Lifted Probabilistic Inference in Relational Models

Guy Van den Broeck
UCLA

Dan Suciu
U. of Washington

IJCAI Tutorial
July 10, 2016
About the Tutorial

Extensive bibliography at the end.
Your speakers:

http://web.cs.ucla.edu/~guyvdb/

https://homes.cs.washington.edu/~suciu/

I work in AI

I work in DB
About the Tutorial

• The tutorial is about
  – deep connections between AI and DBs
  – a unified view on probabilistic reasoning
  – a logical approach to prob. reasoning

• The tutorial is NOT an exhaustive overview of lifted algorithms for graphical models (see references at the end)
If you want more…

• Books
  – Probabilistic Databases
  – Statistical Relational AI
  – (Lifted Inference Book)

• StarAI workshop on Monday
  http://www.starai.org

• Main conference papers
Outline

• Part 1: Motivation
• Part 2: Probabilistic Databases
• Part 3: Weighted Model Counting
• Part 4: Lifted Inference for WFOMC

• Part 5: Completeness of Lifted Inference
• Part 6: Query Compilation
• Part 7: Symmetric Lifted Inference Complexity
• Part 8: Open-World Probabilistic Databases
• Part 9: Discussion & Conclusions
Outline

• Part 1: Motivation
• Part 2: Probabilistic Databases
• Part 3: Weighted Model Counting
• Part 4: Lifted Inference for WFOMC
• Part 5: Completeness of Lifted Inference
• Part 6: Query Compilation
• Part 7: Symmetric Lifted Inference Complexity
• Part 8: Open-World Probabilistic Databases
• Part 9: Discussion & Conclusions
Part 1: Motivation

• Why do we need relational representations of uncertainty?

• Why do we need probabilistic queries?

• Why do we need lifted inference algorithms?
Why Relational Data?

• Our data is already relational!
  – Companies run relational databases
  – Scientific data is relational:
    • Large Hadron Collider generated 25PB in 2012
    • LSST Telescope will produce 30TB per night

• Big data is big business:
  – Oracle: $7.1BN in sales
  – IBM: $3.2BN in sales
  – Microsoft: $2.6BN in sales

[Gartner’06]
Why Probabilistic Relational Data?

• Relational data is increasingly probabilistic
  – NELL machine reading (>50M tuples)
  – Google Knowledge Vault (>2BN tuples)
  – DeepDive (>7M tuples)

• Data is inferred from unstructured information using statistical models
  – Learned from the web, large text corpora, ontologies, etc.
  – The learned/extracted data is relational

[Carlson’10, Dong’14, Niu’12]
Information Extraction

PhD Students Luc De Raedt
- Laura-Andrea Antanas (co-promotor Tinne Tuytelaars)
- Dries Van Daele (co-promotor Kathleen Marchal)
- Thanh Le Van (co-promotor Kathleen Marchal)
- Bogdan Moldovan
- Davide Nitti (co-promotor Tinne De Laet)
- José Antonio Oramas Mogrocola (key supervisor Tinne Tuytelaars)
- Francesco Orsini (co-supervisor Paolo Frasconi)
- Sergey Paramonov
- Joris Renkens
- Mathias Verbeke (with Bettina Berendt)
- Jonas Vlasselaer

Alumni Luc De Raedt

PublishedWith

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luc</td>
<td>Laura</td>
<td>0.7</td>
</tr>
<tr>
<td>Luc</td>
<td>Hendrik</td>
<td>0.6</td>
</tr>
<tr>
<td>Luc</td>
<td>Kathleen</td>
<td>0.3</td>
</tr>
<tr>
<td>Luc</td>
<td>Paol</td>
<td>0.3</td>
</tr>
<tr>
<td>Luc</td>
<td>Paolo</td>
<td>0.1</td>
</tr>
</tbody>
</table>
Extraction is so Noisy!
Representation: Probabilistic Databases

