Constructing Markov Logic Networks from First-Order Default Rules

Ondřej Kuželka[1], Jesse Davis[2] and Steven Schockaert[1]

[1]Cardiff University, [2]KU Leuven

Introduction

- Expert knowledge can often be represented using default rules of the form “if A then typically B”.

- **Example (Penguins):**
  - “Birds typically fly” - If X is a bird then X typically flies.
  - “Antarctic birds typically do not fly” - If X is an antarctic bird then X typically does not fly.

- We show how to construct Markov logic networks from rules of this type.

Preliminaries

(Preliminaries)

Markov Logic Networks

[Richardson and Domingos, 2006]

A set of rules (F_i, w_i) and a universe —> prob. distribution

$$P(X = x) = \frac{1}{Z} \exp\left(\sum_{(F_i, w_i) \in \mathcal{M}} w_i n_{F_i}(x) \right)$$

(Preliminaries)

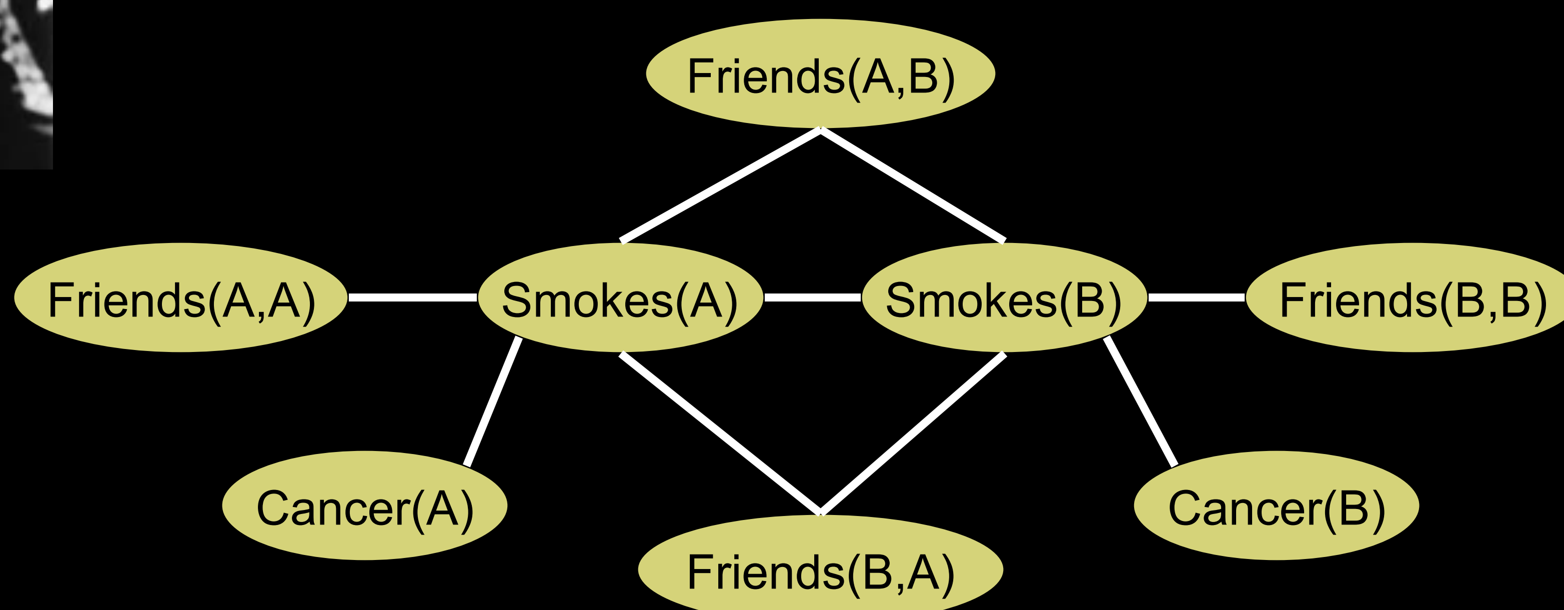
Example (Smokers)

