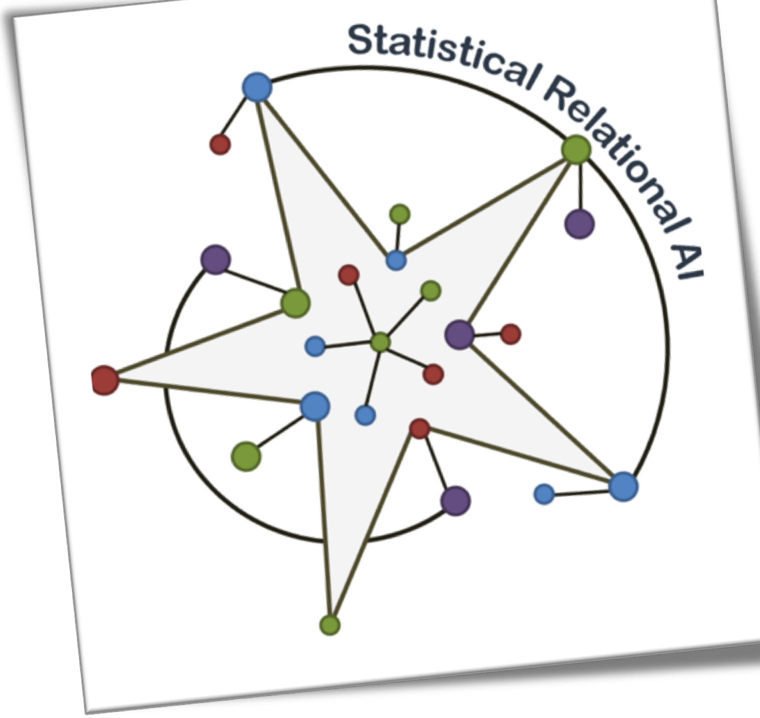
Dynamic StarAI

Dynamic Models and Statistical Relational AI

Tutorial at KI 2019



Tanya Braun, Marcel Gehrke, Ralf Möller
Universität zu Lübeck

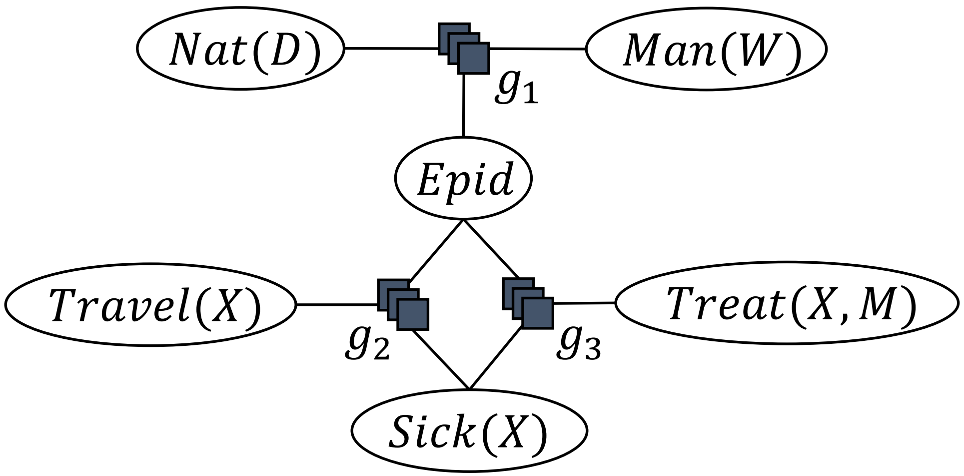
UNIVERSITÄT ZU LÜBECK

Wrap Up

- Probabilistic relational models (PRMs) (Ralf)
- Answering static queries (Tanya)
- Answering continuous queries (Marcel)
- Take home messages (Ralf)
  - Putting PRMs in context
  - LJT and LDJT research
    relevant for all variants of PRMs

