論理に基づく確率モデリングのこれまで，これから

亀谷由隆（東京工業大学）

発表内容

- はじめに
- “論理 + 確率” における諸論点
- 論理に基づく確率モデリングのこれまで
  - – 命題論理に基づく確率推論の高速化
  - – 統計的関係学習 (SRL) / 確率論理学習 (PLL)
  - – 確率モデリング言語処理系
    - PRISM
- 論理に基づく確率モデリングのこれから
- まとめ

発表内容

- はじめに
- “論理 + 確率” における諸論点
- 論理に基づく確率モデリングのこれまで
  - 命題論理に基づく確率推論の高速化
  - 統計的関係学習 (SRL) / 確率論理学習 (PLL)
  - 確率モデリング言語処理系
    - PRISM
- 論理に基づく確率モデリングのこれから
- まとめ

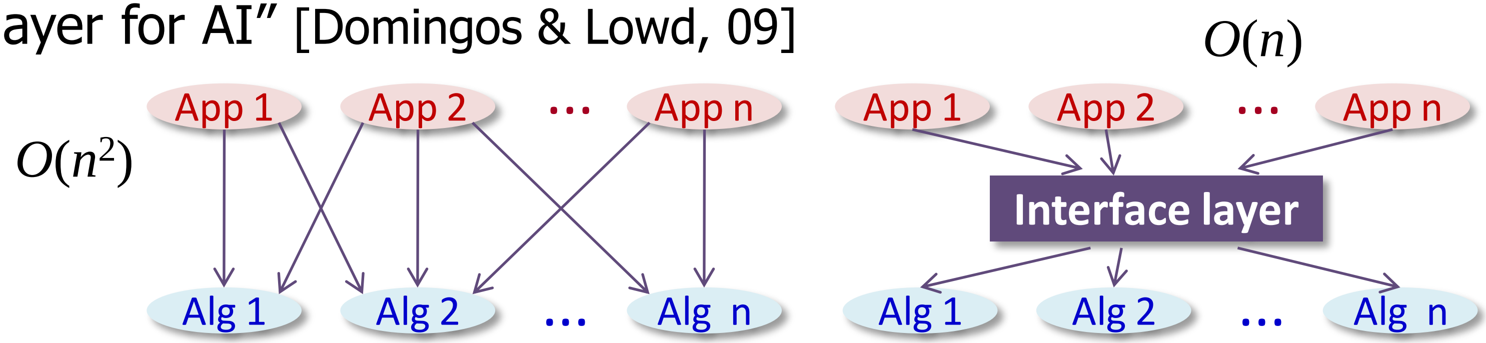
はじめに：なぜ"論理 + 確率"なのか?

- 複雑かつ不確定な現象をうまくモデリングしたい
- 論理側から
  - 不確定性の系統的な取り扱い，複数の解候補に対するランク付け
    - 確率的 RDB，確率的プランニング，
      アブダクションにおける仮説評価，単一化文法の確率モデル化
- 確率側から
  - 論理の持つ記述力と理解/説明可能性を取り入れる
  - 確率モデルの命題論理化 → コンパイル技法による確率推論の高速化
  - 確率モデルの述語論理化

    → コンパクトな表現，一般法則化，
    構造的データ（系列，木，RDB etc.）の系統的取り扱い
- どちらも確立された研究対象
  - 原則に基づき方法論を展開できる／実問題に応用できる

はじめに：AI における交差点

- “Interface layer for AI” [Domingos & Lowd, 09]



- 論理に基づく確率モデリングで考えてみると …
  - 命題論理
    - 意味論（確立している）
      - 論理：真理値表（可能世界 = 真理割り当て）
      - 確率：有限確率測度（可能世界 = 標本点）
    - 関連手法：SAT，BDD/ZDD，ベイズネット，決定木，Apriori，…
  - 一階述語論理
    - 意味論（どう融合する?）
      - 論理：最小モデル意味論，安定モデル意味論，…
      - 確率：一般の確率測度
    - 関連手法：定理証明，機械学習？（統計的関係学習，確率論理学習），…

