Reasoning about Large Populations with Lifted Probabilistic Inference

Kristian Kersting   Brian Milch   Luke S. Zettlemoyer   Michael Haimes   Leslie Pack Kaelbling
MIT Computer Science and Artificial Intelligence Laboratory, 32 Vassar St. Room 32-G480, Cambridge, MA 02139, USA
{kersting,milch,lsz,mhaimes,lpk}@csail.mit.edu

Lifted Inference

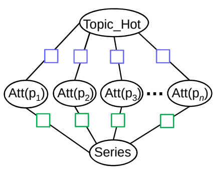
- Problem: Graphical model inference algorithms do not scale well to large populations
- Lifted inference idea:
  - Many individuals are interchangeable in model
  - Exploit that symmetry to speed up inference

[Pfeffer *et al*. 1999; Poole 2003; de Salvo Braz *et al*. 2005]

Example 1: Will My Workshop Start a Series?

Ground Representation — Factor graph

Lifted Representation — Parfactors (parameterized factors)



ϕ_1(topic_hot, attends(X))

ϕ_2(attends(X), series)

Variable Elimination (VE)

- For each invitee i = 1 to n:
  - Eliminate Att(p_i), yielding factor on {Hot, Series}
- Multiply these factors together
- Eliminate Hot to get factor on Series alone

Time: Linear in |invitees|

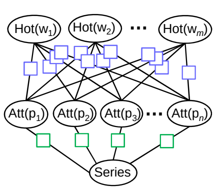
First-Order Variable Elimination (FOVE) [Poole 2003]

Eliminate attends(X) for arbitrary X, get factor on {hot, series}

Raise entries in factor to power n

Time: Constant w.r.t |invitees|

Example 2: Competing Workshops



ϕ_1(attends(X), hot(W))

ϕ_2(attends(X), series)

Assuming |invitees| > |workshops|,

VE time: O(|invitees| x 2^|workshops|)

Counting Elimination

[de Salvo Braz *et al*. 2005]

- Product of ϕ_1 factors in an outcome depends only on how many invitees attend, how many workshops are hot:

$$\prod_{X,W} \phi_1(\text{att}(X), \text{hot}(W)) = \phi_1(0,0)^{n_0 m_0}\phi_1(0,1)^{n_0 m_1}\phi_1(1,0)^{n_1 m_0}\phi_1(1,1)^{n_1 m_1}$$

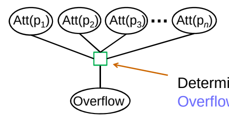
where n_a= |{X : att(X)=a}|, m_h= |{W : hot(W)=h}|

- So: Sum out all attends and hot vars simultaneously, sum over histograms rather than instantiations
- Exponential speed-up over ground VE

Time: O(|invitees| x |workshops|)

Models with Cardinality Potentials

Will the attendees overflow a 40-person room?



Deterministic potential saying Overflow ≡ (|{X : att(X)=1}| > 40)

- Tabular representation of potential is exponential in |invitees|
- But it's a cardinality potential [Gupta et al. 2007]: depends only on how many random variables have a particular value

Cardinality Potentials in Lifted Inference

- We allow parfactors to include not just ordinary atoms, but also counting terms

ϕ(overflow, $\#_X$[attends(X)])) ← counting term

- Possible values of $\#_X$[attends(X)] are histograms over attends(X) variables

- Do inference by summing over histograms

Counting Terms in General

$\#_{(X_1,\ldots,X_s\,:\,C)}$[atom$_1$,…, atom$_r$]

- Examples: $\#_X$[attends(X), position(X)]

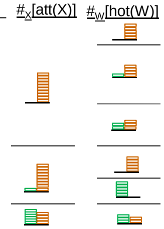
  $\#_{(X,Y:\,X\neq Y)}$[knows(X, Y)]
- Possible values are histograms
  - Bucket for each tuple of values $v_1,\ldots,v_r$ of atom$_1$,…, atom$_r$
  - Sum of entries is number of object tuples $o_1,\ldots,o_s$ that satisfy constraint C when bound to $X_1,\ldots X_s$
- Counting terms can contain free variables, e.g.

