Research

Sagad Hamid

Context

- Exact probabilistic inference using Variable Elimination
- Focus on asymmetrical graphical models
  - Factor graphs with discrete (currently boolean) RVs

Research Interests

Incorporate Lifting Ideas into Exact Inference in Asymmetrical Models

Increase Gap between Lifted Model and Grounded Model

Model Transformation: Approach Probabilistic Inference From Different Perspectives

Lifting in this context: Compact representation + calculations (i.e., currently not necessarily within relational context)

A First Naive Approach

- Expand the model to introduce more structure
  - Add artificial random variables to the model
  - Preserve full joint distribution

A	B	ϕ_1
1	1	a
1	0	b
0	1	c
0	0	d

$\longrightarrow$

A	B	I	ϕ_1^I
1	1	1	a_1
1	1	0	a_2
1	0	1	b_1
1	0	0	b_2
0	1	1	c_1
0	1	0	c_2
0	0	1	d_1
0	0	0	d_2

1. Added random variable I artificially
2. Summing out I yields the original factor ϕ_1

$$a_1 + a_2 = a$$
$$b_1 + b_2 = b$$
$$c_1 + c_2 = c$$
$$d_1 + d_2 = d$$

A First Naive Approach

Introduce **new factorisation**

A	B	ϕ_1
1	1	a
1	0	b
0	1	c
0	0	d

Added random variable I artificially.

A	B	I	ϕ_1^I
1	1	1	a_1
1	1	0	a_2
1	0	1	b_1
1	0	0	b_2
0	1	1	c_1
0	1	0	c_2
0	0	1	d_1
0	0	0	d_2

$\longrightarrow$

1. Replace B with new variable I
2. Introduce new Factor with B and I as arguments

A	I	ϕ_{11}^I
1	1	x_1
1	0	x_2
0	1	x_3
0	0	x_4

B	I	ϕ_{12}^I
1	1	λ_1
1	0	λ_2
0	1	λ_3
0	0	λ_4

$$x_1 \cdot \lambda_1 + x_2 \cdot \lambda_2 = a$$
$$x_1 \cdot \lambda_3 + x_2 \cdot \lambda_4 = b$$
$$x_3 \cdot \lambda_1 + x_4 \cdot \lambda_2 = c$$
$$x_3 \cdot \lambda_3 + x_4 \cdot \lambda_4 = d$$

A First Naive Approach

Introduce new factorisation using artificially added random variables

- $\phi \rightarrow \phi^I$: „Replace“ A and B by A_1 and B_1
- Introduce new factors ϕ^{A_1}, ϕ^{B_1} for „replacement“
- $x_1, \ldots, x_4$ determined by $\lambda_1, \ldots, \lambda_4$(don‘t have to be the same for ϕ^{A_1}, ϕ^{B_1})

A	B	ϕ
1	1	a
1	0	b
0	1	c
0	0	d

A_1	B_1	ϕ^I
1	1	x_1
1	0	x_2
0	1	x_3
0	0	x_4

B	B_1	ϕ^{B_1}
1	1	λ_1
1	0	λ_2
0	1	λ_3
0	0	λ_4

A	A_1	ϕ^{A_1}
1	1	λ_1
1	0	λ_2
0	1	λ_3
0	0	λ_4

A First Naive Approach

Introduce new factorisation using artificially added random variables

A	B	ϕ
1	1	a
1	0	b
0	1	c
0	0	d

$\rightarrow$

A_1	B_1	ϕ^I
1	1	x_1
1	0	x_2
0	1	x_3
0	0	x_4

B	B_1	ϕ^{B_1}
1	1	λ_1
1	0	λ_2
0	1	λ_3
0	0	λ_4

A	A_1	ϕ^{A_1}
1	1	λ_1
1	0	λ_2
0	1	λ_3
0	0	λ_4

$$x_1 = \frac{b - \frac{\lambda_4}{\lambda_2} \cdot a}{\lambda_3 - \frac{\lambda_4}{\lambda_2} \cdot \lambda_1}, \quad x_2 = \frac{b - \frac{\lambda_3}{\lambda_1} \cdot a}{\lambda_4 - \frac{\lambda_3}{\lambda_1} \cdot \lambda_2}, \quad x_3 = \frac{d - \frac{\lambda_4}{\lambda_2} \cdot c}{\lambda_3 - \frac{\lambda_4}{\lambda_2} \cdot \lambda_1}, \quad x_4 = \frac{d - \frac{\lambda_3}{\lambda_1} \cdot c}{\lambda_4 - \frac{\lambda_3}{\lambda_1} \cdot \lambda_2}$$

