Lifted Message Passing

Rorschach Test

K. Kersting
Lifted Message Passing
NeSy 2010 @ AAAI 2010, Atlanta, USA July 11, 2010
Etzioni’s Rorschach Test for Computer Scientists

Moore’s Law?
Storage Capacity?

Number of Facebook Users?
Number of Scientific Publications?

Number of Web Pages?
Computing 2020: Science in an Exponential World

“The amount of scientific data is doubling every year”
[Szalay, Gray; Nature 440, 413-414 (23 March 2006)]

How to deal with millions of images?
How to deal with millions of inter-related research papers?
How to accumulate general knowledge automatically from the Web?
How to deal with billions of shared users’ perceptions stored at massive scale?
How do realize the vision of social search?
Artificial Intelligence in an Exponential World

AI = Structured (Data + Model + Reasoning)

- Real world is structured in terms of objects and relations
- Relational knowledge can reveal additional correlations between variables of interest. Abstraction allows one to compactly model general knowledge and to move to complex inference

[Etzioni et al. ACL08] [Fergus et al. 30(11) 2008; Halevy et al., IEEE Intelligent Systems, 24 2009]

- Most effort has gone into the modeling part
- How much can the data itself help us to solve a problem?
(First-order) Logic handles Complexity

E.g., rules of chess (which is a tiny problem):
1 page in first-order logic,
~100000 pages in propositional logic,
~10000000000000000000000000000000 pages as atomic-state model

- Many types of entities
- Relations between them
- Arbitrary knowledge

Logic
true/false

Explicit enumeration

5th C B.C.  19th C
atomic  propositional  first-order/relational

Probability handles Uncertainty

- Sensor noise
- Human error
- Inconsistencies
- Unpredictability

Explicit enumeration

5th C B.C.  19th C
atomic  propositional  first-order/relational

- Many types of entities
- Relations between them
- Arbitrary knowledge

Probability

17th C  20th C

Logic
true/false
The real world is complex and uncertain

Let’s deal with uncertainty, objects, and relations jointly

... unifies logical and statistical AI,
... solid formal foundations,
... is of interest to many communities.
The real world is complex and uncertain

Today, we can …

- … learn probabilistic relational models automatically from millions of inter-related objects
- … generate optimal plans and learn to act optimally in uncertain environments involving millions of objects and relations among them
- … perform lifted probabilistic inference avoiding explicit state enumeration by manipulating first-order state representations directly
- … exploit shared factors to speed up message-passing algorithms for relational inference but also for classical propositional inference such as solving SAT problems
Lifted Inference
[Peiffer et al. 1999; Poole 2003; de Salvo Braz et al. 2005]
- Example: Inviting $n$ people to a workshop

**Factor graph**

- \( \forall X. \phi_1(\text{popular, attends}(X)) \)
- \( \forall X. \phi_2(\text{attends}(X), \text{series}) \)

- Lifted inference exploits symmetries revealed by relational model

**Variable Elimination**

Sum out non-query variables one by one

\[
\sum_{\text{att}(p_1)} \phi_1(\text{pop, att}(p_1)) \phi_2(\text{att}(p_1), \text{ser})
\]

\( \phi'(\text{pop, ser}) \)

Time is linear in number of invitees $n$
First-Order Variable Elimination

[Poole 2003; de Salvo Braz et al. 2005]

∀X. \( \phi_1(\text{popular}, \text{attends}(X)) \)

∀X. \( \phi_2(\text{attends}(X), \text{series}) \)

∀X. \( \phi'(\text{popular}, \text{series}) \)

\( \phi'(\text{popular, series})^n \)

Sum out all \( \text{attends}(X) \) variables at once

Time is constant in \( n \)

Symmetry Within Factors

[cf Zhang & Poole 1996; Gupta et al. 2007, Milch et al. 2008]

\( \text{attends}(p_1) \)

\( \text{attends}(p_2) \)

