Foundations for Inference in Probabilistic Relational Models

Paulius Dilkas

27th May 2020

Outline

Outline

Probabilistic Relational Models

ProbLog (De Raedt, Kimmig and Toivonen 2007)

$$1.0 :: \texttt{likes}(X, Y) :\!- \texttt{friendOf}(X, Y).$$

$$0.8 :: \texttt{likes}(X, Y) :\!- \texttt{friendOf}(X, Z), \texttt{likes}(Z, Y).$$

$$0.5 :: \texttt{friendOf}(\mathit{john}, \mathit{mary}).$$

$$0.5 :: \texttt{friendOf}(\mathit{mary}, \mathit{pedro}).$$

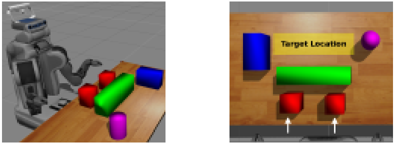
$$0.5 :: \texttt{friendOf}(\mathit{mary}, \mathit{tom}).$$

$$0.5 :: \texttt{friendOf}(\mathit{pedro}, \mathit{tom}).$$

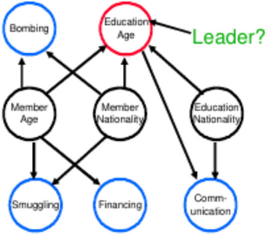
Probabilistic Relational Models

- What do these models have in common?
- When performing inference...
  - what makes inference challenging?
  - do we have to consider every detail?
  - can we do any better?
- How can we learn PRMs from data?

Applications



Moldovan and De Raedt 2014



Delaney et al. 2010

Predicate	Instance	Source(s)
ethnicGroup	Cubans	CSEAL
arthropod	spruce beetles	CPL, CSEAL
female	Kate Mara	CPL, CMC
sport	BMX bicycling	CSEAL, CMC
profession	legal assistants	CPL
magazine	Thrasher	CPL
bird	Buff-throated Warbler	CSEAL
river	Fording River	CPL, CMC
mediaType	chemistry books	CPL, CMC
cityInState	(troy, Michigan)	CSEAL
musicArtistGenre	(Nirvana, Grunge)	CPL
tvStationInCity	(WLS-TV, Chicago)	CPL, CSEAL
sportUsesEquip	(soccer, balls)	CPL
athleteInLeague	(Dan Fouts, NFL)	RL
starredIn	(Will Smith, Seven Pounds)	CPL
productType	(Acrobat Reader, FILE)	CPL
athletePlaysSport	(scott shields, baseball)	RL
cityInCountry	(Dublin Airport, Ireland)	CPL

Table 1: Example beliefs promoted by NELL.

Carlson et al. 2010

```
is_malignant(Case):-
        biopsyProcedure(Case,usCore),
        changes_Sizeinc(Case,missing),
        feature_shape(Case).
is_malignant(Case):-
        assoFinding(Case,asymmetry),
        breastDensity(Case,scatteredFDensities),
        vacuumAssisted(Case,yes).
is_malignant(Case):-
        needleGauge(Case,9),
        offset(Case,14),
        vacuumAssisted(Case,yes).
```

Côrte-Real, Dutra and Rocha 2017

Outline

$\texttt{Husband}(\mathit{joffrey}, \mathit{margaery})$
$\texttt{Husband}(\mathit{tommen}, \mathit{margaery})$
$\texttt{Husband}(\mathit{renly}, \mathit{margaery})$
$\texttt{Parent}(\mathit{cersei}, \mathit{joffrey})$
$\texttt{Parent}(\mathit{cersei}, \mathit{myrcella})$
$\texttt{Parent}(\mathit{cersei}, \mathit{tommen})$
$\texttt{Parent}(\mathit{tywin}, \mathit{cersei})$

$\texttt{Female}(\mathit{cersei})$,
$\texttt{Female}(\mathit{margaery})$,
$\texttt{Female}(\mathit{myrcella})$

$\texttt{Female}(X) :- \texttt{Husband}(\mathit{joffrey}, X).$
$\texttt{Female}(X) :- \texttt{Parent}(X, \mathit{joffrey}).$
$\texttt{Female}(X) :- \texttt{Parent}(\mathit{cersei}, X), \neg\texttt{Husband}(X, \mathit{margaery}).$

Main Results

Definition (Equivalence)

Two n-tuples of constants a and b are equivalent if

$$\Delta \models (\mathtt{P} \circ \rho)(a) \iff \Delta \models (\mathtt{P} \circ \rho)(b)$$

for all atoms $\mathtt{P} \circ \rho$ acting on n variables.