Encoding a Joint Distribution

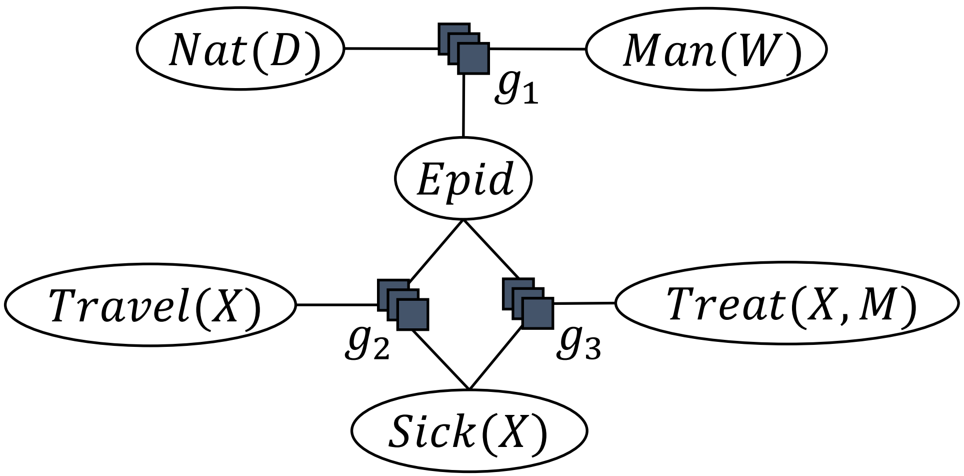
$Travel(X)$	$Epid$	$Sick(X)$	g_2
false	*false*	*false*	5
~~*false*~~	~~*false*~~	~~*true*~~	~~0~~
false	*true*	*false*	4
false	*true*	*true*	6
true	*false*	*false*	4
true	*false*	*true*	6
true	*true*	*false*	2
true	*true*	*true*	9



Nat(D) = natural disaster (D)
Man(W) = man-made disaster (W)

Semantics of a PRM

- Joint probability distribution
  P_G by grounding



$$P_G = \frac{1}{Z} \prod_{f \in gr(G)} f$$

$$Z = \sum_{v \in r(rv(gr(G)))} \prod_{f \in gr(G)} f_i(\pi_{rv(f_i)}(v))$$

$\pi_{variables}(v)$ = projection of v onto *variables*

Probabilistic Relational Models

- Distribution semantics
  (aka grounding or Herbrand semantics) [Sato 95]
  Completely define discrete joint distribution by "factorization"
  Logical atoms treated as random variables
  - Probabilistic extensions to Datalog [Schmidt et al. 90, Dantsin 91, Ng & Subramanian 93, Poole et al. 93, Fuhr 95, Rölleke & Fuhr 98]
  - Primula [Jaeger 95]
  - ProbLog [De Raedt et al. 07]

LJT/LDJT

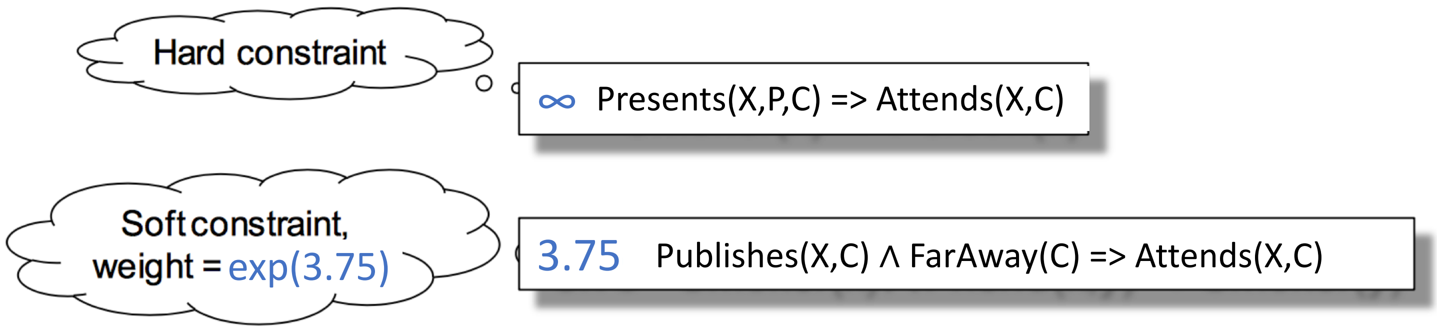
- Define junction tree and relational snapshot interface such that grounding is avoided (e.g., with “node merging” strategies)
- Anwer queries w.r.t. single nodes
(fewer atoms to consider)
- Do we need to compute Z ?
- For dynamic (sequential) reasoning,
copy junction tree pattern, not the model
  - Tradeoff: memory requirements vs. performance
  - Generalization required for reasoning (“taming evidence”)

