})$

In that case $P(\mathcal{A} \cap \mathcal{B}|\mathcal{C}) = P(\mathcal{A}|\mathcal{C}) \cdot P(\mathcal{B}|\mathcal{C})$.

Probability calculus for variables

A is a variable with states $a_1, \ldots, a_n$; B is a variable with states $b_1, \ldots, b_m$.

$P(A) = (x_1, \ldots, x_n)$ is a probability distribution ; $x_i \geq 0; \sum_{i=1}^{n} x_i = 1$.

$P(A|B)$ is a $n \times m$ table containing the numbers $P(a_i|b_j)$.

Note: $\sum_A P(A|b_j) = 1$ for all b_j.

		B		
		b_1	b_2	b_3
A	a_1	0.4	0.3	0.6
	a_2	0.6	0.7	0.4

$P(A, B)$ is a $n \times m$ table too; $\sum_{A,B} P(A, B) = 1$.

		B		
		b_1	b_2	b_3
A	a_1	0.16	0.12	0.12
	a_2	0.24	0.28	0.08

The fundamental rule for variables

$P(A|B)P(B)$: $n \times m$ multiplications $P(a_i|b_j)P(b_j) = P(a_i, b_j)$

	b_1	b_2	b_3
a_1	0.4	0.3	0.6
a_2	0.6	0.7	0.4

$P(A|B)$

b_1	b_2	b_3
0.4	0.4	0.2

$P(B)$

=

	b_1	b_2	b_3
a_1	0.16	0.12	0.12
a_2	0.24	0.28	0.08

$P(A, B)$

Notation

A potential ϕ is a real-valued non-negative function over a set of variables, dom(ϕ).

A table of probabilities is a probability potential.

Notation

A potential ϕ is a real-valued non-negative function over a set of variables, dom(ϕ).

A table of probabilities is a probability potential.

Multiplication

	b_1	b_2
a_1	1	3
a_2	4	5

	b_1	b_2
c_1	6	7
c_2	8	9

=

	b_1	b_2
a_1	$(\_,\_)$	$(\_,\_)$
a_2	$(\_,\_)$	$(\_,\_)$

$\phi_1(A, B)$ $\phi_2(C, B)$ $\phi_3(A, B, C) = \phi_1(A, B) \cdot \phi_2(C, B)$

Notation

A potential ϕ is a real-valued non-negative function over a set of variables, dom(ϕ).

A table of probabilities is a probability potential.

Multiplication

	b_1	b_2
a_1	1	3
a_2	4	5

	b_1	b_2
c_1	6	7
c_2	8	9

=

	b_1	b_2
a_1	$(6_{c_1}, 8_{c_2})$	$(\_, \_)$
a_2	$(\_, \_)$	$(\_, \_)$

$\phi_1(A, B)$ $\phi_2(C, B)$ $\phi_3(A, B, C) = \phi_1(A, B) \cdot \phi_2(C, B)$

Notation

A potential ϕ is a real-valued non-negative function over a set of variables, dom(ϕ).

A table of probabilities is a probability potential.

Multiplication

	b_1	b_2
a_1	1	3
a_2	4	5

	b_1	b_2
c_1	6	7
c_2	8	9

=

	b_1	b_2
a_1	$(6_{c_1}, 8_{c_2})$	$(21_{c_1}, 27_{c_2})$
a_2	$(\_, \_)$	$(\_, \_)$

$\phi_1(A, B)$ $\qquad$ $\phi_2(C, B)$ $\qquad$ $\phi_3(A, B, C) = \phi_1(A, B) \cdot \phi_2(C, B)$

Notation

A potential ϕ is a real-valued non-negative function over a set of variables, dom(ϕ).

A table of probabilities is a probability potential.

