The Model-based Approach to Computer-aided Medical Decision Support

Lecture 5: PGMs meet Logic

Peter Lucas

peterl@cs.ru.nl

Institute for Computing and Information Sciences
Radboud University Nijmegen
The Netherlands

Motivation

- First-order logic: good for relational reasoning in various ways about classes of objects
- Probabilistic graphical models: good for reasoning with uncertainty

⇒ *why not combining them?*

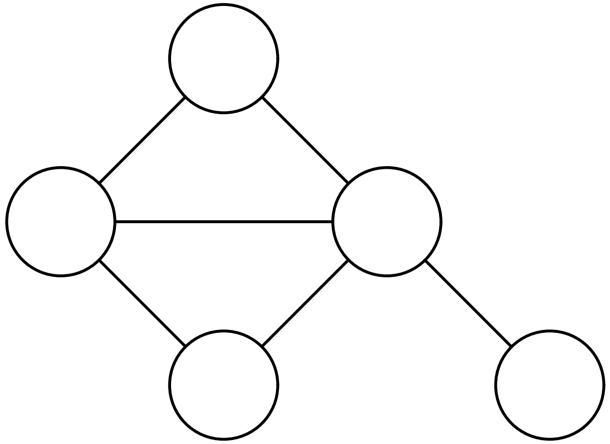
- Markov logic (generates Markov networks)
- Bayesian logic programs (generates Bayesian networks)
- Probabilistic Horn logic (abductive Bayesian-network reasoning)
- Chain logic (with Arjen Hommersom and Nivea Ferreira)

Probabilistic Graphical Models

- Family of probability distributions defined in terms of a directed, an undirected, or hybrid graph
- In general, undirected, and directed graphs make different assumptions regarding conditional independence
- Some independences are captured by directed and not by undirected graphs, and vice-versa

Markov Networks

An undirected graph



Basic idea:

- Each variable X corresponds to a vertex v
- Independence relation $\perp\!\!\!\perp$ is encoded as the absence of edges
- A missing edge (blockage of all paths) between vertices u and v indicates that X_u and X_v are (conditionally) independent

Markov Logic Network (MLN)

- a Markov logic net (MLN) set of pairs:

$$L = \{(F_k, w_k) \mid k = 1, \ldots, n\}$$

  with F_k a formula in first-order logic and w_k a real number

- Example (smoking causes cancer; if one friend smokes, the other smokes as well):

$$0.8 \;\; \forall x (S(x) \rightarrow C(x))$$
$$0.3 \;\; \forall x \forall y (F(x,y) \rightarrow (S(x) \leftrightarrow S(y)))$$

  with

  - S: Smoking; C: Cancer; F: Friends

Semantics of MLN

$C = \{c_1, \ldots, c_n\}$ is a set of constants, then:
corresponding Markov network $M_{L,C}$:

- $M_{L,C}$ includes a vertex with corresponding binary variable for each ground atom
- $M_{L,C}$ includes a complete graph with feature f_k for each instance of formula F_k

Associated probability distribution:

$$P(X) = \frac{1}{Z} \prod_k \phi_k(X_{\{k\}})^{n_k(X)} = \frac{1}{Z} \exp \sum_k w_k n_k(X)$$

with $n_k(X)$ number of instances of F_k based on X

Example

- Formula $F \equiv w \ \ \forall x(S(x) \rightarrow C(x))$
  - with S 'smoking' and C 'cancer'
  - weight w
- Constants $C = \{a, b\}$ (interpretations of x)

Interpretations of F (worlds/models):