- Tuple-independent probabilistic databases

<table>
<thead>
<tr>
<th>Name</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brando</td>
<td>0.9</td>
</tr>
<tr>
<td>Cruise</td>
<td>0.8</td>
</tr>
<tr>
<td>Coppola</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Actor</th>
<th>Director</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brando</td>
<td>Coppola</td>
<td>0.9</td>
</tr>
<tr>
<td>Coppola</td>
<td>Brando</td>
<td>0.2</td>
</tr>
<tr>
<td>Cruise</td>
<td>Coppola</td>
<td>0.1</td>
</tr>
</tbody>
</table>

- Query: SQL or First-order logic

```sql
SELECT Actor.name
FROM Actor, WorkedFor
WHERE Actor.name = WorkedFor.actor
```

Q(x) = ∃y Actor(x) ∧ WorkedFor(x,y)
Why Probabilistic Queries?

> 570 million entities
> 18 billion tuples
What we’d like to do...

Has anyone published a paper with both Erdos and Einstein

Erdős number - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Erd%C5%91s_number  Wikipedia
He published more papers during his lifetime (at least 1,525) than any other ...
Anybody else's Erdős number is k + 1 where k is the lowest Erdős number of any coauthor. ... Albert Einstein and Sheldon Lee Glashow have an Erdős number of 2. ... and mathematician Ruth Williams, both of whom have an Erdős number of 2.

Erdős–Bacon number - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Erd%C5%91s%E2%80%93Bacon_number  Wikipedia
This article possibly contains previously unpublished synthesis of published ... Her paper gives her an Erdős number of 4, and a Bacon number of 2, both of ...
Erdős is in the Knowledge Graph

Paul Erdős - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Paul_Erdős
Paul Erdős was a Hungarian Jewish mathematician. He was one of the most prolific mathematicians of the 20th century. He was known both for his social ...
Fan Chung - Ronald Graham - Béla Bollobás - Category:Paul Erdős

The Man Who Loved Only Numbers - The New York Times
https://www.nytimes.com/books/.../hoffman-man.ht...
Paul Erdős was one of those very special geniuses, the kind who comes along only once in a very long while yet he chose, quite consciously I am sure, to share ...
This guy is in the Knowledge Graph

... and he published with both Einstein and Erdos!
Desired Query Answer

Has anyone published a paper with both Erdos and Einstein

⭐⭐⭐⭐⭐
Ernst Straus

⭐⭐⭐⭐⭐
Kristian Kersting, ...

⭐⭐⭐⭐⭐
Justin Bieber, ...
Observations

- Cannot come from labeled data
- Fuse uncertain information from many pages
- Expose uncertainty in query answers
  … and risk incorrect answers
- Embrace probability!
Siri, Alexa and Other Virtual Assistants Put to the Test

Tech Fix
By BRIAN X. CHEN  JAN. 27, 2016

WHEN I asked Alexa earlier this week who was playing in the Super Bowl, she responded, somewhat monotonously, “Super Bowl 49’s winner is New England Patriots.”

“Come on, that’s last year’s Super Bowl,” I said. “Even I can do better than that.”

At the time, I was actually alone in my living room. I was talking to the virtual companion inside Amazon’s wireless speaker, Echo, which was released last June. Known as Alexa, she has gained raves from Silicon Valley’s tech-obsessed digerati and has become one of the newest members of the virtual assistants club.

All the so-called Frightful Five tech
Summary

Statistical

Logical

Propositional

Relational

Databases

Prob. DBs
Representations in AI and ML

Rain $\Rightarrow$ Cloudy
### Graphical Model Learning

Medical Records

<table>
<thead>
<tr>
<th>Name</th>
<th>Cough</th>
<th>Asthma</th>
<th>Smokes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Bob</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Charlie</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Dave</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Eve</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frank</th>
<th>1</th>
<th>?</th>
<th>?</th>
</tr>
</thead>
</table>

**Big data**

| Frank | 1 | 0.3 | 0.2 |

Bayesian Network

- Asthma
- Smokes
- Cough

**Medical Records**

**Bayesian Network**

**Graphical Model Learning**
Representations in AI and ML

| Cloudy | Rain | P(Rain|Cloudy) |
|--------|------|-------------|
| T      | T    | 0.80        |
| T      | F    | 0.20        |
| F      | T    | 0.01        |
| F      | F    | 0.99        |