Main Results

Definition (Equivalence)

Two n-tuples of constants a and b are equivalent if

$$\Delta \models (\mathtt{P} \circ \rho)(a) \iff \Delta \models (\mathtt{P} \circ \rho)(b)$$

for all atoms $\mathtt{P} \circ \rho$ acting on n variables.

Theorem

There is a logic program $\mathcal{L} \colon \mathcal{KB}(P_1, C) \to \mathcal{KB}(P_2, C)$ *such that* $\mathcal{L}(\Delta_1) = \Delta_2$ *if and only if* $\sim_{\Delta_2}$ *is coarser than* $\sim_{\Delta_1}$.

Outline

What Characterises a (Probabilistic) Logic Program?

$$1.0 :: \texttt{likes}(X, Y) \mathrel{\texttt{:-}} \texttt{friendOf}(X, Y).$$

$$0.8 :: \texttt{likes}(X, Y) \mathrel{\texttt{:-}} \texttt{friendOf}(X, Z), \texttt{likes}(Z, Y).$$

$$0.5 :: \texttt{friendOf}(\mathit{john}, \mathit{mary}).$$

$$0.5 :: \texttt{friendOf}(\mathit{mary}, \mathit{pedro}).$$

$$0.5 :: \texttt{friendOf}(\mathit{mary}, \mathit{tom}).$$

$$0.5 :: \texttt{friendOf}(\mathit{pedro}, \mathit{tom}).$$

What Characterises a (Probabilistic) Logic Program?

$$1.0 :: \texttt{likes}(X, Y) :- \texttt{friendOf}(X, Y).$$
$$0.8 :: \texttt{likes}(X, Y) :- \texttt{friendOf}(X, Z), \texttt{likes}(Z, Y).$$
$$0.5 :: \texttt{friendOf}(\mathit{john}, \mathit{mary}).$$
$$0.5 :: \texttt{friendOf}(\mathit{mary}, \mathit{pedro}).$$
$$0.5 :: \texttt{friendOf}(\mathit{mary}, \mathit{tom}).$$
$$0.5 :: \texttt{friendOf}(\mathit{pedro}, \mathit{tom}).$$

- predicates, arities

What Characterises a (Probabilistic) Logic Program?

$$1.0 :: \mathtt{likes}(X, Y) :\!- \mathtt{friendOf}(X, Y).$$
$$0.8 :: \mathtt{likes}(X, Y) :\!- \mathtt{friendOf}(X, Z), \mathtt{likes}(Z, Y).$$
$$0.5 :: \mathtt{friendOf}(\mathit{john}, \mathit{mary}).$$
$$0.5 :: \mathtt{friendOf}(\mathit{mary}, \mathit{pedro}).$$
$$0.5 :: \mathtt{friendOf}(\mathit{mary}, \mathit{tom}).$$
$$0.5 :: \mathtt{friendOf}(\mathit{pedro}, \mathit{tom}).$$

- predicates, arities
- variables

What Characterises a (Probabilistic) Logic Program?

$$1.0 :: \texttt{likes}(X, Y) \text{ :- } \texttt{friendOf}(X, Y).$$
$$0.8 :: \texttt{likes}(X, Y) \text{ :- } \texttt{friendOf}(X, Z), \texttt{likes}(Z, Y).$$
$$0.5 :: \texttt{friendOf}(\mathit{john}, \mathit{mary}).$$
$$0.5 :: \texttt{friendOf}(\mathit{mary}, \mathit{pedro}).$$
$$0.5 :: \texttt{friendOf}(\mathit{mary}, \mathit{tom}).$$
$$0.5 :: \texttt{friendOf}(\mathit{pedro}, \mathit{tom}).$$

- predicates, arities
- variables
- constants

What Characterises a (Probabilistic) Logic Program?