深化

一般化

発表内容

✓ はじめに

- “論理 + 確率” における諸論点
- 論理に基づく確率モデリングのこれまで
  - 命題論理に基づく確率推論の高速化
  - 統計的関係学習 (SRL) / 確率論理学習 (PLL)
  - 確率モデリング言語処理系
    - PRISM
- 論理に基づく確率モデリングのこれから
- まとめ

“論理 + 確率”の諸論点：意味論

- 論理：disconnected by default
  - 推論機構が公理を組み上げることにより結論を導く
- 確率：connected by default
  - 独立性の仮定を導入することで同時分布を積の形に分解する

[Sato, ILP-09]

- “論理 + 確率” における典型的なアプローチ：
  - 論理表現で記述された事実や規則に確率（重み）を付与
  - それらの事実・規則を組み上げて確率モデル（同時分布）を構築

- **問題**：同時分布（同時確率測度）は本当に存在するのか?

  → 確率的な意味論の構築が必要
  （e.g. 分布意味論 [Sato, ICLP-95][Sato et al., JAIR-01][Sato et al., 08]）

"論理 + 確率"の諸論点：パラメータの付与

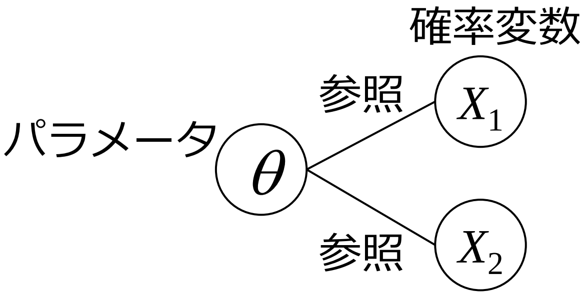
- 論理式のどこにパラメータを付与するか?
  1. 個々の基本命題にパラメータを付ける (PHA, PRISM, ProbLog)
  2. 規則 $x \Rightarrow y$ にパラメータ θ を付ける（その他多数）
     - θ を条件付き確率 $p(y \mid x)$ と見なす
       → ベイズネットのマクロ記法
       （無限サイズのベイズネットは単純に定義できない）
     - θ を規則 $x \Rightarrow y$ が成り立つ確率と見なす
       → $\theta = p(x \Rightarrow y) = p(\neg x \vee y) = p(\neg x \vee (x \wedge y))$
       $= p(\neg x) + p(x \wedge y) = p(\neg x) + p(x)\, p(y \mid x)$

- θ が更に細かい確率に分解
- $p(\neg x)$ が大きくなると θ も増大

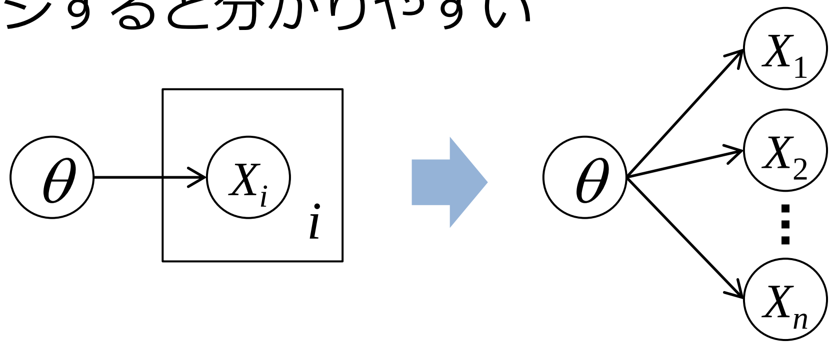
- 条件付き確率は取り扱いにくい概念
  – 基本的に $p(y \mid x)$ と $p(y \mid x')$ に直接の関係はない（$x \neq x'$）
    ➔ 規則 $x \Rightarrow y$ と $x' \Rightarrow y$ が確率的にうまく結び付かない

“論理 + 確率”の諸論点：パラメータ共有

- パラメータ共有 (parameter tying)
  - 複数の確率変数において，分布がたまたま一致するのではなく常に一致する（e.g. 時不変なマルコフ過程）

