



離散構造と離散分布

NTT コミュニケーション科学基礎研究所
石畠正和

第17回情報論的学習理論ワークショップ(IBIS2014)
離散アルゴリズムの機械学習応用

経歴

2006 石川高専 修了
2008 東工大 学部 修了
2010 東工大 修士 修了
2013 東工大 博士 修了
2013 NTT CS 研 入社

指導教官

佐藤泰介 教授

研究内容

論理に基づく確率モデリング 論理と確率を合言葉に！

- 離散分布の複雑さ
- 離散分布と離散構造 (主に論理) の関係
- Lifted Inference のアイデア

論理って
いつの話だよ(笑)

離散って
足すだけじゃん(笑)



こういう人に伝えたい！！

今日、伝えないこと



- 具体的な 論理を用いた確率計算
- 具体的な Lifted Inference の計算方法

どうしても論理の基礎知識が必要になるので...

想定する聴衆

確率	→ 使ったことある
論理	→ 詳しくない
動的計画法	→ あー、あるよね
Bayesian Network	→ あー、グラフィカルモデルね？
Lifted Inference	→ あー、どっかの会議で見たかも

当初: **述語論理** を利用した効率的な離散分布の計算法 [Poole 03]
現在: **Partial Exchangeability** を効率的に扱う計算法 [Niepert+ 14a]

実は流行ってます！！

当初: **述語論理** を利用した効率的な離散分布の計算法 [Poole 03]

現在: **Partial Exchangeability** を効率的に扱う計算法 [Niepert+ 14a]

Tutorials

- “*Lifted inference in Probabilistic logical Models*”, IJCAI11
- “*Probabilistic Inference in Relational Models*”, UAI14
- “*Lifted Approximate Inference: Methods and Theory*”, AAAI14

Workshops

- *StarAI: Statistical Relational AI*, AAAI10, UAI12, AAAI13-14
- *LTPM: Learning Tractable Probabilistic Models*, ICML14
- *BUDA: Big Uncertain Data*, SIGMOD14

1 ページ要約

AI

離散分布は実は複雑 ☹
組み合わせ爆発



論理を使うと
効率的に定義できる ☺



論理をグラフに変換し
効率的に扱う



論理＋グラフで効率的に
離散分布を計算・学習！



Lifted Inference
この組合せが最も力を
発揮できる場合を追求！

ML

1. 離散分布の複雑さ
2. 離散分布を効率的に定義する
3. 離散分布を効率的に計算する
4. 効率的な確率計算の研究例
5. まとめ

離散確率変数

離散値を取る確率変数
この発表では**有限の値**



$$p(\text{表}) = \theta$$
$$p(\text{裏}) = 1 - \theta$$

Bernoulli 分布 (= コイン)

$$p(x | \theta) = \theta^x (1 - \theta)^{1-x},$$
$$x \in \{0, 1\}, \theta \in [0, 1]$$



$$p(1) = \theta_1$$

$$p(2) = \theta_2$$

...

$$p(6) = \theta_6$$

$$(\sum_{i=1 \dots 6} \theta_i = 1)$$

Categorical 分布 (= サイコロ)

$$p(x | \theta) = \prod_i \theta_i^{[x=i]} = \theta_x$$

$$x \in \{1, \dots, M\}, \theta = (\theta_1, \dots, \theta_M), \theta_i \in [0, 1], \sum_i \theta_i = 1$$

離散確率変数の同時分布

- $X_i =$ 離散確率変数, $x_i \in \{0, 1\}$
- $\mathbf{X} = (X_1, \dots, X_N)$, $\mathbf{x} \in \Omega \equiv \{0, 1\}^N$
- $p(\mathbf{X} = \mathbf{x} \mid \boldsymbol{\theta}) = \theta_{\text{ID}(\mathbf{x})}$
- $\boldsymbol{\theta} = (\theta_1, \dots, \theta_M)$, $\theta_i \in [0, 1]$, $\sum_i \theta_i = 1$

$$M = 2^N$$

ID	x_1	x_2	x_3	$p(\mathbf{x})$
1	0	0	0	θ_1
2	0	0	1	θ_2
3	0	1	0	θ_3
4	0	1	1	θ_4
5	1	0	0	θ_5
6	1	0	1	θ_6
7	1	1	0	θ_7
8	1	1	1	θ_8

$\mathbf{X}$ の確率表

任意の同時分布を定めるには指数個のパラメータが必要☹

条件付き確率

$$p(X_1, X_2, X_3) = p(X_1) p(X_2|X_1) p(X_3|X_1, X_2) \quad \text{ただし} \quad \begin{matrix} p(x_1) > 0 \\ p(x_1, x_2) > 0 \end{matrix}$$

ID	$x_1=0$	1
1	θ_{111}	θ_{112}

X_1 の CPT

ID	x_1	$x_2=0$	1
1	0	θ_{211}	θ_{212}
2	1	θ_{221}	θ_{222}

X_2 の CPT

ID	x_1	x_2	$x_3=0$	1
1	0	0	θ_{311}	θ_{312}
2	0	1	θ_{321}	θ_{322}
3	1	0	θ_{331}	θ_{332}
4	1	1	θ_{341}	θ_{342}

X_3 の CPT

条件付き確率の定義より、

X の **確率表** は各 X_i の **条件付き確率表(CPT)** で定義可能

CPT = Conditional Probability Table

条件付き確率

$$p(X_1, X_2, X_3) = p(X_1) p(X_2|X_1) p(X_3|X_1, X_2) \quad \text{ただし } p(x_1) > 0, p(x_1, x_2) > 0$$

ID	$x_1=0$	1
1	θ_{111}	θ_{112}

X_1 の CPT

ID	x_1	$x_2=0$	1
1	0	θ_{211}	θ_{212}
2	1	θ_{221}	θ_{222}

X_2 の CPT

ID	x_1	x_2	$x_3=0$	1
1	0	0	θ_{311}	θ_{312}
2	0	1	θ_{321}	θ_{322}
3	1	0	θ_{331}	θ_{332}
4	1	1	θ_{341}	θ_{342}

X_3 の CPT

$\theta_{ijk} \equiv$ 表 i の行 j の列 k のパラメータ

$$= p(X_i = x_{ik} \mid \text{ID}(x_{1:i-1}) = j)$$

$x_{ik} \equiv X_i$ の k 番目の値

$\mathbf{x}_{1:i} \equiv (x_1, \dots, x_i)$

条件付き確率

$$p(X_1, X_2, X_3) = p(X_1) p(X_2|X_1) p(X_3|X_1, X_2) \quad \text{ただし} \quad \begin{matrix} p(x_1) > 0 \\ p(x_1, x_2) > 0 \end{matrix}$$