Rain ⇒ Cloudy
Relational Representations

- Example: First-Order Logic
  
  Logical variables have domain of constants
  \( x,y \) range over domain People = \{Alice, Bob\}

- Ground formula has no logical variables
  
  Smokes(Alice) \land Friends(Alice, Bob) \Rightarrow Smokes(Bob)
Representations in AI and ML

### Statistical

| Cloudy | Rain | P(Rain|Cloudy) |
|--------|------|-------------|
| T      | T    | 0.80        |
| T      | F    | 0.20        |
| F      | T    | 0.01        |
| F      | F    | 0.99        |


### Logical

- \( \text{Rain} \Rightarrow \text{Cloudy} \)

### Propositional

### Relational

- \( \forall x, \forall y, \text{Smokes}(x) \land \text{Friends}(x,y) \Rightarrow \text{Smokes}(y) \)
Why Statistical Relational Models?

- Probabilistic graphical models
  - Quantify uncertainty and noise
  - Not very expressive
    - Rules of chess in ~100,000 pages

- First-order logic
  - Very expressive
    - Rules of chess in 1 page
  - Good match for abundant relational data
  - Hard to express uncertainty and noise
Graphical Model Learning

Medical Records

<table>
<thead>
<tr>
<th>Name</th>
<th>Cough</th>
<th>Asthma</th>
<th>Smokes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Bob</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Charlie</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Dave</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Eve</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>Cough</th>
<th>Asthma</th>
<th>Smokes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frank</td>
<td>1</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frank</th>
<th>Cough</th>
<th>Friends</th>
<th>Brothers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frank</td>
<td>1</td>
<td>0.3</td>
<td>0.2</td>
</tr>
<tr>
<td>Frank</td>
<td>1</td>
<td>0.2</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Rows are independent during learning and inference!
Augment graphical model with relations between entities (rows).

**Intuition**

- Friends have similar smoking habits
- Asthma can be hereditary

**Markov Logic**

1.9 $\text{Smokes}(x) \land \text{Friends}(x,y) \Rightarrow \text{Smokes}(y)$

1.5 $\text{Asthma}(x) \land \text{Family}(x,y) \Rightarrow \text{Asthma}(y)$

+ Asthma can be hereditary
Classical Machine Learning

<table>
<thead>
<tr>
<th>Name</th>
<th>Age</th>
<th>Product</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dave</td>
<td>40</td>
<td>Android</td>
<td>€249</td>
</tr>
<tr>
<td>Alice</td>
<td>35</td>
<td>iPhone</td>
<td>€799</td>
</tr>
<tr>
<td>Bob</td>
<td>32</td>
<td>iPhone</td>
<td>€799</td>
</tr>
<tr>
<td>Charlie</td>
<td>22</td>
<td>iPhone</td>
<td>€699</td>
</tr>
<tr>
<td>Eve</td>
<td>17</td>
<td>Android</td>
<td>€299</td>
</tr>
<tr>
<td>Frank</td>
<td>15</td>
<td>Android</td>
<td>€199</td>
</tr>
</tbody>
</table>

People *older* than 27 probably buy *iPhone*.

People *younger* than 27 probably buy *Android*.