  ϕ(fame(Y), $\#_{(X:\,X\neq Y)}$[knows(X, Y)]))

  defines a factor for each person Y
- Current work: Handle such general counting terms in lifted inference

Summing over Histograms

att1…att10	h1…h5	product of ϕ_1 factors	$\#_X$[att(X)]	$\#_W$[hot(W)]
0000000000	00000	1.27e-20		
0000000000	10000	4.22e-17		
0000000000	01000	4.22e-17		
…	…	…		
0000000000	00001	4.22e-17		
0000000000	11000	1.40e-13		
…	…	…		
0000000000	00011	1.40e-13		
…	…	…		
1000000000	00000	3.87e-15		
…	…	…		
1000000000	11111	2.56e-04		
…	…	…		
1011001101	10110	5.21e-12		
…	…	…		
2^{10}=1024	2^5=32		11	6



Cardinality Potentials as Intermediate Factors

- Recall parfactor from "competing workshops" model:

  ϕ_1(attends(X), hot(W))
- Couldn't eliminate hot(W) for arbitrary W because it would yield a factor on all the attends(X) nodes (exponentially big table)
- But the resulting factor is a counting potential!

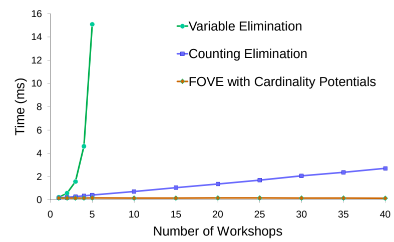
$$\sum_{\text{hot}(W)} \prod_X \phi_1(\text{att}(X), \text{hot}(W)) = \sum_{\text{hot}(W)} \phi_1(0, \text{hot}(W))^{n_0}\phi_1(1, \text{hot}(W))^{n_1}$$

Depends only on histogram [n_0, n_1] over att(X) nodes

- So: iterate over histograms, sum over hot(W) for each one, get factor ϕ'($\#_X$[attends(X)])
- Unlike earlier counting elimination approach:
  - Don't eliminate attends(X) and hot(W) simultaneously
  - No need to iterate over hot(W) histograms

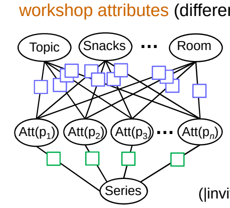
Intermediate cardinality potentials allow more flexibility in elimination ordering

Experiment: Competing Workshops



Example 3: Workshop Attributes

- Flexibility of intermediate cardinality potentials can yield exponential speed-up over standard FOVE
- Example: Attendance depends on a large number of workshop attributes (different dependence on each)



ϕ_1(attends(X), topic)

ϕ_2(attends(X), snacks)

⋮

ϕ_3(attends(X), room)

ψ(attends(X), series)

(|invitees| is large, say 100)

Standard FOVE

- Must eliminate attend(X) first to avoid huge factor
- Thus get factor on all attrs

Time: exponential in number of attributes

With Cardinality Potentials

- Eliminate each attribute in turn, yielding cardinality potentials on attend(X)

Time: linear in num. attributes and |invitees|

Relational Constraints

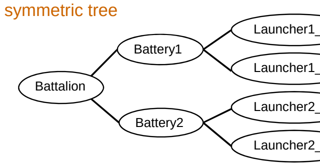
- What if probabilistic dependencies follow a social network over the invitees?

  ϕ(attends(X), attends(Y) : knows(X, Y))

  Fixed (non-random) relational constraint
- Then dependencies are not fully symmetric
  - attends(X) nodes have varying numbers of neighbors
- Current FOVE algorithm does not handle this

Exploiting More Limited Symmetry

- If dependencies have no symmetry, we're stuck
- But Pfeffer *et al*. [1999] use a lifted algorithm on a symmetric tree



- Should also be able to exploit repeated group structures across families, labs, etc.
  - "When you've seen one 2-postdoc, 6-grad-student research group, you've seen them all!"
  - Open problem: make this operational

Summary

- Lifted inference algorithms exploit symmetry to scale better with population size
  - Linear → constant, by summing out vars at abstract level
  - Exponential → polynomial, by summing over histograms
- Cardinality potentials provide additional speed-ups
  - As part of model representation
  - As intermediate factors
- Open problem: exploiting symmetries in networks where objects are distinguished

References

- R. de Salvo Braz, E. Amir, and D. Roth. "Lifted first-order probabilistic inference". In *Proc. 19th IJCAI*, pages 1319–1325, 2005.
- R. Gupta, A.A. Diwan, and S. Sarawagi. "Efficient inference with cardinality-based clique potentials". In *Proc. 24th ICML*, 2007.
- A. Pfeffer, D. Koller, B. Milch, and K. T. Takusagawa. "SPOOK: A System for Probabilistic Object-Oriented Knowledge". In *Proc. 15th UAI*, pages 541-550, 1999.
- D. Poole. "First-order probabilistic inference". In *Proc. 18th IJCAI*, pages 985–991, 2003.
- N. L. Zhang and D. Poole. "Exploiting causal independence in Bayesian network inference". *J. Artif. Intell. Res.*, 5:301–328, 1996.