ID	$x_1=0$	1
1	θ_{111}	θ_{112}

X_1 の CPT

ID	x_1	$x_2=0$	1
1	0	θ_{211}	θ_{212}
2	1	θ_{221}	θ_{222}

X_2 の CPT

ID	x_1	x_2	$x_3=0$	1
1	0	0	θ_{311}	θ_{312}
2	0	1	θ_{321}	θ_{322}
3	1	0	θ_{331}	θ_{332}
4	1	1	θ_{341}	θ_{342}

X_3 の CPT

$\theta_{ijk} \equiv$ 表 i の行 j の列 k のパラメータ

$$= p(X_i = x_{ik} \mid \text{ID}(x_{1:i-1}) = j)$$

$$p(X_1=1, X_2=1, X_3=1) = \theta_{112} \theta_{222} \theta_{342}$$

条件付き確率

$$p(X_1, X_2, X_3) = p(X_1) p(X_2|X_1) p(X_3|X_1, X_2) \quad \text{ただし} \quad \begin{matrix} p(x_1) > 0 \\ p(x_1, x_2) > 0 \end{matrix}$$

ID	$x_1=0$	1
1	θ_{111}	θ_{112}

X_1 の CPT

ID	x_1	$x_2=0$	1
1	0	θ_{211}	θ_{212}
2	1	θ_{221}	θ_{222}

X_2 の CPT

ID	x_1	x_2	$x_3=0$	1
1	0	0	θ_{311}	θ_{312}
2	0	1	θ_{321}	θ_{322}
3	1	0	θ_{331}	θ_{332}
4	1	1	θ_{341}	θ_{342}

X_3 の CPT

$\theta_{ijk} \equiv$ 表 i の行 j の列 k のパラメータ

$$= p(X_i = x_{ik} \mid \text{ID}(x_{1:i-1}) = j)$$

$$p(X_1=1, X_2=1, X_3=1) = \theta_{112} \theta_{222} \theta_{342}$$

条件付き確率

$$p(X_1, X_2, X_3) = p(X_1) p(X_2|X_1) p(X_3|X_1, X_2) \quad \text{ただし} \quad \begin{matrix} p(x_1) > 0 \\ p(x_1, x_2) > 0 \end{matrix}$$

ID	$x_1=0$	1
1	θ_{111}	θ_{112}

X_1 の CPT

ID	x_1	$x_2=0$	1
1	0	θ_{211}	θ_{212}
2	1	θ_{221}	θ_{222}

X_2 の CPT

ID	x_1	x_2	$x_3=0$	1
1	0	0	θ_{311}	θ_{312}
2	0	1	θ_{321}	θ_{322}
3	1	0	θ_{331}	θ_{332}
4	1	1	θ_{341}	θ_{342}

X_3 の CPT

$\theta_{ijk} \equiv$ 表 i の行 j の列 k のパラメータ

$$= p(X_i = x_{ik} \mid \text{ID}(x_{1:i-1}) = j)$$

$$p(X_1=1, X_2=1, X_3=1) = \theta_{112} \theta_{222} \theta_{342}$$

条件付き確率

$$p(X_1, X_2, X_3) = p(X_1) p(X_2|X_1) p(X_3|X_1, X_2) \quad \text{ただし} \quad \begin{matrix} p(x_1) > 0 \\ p(x_1, x_2) > 0 \end{matrix}$$

ID	$x_1=0$	1
1	θ_{111}	θ_{112}

X_1 の CPT

ID	x_1	$x_2=0$	1
1	0	θ_{211}	θ_{212}
2	1	θ_{221}	θ_{222}

X_2 の CPT

ID	x_1	x_2	$x_3=0$	1
1	0	0	θ_{311}	θ_{312}
2	0	1	θ_{321}	θ_{322}
3	1	0	θ_{331}	θ_{332}
4	1	1	θ_{341}	θ_{342}

X_3 の CPT

$\theta_{ijk} \equiv$ 表 i の行 j の列 k のパラメータ

$$= p(X_i = x_{ik} \mid \text{ID}(x_{1:i-1}) = j)$$

$$p(X_1=1, X_2=1, X_3=1) = \theta_{112} \theta_{222} \theta_{342}$$

条件付き確率

$$p(X_1, X_2, X_3) = p(X_1) p(X_2|X_1) p(X_3|X_1, X_2) \quad \text{ただし} \quad \begin{matrix} p(x_1) > 0 \\ p(x_1, x_2) > 0 \end{matrix}$$

ID	$x_1=0$	1
1	θ_{111}	θ_{112}

X_1 の CPT

ID	x_1	$x_2=0$	1
1	0	θ_{211}	θ_{212}
2	1	θ_{221}	θ_{222}

X_2 の CPT

ID	x_1	x_2	$x_3=0$	1
1	0	0	θ_{311}	θ_{312}
2	0	1	θ_{321}	θ_{322}
3	1	0	θ_{331}	θ_{332}
4	1	1	θ_{341}	θ_{342}

X_3 の CPT

各行が **Categorical 分布** ←

条件付き確率に分解してもパラメータは指数個 ☹

同時確率 $p(\mathbf{x} \mid \boldsymbol{\theta}) = \theta_{\text{ID}(\mathbf{x})}, \mathbf{x} \in \Omega \equiv \{0,1\}^N$

周辺確率 $p(\mathbf{e} \mid \boldsymbol{\theta}) = \sum_{\mathbf{x} \in \mathbf{e}} p(\mathbf{x} \mid \boldsymbol{\theta}), \mathbf{e} \subseteq \Omega$

期待値 $E[F]_{p(\mathbf{x}|\mathbf{e},\boldsymbol{\theta})} \equiv \sum_{\mathbf{x} \in \Omega} F(\mathbf{x})p(\mathbf{x} \mid \mathbf{e},\boldsymbol{\theta})$

Viterbi 値 $\mathbf{x}^* \equiv \operatorname{argmax}_{\mathbf{x}} p(\mathbf{x} \mid \mathbf{e},\boldsymbol{\theta})$

最尤推定量 $\boldsymbol{\theta}^* \equiv \operatorname{argmax}_{\boldsymbol{\theta}} \prod_{\mathbf{e} \in E} p(\mathbf{e}|\boldsymbol{\theta}), E = \{\mathbf{e}_1, \mathbf{e}_2, \dots\}$

事後分布 $p(\boldsymbol{\theta} \mid E)$

愚直に計算すると指数的な時間を要する(NP-hard) ☹

論理式 F

命題変数と論理演算と括弧の組合せ
論理関数を表現

$$F = (X_1 \wedge X_2) \vee X_3$$

論理関数 $F(X)$

$$F : \{0,1\}^N \rightarrow \{0,1\}$$

$$F(1,1,1) = 1$$

真理値表

x と $F(x)$ の対応表

ID	x_1	x_2	x_3	F
1	0	0	0	0
2	0	0	1	1
3	0	1	0	0
4	0	1	1	1
5	1	0	0	0
6	1	0	1	1
7	1	1	0	1
8	1	1	1	1