**Inference:** Does Guy buy an iPhone?

**Answer:** Yes, with probability 66%
### Statistical Relational Learning

<table>
<thead>
<tr>
<th>Name</th>
<th>Age</th>
<th>Product</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dave</td>
<td>40</td>
<td>Android</td>
<td>€249</td>
</tr>
<tr>
<td>Alice</td>
<td>35</td>
<td>iPhone</td>
<td>€799</td>
</tr>
<tr>
<td>Bob</td>
<td>32</td>
<td>iPhone</td>
<td>€799</td>
</tr>
<tr>
<td>Charlie</td>
<td>22</td>
<td>iPhone</td>
<td>€699</td>
</tr>
<tr>
<td>Eve</td>
<td>17</td>
<td>Android</td>
<td>€299</td>
</tr>
<tr>
<td>Frank</td>
<td>15</td>
<td>Android</td>
<td>€199</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Person A</th>
<th>Person B</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>Bob</td>
<td>Spouse</td>
</tr>
<tr>
<td>Alice</td>
<td>Charlie</td>
<td>Mother</td>
</tr>
<tr>
<td>Bob</td>
<td>Charlie</td>
<td>Father</td>
</tr>
<tr>
<td>Dave</td>
<td>Eve</td>
<td>Father</td>
</tr>
<tr>
<td>Dave</td>
<td>Frank</td>
<td>Father</td>
</tr>
<tr>
<td>Eve</td>
<td>Frank</td>
<td>Siblings</td>
</tr>
</tbody>
</table>

**Family members** probably buy the **same type** of phone.
Example: Markov Logic

- Weighted First-Order Logic

Weight or Probability

FOL Formula

3.14 \( \text{Smokes}(x) \land \text{Friends}(x,y) \Rightarrow \text{Smokes}(y) \)

- Ground atom/tuple = random variable in \{true,false\} e.g., Smokes(Alice), Friends(Alice,Bob), etc.
- Ground formula = factor in propositional factor graph

[Richardson’06]
Representations in AI and ML

| Cloudy | Rain | P(Rain|Cloudy) |
|--------|------|-------------|
| T      | T    | 0.80        |
| T      | F    | 0.20        |
| F      | T    | 0.01        |
| F      | F    | 0.99        |

Rain ⇒ Cloudy

∀x, ∀y, Smokes(x) ∧ Friends(x,y)⇒ Smokes(y)
Collective Classification

Can we predict the type of the nodes given information on its links and attributes?

E.g., the type of a webpage given its links and the words on the page?

[Getoor’07] [DeRaedt’16]
Entity Resolution

Automatically extracted co-author network: which nodes refer to the same person?

[Getoor’07] [DeRaedt’16]
Viral Marketing

Which advertising strategy maximizes expected profit?

[VdBroeck’10]
Can we predict preferences?

[Bach’15]
Summary

Uncertainty in AI

Statistical

Logical

Propositional

Relational

StarAI
Summary

Uncertainty in AI

Databases

Statistical

Logical

Propositional

Relational
Why Lifted Inference?

• Main idea: exploit high level relational representation to speed up reasoning

• Let’s see an example…
A Simple Reasoning Problem

- 52 playing cards
- Let us ask some simple questions

[VdB’15]
A Simple Reasoning Problem

Probability that Card1 is Hearts? 1/4
A Simple Reasoning Problem

Probability that Card1 is Hearts given that Card1 is red? 1/2
A Simple Reasoning Problem

Probability that Card52 is Spades given that Card1 is QH? 13/51

[VdB’15]
A Simple Reasoning Problem

Probability that Card1 is Hearts? \( \frac{1}{4} \)
A Simple Reasoning Problem

Probability that Card52 is Spades given that Card1 is QH?

13/51
Automated Reasoning

Let us automate this:

1. Probabilistic graphical model (e.g., factor graph)

2. Probabilistic inference algorithm (e.g., variable elimination or junction tree)
Classical Reasoning

- Tree
- Sparse Graph
- Dense Graph

- Higher treewidth
- Fewer conditional independencies
- Slower inference
Is There Conditional Independence?

$$\Pr(\text{Card52} \mid \text{Card1}) \neq \Pr(\text{Card52} \mid \text{Card1, Card2})$$

$$\frac{13}{51} \neq \frac{12}{50}$$

$$\Pr(\text{Card52} \mid \text{Card1, Card2}) \neq \Pr(\text{Card52} \mid \text{Card1, Card2, Card3})$$

$$\frac{12}{50} \neq \frac{12}{49}$$
Automated Reasoning

Let us automate this:

1. Probabilistic graphical model (e.g., factor graph) is fully connected!

2. Probabilistic inference algorithm (e.g., variable elimination or junction tree) builds a table with $52^{52}$ rows

[VD’15]
What's Going On Here?