$1.0 :: \texttt{likes}(X, Y) \coloneq \texttt{friendOf}(X, Y).$

$0.8 :: \texttt{likes}(X, Y) \coloneq \texttt{friendOf}(X, Z), \texttt{likes}(Z, Y).$

$0.5 :: \texttt{friendOf}(\mathit{john}, \mathit{mary}).$

$0.5 :: \texttt{friendOf}(\mathit{mary}, \mathit{pedro}).$

$0.5 :: \texttt{friendOf}(\mathit{mary}, \mathit{tom}).$

$0.5 :: \texttt{friendOf}(\mathit{pedro}, \mathit{tom}).$

- predicates, arities
- variables
- constants
- probabilities

What Characterises a (Probabilistic) Logic Program?

$$\left\{\begin{array}{l}
1.0 :: \texttt{likes}(X, Y) :\!- \texttt{friendOf}(X, Y). \\
0.8 :: \texttt{likes}(X, Y) :\!- \texttt{friendOf}(X, Z), \texttt{likes}(Z, Y). \\
0.5 :: \texttt{friendOf}(\mathit{john}, \mathit{mary}). \\
0.5 :: \texttt{friendOf}(\mathit{mary}, \mathit{pedro}). \\
0.5 :: \texttt{friendOf}(\mathit{mary}, \mathit{tom}). \\
0.5 :: \texttt{friendOf}(\mathit{pedro}, \mathit{tom}).
\end{array}\right.$$

- predicates, arities
- variables
- constants
- probabilities
- length

What Characterises a (Probabilistic) Logic Program?

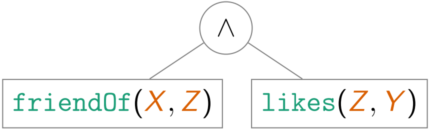
$$\left\{ \begin{array}{l} 1.0 :: \texttt{likes}(X, Y) :\!- \texttt{friendOf}(X, Y). \\ 0.8 :: \texttt{likes}(X, Y) :\!- \underbrace{\texttt{friendOf}(X, Z), \texttt{likes}(Z, Y)}. \\ 0.5 :: \texttt{friendOf}(\mathit{john}, \mathit{mary}). \\ 0.5 :: \texttt{friendOf}(\mathit{mary}, \mathit{pedro}). \\ 0.5 :: \texttt{friendOf}(\mathit{mary}, \mathit{tom}). \\ 0.5 :: \texttt{friendOf}(\mathit{pedro}, \mathit{tom}). \end{array} \right.$$

- predicates, arities
- variables
- constants
- probabilities
- length
- complexity

What Characterises a (Probabilistic) Logic Program?

```
1.0 :: likes(X, Y) :- friendOf(X, Y).
0.8 :: likes(X, Y) :- friendOf(X, Z), likes(Z, Y).
0.5 :: friendOf(john, mary).
0.5 :: friendOf(mary, pedro).
0.5 :: friendOf(mary, tom).
0.5 :: friendOf(pedro, tom).
```

- predicates, arities
- variables
- constants
- probabilities
- length
- complexity



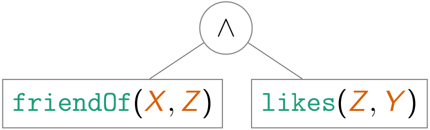
What Characterises a (Probabilistic) Logic Program?

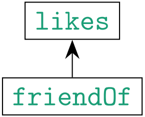
```
1.0 :: likes(X, Y) :- friendOf(X, Y).
0.8 :: likes(X, Y) :- friendOf(X, Z), likes(Z, Y).
0.5 :: friendOf(john, mary).
0.5 :: friendOf(mary, pedro).
0.5 :: friendOf(mary, tom).
0.5 :: friendOf(pedro, tom).
```

- predicates, arities
- variables
- constants
- probabilities
- length
- complexity

Also:

- cyclicity
- independence





Outline

1. Introduction
2. Equivalence
3. Random Programs
4. WMC
5. Future Work

WMC Requires Independent Literals

Theorem

Let $\mathbf{B}$ *be a free Boolean algebra over* $\{l_i\}_{i=1}^n$ *with measure*

$$m\colon \mathbf{B} \to \mathbb{R}_{\geq 0},$$

and let

$$L = \{l_i\}_{i=1}^n \cup \{\neg l_i\}_{i=1}^n.$$

Then there exists a weight function $w\colon L \to \mathbb{R}_{\geq 0}$ *such that* $m = \mathrm{WMC}_w$ *if and only if*

$$m(l \wedge l') = m(l)m(l')$$

for all distinct $l, l' \in L$ *such that* $l \neq \neg l'$*.*

Extending the Algebra

$$
\begin{tikzcd}
\mathbb{R}_{\geq 0} & & \\
\mathbf{B} \arrow[u, "m"] \arrow[r, hook, "\iota"] & \mathbf{B}' \arrow[ul, "\mathrm{WMC}_w"'] & \\
L \arrow[u, hook] \arrow[r, hook] & L' \arrow[u, hook] \arrow[r, "w"] & \mathbb{R}_{\geq 0}
\end{tikzcd}
$$

How Can This Benefit Inference?