PRMs and variants

- Probabilistic Relational Models (raw PRMs)
  [Poole 03, Tagipour et al. 13]
- Markov Logic Networks (MLNs) [Richardson and Domingos 06]
  - Use logical formulas to specify potential functions
- Probabilistic Soft Logic (PSL) [Bach, Broecheler, Huang Getoor et al. 17]
  - Use density functions to specify potential functions

Markov Logic Networks (MLNs)

- Weighted formulas for modelling constraints [Richardson & Domingos 06]



- An MLN is a set of constraints $(w, \Gamma(\boldsymbol{x}))$
  - w = weight
  - $\Gamma(\boldsymbol{x})$ = FO formula
- $weight$ of a world = product of $\exp(w)$ terms
  - for all MLN rules $(w, \Gamma(\boldsymbol{x}))$ and groundings $\Gamma(\boldsymbol{a})$ that hold in that world
- Probability of a world $= \frac{weight}{Z}$
  - Z = sum of weights of all worlds (no longer a simple expression!)

Why exp?

- Log-linear models
- Let D be a set of constants and $\omega \in \{0,1\}^m$ a world with m atoms w.r.t. D

$$weight(\omega) = \prod_{\{(w,\Gamma(\boldsymbol{x}))\in MLN \mid \exists \boldsymbol{a}\in D^n : \omega\vDash\Gamma(\mathbf{a})\}} \exp(w)$$

$$\ln(weight(\omega)) = \sum_{\{(w,\Gamma(\boldsymbol{x}))\in MLN \mid \exists \boldsymbol{a}\in D^n : \omega\vDash\Gamma(\mathbf{a})\}} w$$

  - Sum allows for component-wise optimization during weight learning
- $Z = \sum_{\omega\in\{0,1\}^m} \ln(weight(\omega))$
- $P(\omega) = \frac{\ln(weight(\omega))}{Z}$

Factorization

Log-linear representation

MLN Reasoning

- Answering queries for *m* discrete atoms

$$P(Attends(alice)) = \sum_{\omega \in \{FALSE, TRUE\}^m \wedge \omega \vDash Attends(alice)} P(\omega)$$

- Approximated lifted reasoning for MLNs

  Lifted Weighted Model Counting [Gogate and Domingos 11, 16]
  Lifted MAP [Sarkhel, Venugopal et al. 14]

Transforming MLNs into PRMs …

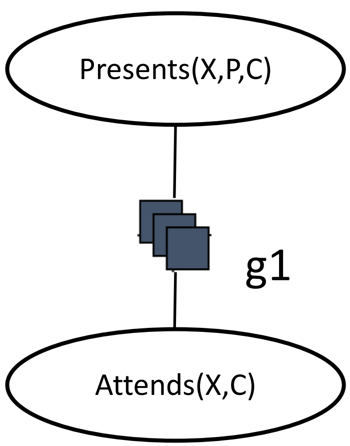
[van den Broeck 13]

Hard constraint

∞ Presents(X,P,C) => Attends(X,C)

Soft constraint, weight = exp(3.75)

3.75 Publishes(X,C) ∧ FarAway(C) => Attends(X,C)



Presents(X,P,C)	Attends(X,C)	g1
TRUE	TRUE	1000
TRUE	FALSE	1
FALSE	TRUE	1000
FALSE	FALSE	1000

