Representing Aggregators in Relational Probabilistic Models

David Buchman[1] David Poole[1]

[1] Department of Computer Science, University of British Columbia, Vancouver, B.C., Canada

29th AAAI Conference on Artificial Intelligence (AAAI 2015)

Aggregators

Aggregators: dependency on unbounded #vars.

Adding aggregators to MLNs – the simplest case:

MLN over $\{A(x)\}$, defining $\mathbf{P}_{\text{old}}(A_{1..n})$ $\xrightarrow{\text{Add } B \text{ \& weighted formulae } \mathbb{F}}$ MLN over $\{A(x), B\}$, defining $\mathbf{P}_{\text{new}}(A_{1..n}, B)$.

$$\sum_{B\in\{0,1\}} \mathbf{P}_{\text{new}}(A_{1..n}, B) = \mathbf{P}_{\text{new}}(A_{1..n})$$

"Adding aggregators **as dependent vars**" = the original distribution does not change:

$$\mathbf{P}_{\text{new}}(A_{1..n}) \quad = \quad \mathbf{P}_{\text{old}}(A_{1..n})$$

Theorem 2: If $\mathbb{F}$ has no quantifiers ($\exists$ and $\forall$), then B is independent of $A_{1..n}$ in all possible groundings.

$\Longrightarrow$ No useful aggregator can be represented!

- ◇ $n_0 :=$ #vars among $A_{1..n}$ that are $\mathbf{0}$ ("*false*")
- ◇ $n_1 :=$ #vars among $A_{1..n}$ that are $\mathbf{1}$ ("*true*")
- ◇ An aggregator is **saturated** if for any n, $\mathbf{P}(B \mid A_{1..n})$ is equal for all assignments to $A_{1..n}$ for which $n_0, n_1 \geq$ threshold.

Theorem 4: General case – $\mathbb{F}$ may contain quantifiers: then $\mathbf{P}(B \mid A_{1..n})$ is a saturated aggregator.

DI-MLNs: A Normal Form for MLNs

DI-MLN (Distinct-Individuals MLN): A model similar to MLNs, but:

1. Simplified formulae:
   syntax: no "="
   semantics: simplified semantics for $\exists$ & $\forall$, denoted by $\exists_{\neq}$ and $\forall_{\neq}$
2. Its formulae are only instantiated when all free variables are assigned distinct individuals.

Theorem 1:

MLNs $\xLeftrightarrow{\text{equivalent (equally expressive)}}$ DI-MLNs

MLNs with no $\exists, \forall$ $\xLeftrightarrow{\text{equivalent (equally expressive)}}$ DI-MLNs with no $\exists_{\neq}, \forall_{\neq}$

Quantifiers = "Exceptions"

Theorem 3:

For any [DI-]formula φ over $\{A(x), B\}$ there is a [DI-]formula φ_{unq} **with no quantifiers**, such that

$$\text{whenever} \quad n_0, n_1 \geq \text{threshold}_{\varphi} \quad \Longrightarrow \quad \varphi_{unq} \equiv \varphi$$

(I.e.: $\exists$ and $\forall$ only allow "exceptions" when n_0 <threshold or n_1 <threshold.)

Example 3: $\underbrace{\forall_{\neq} x A(x)}_{\varphi} \equiv \underbrace{\mathit{false}}_{\varphi_{unq}}$ whenever $n_0, n_1 \geq \underbrace{1}_{\text{threshold}_{\varphi}}$.

Overview of Theorem 2 & 4's Proofs

- ◇ Ignore the original MLN's formulae; treat $\mathbb{F}$ as an MLN over $\{A(x), B\}$.
- ◇ Theorems 2 & 4 are **negative** results $\Longrightarrow$ must show **for all possible MLNs** $\mathbb{F}$.
- ◇ Approach: **simplify representation until results can be proved analytically.**

Theorem 2:

MLN $\mathbb{F}$ w/o quantifiers $\xLeftrightarrow[\text{(Thm 1)}]{\text{equivalent}}$ DI-MLN w/o quantifiers $\xLeftrightarrow{\text{equivalent}}$ parametric representation $\xRightarrow{\text{prove thm 2 using limits } n\to\infty \text{ (not trivial)...}}$ Q.E.D.

syntax: $\neg, \wedge, \vee, =$

syntax: $\neg, \wedge, \vee$
semantics: simpler

Theorem 4:

MLN $\mathbb{F}$ $\xRightarrow[\text{(based on Thm 3)}]{\text{can be represented as (when } n_0, n_1 \geq \text{threshold)}}$ MLN w/o quantifiers $\xRightarrow{\text{continue similarly to Thm 2's proof}}$ Q.E.D.

syntax: $\neg, \wedge, \vee, =, \exists, \forall$

syntax: $\neg, \wedge, \vee, =$

Implications

Our Results for Common Binary Aggregators:

	Aggregator	$\mathbf{P}(B=1 \mid n_0, n_1)$	Parameters	Thm 2	Thm 4
	AND	$\mathbb{I}_{n_0=0}$		no	yes
	OR	$\mathbb{I}_{n_1\geq 1}$		no	yes
Deterministic:	At least t 1's	$\mathbb{I}_{n_1\geq t}$	$t \geq 0$	no	yes
	XOR	$\frac{1+(-1)^{n_1}}{2}$		no	no
	At least t% 1's	$\mathbb{I}_{n_1/(n_0+n_1) \geq t}$	$0 < t < 1$	no	no
"Hybrid":	Majority	$\frac{1}{2}\mathbb{I}_{n_1=n_0} + \mathbb{I}_{n_1>n_0}$		no	no
	Noisy AND	α^{n_0}	$\alpha \in [0,1]$	no	approximated
	Noisy OR	$1-\alpha^{n_1}$	$\alpha \in [0,1]$	no	approximated
Probabilistic:	Random mux (or "average")	$\frac{n_1}{n_0+n_1}$		no	no
	Logistic regression	$1/\left(1+e^{-(w+w_0n_0+w_1n_1)}\right)$	w, w_0, w_1	no	no(approx if $w_0w_1\geq 0$)
	Relational logistic regr. (RLR)	$1/\left(1+e^{-\sum_{\langle L,F,w\rangle} w \sum_L F_{\Pi,x\to X}}\right)$	$\{\langle L, F, w\rangle\}$	no	no(approx if ...)

Our mathematical tools may facilitate further theoretic results. E.g.,

Theorem 5: Relational logistic regression (RLR) with quantifiers for $\mathbf{P}(B \mid A_{1..n})$ represents a "sigmoid of a polynomial of counts" when $n_0, n_1 \geq$ threshold.

◇ Kazemi, S. M.; Buchman, D.; Kersting, K.; Natarajan, S.; Poole, D. *Relational logistic regression.* In 14th International Conference on Principles of Knowledge Representation and Reasoning (KR 2014).

Results vs. Literature

Paper/Setting:		Poole et al. (2012)	Thm 2	Thm 4	Natarajan et al. (2010)
	Basic Model:	MLN	MLN	MLN	MLN
	Population size:	unbounded	unbounded	unbounded	unbounded
	Hard constraints allowed?	yes	yes	yes	yes
Model:	# free vars per formula:	$0-1$	$0-\infty$	$0-\infty$	$0-\infty$
	"=" allowed?	**no** (1 lo. var)	yes	yes	yes
	$\exists, \forall$ allowed?	**no**	**no**	**yes**	**yes**
	Weights depend on n?	no	no	no	no
Auxiliary vars:		**0**	**0**	**0**	$\Theta(n)$
Summary:	Maximal factor scope:	$\{B, A_i\}$	**fixed, arbitrarily large**	$\{B, A_{1..n}\}$	$\{B, A_{1..n}$, **aux vars**$\}$
	Flexibility:	**very limited flex.**	**flexible**	**highly flexible**	**most flexible**
Result:	Result type[1]:	negative	negative	negative	**positive**
	Which aggregators can be added as dependent PRVs?	none[2]	**none**[2]	**only saturated aggregators**	at least some

[1] A positive result (e.g., "some") is stronger when allowing less flexibility. A negative result (e.g., "none" or "only saturated") is stronger when allowing more flexibility.

[2] Some dependencies on n itself, but not on $A_{1..n}$, can be modeled.

◇ Poole, D.; Buchman, D.; Natarajan, S.; Kersting, K. *Aggregation and population growth: The relational logistic regression and Markov logic cases.* In Proc. UAI Workshop on Statistical Relational AI, 2012.

◇ Natarajan, S.; Khot, T.; Lowd, D.; Tadepalli, P.; Kersting, K. *Exploiting causal independence in Markov logic networks: Combining undirected and directed models.* In European Conference on Machine Learning (ECML), 2010.