F の真理値表

F を X 上の論理式、 $F(X)$ を F が表す論理関数とする

Satisfiability (SAT) problem [Cook 71]

$F(x) = 1$ なる x が存在する(充足可能)か判定する

→ NP-complete

Model Counting (MC) problem

$F(x) = 1$ なる x の数 $MC(F)$ を求める

→ #P-complete

論理式の確率

$$p(F \mid \theta) \equiv \sum_{\mathbf{x} \in \Omega} F(\mathbf{x})p(\mathbf{x} \mid \theta) = E[F]_{p(\mathbf{x}|\theta)}$$

離散分布と SAT

$p(\mathbf{x} \mid \theta)$ を $\forall \mathbf{x} (p(\mathbf{x} \mid \theta) > 0)$ なる $\mathbf{x}$ の同時分布とする
このとき、 $p(F \mid \theta) > 0 \Leftrightarrow$ 論理式 F は充足可能

→ NP-complete

離散分布と MC

$p(\mathbf{x} \mid \theta)$ を $\forall \mathbf{x} (p(\mathbf{x} \mid \theta) = 1/2^N)$ なる $\mathbf{x}$ の一様分布とする
このとき、 $p(F \mid \theta) = M / 2^N \Leftrightarrow MC(F) = M$

→ #P-complete

X が連続のとき

$$E[F]_{p(X)} \equiv \int F(x)p(x)dx$$

が**解析的**に計算できると嬉しい

X が離散のとき

$$E[F]_{p(X)} \equiv \sum_x F(x)p(x)$$

が **？ ？ ？ ？** で計算できると嬉しい

X が連続のとき

$$E[F]_{p(X)} \equiv \int F(x)p(x)dx$$

が**解析的**に計算できると嬉しい

X が離散のとき

$$E[F]_{p(X)} \equiv \sum_x F(x)p(x)$$

が**多項式時間**で計算できると嬉しい

→ Model counting が多項式時間で解ける

離散における積分 = Model Counting (Model Enumeration)

離散分布の複雑さ (まとめ)

離散分布は
愚直に定義、愚直に計算すると指数 ☹



離散分布を
効率的に定義、効率的に計算したい ☺



効率的な定義 = パラメータ数が多項式
効率的な計算 = 計算量が多項式

1. 離散分布の複雑さ
- 2. 離散分布を効率的に定義する by 論理**
3. 離散分布を効率的に計算する
4. 効率的な確率計算の研究例
5. まとめ

$p(x|\theta)$ を効率的に定義する

パラメータを共有する

$p(X_i | X_{1:i-1}, \theta)$

ID	x_1	x_2	...	x_{i-1}	$x_i=0$	1
1	0	0	...	0	θ_{i11}	θ_{i12}
2	0	0	...	1	θ_{i21}	θ_{i22}
...	...	...	...	...	...	...
...	...	...	...	...	...	...
...	...	...	...	...	...	...
$2^{i-1}-1$	1	1	...	0	...	...
2^{i-1}	1	1	...	1	$\theta_{i2^{i-1}1}$	$\theta_{i2^{i-1}2}$

写像
 T_i

$p(X_i | X_{1:i-1}, \theta)$

ID					$x_i=0$	1
1					θ_{i11}	θ_{i12}
2	???				θ_{i21}	θ_{i22}
...					...	...
M_i					θ_{iM_i1}	θ_{iM_i2}

X_i のコンパクトな CPT
 $M_i = \text{多項式サイズ}$

X_i の条件付き確率表 (CPT) = 指数サイズ

$p(x|\theta)$ を効率的に定義する

パラメータを共有する

Parameter Tying (Sharing)

$$T_i : \{0,1\}^{i-1} \rightarrow \{1,\dots,M_i\}$$

$$p(X_i = x_{ik} \mid T_i(\mathbf{x}_{1:i-1}) = j, \theta) \equiv \theta_{ijk}$$

パラメータ数 $2^{i-1} \rightarrow M_i$

問題

T_i をどう定める？

→ T_i を決める = 独立性を仮定

$$p(X_i \mid \mathbf{X}_{1:i-1}, \theta)$$

ID	$T_i(\mathbf{x}_{1:i-1})$	$x_i=0$	1
1	1	θ_{i11}	θ_{i12}
2	2	θ_{i21}	θ_{i22}
...	...	...	...
M_i	M_i	θ_{iMi1}	θ_{iMi2}

X_i のコンパクトな CPT

$$\mathbf{x}_{1:i-1} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{i-1})$$

$$\mathbf{x}_{1:i-1} \in \{0,1\}^{i-1}$$

強い仮定

1. Independence
2. Conditional Independence
3. Context Specific Independence [Boutilier+ 96]
4. Partial Exchangeability [Niepert+ 14a]

弱い仮定

Independence

$$p(X_1, X_2, X_3) = p(X_1)p(X_2)p(X_3) : X_1 \perp\!\!\!\perp X_2, X_2 \perp\!\!\!\perp X_3, X_1 \perp\!\!\!\perp X_3$$

ID	$x_1=0$	1
1	θ_{111}	θ_{112}

X_1 の CPT

ID	x_1	$x_2=0$	1
1	0	θ_{211}	θ_{212}
2	1	θ_{211}	θ_{212}

X_2 の CPT

ID	x_1	x_2	$x_3=0$	1
1	0	0	θ_{311}	θ_{312}
2	0	1	θ_{311}	θ_{312}
3	1	0	θ_{311}	θ_{312}
4	1	1	θ_{311}	θ_{312}

X_3 の CPT

独立性 = すべての条件列を無視する

Independence

$$p(X_1, X_2, X_3) = p(X_1)p(X_2)p(X_3) : X_1 \perp\!\!\!\perp X_2, X_2 \perp\!\!\!\perp X_3, X_1 \perp\!\!\!\perp X_3$$

ID	$x_1=0$	1
1	θ_{111}	θ_{112}

ID	x_1	$x_2=0$	1
1	*	θ_{211}	θ_{212}

ID	x_1	x_2	$x_3=0$	1
1	*	*	θ_{311}	θ_{312}

$$T_2(x_1) = 1, \quad T_3(x_1, x_2) = 1$$

独立性 = すべての条件列を無視する

Conditional Independence

$$p(X_3|X_1, X_2) = p(X_3|X_1) : X_2 \perp\!\!\!\perp X_3 \mid X_1$$

ID	$x_1=0$	1
1	θ_{111}	θ_{112}

ID	x_1	$x_2=0$	1
1	0	θ_{211}	θ_{212}
2	1	θ_{221}	θ_{222}

ID	x_1	x_2	$x_3=0$	1
1	0	0	θ_{311}	θ_{312}
2	0	1	θ_{311}	θ_{312}
3	1	0	θ_{321}	θ_{322}
4	1	1	θ_{321}	θ_{322}