Probability that Card52 is Spades given that Card1 is QH? 13/51
What's Going On Here?

Probability that Card52 is Spades given that Card2 is QH? 13/51

[VdB’ 15]
What's Going On Here?

Probability that Card52 is Spades given that Card3 is QH? \[ \frac{13}{51} \]
Tractable Reasoning

What's going on here?
Which property makes reasoning tractable?

- High-level (first-order) reasoning
- Symmetry
- Exchangeability

⇒ Lifted Inference
Automated Reasoning

Let us automate this:

- **Relational** model

\[
\forall p, \exists c, \text{Card}(p,c) \\
\forall c, \exists p, \text{Card}(p,c) \\
\forall p, \forall c, \forall c', \text{Card}(p,c) \land \text{Card}(p,c') \Rightarrow c = c'
\]

- **Lifted** probabilistic inference algorithm
Other Examples of Lifted Inference

- First-order resolution

\[
\forall x, \text{Human}(x) \Rightarrow \text{Mortal}(x) \\
\forall x, \text{Greek}(x) \Rightarrow \text{Human}(x)
\]

implies

\[
\forall x, \text{Greek}(x) \Rightarrow \text{Mortal}(x)
\]
Other Examples of Lifted Inference

- First-order resolution
- Reasoning about populations

We are investigating a rare disease. The disease is more rare in women, presenting only in one in every two billion women and one in every billion men. Then, assuming there are 3.4 billion men and 3.6 billion women in the world, the probability that more than five people have the disease is

\[
1 - \sum_{n=0}^{5} \sum_{f=0}^{n} \binom{3.6 \times 10^9}{f} \left(1 - 0.5 \times 10^{-9}\right)^{3.6 \times 10^9 - f} \left(0.5 \times 10^{-9}\right)^f \times \binom{3.4 \times 10^9}{(n-f)} \left(1 - 10^{-9}\right)^{3.4 \times 10^9 - (n-f)} \left(10^{-9}\right)^{(n-f)}
\]

[VdB’ 13]
Lifted Inference in SRL

- Statistical relational model (e.g., MLN)

3.14 FacultyPage(x) ∧ Linked(x, y) ⇒ CoursePage(y)

- As a probabilistic graphical model:
  - 26 pages; 728 variables; 676 factors
  - 1000 pages; 1,002,000 variables; 1,000,000 factors

- Highly intractable?
  - Lifted inference in milliseconds!
Statistical Properties

1. Independence

\[
P(\begin{bmatrix}
1 & 1 & 0 \\
0 & 0 & 0 \\
0 & 1 & 0
\end{bmatrix}) = P(\begin{bmatrix}
1 & 1 & 0 \\
0 & 0 & 0 \\
0 & 1 & 0
\end{bmatrix})
\]

2. Partial Exchangeability

\[
P(\begin{bmatrix}
1 & 1 & 0 \\
0 & 0 & 0 \\
0 & 1 & 0
\end{bmatrix}) = P(\begin{bmatrix}
1 & 1 & 0 \\
0 & 0 & 0 \\
0 & 1 & 0
\end{bmatrix})
\]

3. Independent and identically distributed (i.i.d.)

= Independence + Partial Exchangeability
Statistical Properties for Tractability

• Tractable classes independent of representation

• Traditionally:
  – Tractable learning from i.i.d. data
  – Tractable inference when cond. independence

• New understanding:
  – Tractable learning from exchangeable data
  – Tractable inference when
    • Conditional independence
    • Conditional exchangeability
    • A combination

[Niepert’14]
Summary of Motivation

• Relational data is everywhere:
  – Databases in industry and sciences
  – Knowledge bases
  – Probabilistically extracted/learned/queried

• Lifted inference:
  – Use relational structure during reasoning
  – Very efficient where traditional methods break

This tutorial: Lifted Inference in Relational Models