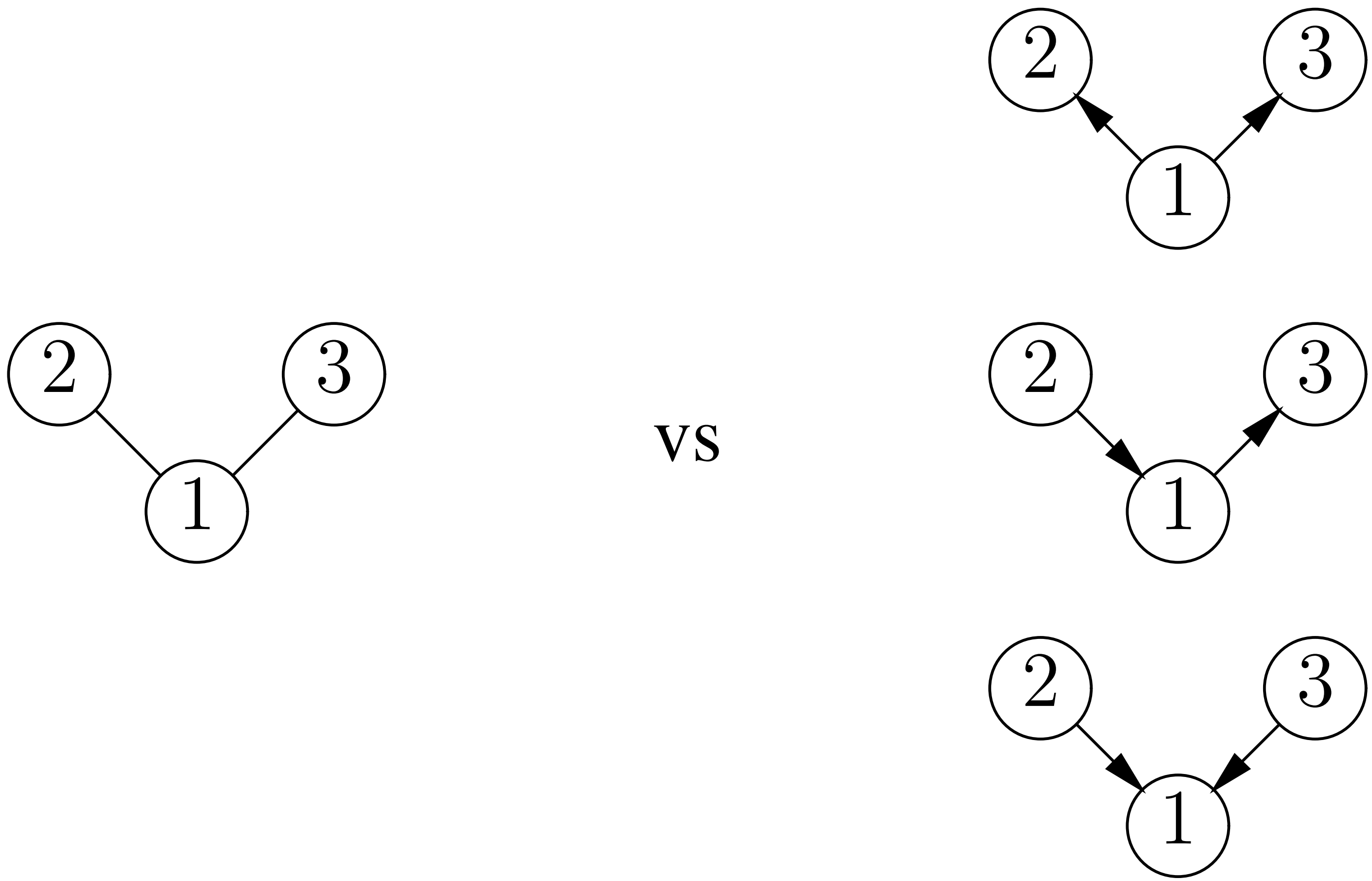
$\{S(a), C(a), S(b), C(b)\}$	2 models
$\{S(a), \neg C(a), S(b), \neg C(b)\}$	0 models
$\{S(a), \neg C(a), S(b), C(b)\}$	1 model
$\vdots$	

$$P(S(a), C(a), S(b), C(b)) = \frac{1}{Z} e^{w2} \qquad \text{2 models}$$

Markov network: $S(a)$——$C(a)$ $S(b)$——$C(b)$

Expressiveness

Directed graphs are more subtle when it comes to expressing independence information than undirected graphs:



Chain Graphs

- Graphical representation associated with a Bayesian network is not unique
  - different graphs may represent the same independence information
- Markov networks can be seen as the weakest type of graphical models
  - much of the subtleties of representing conditional dependence and independence cannot be handled
- Unique chain graph representatives of Bayesian networks (essential graphs)
  - Bayesian networks and Markov networks as special cases

Chain Graph Definition

- A chain graph is a hybrid graph with the restriction that no directed cycles exist
- Factorisation: chain graphs can be interpreted as an acyclic directed graph of chain components