$(X, P, C) \in Dom(X) \times Dom(P) \times Dom(C)$

Possibly shattering required

Transforming MLNs into PRMs …

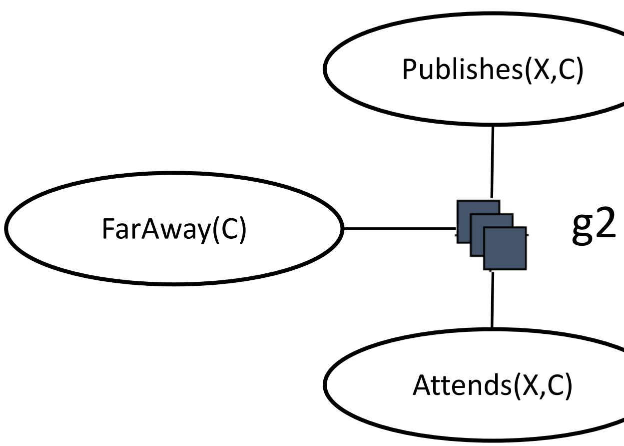
[van den Broeck 13]

Hard constraint

∞ Presents(X,P,C) => Attends(X,C)

Soft constraint, weight = exp(3.75)

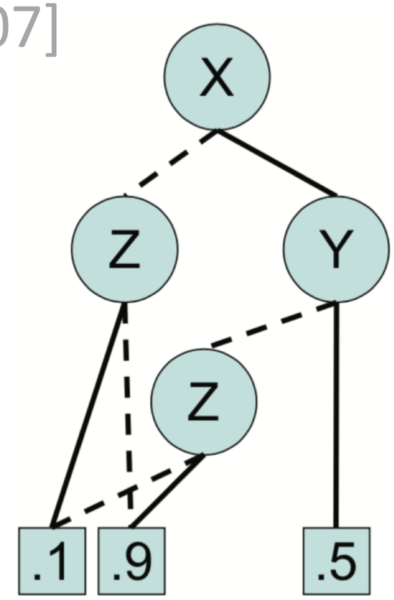
3.75 Publishes(X,C) ∧ FarAway(C) => Attends(X,C)



Publishes(X,C)	FarAway(C)	Attends(X,C)	g2
TRUE	TRUE	TRUE	exp(3.75)
TRUE	FALSE	FALSE	exp(3.75)
FALSE	TRUE	FALSE	exp(3.75)
FALSE	FALSE	TRUE	exp(3.75)
~~TRUE~~	~~TRUE~~	~~FALSE~~	~~1~~
TRUE	FALSE	TRUE	exp(3.75)
FALSE	TRUE	TRUE	exp(3.75)
FALSE	FALSE	FALSE	exp(3.75)

… and back?

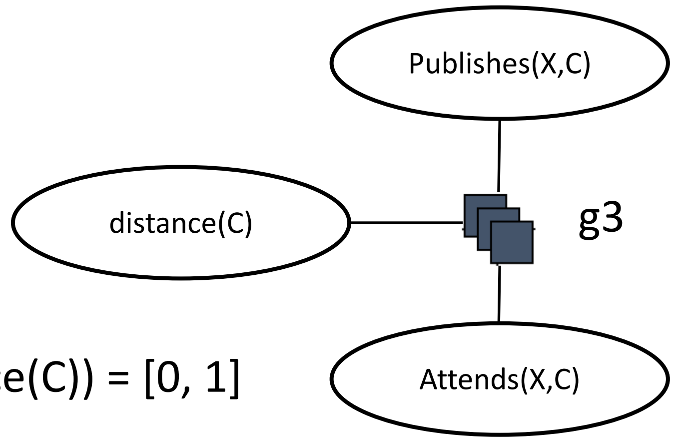
- Represent/approximate potential functions using formulas with parfactor input atoms, e.g., for (approximate) explanation purposes
- Alternative: Represent potential functions using algebraic decision diagrams (ADDs) [Chavira and Darwiche 07]
  - Requires "multiplications" used for elimination to be defined over ADDs



X	Y	Z	$f(.)$
x_1	y_1	z_1	0.9
x_1	y_1	z_2	0.1
x_1	y_2	z_1	0.9
x_1	y_2	z_2	0.1
x_2	y_1	z_1	0.1
x_2	y_1	z_2	0.9
x_2	y_2	z_1	0.5
x_2	y_2	z_2	0.5

Probabilistic Soft Logic (PSL)