\( \cdots \)

\( \text{attends}(p_n) \)

\( \phi(\text{overflow}, \#_X[\text{attends}(X)]) \)

Size of naïve factor representation: \( 2 \times 2^n \)

- Values of counting formula are histograms counting how many objects \( X \) yield each possible value of \( \text{attends}(X) \)
  - Only \( n+1 \) histograms, e.g., [50, 0], [49, 1], ..., [0, 50]
  - Factor size now \( 2 \times (n+1) \): linear in \( n \)
Example: Competing Workshops

∀W ∀X. φ(hot(W), att(X))

Can’t sum out attends(X) without joining all the hot(W) variables

Create counting formula on hot(W), then sum out attends(X) at lifted level

Conversion to counting formulas creates new opportunities for lifted elimination

Results: Competing Workshops

- These exact inference approaches are rather complex
- so far do not easily scale to realistic domains,
- and hence have only been applied to rather small artificial problems
How do you spend your spare time?

YouTube like media portals have changed the way users access media content in the Internet.

Every day, millions of people visit social media sites such as Flickr, YouTube, and Jumpcut, among others, to share their photos and videos, …

while others enjoy themselves by searching, watching, commenting, and rating the photos and videos; what your friends like will bear great significance for you.

How do you efficiently broadcast information?

Facebook, YouTube, Google, BitTorrent
Content Distribution using Belief Propagation
[Bickson et al. 04]

- Approximate inference
- Compute the marginal $P(x_i|X)$ for each $x_i$ with local computations only
- Computer vision, combinatorial problems, SAT, NLP, ...

\[
\begin{align*}
\mu_{X \rightarrow f}(x) &= \prod_{h \in \text{ab}(X) \setminus \{f\}} \mu_{h \rightarrow X}(x) \\
\mu_{f \rightarrow X}(x) &= \sum_{\bar{x}} \left( f(x) \prod_{y \in \text{ab}(f) \setminus \{X\}} \mu_{y \rightarrow f}(y) \right)
\end{align*}
\]

Content Distribution using Belief Propagation
[Bickson et al. WDAS04]

<table>
<thead>
<tr>
<th>$\psi_{AB}(X_B)$</th>
<th>3 from A</th>
<th>1 from C</th>
<th>1 from D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform policy</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\psi_{CD}(X_C)$</th>
<th>3 from A</th>
<th>2 from C</th>
<th>3 from D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 from B</td>
<td>1/2</td>
<td>1/2</td>
<td>1/2</td>
</tr>
<tr>
<td>2 from C</td>
<td>1/2</td>
<td>1/2</td>
<td>1/2</td>
</tr>
<tr>
<td>3 from C</td>
<td>1/2</td>
<td>1/2</td>
<td>1/2</td>
</tr>
</tbody>
</table>

Table 1: Possible actions for the node B in the example graph shown in Figure 1

A lot of shared factors, so use lifted belief propagation
Lifted Belief Propagation

Counting shared factors can result in great efficiency gains for (loopy) belief propagation.

Shared factors appear more often than you think in relevant real world problems.

Step 1: Compression
Step 2: Modified Belief Propagation

\[ \mu_{T \rightarrow A}(x) = \frac{1}{Z} \prod_{b \in \mathcal{F} \setminus \{T\}} \mu_{b \rightarrow A}(x) \]

\[ \mu_{T \rightarrow x}(x) = \prod_{r \in \mathcal{R}(x)} \mu_{r \rightarrow x}(x) \]

Social Network Analysis
Social Network Analysis

Social Network Analysis
Lifted First-order Factored Frontier

- Apriori: most people do not smoke
- Apriori: most people do not have cancer
- Apriori: most people are not friends
- Smoking causes cancer
- Friends have similar smoking habits
- Most friends stay friends
- Most smokers stay smokers

20 people over 10 time steps. Max number of friends 5. Cancer never observed.

Time step randomly selected.