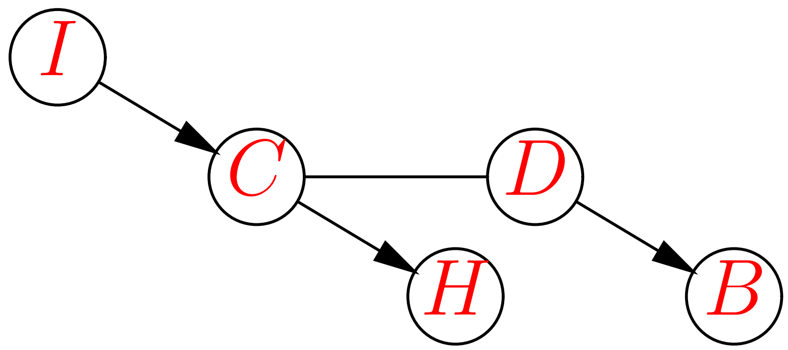
$$P(X_V) = \prod_{C \in \mathcal{C}} P(X_C \mid X_{\text{pa}(C)})$$

with $V = \bigcup_{C \in \mathcal{C}} C$, and where each $P(X_C \mid X_{\text{pa}(C)})$ factorises according to

$$P(X_C \mid X_{\text{pa}(C)}) = Z^{-1}(X_{\text{pa}(C)}) \prod_{M \in M(C)} \varphi_M(X_M)$$

Chain Graph Example

Influenza (I) causes coughing (C), where coughing is known as a possible cause for hoarseness (H). In addition, coughing is known to be associated with dyspnoea (shortness of breath) (D). Dyspnoea restricts the oxygen supply to the blood circulation; the resulting low oxygen saturation of the blood will turn the skin to colour blue (B)



Horn Clauses

- A formula in first-order logic
- A Horn-clause has a general form given by

  $$A \leftarrow B_1, \ldots, B_n$$

  where A is the head and $B_1, \ldots, B_n$ the body of the clause.
- Reasoning:
  - standard model-theoretic semantic, defined in terms of the logical consequence operator $\models$
  - procedural semantics, defined in terms of the deduction relation $\vdash$

Abduction Logic

Horn clauses of the form:

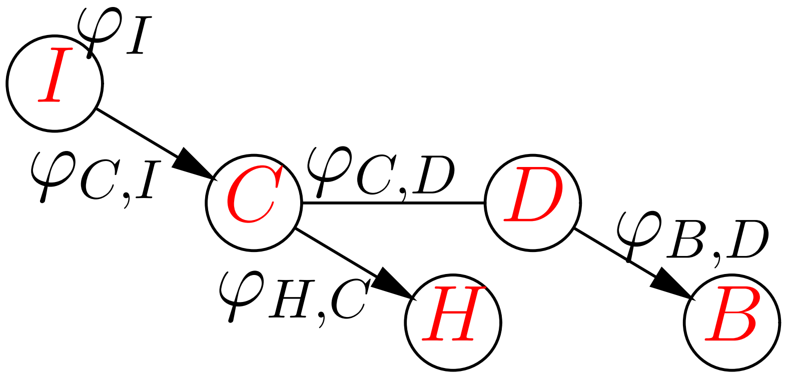
$$D \leftarrow B_1, \ldots, B_n : R_1, \ldots, R_m$$

where

- D: head of the clause, a predicate or $\perp$
- $B_1, \ldots, B_n$: body of the clause, a set of predicates (will become 'random variables')
- R_i: templates, to express relations between variables

Both the ',' as well as the ':' are interpreted as a conjunction

Influenza: Logical Specification



$$I(x) \leftarrow : \varphi_I(x)$$
$$C(x) \leftarrow I(y) : \varphi_{C,I}(x,y), \varphi_{C,D}(x,z)$$
$$D(x) \leftarrow I(y) : \varphi_{C,I}(z,y), \varphi_{C,D}(z,x)$$
$$H(x) \leftarrow C(y) : \varphi_{H,C}(x,y)$$
$$B(x) \leftarrow D(y) : \varphi_{B,D}(x,y)$$
$$\bot \leftarrow \varphi_{C,I}(x,y), \varphi_{C,D}(\bar{x},z)$$

where the φs are relations R_k

Reasoning: Explanations

Let:

- T: an abductive theory, which is a set of formulae
- A: the set of all assumables
- A': denote the set of ground instances of A

An explanation E of a set of observations O based on the pair $\langle T, A\rangle$ is defined as a set of ground assumables $E \subseteq A'$ satisfying the following conditions:

- $T \cup E \models O$, and
- $T \cup E$ is consistent, i.e., $T \cup E \not\models \perp$.