- Allows for choosing $\lambda_1, \dots, \lambda_4$ arbitrarily as long as division by zero is avoided
  - $x_1, \dots, x_4, \lambda_1, \dots, \lambda_4 \in \mathbb{C}$
- As long as we sum out A_1, B_1 (which will happen since artificially added RVs do not appear in query nor evidence) the full joint over A, B is preserved

A First Naive Approach

A different view on what we achieve by this procedure: Representing the potentials of ϕ as a Matrix-Vector-Multiplication, i.e.,

A	B	ϕ
1	1	a
1	0	b
0	1	c
0	0	d

$$\begin{bmatrix} \lambda_1' & \lambda_2' & \lambda_3' & \lambda_4' \\ \lambda_5' & \lambda_6' & \lambda_7' & \lambda_8' \\ \lambda_9' & \lambda_{10}' & \lambda_{11}' & \lambda_{12}' \\ \lambda_{13}' & \lambda_{14}' & \lambda_{15}' & \lambda_{16}' \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

$$\parallel$$

$$\begin{bmatrix} \lambda_1 & \lambda_2 \\ \lambda_3 & \lambda_4 \end{bmatrix} \otimes \begin{bmatrix} \lambda_1 & \lambda_2 \\ \lambda_3 & \lambda_4 \end{bmatrix}$$

A First Naive Approach

A different view on what we achieve by this procedure: Representing the potentials of ϕ as a Matrix-Vector-Multiplication, i.e.,

$$\begin{bmatrix} \lambda_1{}' & \lambda_2{}' & \lambda_3{}' & \lambda_4{}' \\ \lambda_5{}' & \lambda_6{}' & \lambda_7{}' & \lambda_8{}' \\ \lambda_9{}' & \lambda_{10}{}' & \lambda_{11}{}' & \lambda_{12}{}' \\ \lambda_{13}{}' & \lambda_{14}{}' & \lambda_{15}{}' & \lambda_{16}{}' \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \underbrace{\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}}_{\text{...instead of this representation}}$$

$$\| $$

$$\underbrace{\begin{bmatrix} \lambda_1 & \lambda_2 \\ \lambda_3 & \lambda_4 \end{bmatrix} \otimes \begin{bmatrix} \lambda_1 & \lambda_2 \\ \lambda_3 & \lambda_4 \end{bmatrix}}_{\text{Work with this representation...}}$$

A First Naive Approach

A different view on what we achieve by this procedure: Representing the potentials of ϕ as a Matrix-Vector-Multiplication, i.e.,

$$\begin{bmatrix} \lambda_1{}' & \lambda_2{}' & \lambda_3{}' & \lambda_4{}' \\ \lambda_5{}' & \lambda_6{}' & \lambda_7{}' & \lambda_8{}' \\ \lambda_9{}' & \lambda_{10}{}' & \lambda_{11}{}' & \lambda_{12}{}' \\ \lambda_{13}{}' & \lambda_{14}{}' & \lambda_{15}{}' & \lambda_{16}{}' \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

$$=$$

$$\underbrace{\begin{bmatrix} \lambda_1 & \lambda_2 \\ \lambda_3 & \lambda_4 \end{bmatrix} \otimes \begin{bmatrix} \lambda_1 & \lambda_2 \\ \lambda_3 & \lambda_4 \end{bmatrix}}$$

Work with this representation

1. Allows for introducing new structure into a factor
   - Structured Matrix
   - Sparse Vector
2. Allows for approaching probabilistic inference differently
3. Allows for exploiting existing structure differently

A First Naive Approach – Overview Introducing Artificial RVs

A	B	ϕ
1	1	a
1	0	b
0	1	c
0	0	d

A_1	B_1	ϕ^I
1	1	x_1
1	0	x_2
0	1	x_3
0	0	x_4

B	B_1	ϕ^{B_1}
1	1	λ_1
1	0	λ_2
0	1	λ_3
0	0	λ_4

A	A_1	ϕ^{A_1}
1	1	λ_1
1	0	λ_2
0	1	λ_3
0	0	λ_4

$$\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

$$\begin{bmatrix} \lambda_1' & \lambda_2' & \lambda_3' & \lambda_4' \\ \lambda_5' & \lambda_6' & \lambda_7' & \lambda_8' \\ \lambda_9' & \lambda_{10}' & \lambda_{11}' & \lambda_{12}' \\ \lambda_{13}' & \lambda_{14}' & \lambda_{15}' & \lambda_{16}' \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

$$\begin{bmatrix} \lambda_1' & \lambda_2' & \lambda_3' & \lambda_4' \\ \lambda_5' & \lambda_6' & \lambda_7' & \lambda_8' \\ \lambda_9' & \lambda_{10}' & \lambda_{11}' & \lambda_{12}' \\ \lambda_{13}' & \lambda_{14}' & \lambda_{15}' & \lambda_{16}' \end{bmatrix} = \begin{bmatrix} \lambda_1 & \lambda_2 \\ \lambda_3 & \lambda_4 \end{bmatrix} \otimes \begin{bmatrix} \lambda_1 & \lambda_2 \\ \lambda_3 & \lambda_4 \end{bmatrix}$$