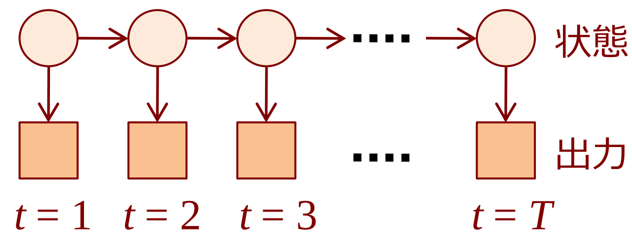


  - SRL/PLL 分野で（暗黙的に）多用される
    - マクロ表記上でパラメータが付与される
    - 命題レベルの確率モデルにマクロ展開されたとき，そのパラメータは確率モデル中の各所で共有される
    - 確率計算では違いは出ないが，パラメータ学習時に違いが起こる
    - プレート表記でイメージすると分かりやすい



“論理 + 確率”の諸論点：独立性 (1 of 3)

- 独立性を仮定 → 確率推論の高速化が可能
  - 隠れマルコフモデルの確率推論アルゴリズム（Viterbi 等）



- 条件付き独立性と文脈依存独立性 (context-sensitive independence, CSI) [Boutilier et al. UAI-96]
  - $\boldsymbol{Z}$ が与えられたとき $\boldsymbol{X}$ と $\boldsymbol{Y}$ は**条件付き独立**(conditionally independent) ⇔

$$\forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \;\; p(\boldsymbol{x} \mid \boldsymbol{z}, \boldsymbol{y}) = p(\boldsymbol{x} \mid \boldsymbol{z}) \;\; \big(p(\boldsymbol{y}, \boldsymbol{z}) > 0\big)$$

変数単位の関係

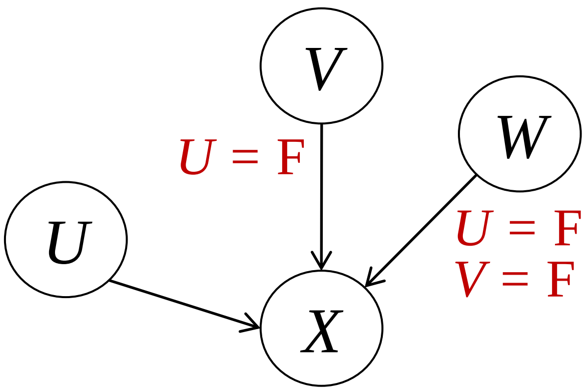
  - $\boldsymbol{Z}$ および文脈 $\boldsymbol{c}$ が与えられたとき $\boldsymbol{X}$ と $\boldsymbol{Y}$ は**文脈的に独立**(contextually independent) ⇔

$$\forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \;\; p(\boldsymbol{x} \mid \boldsymbol{z}, \boldsymbol{c}, \boldsymbol{y}) = p(\boldsymbol{x} \mid \boldsymbol{z}, \boldsymbol{c}) \;\; \big(p(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{c}) > 0\big)$$

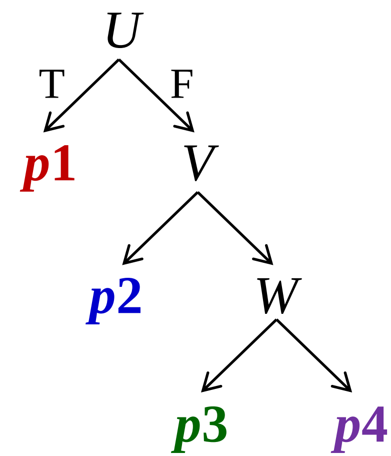
値単位の関係

"論理 + 確率"の諸論点：独立性 (2 of 3)

- Parameter equality [Chavira & Darwiche, IJCAI-05]
  - CSI が成り立つとき，条件付き確率表中に値の等しいエントリが出現



U	V	W	$p(X$=T $\mid U, V, W)$
T	T	T	***p*1**
T	T	F	***p*1**
T	F	T	***p*1**
T	F	F	***p*1**
F	T	T	***p*2**
F	T	F	***p*2**
F	F	T	***p*3**
F	F	F	***p*4**



決定木表現

- Contingent Bayesian networks [Milch, 06]
  - 文脈依存独立性 (CSI) をベイズネット上で表現
  - リンクが有効な場合をラベル付け