Lower Bound on Model Count of CNF

- BPCount [Kroc et al 08]
  - BP used to estimate marginals
  - Provable bound

Idea:
- Identify a “balanced” row split or column split (roughly equal number of solutions on each side)
  - Use marginals for estimate
  - Pick one side at random
  - Count on that side recursively
  - Multiply result by 2

[similar to decimation]
Model Counting

Satisfied by Lifted Message Passing?

Message Passing for Satisfiability

- Warning and survey propagation can also be lifted
- Enables lifted treatment of both prob. and det. knowledge

Gaussian Belief Propagation can also be lifted! Lifted Solvers for Systems of Linear Equations? Lifted Page Rank? Lifted HITS? Lifted Kalman Filter? ...
Content Distribution (Gnutella): Lifted BP vs. BP

Message Errors to the Rescue!

- **Ihler et al. 05**: BP message errors decay along paths
- LBP may spuriously assume some nodes send and receive different messages and, hence, produce **pessimistic lifted network**

Make use of decaying message errors already at lifting time
Informed Lifted Belief Propagation [El Massaoudi et al. AAAI10]

Algorithm 1: ILBP — informed Lifted BP. We use \( b_i(x_i) \) resp. \( m_i(x_i) \) to denote the unnormalized beliefs resp. messages of both variable node \( X_i \) and variable nodes covered by supernodes \( X_s \).

Data: A factor graph \( G \) with variable nodes \( X \) and factors \( f \), Evidence \( E \)
Result: Unnormalized marginals \( b_i(x_i) \) for all supernodes and, hence, for all variable nodes

1. Colorize \( X \) and \( f \) w.r.t. \( E \);
2. \( \emptyset \) — one iteration CP;
3. Initialize messages for \( \emptyset \);
4. \((b_i(x_i), m_i(x_i)) \) — one iteration MBP on \( \emptyset \);
5. Colorize all \( X_s \) according to \( m_i(x_i) \);
6. While \( b_i(x_i) \) have not converged do
7. \( \emptyset' \) — one iteration CP (based on new colors);
8. Refine BPs by supernodes using \( b_i(x_i) \) and \( m_i(x_i) \);
9. \((b_i(x_i), m(x_i)) \) — one iteration of MBP on \( \emptyset' \);
10. For each supernode \( X \) in \( \emptyset' \) do
11. If the \( m(x_i) \)s of the \( X_s \) in \( X \) differ then
12. Colorize all \( X_i \) in \( X \) according to \( m(x_i) \)
13. Return \( b_i(x_i) \) for all supernodes

Social Networks

![Graph showing average # (super) nodes vs. # Iterations for different algorithms (BP, ILBP, LBP) with varying number of friends (5, 10, no friends).]
Lifted Content Distribution

- 1 file, Gnutella snapshot
  - 10876 nodes
  - 39994 edges
- iLBP 4.272.164 mess.
- < BP 5.761.952 mess.
- < LBP 6.381.516 mess.

- On a different network:
  - iLBP 1.972.662 < LBP 2.962.311 < BP 5.761.952

Conclusions

- SRL $\geq$ Objects&Relations + Probabilities + Machine Learning
- SRL works and covers the whole AI spectrum
  - Lifted Reasoning / SAT / Message Passing
  - Relational Machine Learning
  - Relational POMDPs [Sanner, K, AAAI10]
- Relational / Symbolic Neural Networks?
  - Deep Relational Networks?
  - Relational LPs?
  - Lifted Boltzmann Machines?
  - ...
Let's explore the minimal perturbations required for each of the AI areas to start using SR techniques

- Planning: from PDDL to SRPDDL?
- MDPs: from 2-TBN models to DSRL models?
- Vision: from graphical models and scene grammars to SR generative models?
- NLP: from unification grammars to SR-unification grammars?
- KR: ontologies and event calculus to SR models?
- ...

Thanks for your attention