条件付き独立性 = 特定の条件列を無視する

Conditional Independence

$$p(X_3|X_1, X_2) = p(X_3|X_1) : X_2 \perp\!\!\!\perp X_3 \mid X_1$$

ID	$x_1=0$	1
1	θ_{111}	θ_{112}

ID	x_1	$x_2=0$	1
1	0	θ_{211}	θ_{212}
2	1	θ_{221}	θ_{222}

ID	x_1	x_2	$x_3=0$	1
1	0	*	θ_{311}	θ_{312}
2	1	*	θ_{321}	θ_{322}

$$T_3(x_1, x_2) = 1 + x_1$$

条件付き独立性 = 特定の条件列を無視する

Context Specific Independence

$$p(X_3 | X_1 = x_{1k}, X_2) = p(X_3 | X_1 = x_{1k}) : X_2 \perp\!\!\!\perp X_3 \mid X_1 = x_{1k}$$

ID	$x_1=0$	1
1	θ_{111}	θ_{112}

ID	x_1	$x_2=0$	1
1	0	θ_{211}	θ_{212}
2	1	θ_{221}	θ_{222}

ID	x_1	x_2	$x_3=0$	1
1	0	0	θ_{311}	θ_{312}
2	0	1	θ_{321}	θ_{322}
3	1	0	θ_{331}	θ_{332}
4	1	1	θ_{331}	θ_{332}

文脈依存独立性 = 特定の条件下で特定の条件列を無視する

Context Specific Independence

$$p(X_3 | X_1 = x_{1k}, X_2) = p(X_3 | X_1 = x_{1k}) : X_2 \perp\!\!\!\perp X_3 \mid X_1 = x_{1k}$$

ID	$x_1=0$	1
1	θ_{111}	θ_{112}

ID	x_1	$x_2=0$	1
1	0	θ_{211}	θ_{212}
2	1	θ_{221}	θ_{222}

ID	x_1	x_2	$x_3=0$	1
1	0	0	θ_{311}	θ_{312}
2	0	1	θ_{321}	θ_{322}
3	1	*	θ_{331}	θ_{332}

$$T_3(x_1, x_2) = \begin{cases} x_2 + 1 & : x_1 = 0 \\ 3 & : x_1 = 1 \end{cases}$$

文脈依存独立性 = 特定の条件下で特定の条件列を無視する

Partial Exchangeability

$$p(X_3|X_1, X_2) = ???$$

ID	$x_1=0$	1
1	θ_{111}	θ_{112}

ID	x_1	$x_2=0$	1
1	0	θ_{211}	θ_{212}
2	1	θ_{221}	θ_{222}

ID	x_1	x_2	$x_3=0$	1
1	0	0	θ_{311}	θ_{312}
2	0	1	θ_{321}	θ_{322}
3	1	0	θ_{321}	θ_{322}
4	1	1	θ_{331}	θ_{332}

逆にこのように Parameter Tying するには
どうすればよいか？

→ 今までの独立性では無理 ☹

Partial Exchangeability

$$p(X_3 = x_{3k} \mid X_1 = x_1, X_2 = x_2) = \theta_3 T_3(x_1, x_2)_k$$

ID	$x_1=0$	1
1	θ_{111}	θ_{112}

ID	x_1	$x_2=0$	1
1	0	θ_{211}	θ_{212}
2	1	θ_{221}	θ_{222}

ID	x_1	x_2	$x_3=0$	1
1	0	0	θ_{311}	θ_{312}
2	0	1	θ_{321}	θ_{322}
3	1	0	θ_{321}	θ_{322}
4	1	1	θ_{331}	θ_{332}

$$T_3(x_1, x_2) = ???$$

Partial Exchangeability

$$p(X_3 = x_{3k} \mid X_1 = x_1, X_2 = x_2) = \theta_3 T_3(x_1, x_2)_k$$

ID	$x_1=0$	1
1	θ_{111}	θ_{112}

ID	x_1	$x_2=0$	1
1	0	θ_{211}	θ_{212}
2	1	θ_{221}	θ_{222}

ID	$1+x_1+x_2$	$x_3=0$	1
1	1	θ_{311}	θ_{312}
2	2	θ_{321}	θ_{322}
3	3	θ_{331}	θ_{332}

$$T_3(x_1, x_2) = 1 + x_1 + x_2$$

部分交換可能性 = 特定の条件 T でパラメータを共有

$$p(X_3 = x_{3k} \mid X_1 = x_1, X_2 = x_2) = \theta_3 T_3(x_1, x_2)_k$$

ID	$x_1=0$	1
1	θ_{111}	θ_{112}

ID	x_1	$x_2=0$	1
1	0	θ_{211}	θ_{212}
2	1	θ_{221}	θ_{222}

ID	x_1	x_2	$x_3=0$	1
1	0	0	θ_{311}	θ_{312}
2	0	1	θ_{321}	θ_{322}
3	1	0	θ_{321}	θ_{322}
4	1	1	θ_{311}	θ_{312}

更に！このように Parameter Tying するには
どうすればよいか？

$$p(X_3 = x_{3k} \mid X_1 = x_1, X_2 = x_2) = \theta_3 T_3(x_1, x_2)_k$$

ID	$x_1=0$	1
1	θ_{111}	θ_{112}

ID	x_1	$x_2=0$	1
1	0	θ_{211}	θ_{212}
2	1	θ_{221}	θ_{222}

ID	x_1	x_2	$x_3=0$	1
1	0	0	θ_{311}	θ_{312}
2	0	1	θ_{321}	θ_{322}
3	1	0	θ_{321}	θ_{322}
4	1	1	θ_{331}	θ_{332}

$$T_3(x_1, x_2) = \begin{cases} 1 & : x_1 \oplus x_2 \\ 2 & : \text{otherwise} \end{cases}$$

あらゆる条件は論理で記述可能！！

論理を用いたパラメータ共有

$T_{ij} \equiv “T_i(X_{1:i-1}) = j”$ と等価な論理式

→ T_{ij} を定義 = パラメータ共有を定義

ID	x_1	x_2	$x_3=0$	1	T_{31}	T_{32}
1	0	0	θ_{311}	θ_{312}	1	0
2	0	1	θ_{321}	θ_{322}	0	1
3	1	0	θ_{321}	θ_{322}	0	1
4	1	1	θ_{331}	θ_{332}	1	0

パラメータが共有される
場所を論理式で記述

$p(x | \theta)$ を効率的に定義する

論理式 T_{ij} によりパラメータを共有する

任意のパラメータ共有を表現可能 ☺

通常 of 独立性では表現できない共有も可能 ☺

$p(x | \theta)$ を効率的に計算する

T_{ij} の Model counting を効率的に行う

Model counting は一般には NP-hard ☹

How?