Factor Graph -> Tensor Network

A	B	ϕ_1
1	1	a_1
1	0	b_1
0	1	c_1
0	0	d_1

A	C	ϕ_2
1	1	a_2
1	0	b_2
0	1	c_2
0	0	d_2

B	C	ϕ_3
1	1	a_3
1	0	b_3
0	1	c_3
0	0	d_3

A_1	B_1	ϕ_1^I
1	1	x_{11}
1	0	x_{12}
0	1	x_{13}
0	0	x_{14}

A_2	C_1	ϕ_2^I
1	1	x_{21}
1	0	x_{22}
0	1	x_{23}
0	0	x_{24}

B_2	C_2	ϕ_3^I
1	1	x_{31}
1	0	x_{32}
0	1	x_{33}
0	0	x_{34}

A	A_1	ϕ^{A_1}
1	1	λ_1
1	0	λ_2
0	1	λ_3
0	0	λ_4

B	B_1	ϕ^{B_1}
1	1	λ_1
1	0	λ_2
0	1	λ_3
0	0	λ_4

C	C_1	ϕ^{C_1}
1	1	λ_1
1	0	λ_2
0	1	λ_3
0	0	λ_4

A	A_2	ϕ^{A_2}
1	1	λ_1
1	0	λ_2
0	1	λ_3
0	0	λ_4

B	B_2	ϕ^{B_2}
1	1	λ_1
1	0	λ_2
0	1	λ_3
0	0	λ_4

C	C_2	ϕ^{C_2}
1	1	λ_1
1	0	λ_2
0	1	λ_3
0	0	λ_4

Factor Graph -> Tensor Network

A_1	B_1	ϕ_1^I
1	1	x_{11}
1	0	x_{12}
0	1	x_{13}
0	0	x_{14}

A_2	C_1	ϕ_2^I
1	1	x_{21}
1	0	x_{22}
0	1	x_{23}
0	0	x_{24}

B_2	C_2	ϕ_3^I
1	1	x_{31}
1	0	x_{32}
0	1	x_{33}
0	0	x_{34}

A	A_1	ϕ^{A_1}
1	1	λ_1
1	0	λ_2
0	1	λ_3
0	0	λ_4

B	B_1	ϕ^{B_1}
1	1	λ_1
1	0	λ_2
0	1	λ_3
0	0	λ_4

C	C_1	ϕ^{C_1}
1	1	λ_1
1	0	λ_2
0	1	λ_3
0	0	λ_4

A	A_2	ϕ^{A_2}
1	1	λ_1
1	0	λ_2
0	1	λ_3
0	0	λ_4

B	B_2	ϕ^{B_2}
1	1	λ_1
1	0	λ_2
0	1	λ_3
0	0	λ_4

C	C_2	ϕ^{C_2}
1	1	λ_1
1	0	λ_2
0	1	λ_3
0	0	λ_4

Sum Out A

Sum Out B

Sum Out C

Factor Graph -> Tensor Network

A_1	B_1	ϕ_1^I
1	1	x_{11}
1	0	x_{12}
0	1	x_{13}
0	0	x_{14}

A_2	C_1	ϕ_2^I
1	1	x_{21}
1	0	x_{22}
0	1	x_{23}
0	0	x_{24}

B_2	C_2	ϕ_3^I
1	1	x_{31}
1	0	x_{32}
0	1	x_{33}
0	0	x_{34}

A_1	A_2	$\phi^{A_{12}}$
1	1	λ_1^*
1	0	λ_2^*
0	1	λ_3^*
0	0	λ_4^*

B_1	B_2	$\phi^{B_{12}}$
1	1	λ_1^*
1	0	λ_2^*
0	1	λ_3^*
0	0	λ_4^*

C_1	C_2	$\phi^{C_{12}}$
1	1	λ_1^*
1	0	λ_2^*
0	1	λ_3^*
0	0	λ_4^*

1. Each random variable appears in exactly 2 factors
2. Factors on the left / right do not share any random variable
3. Sum out of arbitrary RV requires single multiplication -> factor multiplication + sum out combined