[Richardson and Domingos, 2006]

$$(friends(A, B) \Rightarrow (smokes(A) \Leftrightarrow smokes(B)), 1)$$

$$(smokes(A) \Rightarrow cancer(A), 1)$$

Alice



Bob

(Preliminaries)

MAP Inference

[Richardson and Domingos, 2006]

- Given an MLN **M** and evidence **E**, find a possible world **x** which maximises *P(X=x)* - i.e. a *most probable world* of **(M,E)**

- **MAP-entailment relation** [Dupin de Saint-Cyr et al., 1994]

$$(\mathcal{M}, E) \vdash_{MAP} \alpha \quad \text{iff} \quad \forall \omega \in \max(\mathcal{M}, E) : \omega \models \alpha$$

- **Example:**

$$(\mathcal{M}, \{smokes(alice), \neg smokes(bob)\}) \vdash_{MAP} \neg friends(alice, bob)$$

(Preliminaries)

Default Rules and System P

[Kraus et al., 1990]

- Rules of the form:

$$\text{if } \alpha \text{ then typically } \beta \text{ (written } \alpha \mathrel{|\!\sim} \beta\text{)}$$

- Example:

$$bird(X) \mathrel{|\!\sim} flies(X)$$

$$bird(X) \wedge antarctic(X) \mathrel{|\!\sim} \neg flies(X)$$

- From evidence *bird(tweety),* we should be able to derive *flies(tweety)*, but if *antarctic(tweety)* is added we should withdraw flies(tweety).

(Preliminaries)

System P and Rational Monotony

[Kraus et al., 1990]

- Axioms of System P

Reflexivity	$\alpha \mid\!\sim \alpha$
Left logical equivalence	If $\alpha \equiv \alpha'$ and $\alpha \mid\!\sim \beta$ then $\alpha' \mid\!\sim \beta$
Right weakening	If $\beta \models \beta'$ and $\alpha \mid\!\sim \beta$ then $\alpha \mid\!\sim \beta'$
Or	If $\alpha \mid\!\sim \gamma$ and $\beta \mid\!\sim \gamma$ then $\alpha \vee \beta \mid\!\sim \gamma$
Cautious monotonicity	If $\alpha \mid\!\sim \beta$ and $\alpha \mid\!\sim \gamma$ then $\alpha \wedge \beta \mid\!\sim \gamma$
Cut	If $\alpha \wedge \beta \mid\!\sim \gamma$ and $\alpha \mid\!\sim \beta$ then $\alpha \mid\!\sim \gamma$

- The following rational monotonicity axiom is often added:

If $\alpha \mid\!\sim \beta$ and we cannot derive $\alpha \mid\!\sim \neg\gamma$ then $\alpha \wedge \gamma \mid\!\sim \beta$

(Preliminaries)

System P and Rational Monotony

(representation)

[Kraus et al., 1990]

- A set of defaults closed under axioms of System P and rational monotony corresponds to a linear ranking κ of possible worlds such that

$$\alpha \mid\!\sim \beta \text{ iff } \max\{\kappa(\omega) : \omega \models \alpha \wedge \beta\} > \max\{\kappa(\omega) : \omega \models \alpha \wedge \neg\beta\}$$

- The MAP-entailment relation also satisfies the axioms of System P and the rational monotony axiom.

(Preliminaries)

Closures of Default Theories

- A non-exhaustive set of default rules under System P + rational monotony may be completed in many ways.

- Rational closure [Lehmann & Magidor, 92]
- Lexicographic closure [Lehmann, 95]
- Maximum-entropy closure [Goldzmidt, Morris, Pearl, 93]

Correspond to different rankings of possible worlds based on rules.