1. 離散分布の複雑さ
2. 離散分布を効率的に定義する
- 3. 離散分布を効率的に計算する by 動的計画法**
4. 効率的な確率計算の研究例
5. まとめ

$p(x|\theta)$ 上で効率的に計算する

Model Counting (MC)

論理式 F に対し、 $F(x) = 1$ なる x の数 $MC(F)$ を計算

→ #P-complete

離散分布とMC

MC は離散分布の積分に対応

MC(F) を計算可能
→ $E[F]_{p(X|\theta)}$ を計算可能

2種類の論理式

同時分布を定義する論理式 T_{ij}

確率事象を定義する論理式 F

確率事象

T_{ij} でグループ化

$$p(F | \theta)$$

タスクに応じて論理式を適切な形式に変換
変換後の形式上で効率的にタスクを実行

||

変換後の形式サイズに対して多項式時間

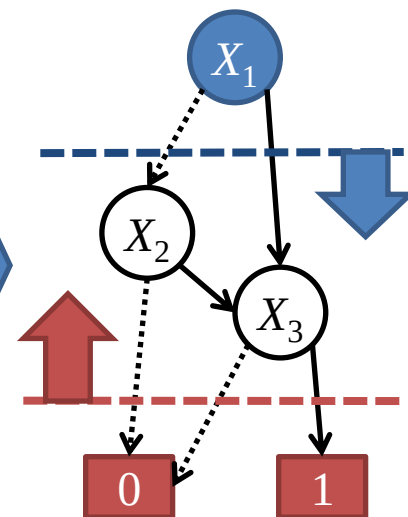
タスクの例

Consistency
Validity
Implicant
Equivalence
Model Counting
Model Enumeration
...



形式の例

Negation Normal Form (NNF)
Decomposable NNF (DNNF)
Deterministic NNF (d-NNF)
d-DNNF
Binary Decision Diagram (BDD)
Disjunctive Normal Form (DNF)
Conjunctive Normal Form (CNF)
...



Binary Decision Diagram [Bryant 86]

論理関数を表現する DAG (有向非循環グラフ)

様々な論理演算をサポート

Zero-suppressed BDD [Minato 93]

アイテムセット(集合)を効率的に表現する BDD の一種

d-DNNF [Darwiche 01]

(多くの場合で) BDD よりもコンパクトな論理関数の DAG 表現

コンパクトに圧縮される → 効率的に計算できる

Compiling Probabilistic Models

1. 離散同時分布を論理式 F, T_{ij} を用いて定義
2. 適切なデータ構造に Compile (圧縮)
3. データ構造上で Dynamic Programming を実行

1. Modeling

Model

$$T_{i1} = (X_1 \vee X_2) \wedge X_3$$

$$T_{i2} = \neg T_{i1} \wedge X_4$$

$$T_{i3} = \neg (T_{i1} \vee T_{i2})$$

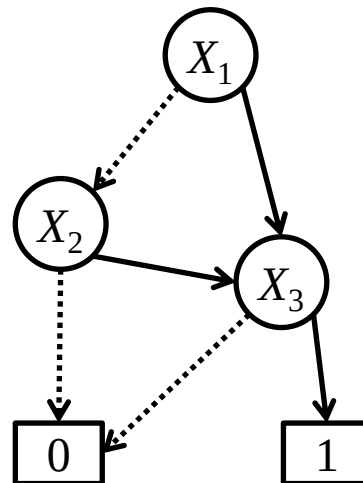
....

Observation

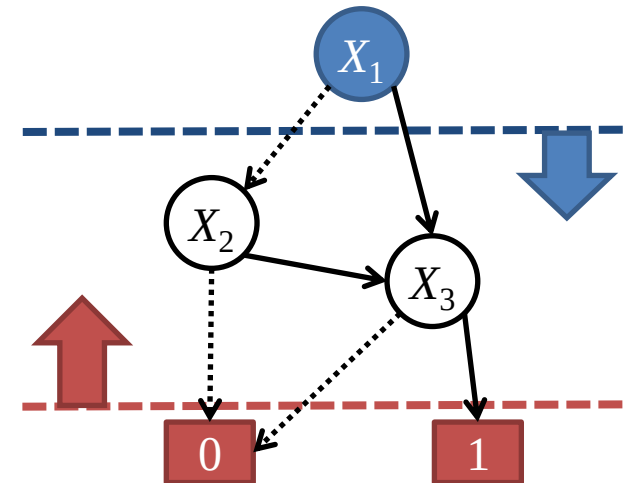
$$F = X_1 \vee X_2$$



2. Compiling



3. Dynamic Programming



1. 離散分布の複雑さ
2. 離散分布を効率的に定義する by 論理
3. 離散分布を効率的に計算する by 動的計画法
4. 効率的な確率計算の研究例
 1. Compiling Bayesian Network
 2. Probabilistic Logic Programming
 3. Lifted Inference
 4. 発想の転換
5. まとめ

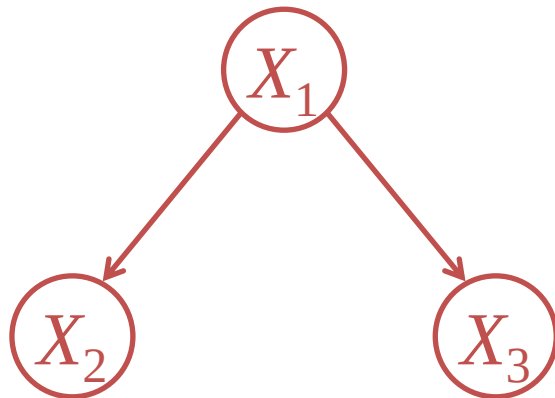
Compiling Bayesian Network

Bayesian Network (BN) [Pearl 85]

DAG + CPT で同時分布を定義

DAG = 条件付き独立 を定義

CPT = 条件付き確率 を定義



$$X_2 \perp\!\!\!\perp X_3 \mid X_1$$

ID	$x_1=0$	1
1	θ_{111}	θ_{112}

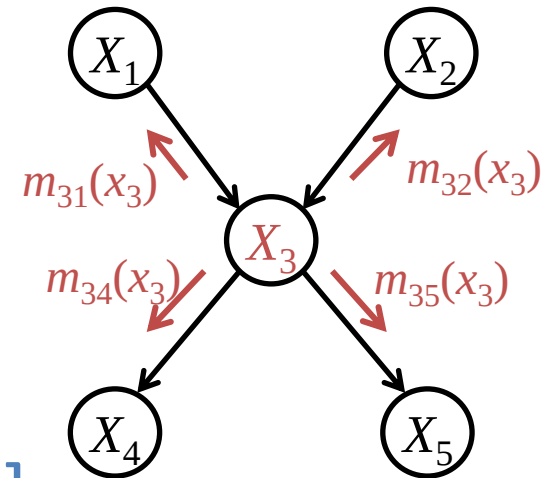
ID	x_1	$x_2=0$	1
1	0	θ_{211}	θ_{212}
2	1	θ_{221}	θ_{222}