Theorem (Sikorski 1969)

If $\mathbf{B} = \mathcal{F}\{a\} + \mathcal{F}\{b\}$, *then* $\Pr(a \wedge b) = \Pr(a)\Pr(b)$.

How Can This Benefit Inference?

Theorem (Sikorski 1969)

If $\mathbf{B} = \mathcal{F}\{a\} + \mathcal{F}\{b\}$, *then* $\Pr(a \land b) = \Pr(a)\Pr(b)$.

Conjecture

If $\mathbf{B} = \mathcal{F}\{a\} +_{\mathcal{F}\{c\}} \mathcal{F}\{b\}$, *then* $\Pr(a \land b \land c) = \Pr(a \land c)\Pr(b \land c)$.

How Can This Benefit Inference?

Theorem (Sikorski 1969)

If $\mathbf{B} = \mathcal{F}\{a\} + \mathcal{F}\{b\}$, *then* $\Pr(a \land b) = \Pr(a)\Pr(b)$.

Conjecture

If $\mathbf{B} = \mathcal{F}\{a\} +_{\mathcal{F}\{c\}} \mathcal{F}\{b\}$, *then* $\Pr(a \land b \land c) = \Pr(a \land c)\Pr(b \land c)$.

Conjecture

Using coproducts and pushouts, one can encode a Bayesian network into WMC with fewer literals and a shorter theory than before.

How Can This Benefit Inference?

Theorem (Sikorski 1969)

If $\mathbf{B} = \mathcal{F}\{a\} + \mathcal{F}\{b\}$, *then* $\Pr(a \wedge b) = \Pr(a)\Pr(b)$.

Conjecture

If $\mathbf{B} = \mathcal{F}\{a\} +_{\mathcal{F}\{c\}} \mathcal{F}\{b\}$, *then* $\Pr(a \wedge b \wedge c) = \Pr(a \wedge c)\Pr(b \wedge c)$.

Conjecture

Using coproducts and pushouts, one can encode a Bayesian network into WMC with fewer literals and a shorter theory than before.

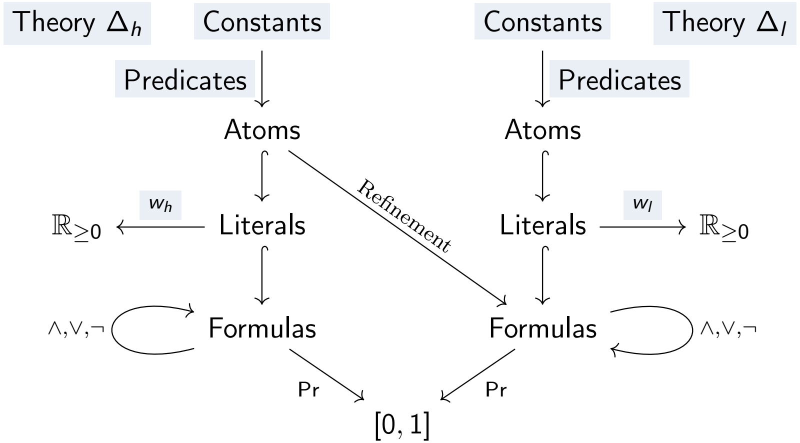
Conjecture

A #SAT algorithm can be adapted without sacrificing efficiency.

Outline

"I'm a fiction writer in the grant-proposal genre."

Abstraction: Before (Belle 2020)



Abstraction: After

$$\begin{tikzcd}
S_h \arrow[d, hook] \arrow[dr] & S_l \arrow[d, hook] \\
\mathbf{B}_h \arrow[r, dashed] \arrow[d] & \mathbf{B}_l \arrow[d] \\
\mathbf{B}_h/(\neg\Delta_h) \arrow[r, dashed] & \mathbf{B}_l/(\neg\Delta_l)
\end{tikzcd}$$

Plan for the Future

1. Rework the equivalence paper (2 months)
2. Improve and resubmit the random programs paper (done)
3. WMC 2.0
   - Formalise a new encoding for Bayesian networks (2 months)
   - Experimentally compare with other encodings (2 months)
4. Abstractions as homomorphisms
   - Find algebraic counterparts for logic-based concepts (1 month)
   - Establish ‘iff’ results for their preservation (2 months)
   - Develop algorithms for constructing abstractions (2 months)
   - Theorems for the preservation of independence (3 months)
5. And lots of writing, editing, and rewriting ($[9, \infty)$ months)

Probabilistic Relational Models

Markov Logic Network (Richardson and Domingos 2006)