- What if ranges of atoms are real-valued?
- Need to define potential function as density function



range(distance(C)) = [0, 1]
0 = “close”
1 = “far away”

[Bach, Broecheler, Huang Getoor et al. 17]

Probabilistic Soft Logic (PSL): Example

- First Order Logic weighted rules

$$0.3 : friend(B, A) \wedge votesFor(A, P) \rightarrow votesFor(B, P)$$

$$0.8 : spouse(B, A) \wedge votesFor(A, P) \rightarrow votesFor(B, P)$$

- Evidence

I(friend(John,Alex)) = 0.5 I(spouse(John,Mary)) = 0.8

I(votesFor(Alex,Romney)) = 0.6 I(votesFor(Mary,Obama)) = 0.1

- Inference
  - I(votesFor(John, Obama)) = ?
  - I(votesFor(John, Romney)) = ?

[Bach, Broecheler, Huang Getoor et al. 17]

PSL's Interpretation of Logical Connectives

- Continuous truth assignment
  Łukasiewicz relaxation of AND, OR, NOT
  - $I(\ell_1 \wedge \ell_2) = \max\{0, I(\ell_1) + I(\ell_2) - 1\}$
  - $I(\ell_1 \vee \ell_2) = \min\{1, I(\ell_1) + I(\ell_2)\}$
  - $I(\neg \ell_1) = 1 - I(\ell_1)$
- Distance to satisfaction d
  - Implication: $\ell_1 \rightarrow \ell_2$ is satisfied iff $I(\ell_1) \leq I(\ell_2)$
  - $d = \max\{0, I(\ell_1) - I(\ell_2)\}$
- Example
  - $I(\ell_1) = 0.3, I(\ell_2) = 0.9 \Rightarrow d = 0$
  - $I(\ell_1) = 0.9, I(\ell_2) = 0.3 \Rightarrow d = 0.6$

Satisfied

Rather not satisfied

Digression: Maximum Entropy Principle

- Given:
  - States $s = s_1, s_2, \ldots, s_n$
  - Density $p_s = (p_1, p_2, \ldots, p_n)$
- Maximum Entropy Principle:
  - W/o further information, select p_s , s.t. extropy H is maximized

$$H = -\sum_s p_s(s) \, log \, p_s(s) = -p_s \log p_s$$

  - w.r.t. k constraints (expected values of features fi)

$$\sum_s p_s(s) \, f_i(s) = D_i \quad i \in \{1..k\}$$

Minimize component-wise weighted difference:

$$\sum_i \lambda_i \left(\sum_s p_s(s) \, f_i(s) - D_i \right)$$

Maximum Entropy Principle

- Maximize Lagrange functional for determining p_s

$$L = -p_s \log p_s - \sum_i \lambda_i \left(\sum_s p_s(s)\, f_i(s) - D_i\right) - \mu\left(\sum_s p_s(s) - 1\right)$$

Probability values should sum up to 1

- Partial derivatives of L w.r.t. p_s → roots:

$$p_s(s) = \frac{\exp\left(-\sum_i \lambda_i f_i(s)\right)}{Z}$$

Features

Weights

where Z is for normalization

(Boltzmann-Gibbs distribution)

- "Global" modeling: additions/changes to constraints/rules influence the whole joint probability distribution

PSL Probability Distribution

Density function *f(I)* for assignment *I*
in the style of Boltzmann-Gibbs distributions:

$$f(I) = \frac{1}{Z}\exp[-\sum_{r \in R} \lambda_r (d_r(I))]$$

- a possible <u>continuous</u> truth assignment
- Normalization constant
- For all rules
- Weight of formula *r*
- Distance to satisfaction of rule r

Suggestion:

- Compile model into PRM with junction tree representation
- Variable elimination on subsets of model
  - Algebraic convolution of density functions *f*
  - Compilation into convex optimization problem

PRMs: Take Home Messages

- Knowledge representation with PRMs
  - Real-valued range values (PSL)
  - Discrete range values, adopt propositional logic (MLNs)
  - Discrete range values, drop propositional logic (raw PRMs)
- Our PRM research might be relevant for all formalisms
  - (Dynamic) junction tree provides for smaller submodels that need to be handled (also holds for knowledge compilation)
  - Organization of junction tree w.r.t. grounding avoidance
  - Taming evidence in the dynamic case
- PRMs are an active area of research
  - Generalized sequential structures
  - Compilation approaches
  - Varying domains (towards first-order structures)
  - Transfer, Explanation