ID	x_1	$x_3=0$	1
1	0	θ_{311}	θ_{312}
2	1	θ_{321}	θ_{322}

Compiling Bayesian Network

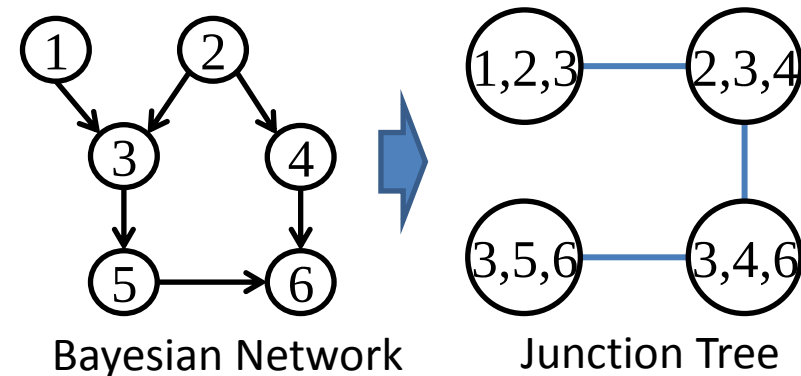
Belief Propagation (BP) [Pearl 82]

木構造 BN の周辺確率を計算
計算量は BN のサイズに比例



Junction Tree Algorithm [Lauritzen88]

一般のBNを木に変換
変換後の木で BP を実行



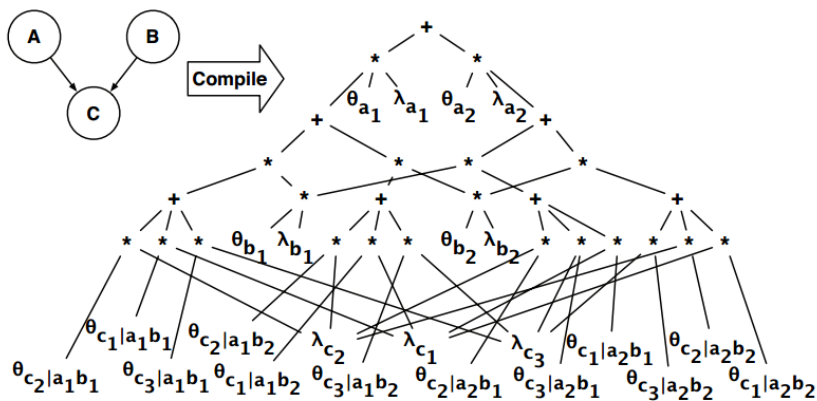
Context Specific Independence
を効率に扱えない ☹

Compiling Bayesian Network

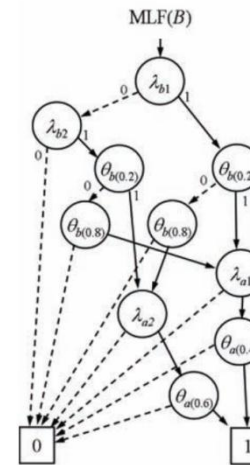
DAG + CPT を合わせて別の構造に Compile (圧縮)

変換後の構造上の DP で確率計算

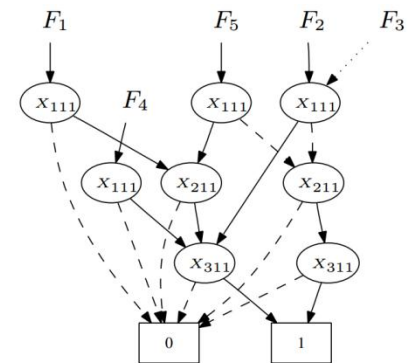
→ Context Specific Independence を効率的に扱える



d-DNNF に変換
[Chavira+ 05, 07]



ZDD に変換
[Minato+ 07]



BDD に変換
[Ishihata+ 11a]

述語論理を利用して確率モデルを定義 [DeRaedt+ 08]

Statistical Relational Learning と呼ばれる [Getoor+ 07]

推論、学習を行う処理系を含む

Model (Language)	Structure
PRISM [Sato+ 01]	Explanation Graph / BDD
LPAD [Vennekens+ 04]	BDD
Markov Logic [Richardson+06]	d-DNNF
ProbLog [DeRaedt+07]	BDD / d-DNNF

詳しくは亀谷先生のチュートリアル参照
[亀谷 11]

EM algorithm on BDDs [Ishihata+ 08]
Variational Bayes on BDDs [Ishihata+ 11b]
MCMC-Bayes on BDDs [Ishihata+ 11c]

当初 : Lifted Inference [Poole 03]

述語論理を利用して効率的に確率を定義・計算する

述語 : $\text{Friend}(X, Y)$, $\text{Smoke}(X)$, $X, Y \in D$

ルール : $0.9 \quad \forall X, Y \text{ Smoke}(X) \wedge \text{Friend}(X, Y) \Rightarrow \text{Smoke}(Y)$

ドメイン : $D \equiv \{\text{Alice}, \text{Bob}, \text{Charlie}, \dots\}$

確率変数: $\text{Friend}(\text{Alice}, \text{Bob})$, $\text{Friend}(\text{Alice}, \text{Charlie})$, ...

述語ルールで複数の確率変数の関係をまとめて記述

当初 : Lifted Inference [P] “効率的” = 曖昧な定義

述語論理を利用して効率的に確率を定義・計算する

述語 : $\text{Friend}(X, Y)$, $\text{Smoke}(X)$, $X, Y \in D$

ルール : $0.9 \quad \forall X, Y \text{ Smoke}(X) \wedge \text{Friend}(X, Y) \Rightarrow \text{Smoke}(Y)$

ドメイン : $D \equiv \{\text{Alice}, \text{Bob}, \text{Charlie}, \dots\}$

確率変数 : $\text{Friend}(\text{Alice}, \text{Bob})$, $\text{Friend}(\text{Alice}, \text{Charlie})$, ...

現在 : Domain-lifted Inference [Broeck 11b]

確率推論がドメインサイズ $|D|$ の多項式時間 $O(\text{Poly}(|D|))$

変数の数 = $O(\text{Poly}(|D|))$

通常の計算量 = $O(2^{\text{Poly}(|D|)})$

Lift!!