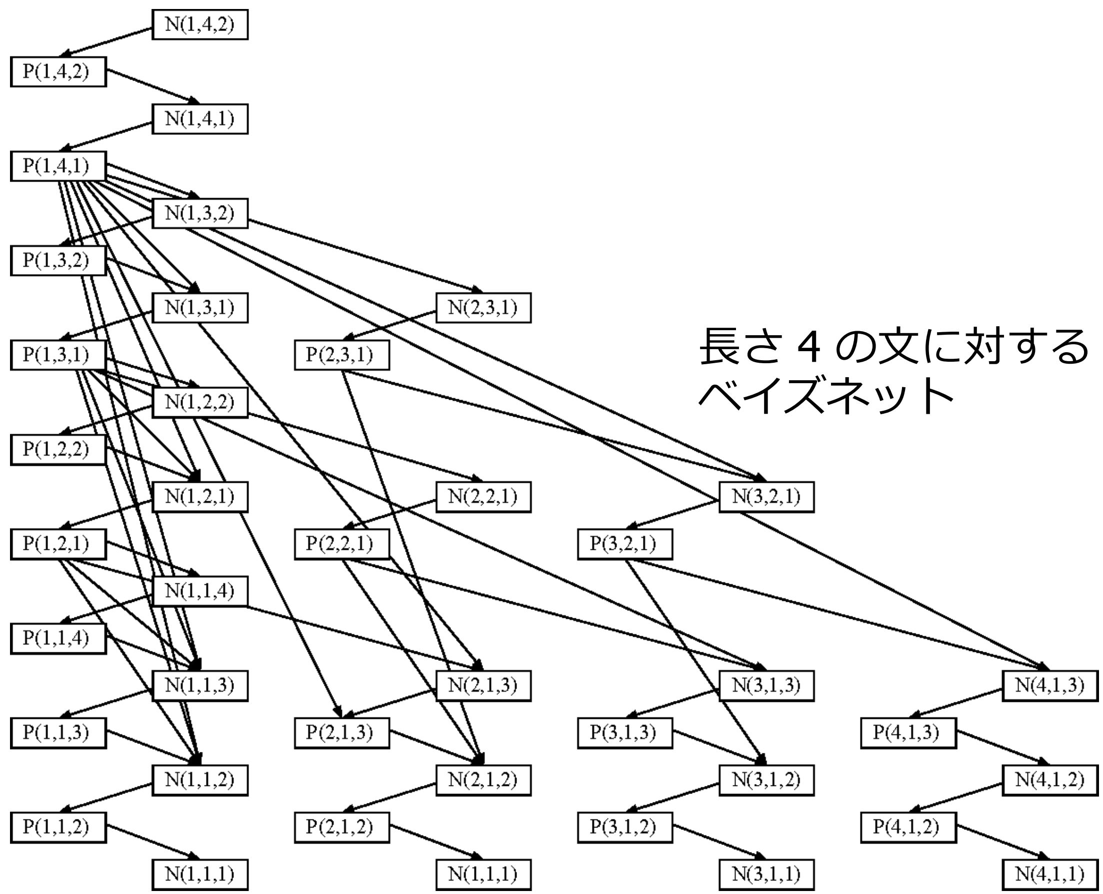
他の確率変数の値によっては
リンクが切れる → より簡単な構造に

“論理 + 確率”の諸論点：独立性 (3 of 3)

- Pynadath & Wellman (1998) の試み：

ベイズネットで確率文脈自由文法を表現する

➔ 確率推論：文長に対して指数オーダの計算量



長さ 4 の文に対する
ベイズネット

依存関係があるからといって
全部リンクでつなぐと確率推論
が破綻することがある
（CSI 等を考慮した細粒度な
モデルが必要）

"論理 + 確率"の諸論点：確率変数の命題化

- 命題化 (propositionalization)
  - 元々は ILP 分野の一手法に付けられた名前 [Kramer et al. 02]
  - 命題論理で確率モデルを表現した場合に用いる
  - 述語論理レベルの記述を代入展開して命題論理レベルのデータ構造を得る際にも用いる