Bibliography

- [Chavira and Darwiche 07]
  Mark Chavira, Adnan Darwiche. Compiling Bayesian Networks Using Variable Elimination. In: Proc. IJCAI 2007: 2443-2449
- [De Raedt et al. 07]
  Luc De Raedt, Angelika Kimmig, and Hannu Toivonen. ProbLog: A Probabilistic Prolog and its Application in Link Discovery. In: Proc. IJCAI 2007: 2062–2467
- [Domingos et al. 06]
  Matthew Richardson, Pedro Domingos. Markov logic networks. In: J. Machine Learning. Band 62. Nr. 1-2. 2006. 107–136
- [Gogate and Domingos 11]
  Vibhav Gogate and Pedro Domingos. Probabilistic Theorem Proving. In: Proc. UAI 2011: 256–265
- [Jaeger 95]
  Manfred Jaeger. Minimum Cross-Entropy Reasoning: A Statistical Justification. IJCAI 1995: 1847-1852
- [Bach, Broecheler, Huang Getoor et al. 17]
  Stephen H. Bach, Matthias Broecheler, Bert Huang, Lise Getoor. Hinge-Loss Markov Random Fields and Probabilistic Soft Logic. In: J. Mach. Learn. Res. 18:. 109:1-109:67, 2017
- [Poole 03]
  David Poole. First-order probabilistic inference. IJCAI 2003: 985-991
- [Tagipour et al. 13]
  Nima Taghipour, Daan Fierens, Jesse Davis, and Hendrik Blockeel. Lifted Variable Elimination: Decoupling the Operators from the Constraint Language. Journal of Artificial Intelligence Research, 47(1):393–439, 2013

Bibliography

- [Schmidt et al. 90]
  Schmidt, H.; Steger, N.; Güntzer, U.; Kiessling, W.; Azone, R.; Bayer, R., Combining Deduction by Certainty with the Power of Magic. In: Deductive and Object-Oriented Databases, pages 103–122. Elsevier Science Publishers, North-Holland, 1990
- [Dantsin 91]
  Dantsin, E., Probabilistic logic programs and their semantics. In: Russian Conference on Logic Programming, volume 592 of LNCS, pages 152–164. Springer, 1991
- [Ng & Subramanian 93]
  Ng, R.; Subrahmanian, V. S., A Semantical Framework for Supporting Subjective and Conditional Probabilities in Deductive Databases. Journal of Automated Reasoning 10, pages 191–235, 1993
- [Poole et al. 93]
  Poole, D., Logic Programming, Abduction and Probability. New Generation. Computing 11(3), pages 377–400, 1993
- [Fuhr 95]
  N. Fuhr, Probabilistic Datalog - A Logic For Powerful Retrieval Methods. Prod. SIGIR, pp. 282-290, 1995
- [Rölleke & Fuhr 98]
  Thomas Rölleke; Norbert Fuhr, Information Retrieval with Probabilistic Datalog. In: Logic and Uncertainty in Information Retrieval: Advanced models for the representation and retrieval of information, 1998
- [Sarkhel, Venugopal et al. 14]
  Somdeb Sarkhel, Deepak Venugopal, Parag Singla, Vibhav Gogate:. Lifted MAP Inference for Markov Logic Networks. AISTATS 2014: 859-867
- [Sato 95]
  Sato, T., A statistical learning method for logic programs with distribution semantics,
  In: International Conference on Logic Programming, pages 715–729, 1995
- [van den Broeck 13]
  Guy Van den Broeck. Lifted Inference and Learning in Statistical Relational Models, PhD thesis, KU Leuven, 2013

The End *

*PRMs are a true backbone of AI, and this tutorial emphasized only some central topics. We definitely have not cited all publications relevant to the whole field of PRMs here. We would like to thank all our colleagues for making their slides available (see some of the references to papers for respective credits). Slides or parts of it are almost always modified.