Factor Graph -> Tensor Network

Represents a „2-sided-model“

A_1	B_1	ϕ_1^I
1	1	x_{11}
1	0	x_{12}
0	1	x_{13}
0	0	x_{14}

A_2	C_1	ϕ_2^I
1	1	x_{21}
1	0	x_{22}
0	1	x_{23}
0	0	x_{24}

B_2	C_2	ϕ_3^I
1	1	x_{31}
1	0	x_{32}
0	1	x_{33}
0	0	x_{34}

A_1	A_2	$\phi^{A_{12}}$
1	1	λ_1^*
1	0	λ_2^*
0	1	λ_3^*
0	0	λ_4^*

B_1	B_2	$\phi^{B_{12}}$
1	1	λ_1^*
1	0	λ_2^*
0	1	λ_3^*
0	0	λ_4^*

C_1	C_2	$\phi^{C_{12}}$
1	1	λ_1^*
1	0	λ_2^*
0	1	λ_3^*
0	0	λ_4^*

$*$

A_1	A_2	B_1	B_2	C_1	C_2	ϕ
1	1	1	1	1	1	…
…	…	…	…	…	…	…

Factor Graph -> Tensor Network

Represents a „2-sided-model“

A_1	B_1	ϕ_1^I
1	1	x_{11}
1	0	x_{12}
0	1	x_{13}
0	0	x_{14}

A_2	C_1	ϕ_2^I
1	1	x_{21}
1	0	x_{22}
0	1	x_{23}
0	0	x_{24}

B_2	C_2	ϕ_3^I
1	1	x_{31}
1	0	x_{32}
0	1	x_{33}
0	0	x_{34}

A_1	A_2	$\phi^{A_{12}}$
1	1	λ_1^*
1	0	λ_2^*
0	1	λ_3^*
0	0	λ_4^*

B_1	B_2	$\phi^{B_{12}}$
1	1	λ_1^*
1	0	λ_2^*
0	1	λ_3^*
0	0	λ_4^*

C_1	C_2	$\phi^{C_{12}}$
1	1	λ_1^*
1	0	λ_2^*
0	1	λ_3^*
0	0	λ_4^*

$*$

A_1	A_2	B_1	B_2	C_1	C_2	ϕ
1	1	1	1	1	1	…
…	…	…	…	…	…	…

1. Inference (VE) does not primarily focus on summing out RVs one by one
2. Structure / Symmetries on one side sufficient for efficient probabilistic inference
   - Interesting for handling evidence (evidence only affects one side)
3. Potentials of both sides can be affected since we can choose $\lambda_1, \dots, \lambda_4$ arbitrarily.
4. Parallelisation / efficient implementation of tensor operations (GPU/CPU)
5. …

Factor Graph -> Tensor Network

Represents a „2-sided-model“

A_1	B_1	ϕ_1^I
1	1	x_{11}
1	0	x_{12}
0	1	x_{13}
0	0	x_{14}

A_2	C_1	ϕ_2^I
1	1	x_{21}
1	0	x_{22}
0	1	x_{23}
0	0	x_{24}

B_2	C_2	ϕ_3^I
1	1	x_{31}
1	0	x_{32}
0	1	x_{33}
0	0	x_{34}

A_1	A_2	B_1	B_2	C_1	C_2	ϕ
1	1	1	1	1	1	…
…	…	…	…	…	…	…

Example

–	–	ϕ_1
1	1	5
1	0	6
0	1	7
0	0	8

–	–	–	ϕ
1	1	1	1
1	1	0	2
1	0	1	2
1	0	0	3
0	1	1	2
0	1	0	3
0	0	1	3
0	0	0	4

Sum out all RVs in ϕ_1

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \end{pmatrix} \cdot \begin{pmatrix} 5 \\ 6+7 \\ 8 \end{pmatrix}$$

Introducing Structure – Simple Example

A	B	C	D	ϕ
1	1	1	1	x_1
1	1	1	0	x_2
1	1	0	1	x_3
1	1	0	0	x_4
1	0	1	1	x_5
1	0	1	0	x_6
1	0	0	1	x_7
1	0	0	0	x_8
0	1	1	1	x_9
0	1	1	0	x_{10}
0	1	0	1	x_{11}
0	1	0	0	x_{12}
0	0	1	1	x_{13}
0	0	1	0	x_{14}
0	0	0	1	x_{15}
0	0	0	0	x_{16}