Chain Logic Syntax

Syntax of chain logic consists of:

- Formulae in abduction logic
- Weight declarations, which are of the form

  $$weight(a_1 : w_1, \ldots, a_n : w_n)$$

  where a_i represents an atom and w_i real, such that a weight declaration contains all instances of a predicate

Then, we define:

- Assumables A: atoms that occur in $weight$
- Hypothesis H: consistent set of *ground* atoms in $weight$ (one per $weight$)

Influenza

Potential functions:

φ_{CI}	i	$\bar{\imath}$
c	8	2
$\bar{c}$	1	10

φ_{CD}	d	$\bar{d}$
c	18	2
$\bar{c}$	5	2

φ_{HC}	c	$\bar{c}$
h	0.6	0.1
$\bar{h}$	0.4	0.9

The abduction clauses:

$$I(x) \leftarrow: \varphi_I(x)$$

$$C(x) \leftarrow I(y) : \varphi_{C,I}(x,y), \varphi_{C,D}(x,z)$$

$$D(x) \leftarrow I(y) : \varphi_{C,I}(z,y), \varphi_{C,D}(z,x)$$

$$H(x) \leftarrow C(y) : \varphi_{H,C}(x,y)$$

$$B(x) \leftarrow D(y) : \varphi_{B,D}(x,y)$$

$$\bot \leftarrow \varphi_{C,I}(x,y), \varphi_{C,D}(\bar{x},z)$$

Weights of the assumables $weight(\varphi_{CD}(t,t) : 18, \varphi_{CD}(t,f) : 2, \varphi_{CD}(f,t) : 5, \varphi_{CD}(f,f) : 2)$

Chain Logic Semantics

Abductive theory:

$$
\begin{aligned}
T = \; \{ & I(x) \leftarrow: \varphi_I(x), \\
& C(x) \leftarrow I(y) : \varphi_{C,I}(x,y), \varphi_{C,D}(x,z), \\
& D(x) \leftarrow I(y) : \varphi_{C,I}(z,y), \varphi_{C,D}(z,x), \\
& H(x) \leftarrow C(y) : \varphi_{H,C}(x,y), \\
& B(x) \leftarrow D(y) : \varphi_{B,D}(x,y), \\
& \bot \leftarrow \varphi_{C,I}(x,y), \varphi_{C,D}(\bar{x},z)\}
\end{aligned}
$$

where each of the variables has $\{f,t\}$ as domain
It now holds that:

$T \cup E \vDash H(t)$ and $T \cup E \nvDash \bot$, with

$$E = \{\varphi_I(t), \varphi_{H,C}(t,t), \varphi_{C,I}(t,t), \varphi_{C,D}(t,t)\}$$

Minimal Explanations

A minimal explanation E of O is an explanation whose proper subsets are not explanations of O. The set of all minimal explanations is denoted by $\mathcal{E}_T(O)$

Suppose we would like to calculate if a person is blue, i.e., $P(B(t))$; we obtain the minimal explanations for $B(t)$, i.e., $\mathcal{E}_T(B(t))$, as the set with the following 8 members:

$$\{\varphi_{B,D}(t,t), \varphi_{C,D}(t,t), \varphi_{C,I}(t,t), \varphi_I(t)\}$$
$$\{\varphi_{B,D}(t,t), \varphi_{C,D}(t,t), \varphi_{C,I}(t,f), \varphi_I(f)\}$$
$$\vdots$$

$$P(B(t)) = \sum_{E \in \mathcal{E}_T(B(t))} P(E) = 27.7/Z \approx 0.24$$

Probabilities of Formulae

Suppose E is a minimal explanation. Then, given T, $P_T(E)$ is obtained by marginalisation:

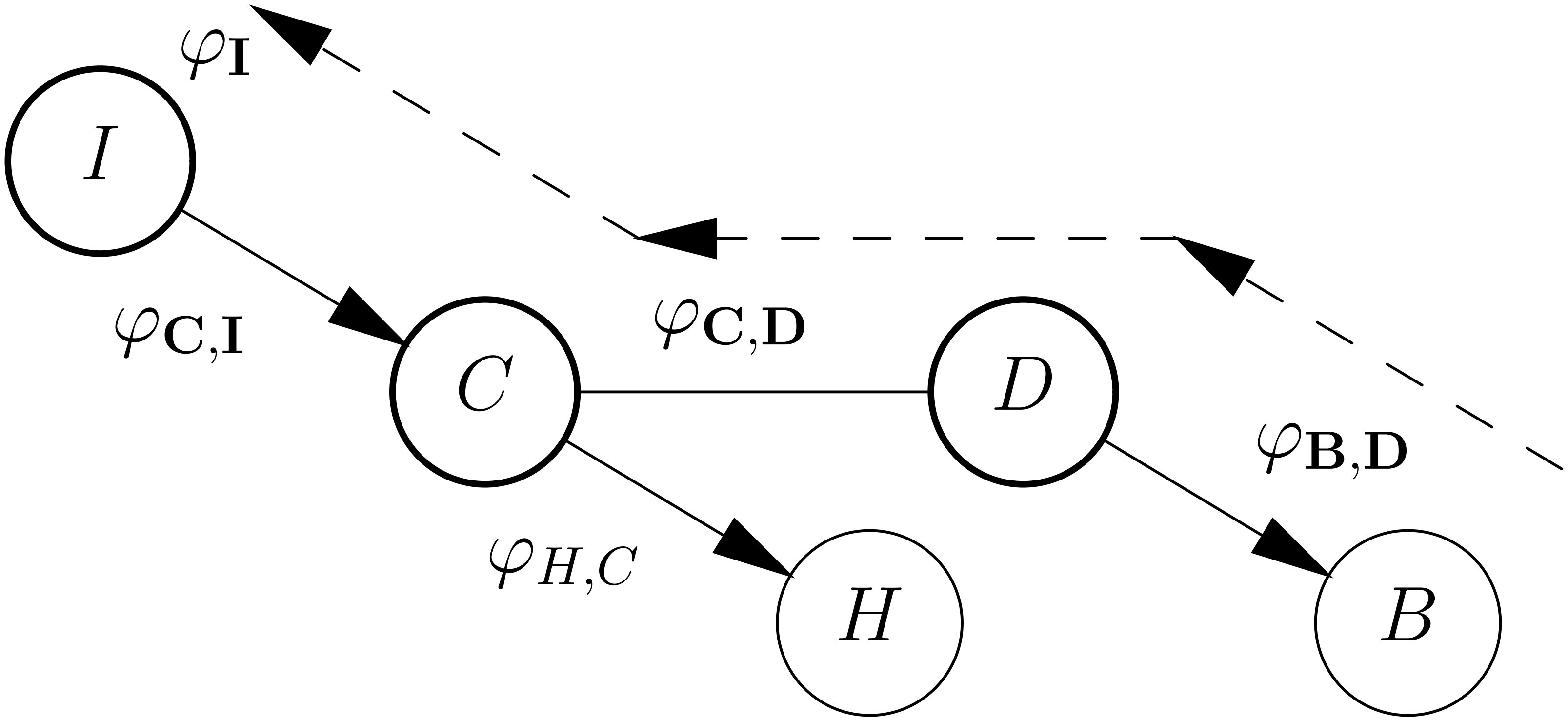
$$P_T(E) = P_T(\bigvee_i H_i) = \sum_i P_T(H_i)$$

as H_i's are mutually exclusive hypotheses (one atom per $weight$)

Theorem If $\mathcal{E}_T(\psi)$ is the set of minimal explanations of the conjunction of atoms ψ from the chain logic theory T, then:

$$P_T(\psi) = \sum_{E \in \mathcal{E}_T(\psi)} P_T(E)$$

Reasoning: Abductively



Direction of reasoning about B: from B to I, but ignoring H

Probabilistic reasoning = logical reasoning

Conditional Probabilities

Probability of influenza given blue colour, $P(I(t) \mid B(t))$:

- Find explanations: $P(I(t), B(t)) = \sum_{E \in \mathcal{E}_T(I(t),B(t))} P_T(E)$; we obtain: $P(I(t), B(t)) \approx 0.04$
- $P(B(t))$ is known (0.24)

$P(I(t) \mid B(t)) = P(I(t), B(t))/P(B(t)) \approx \frac{0.04}{0.24} \approx 0.16$

Note that the prior probability for $P(I(t))$ is 0.1

Final Considerations

- Chain logic is inspired by Poole's probabilistic Horn logic
- Additional integrity constraints guarantee that instantiations of potentials functions appear consistently in each explanation
- We present here a language that can be used for the specification of both Bayesian and Markov network models
- Maintaining a close relation between logical and probabilistic reasoning – without loss of expressiveness