Partial Exchangeability in First-Order Logic

述語論理 = 1ルールで複数変数の関係を記述

→ 似た関係をもつ変数が大量に発生

→ **Partial Exchangeability**

First-Order Model Counting [Broeck 11]

述語論理の MC を **命題論理を経由せずに** 計算

述語論理の **対称性** を利用することで効率化

Lifted Inference の Trick

First-Order d-DNNF [Broeck 11a]

d-DNNF を述語論理に拡張

Partial Exchangeability を非常に効率的に表現

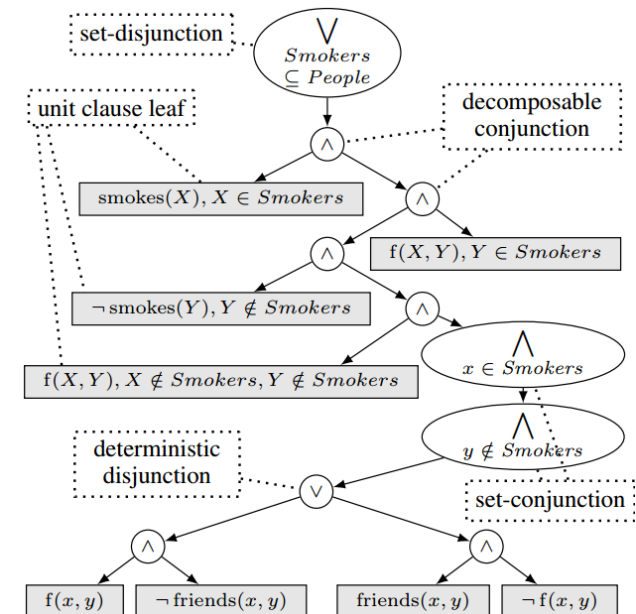
→ $O(2^{\text{Poly}(N)})$ を $O(\text{Poly}(N))$ に削減

FO d-DNNF は万能か?? No!!

モデルの持つ対称性に依存

強い対称性を仮定 [Broack+ 12]

(2項関係の観測があると #P-hard)



従来の考え

論理で確率モデルを記述

論理を別形式 (BDD, d-DNNF,...) に Compile (圧縮)

別形式がコンパクトなら効率的に計算可能

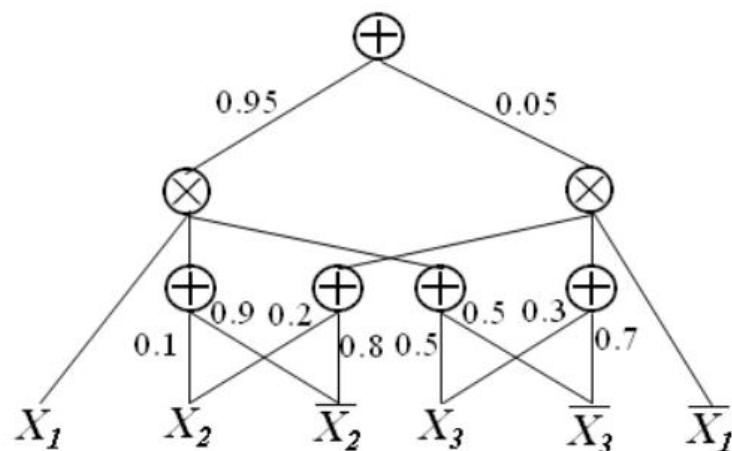
発想の転換

はじめてからコンパクトな形式で確率モデルを定義

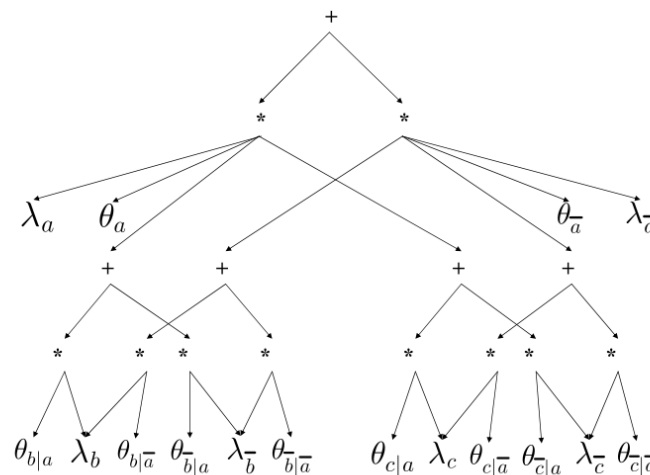
Sum-Product Network

d-DNNF と似た構造を利用して $p(x)$ を定義

- 厳密計算 [Poon+ 11] ← UAI Best Paper
- パラメータ学習 [Gens+ 12] ← NIPS Best Paper
- 構造学習 [Gens+ 13]
- ID-SPN [Rooshenas+ 14]



Sum-Product Network



Compiling BN to d-DNNF

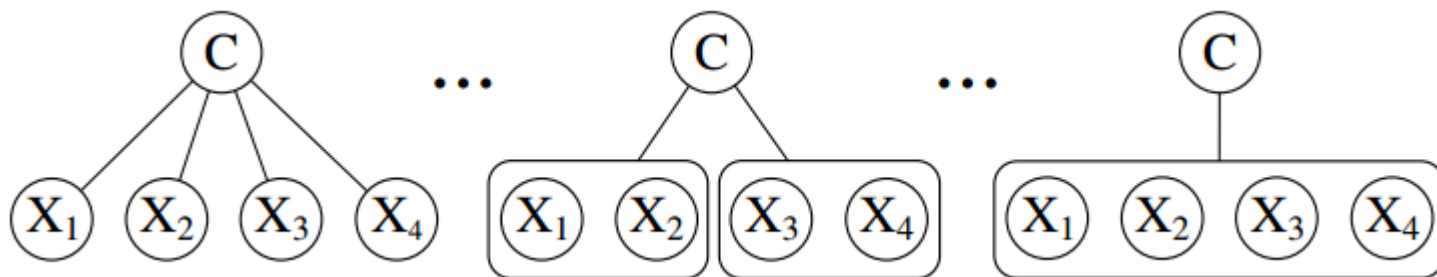
Exchangeable Variable Model [Niepert+ 14b]