$\Rightarrow M \cdot$

A_1	B_1	C_1	D_1	ϕ
1	1	1	1	x_1^*
1	1	1	0	x_2^*
1	1	0	1	x_3^*
1	1	0	0	0
1	0	1	1	x_4^*
1	0	1	0	0
1	0	0	1	0
1	0	0	0	0
0	1	1	1	x_5^*
0	1	1	0	0
0	1	0	1	0
0	1	0	0	0
0	0	1	1	0
0	0	1	0	0
0	0	0	1	0
0	0	0	0	0

- M Walsh-Matrix
  - Recursive Matrix-Definition
  - $M_1 = (1)$
  - $M_k = \begin{pmatrix} M_k & M_k \\ M_k & -M_k \end{pmatrix}$
  - Special case of discrete Fourier-Transformation
- There are factor structures which can be captured by a Walsh-Matrix and a vector consisting of only $(|rv(\phi)| + 1)$ **non-zero values**
  - Same reduction as counting symmetry for boolean RVs
  - Linear instead of exponential
- Factor does not require
  - Decomposability / Independence
  - State-Space Symmetry

Introducing Structure – Simple Example

A_x	B_x	C_x	D_x	ϕ
1	1	1	1	x_1
1	1	1	0	x_2
1	1	0	1	x_3
1	1	0	0	x_4
1	0	1	1	x_5
1	0	1	0	x_6
1	0	0	1	x_7
1	0	0	0	x_8
0	1	1	1	x_9
0	1	1	0	x_{10}
0	1	0	1	x_{11}
0	1	0	0	x_{12}
0	0	1	1	x_{13}
0	0	1	0	x_{14}
0	0	0	1	x_{15}
0	0	0	0	x_{16}

- Another interesting aspect: Factors with completely different potentials but same structure are identical* after this transformation
  - *differ in a single potential

A_1	B_1	C_1	D_1	ϕ
[illegible]	[illegible]	[illegible]	[illegible]	x_1^*
[illegible]	[illegible]	[illegible]	[illegible]	x_2^*
[illegible]	[illegible]	[illegible]	[illegible]	x_3^*
[illegible]	[illegible]	[illegible]	[illegible]	0
[illegible]	[illegible]	[illegible]	[illegible]	x_4^*
1	0	1	0	0
1	0	0	1	0
1	0	0	0	0
0	1	1	1	x_5^*
0	1	1	0	0
0	1	0	1	0
0	1	0	0	0
0	0	1	1	0
0	0	1	0	0
0	0	0	1	0
0	0	0	0	*

A_y	B_y	C_y	D_y	ϕ
1	1	1	1	y_1
1	1	1	0	y_2
1	1	0	1	y_3
1	1	0	0	y_4
1	0	1	1	y_5
1	0	1	0	y_6
1	0	0	1	y_7
1	0	0	0	y_8
0	1	1	1	y_9
0	1	1	0	y_{10}
0	1	0	1	y_{11}
0	1	0	0	y_{12}
0	0	1	1	y_{13}
0	0	1	0	y_{14}
0	0	0	1	y_{15}
0	0	0	0	y_{16}

Introducing Structure – New Operator

We introduce a new factor operation to work more efficiently with this representation (by adding more structure): factor addition

A	B	ϕ
1	1	a
1	0	b
0	1	c
0	0	d

$=$

A	B	ϕ
1	1	a_1
1	0	b_1
0	1	c_1
0	0	d_1

$+$

A	B	ϕ
1	1	a_2
1	0	b_2
0	1	c_2
0	0	d_2

- If full joint given by $\boldsymbol{\phi_1} \cdot \boldsymbol{\phi_2} \cdot \boldsymbol{\phi_3}$ we have, e.g., $\boldsymbol{\phi_1} \cdot \boldsymbol{\phi_2} \cdot (\boldsymbol{\phi_{31}} + \boldsymbol{\phi_{32}})$
    - i.e., $(\boldsymbol{\phi_1} \cdot \boldsymbol{\phi_2} \cdot \boldsymbol{\phi_{31}}) + (\boldsymbol{\phi_1} \cdot \boldsymbol{\phi_2} \cdot \boldsymbol{\phi_{32}})$
    - Can be understood as splitting a factor graph into two factor graphs
        - Inference: Calculate result for each factor graph, add up results

Summary

- Model transformation
  - Artificial Random Variables
  - Factor Graph -> Tensor Network
  - „2-sided-model“
- Probabilistic inference by means of tensor operations
- Structure exploitation in Tensor Networks
- Extending structure exploitation by matrix-vector-representation
- New operator: factor addition
- Incorporating lifting ideas in asymmetrical models