(Preliminaries)

Z-Ranking

- Rational closure and lexicographic closure are based on so-called Z-ranking of default rules.*

Partition $\Delta_1 \cup \ldots \cup \Delta_k$ of Δ, where each Δ_j contains all formulas $\alpha \mid\!\sim \beta$ from Δ for which the following set of classical formulas is consistent:

$$\{\alpha \wedge \beta\} \cup \{\neg\alpha_i \vee \beta_i \mid (\alpha_i \mid\!\sim \beta_i) \in \Delta \setminus (\Delta_1 \cup \ldots \cup \Delta_{j-1})\}$$

$$bird(X) \mid\!\sim \mathit{flies}(X)$$

$$bird(X) \wedge antarctic(X) \mid\!\sim \neg \mathit{flies}(X)$$

$$hasFeather(X) \wedge laysEggs(X) \mid\!\sim bird(X)$$

$\Rightarrow$

Δ_1:

$$bird(X) \mid\!\sim \mathit{flies}(X)$$

$$hasFeather(X) \wedge laysEggs(X) \mid\!\sim bird(X)$$

Δ_2:

$$bird(X) \wedge antarctic(X) \mid\!\sim \neg \mathit{flies}(X)$$

**For simplicity, we do not consider hard rules here, see the paper for the full version.*

(Preliminaries)

Rational and Lexicographic Closures

- **Given:** Z-ranking $\Delta_1 \cup \ldots \cup \Delta_k$
- **Rational Closure:**
  - Let *j* be the smallest index for which $\Delta^{rat}_{\alpha} = \Delta_j \cup \ldots \cup \Delta_k \cup \{\alpha\}$ is consistent. Then $\alpha \mid\!\sim \beta$ is in the **rational closure** of Δ if $\Delta^{rat}_{\alpha} \models \beta$.
- **Lexicographic Closure:**
  - Let $sat(\omega,\Delta_j)$ denote number of defaults from Δ_j satisfied by ω (as classical formulas).
  - ω_1 is lex-preferred over an interpretation ω_2, if there exists a *j* such that $sat(\omega_1,\Delta_j) > sat(\omega_2,\Delta_j)$ while $sat(\omega_1,\Delta_l) = sat(\omega_2,\Delta_l)$ for all i>j.
  - The default $\alpha \mid\!\sim \beta$ is in the lexicographic closure of Δ if β is satisfied in all the most lex-preferred models of α.

Encoding Non-ground Default Theories in MLNs

Rationale

- The ranking function κ may also be probability of possible worlds (*leads to MAP-entailment* as |~).
- Markov logic networks can represent any distribution on a finite set of possible worlds.

=> We can encode closures of default theories in MLNs

But can we do it efficiently? And will the MLNs be compact and intuitive? With non-ground theories? **Yes! (this work)**

Non-Ground Default Theories

- We view non-ground as templates for specifying ground default theories (*similar to MLNs*).
- Ground default rules are interpreted as MAP constraints.

Example:

$$bird(X) \mid\!\sim flies(X)$$
$$bird(X) \wedge antarctic(X) \mid\!\sim \neg flies(X)$$

+

$$\{tweety, donald\}$$

A MLN satisfying:

$$bird(tweety) \vdash_{MAP} flies(tweety)$$
$$bird(tweety) \wedge antarctic(tweety) \vdash_{MAP} \neg flies(tweety)$$
$$bird(donald) \vdash_{MAP} flies(donald)$$
$$bird(donald) \wedge antarctic(donald) \vdash_{MAP} \neg flies(donald)$$

Models of Non-Ground Default Theories

Definition:

Let $\Delta \cup \Theta$ be a default theory where Δ is a set of default rules and Θ is a set of hard rules. A Markov logic network $\mathcal{M}$ is a model of the default logic theory $\Delta \cup \Theta$ if the following holds:

1. $P\left[X = \omega\right] = 0$ whenever $\omega \not\models \Theta$,

2. for any default rule $\alpha \mathrel{|\!\!\sim} \beta \in \Delta$ and any grounding substitution θ of the unquantified (open) variables of $\alpha \mathrel{|\!\!\sim} \beta$, either $\{\alpha\theta\} \cup \Theta \vdash \bot$ or

$$(\mathcal{M}, \alpha\theta) \vdash_{MAP} \beta\theta.$$

This means MLN *M* with evidence $\alpha\theta$

Lifted Z-Ranking

- **Given:** a default theory Δ, universe *U*
- **Procedure:**

1. Group interchangeable constants.
2. Let $G = \{G_1, \ldots, G_k\}$ contain ground nonisomorphic representatives** of rules in Δ (w.r.t. interchangeability of constants).
3. Let $L = \{L_1, \ldots, L_k\}$ contain variabilized variants of the rules from *G* (using typing to distinguish variables corresponding to non-interchangeable constants).
4. Find partition $\Delta_1 \cup \ldots \cup \Delta_k$ of *L*, where each Δ_j contains all formulas $L_i = \alpha \mid\!\sim \beta$ from *L* for which the following set of classical formulas

$$\{G_i\} \cup \{\neg\alpha_i \vee \beta_i : (\alpha_i \mid\!\sim \beta_i) \in L \setminus (L_1 \cup \ldots L_{j-1})\}$$

has a Herbrand model with universe *U*.

**For simplicity, we do not consider hard rules here, see the paper for the full version.*

***This is related to shattering from lifted inference [Poole, 2003].*

“Rational” MLNs

- **Given:** Lifted Z-ranking $\Delta_1 \cup \ldots \cup \Delta_k$
- **Output:**

$$\mathcal{M} = \bigcup_{i=1}^{k} \{(\neg a_i \vee \neg\alpha \vee \beta, \infty) : \alpha \mathrel{|\!\sim} \beta \in \Delta_i^*\} \cup$$
$$\cup \{(a_i, 1)\} \cup \{(\phi, \infty) : \phi \in \Theta\} \cup \{(a_i \vee \neg a_{i-1}, \infty)\}$$

where a_i are auxiliary (ground) literals.

- If (Δ,Θ) is satisfiable then *M* is a Markov logic model of (Δ,Θ).

(Corresponds to reasoning in possibilistic logic.)

“Lexicographic” MLNs

- **Given:** Lifted Z-ranking $\Delta_1 \cup \ldots \cup \Delta_k$, universe U
- **Output:**

$$\bigcup_{i=1}^{k} \{(\neg\alpha \vee \beta, \lambda_i) \colon \alpha \mathrel{|\!\sim} \beta \in \Delta_i\} \cup \{(\phi, \infty) \colon \phi \in \Theta\}$$

where $\lambda_i = 1 + \sum_{j=1}^{i-1} \sum_{\alpha \mathrel{|\!\sim} \beta \in \Delta_j^*} |\mathcal{U}|^{vars(\alpha \mathrel{|\!\sim} \beta)} \cdot \lambda_j \text{ for } i > 1$

and $\lambda_1 = 1$.

- If (Δ,Θ) is satisfiable then M is a Markov logic model of (Δ,Θ).

Example

$$bird(X) \wedge (X \neq tweety) \mid\!\sim \mathit{flies}(X),$$
$$bird(X) \wedge antarctic(X) \mid\!\sim \neg \mathit{flies}(X)$$
$$bird(X) \wedge antarctic(X) \wedge (X \neq Y) \wedge sameSpecies(X, Y) \mid\!\sim antarctic(Y)$$

$$\mathcal{U} = \{tweety, donald, beeper\}$$

$$\phi_1 = (\neg bird(\tau_{tweety} : X) \vee \neg antarctic(\tau_{tweety} : X) \vee \neg \mathit{flies}(\tau_{tweety} : X), 1)$$
$$\phi_2 = (\neg bird(X) \vee \neg (X \neq \tau_{tweety} : tweety) \vee \mathit{flies}(Y), 1)$$
$$\phi_3 = (\neg bird(\tau_{beeper} : X) \vee \neg antarctic(\tau_{beeper} : X) \vee \neg \mathit{flies}(\tau_{beeper} : X), 7)$$
$$\phi_4 = (\neg bird(X) \vee \neg sameSpecies(X, Y) \vee \neg (X \neq Y) \vee \neg antarctic(X) \vee antarctic(Y), 7)$$