Multiplication

	b_1	b_2
a_1	1	3
a_2	4	5

	b_1	b_2
c_1	6	7
c_2	8	9

=

	b_1	b_2
a_1	$(6_{c_1}, 8_{c_2})$	$(21_{c_1}, 27_{c_2})$
a_2	$(24_{c_1}, 32_{c_2})$	$(\_, \_)$

$\phi_1(A, B)$ $\qquad$ $\phi_2(C, B)$ $\qquad$ $\phi_3(A, B, C) = \phi_1(A, B) \cdot \phi_2(C, B)$

Notation

A potential ϕ is a real-valued non-negative function over a set of variables, dom(ϕ).

A table of probabilities is a probability potential.

Multiplication

	b_1	b_2
a_1	1	3
a_2	4	5

	b_1	b_2
c_1	6	7
c_2	8	9

=

	b_1	b_2
a_1	$(6_{c_1}, 8_{c_2})$	$(21_{c_1}, 27_{c_2})$
a_2	$(24_{c_1}, 32_{c_2})$	$(35_{c_1}, 45_{c_2})$

$\phi_1(A, B)$ $\quad$ $\phi_2(C, B)$ $\quad$ $\phi_3(A, B, C) = \phi_1(A, B) \cdot \phi_2(C, B)$

Marginalization of potentials

$$\sum_{B} \left(\begin{array}{c|cc} & b_1 & b_2 \\ \hline a_1 & 2 & 3 \\ a_2 & 1 & 4 \end{array} \right) = \begin{array}{c|c} a_1 & 5 \\ a_2 & 5 \end{array}$$

$$\sum_{A} \left(\begin{array}{c|cc} & b_1 & b_2 \\ \hline a_1 & 2 & 3 \\ a_2 & 1 & 4 \end{array} \right) = \begin{array}{c|c} b_1 & 3 \\ b_2 & 7 \end{array}$$

The algebra of potentials

(i) dom($\phi_1\phi_2$) =dom$\phi_1 \cup dom\phi_2$.

(ii) **The commutative law**: $\phi_1\phi_2 = \phi_2\phi_1$.

(iii) **The associative law**: $(\phi_1\phi_2)\phi_3 = \phi_1(\phi_2\phi_3)$.

(iv) **Existence of unit**: The number 1 is a potential over the empty domain, and $1 \cdot \phi = \phi$ for all potentials ϕ. The unit potential is denoted $\mathbf{1}$.

$\sum_A \phi$ is a potential over dom$(\phi) \setminus \{A\}$.

(v) $\sum_A \sum_B \phi = \sum_B \sum_A \phi$.

(vi) **The unit potential property**: $\sum_A P(A \mid V) = \mathbf{1}$.

(vii) **The distributive law**: If $A \notin$ dom(ϕ_1), then $\sum_A \phi_1\phi_2 = \phi_1 \sum_A \phi_2$.

A simple example of belief updating

The universe $\mathcal{U} = \{A, B, C\}$, evidence : $B = \beta$

We have access to $P(A, B, C)$.

A simple example of belief updating

The universe $\mathcal{U} = \{A, B, C\}$, evidence : $B = \beta$

We have access to $P(A, B, C)$.

Then

$$P(A) = \sum_{B,C} P(A, B, C).$$

We have

$$P(A, B = \beta) = \sum_{C} P(A, B = \beta, C),$$

and

$$P(A|B = \beta) = \frac{P(A, B = \beta)}{\sum_A P(A, B = \beta).}$$

A simple example of belief updating

The universe $\mathcal{U} = \{A, B, C\}$, evidence : $B = \beta$

We have access to $P(A, B, C)$.

Then

$$P(A) = \sum_{B,C} P(A, B, C).$$

We have

$$P(A, B = \beta) = \sum_{C} P(A, B = \beta, C),$$

and

$$P(A|B = \beta) = \frac{P(A, B = \beta)}{\sum_A P(A, B = \beta).}$$

With access to the joint distribution over $\mathcal{U}$, belief updating is mathematically a rather simple task. However, it is intractable with large $\mathcal{U}$.