Exchangeable Component の混合分布

Naïve Bayes を Partial Exchangeability に一般化
単純な計算で SVM 並みの性能

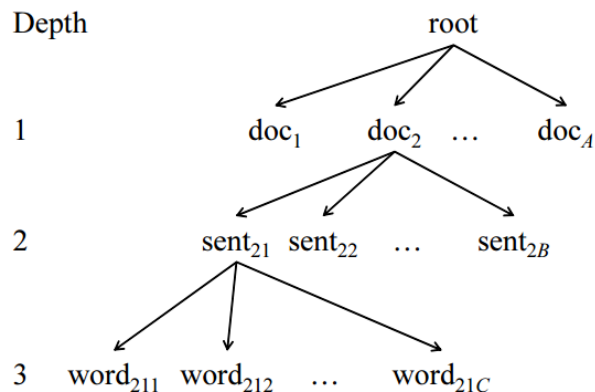
Exchangeable Component

Partial Exchangeable が成り立つ変数集合 X の生成分布
 $T(X)$ の値 j を生成し、 $T(x) = j$ なる x を一様に生成

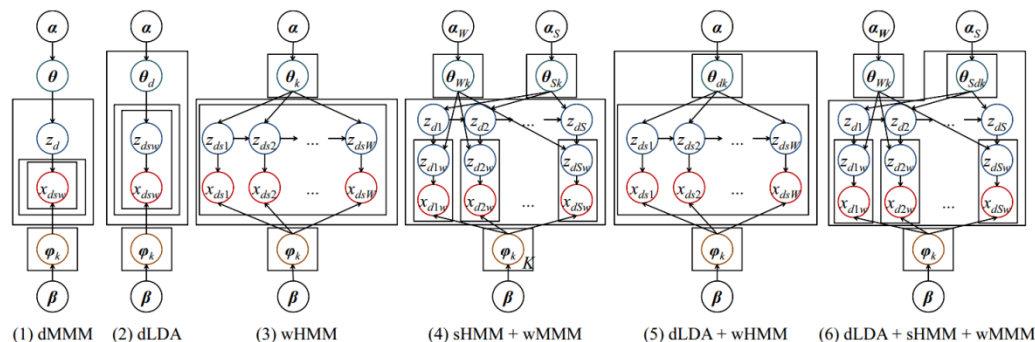


- データは**順序木**で表現される構造を持つと仮定
- 構造を元に**効率的に計算できる範囲**で Graphical Model を生成

Depth



文書データを表す順序木



Mixture, LDA, HMM その組合せを含む
様々なモデルからデータにあったものを生成

1. 離散分布の複雑さ
2. 離散分布を効率的に定義する by 論理
3. 離散分布を効率的に計算する by 動的計画法
4. 論理を用いた確率計算の例
5. まとめ

～ 2010 年

離散分布は複雑

離散分布は 論理 を用いて効率的に定義可能 (Probabilistic Logic)

論理 は グラフ構造 を用いて効率的に計算可能 (Knowledge Compilation)

2010年～

最も効率的に計算できるのはどのようなときか？ (Lifted Inference)

離散分布を直接効率的に計算できるグラフで定義 (Tractable model)

人工知能ブーム

Deep Learning に便乗

論理(やグラフ)という古き良き AI の技を ML 業界で流行らせたい
(という意図が見られる)



Reference

[Darwiche 02] Adnan Darwiche: “A Knowledge Compilation Map”, JAIR, 2002

[Bryant 86] Randal E. Bryant: "Graph-Based Algorithms for Boolean Function Manipulation", IEEE Transactions on Computers, 1986

[Minato 93] Shin-ichi Minato: "Zero-suppressed BDDs for set manipulation in combinatorial problems", DAC '93

[Darwiche 01] Adnan Darwiche: “On the tractability of counting theory models and its application to belief revision and truth maintenance”, Journal of Applied Non-Classical Logics 2001

[Pearl 85] Judea Pearl: “Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning”, CogSci85

[Pearl 82] Judea Pearl: “Reverend Bayes on inference engines: A distributed hierarchical approach”, AAAI Press

[Kim+ 83] Jin H Kim, Judea Pearl: “A computational model for combined causal and diagnostic reasoning in inference systems”, IJCAI83

[Pearl 88] Judea Pearl: “Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference”

[Lauritzen88] Lauritzen, S.L. and Spiegelhalter, D.: “local computations with probabilities on graphical structures and their application to expert systems”, JRSS88

[Boutilier+ 96] C. Boutilier, N. Friedman, M. Goldszmidt, and D. Koller: “Context-Specific Independence in Bayesian Networks”, UAI96

[Chavira+ 05] Mark Chavira, Adnan Darwiche: “Compiling Bayesian Networks with Local Structure”, IJCAI05

[Minato+ 07] S. Minato, K. Satoh, and T. Sato: “Compiling Bayesian Networks by Symbolic Probability Calculation Based on Zero-suppressed BDDs”, IJCAI07

[Chavira+ 07] “Compiling Bayesian Networks Using Variable Elimination ”, IJCAI07

[Ishihata+ 11] M. Ishihata, T. Sato and S. Minato: "Compiling Bayesian Networks for Parameter Learning based on Shared BDDs“, AAI11

- [Getoor+ 07] Lise Getoor and Ben Taskar: “Introduction to Statistical Relational Learning”, MIT Press
- [DeRaedt+ 08] L. DeRaedt and K. Kersting: “Probabilistic inductive logic programming”, Springer
- [Sato+ 01] T. Sato and Y. Kameya: “Parameter learning of logic programs for symbolic-statistical modeling”, JAIR
- [Vennekens+ 04] Vennekens, J., Verbaeten, S., Bruynooghe, M.: “Logic programs with annotated disjunctions”, ICLP04
- [Richardson+ 06] M. Richardson and P. Domingos: “Markov logic networks”, Machine Learning
- [DeRaedt+ 07] L. De Raedt, A. Kimming and H. Toivonen: “ProbLog: a probabilistic Prolog and its application in link discovery”, IJCAI07
- [亀谷 11] 亀谷由隆: “論理に基づく確率モデリングのこれまで, これから”, 第4回 IBISML チュートリアル

- [Poole 03] David Poole “First-order probabilistic inference”, IJCAI03
- [Broeck 11a] Guy Van den Broeck, Nima Taghipour, Wannes Meert, Jesse Davis and Luc De Raedt: “Lifted Probabilistic Inference by First-Order Knowledge Comppilation”, IJCAI11
- [Broeck 11b] Guy Van den Broeck: “On the completeness of first-order knowledge compilation for lifted probabilistic inference”, NIPS11
- [Broeck+ 12] Guy Van den Broeck and Adnan Darwiche: “On the complexity and approximation of binary evidence in lifted inference”, NIPS12
- [Niepert+ 14a] Mathias Niepert and Guy Van den Broeck: “Tractability through Exchangeability: A New Perspective on Efficient Probabilistic Inference”, AAI’14

[Poon+ 11] Hoifung Poon and Pedro Domingos: “Sum-Product Networks: A New Deep Architecture”, UAI11

[Gens+ 12] Robert Gens and Pedro Domingos: “Discriminative Learning of Sum-Product Networks”, NIPS12

[Gens+ 13] Robert Gens and Pedro Domingos: “Learning the Structure of Sum-Product Networks”, ICML13

[Rooshenas+ 14] Amirmohammad Rooshenas and Daniel Lowd: “Learning Sum-Product Networks with Direct and Indirect Variable Interactions”, ICML14

[Niepert+ 14b] Mathias Niepert and Pedro Domingos: “Exchangeable Variable Models”, ICML14

[Ishihata+ 14] Masakazu Ishihata and Tomoharu Iwata: “Generating structure of latent variable models for nested data”, UAI14