“Max-Entropy” MLNs

- Based on a ranking which refines Z-ranking.
- Computationally more expensive than rational and lexicographic transformations, because it requires running MAP-inference to get the ranking.
- *(Covered in the paper in detail.)*

Experiments

- UW-CSE, prediction of *advisedBy* relation
- A hand-crafted set of default rules:

$$
\begin{aligned}
D_1 : \quad & \mathrel{|\!\sim} \neg advisedBy(S, P) \\
D_2 : \quad & advisedBy(S, P_1) \mathrel{|\!\sim} \neg tempAdvisedBy(S, P_2) \\
D_3 : \quad & advisedBy(S, P) \wedge publication(Pub, S) \mathrel{|\!\sim} publication(Pub, P) \\
D_4 : \quad & (P_1 \neq P_2) \wedge advisedBy(S, P_1) \mathrel{|\!\sim} \neg advisedBy(S, P_2) \\
D_5 : \quad & advisedBy(S, P) \wedge ta(C, S, T) \mathrel{|\!\sim} taughtBy(C, P, T) \\
D_6 : \quad & professor(P) \wedge student(S) \wedge publication(Pub, P) \wedge publication(Pub, S) \\
& \mathrel{|\!\sim} advisedBy(S, P) \\
D_7 : \quad & professor(P) \wedge student(S) \wedge publication(Pub, P) \wedge publication(Pub, S) \wedge \\
& tempAdvisedBy(S, P2) \mathrel{|\!\sim} \neg advisedBy(S, P) \\
D_8 : \quad & (S_1 \neq S_2) \wedge advisedBy(S_2, P) \wedge ta(C, S_2, T) \wedge ta(C, S_1, T) \wedge \\
& taughtBy(C, P, T) \wedge student(S_1) \wedge professor(P) \mathrel{|\!\sim} advisedBy(S_1, P) \\
D_9 : \quad & (S_1 \neq S_2) \wedge advisedBy(S_2, P) \wedge ta(C, S_2, T) \wedge ta(C, S_1, T) \wedge \\
& taughtBy(C, P, T) \wedge student(S_1) \wedge professor(P) \wedge tempAdvisedBy(S_1, P_2) \\
& \mathrel{|\!\sim} \neg advisedBy(S_1, P)
\end{aligned}
$$

Experimental Results

	MaxEnt		LEX		ONES		LEARNED	
	TP	**FP**	**TP**	**FP**	**TP**	**FP**	**TP**	**FP**
AI	10 ± 0	7 ± 0	10 ± 0	7 ± 0	8.6 ± 0.7	4.9 ± 0.9	10 ± 0	2 ± 0
GRA.	4 ± 0	5 ± 0	4 ± 0	5 ± 0	3.5 ± 0.7	3.9 ± 0.7	2 ± 0	2 ± 0
LAN.	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	2 ± 0	1 ± 0
SYS.	10.5 ± 0.5	3.5 ± 0.5	11 ± 0	3 ± 0	7.2 ± 1.1	2.4 ± 0.5	4 ± 0	0 ± 0
THE.	3 ± 0	3 ± 0	3 ± 0	3 ± 0	3 ± 0	1.7 ± 0.7	2 ± 0	1 ± 0

Table 1: Experimental results for MLNs obtained by the described methods.

Conclusions

- It is possible to construct MLNs capturing default reasoning with non-ground rules.
- The method is efficient and allows us to construct MLNs with meaningful “MAP-results” even if only expert rules and no training examples are available.
- Next step: Combine this with weight learning (default rules as constraints on MAP inference)

Thank You!