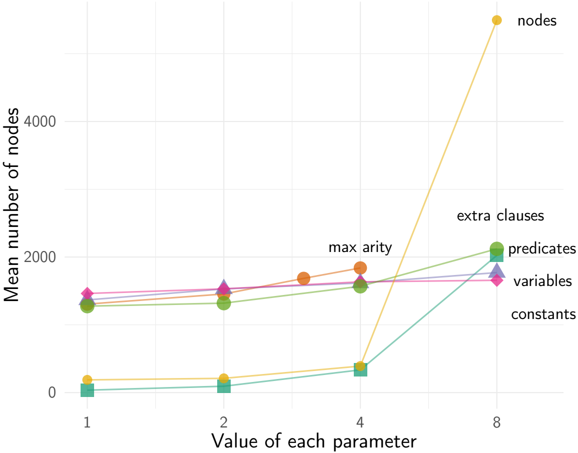
$$0.7 \quad \forall x \forall y \forall z\, \mathtt{Friends}(x, y) \land \mathtt{Friends}(y, z) \Rightarrow \mathtt{Friends}(x, z)$$

$$2.3 \quad \forall x\, \neg\exists y\, \mathtt{Friends}(x, y) \Rightarrow \mathtt{Smokes}(x)$$

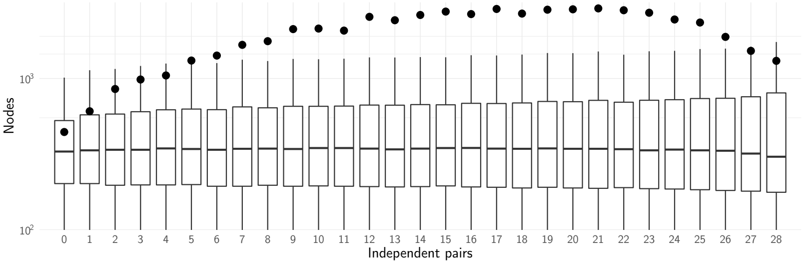
$$1.5 \quad \forall x\, \mathtt{Smokes}(x) \Rightarrow \mathtt{Cancer}(x)$$

$$1.1 \quad \forall x \forall y\, \mathtt{Friends}(x, y) \Rightarrow (\mathtt{Smokes}(x) \Leftrightarrow \mathtt{Smokes}(y))$$

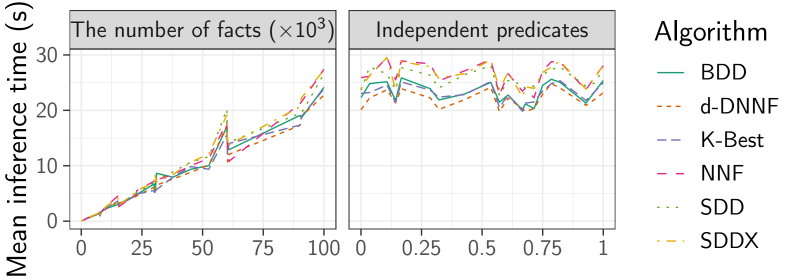
What Programs Are Hard to Generate?



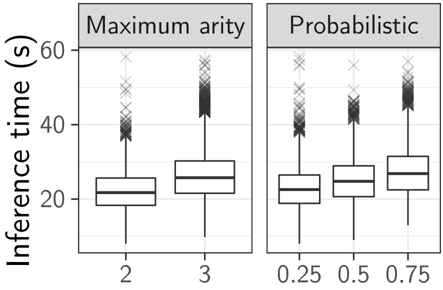
What Programs Are Hard to Generate?



How Program Features Influence Inference Time



How Program Features Influence Inference Time



Defining WMC

Definition

Let **B** be an atomic Boolean algebra. Let $L \subset \mathbf{B}$ be such that every atom m can be uniquely expressed as $m = \bigwedge L'$ for some $L' \subseteq L$, and let $w \colon L \to \mathbb{R}_{\geq 0}$ be arbitrary. The weighted model count $\mathrm{WMC}_w \colon \mathbf{B} \to \mathbb{R}_{\geq 0}$ is defined as

$$\mathrm{WMC}_w(x) = \begin{cases} 0 & \text{if } x = 0 \\ \prod_{l \in L'} w(l) & \text{if } x = \bigwedge L' \text{ is an atom} \\ \sum_{\text{atoms } a \leq x} \mathrm{WMC}_w(a) & \text{otherwise} \end{cases}$$

for any $x \in \mathbf{B}$.