- "Name trick"
  - 命題記号と条件付き確率を対応付けるときに用いる
  - 条件式で名前を修飾

    e.g. $p(Y = y \mid X = x)$ に対して "$Y_x = y$" [Ishihata et al. ACML-10] や

    $\theta_{y|x}$ [Chavira & Darwiche, IJCAI-05] など
  - 条件付き確率表の一行に対応
  - 命題論理の AND 式 → 確率積と見なせる

    e.g. $\pi(\text{“}X = x\text{”} \wedge \text{“}Y_x = y\text{”})$

    $= \pi(\text{“}X = x\text{”})\, \pi(\text{“}Y_x = y\text{”})$

    $= p(X = x)\, p(Y = y \mid X = x) = p(X = x, Y = y)$

$p(\cdot)$: 命題化前の確率分布

$\pi(\cdot)$: 命題化後の確率分布

"論理 + 確率"の諸論点：関係データ

- 従来の機械学習：単一の関係（表，データ行列）からの学習
- 関係データベースからの機械学習：
  - 多重性 (multiplicity) の扱い
  - join による単一表化：
    論理的には OK, 統計的には NG（iid を仮定した場合）

Customer

ID	Gender	Income	Age
...	...	...	...
3478	m	60-70	32
3479	f	80-90	45
...	...	...	...

Order

OrderID	CustomerID	StoreID	Delivery	Payment
...	...	...	...	...
2140267	3478	12	regular	cash
3446778	3478	12	express	check
4728386	3478	17	regular	check
3233444	3479	17	express	credit
3475886	3479	12	regular	credit
...	...	...	...	...

Store

ID	Size	Type	Location
...	...	...	...
12	small	franchise	city
17	large	indep.	rural
...	...	...	...

※ [Wrobel, 02]
の例を簡単化

- 多関係 (multi-relational) データマイニング
  - 帰納論理プログラミング (inductive logic programming, ILP) の一分野
- 関係データベースを扱う確率モデル：
  PRM (probabilistic relational model) [Friedman et al. IJCAI-99] （しかし決定打ではない）

発表内容

- ✓ はじめに
- ✓ "論理 + 確率" における諸論点
- 論理に基づく確率モデリングのこれまで
  - 命題論理に基づく確率推論の高速化
  - 統計的関係学習 (SRL) / 確率論理学習 (PLL)
  - 確率モデリング言語処理系
    - PRISM
- 論理に基づく確率モデリングのこれから
- まとめ

命題論理に基づく確率推論：概要 (1 of 2)

- ベイズネットの条件付き確率表には多くの文脈依存独立性 (CSI), ゼロ確率が存在する
  - 特に人手で作った条件付き確率表では顕著
  - ベイズネットの標準的ベンチマーク [Charvira & Darwiche, IJCAI-05]

Network	....	Vars	Card	....	Total Parms	....	%Det	%DP
alarm		37	2-4		752		0.9	24.6
bm		1005	2-2		6972		99.6	100.0
diabetes		413	3-21		461069		78.2	17.6
hailfinder		56	2-11		3741		15.7	26.9
mildew		35	3-100		547158		93.2	25.1
mm		1220	2-2		8326		98.7	75.0
munin1		189	1-21		19466		66.5	61.2
munin2		1003	2-21		83920		63.3	69.5
munin3		1044	1-21		85855		63.1	71.3
munin4		1041	1-21		98183		64.5	65.3
pathfinder		109	2-63		97851		56.1	5.1
pigs		441	3-3		8427		56.2	23.9
students		376	2-2		2616		90.7	79.3
tcc4f		105	2-2		3236		0.4	35.6
water		32	3-4		13484		54.0	57.0

0/1 であるパラメータの割合

同一 CPTにおいて異なる非 0/1 パラメータの割合

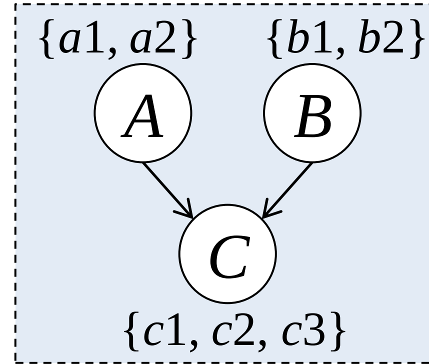
命題論理に基づく確率推論：概要 (2 of 2)

- 命題論理によるベイズネット（確率モデル）の符号化：
  - [Chavira & Darwiche, IJCAI-05]:
    IP 節 + PI 節 → d-DNNF → Arithmetic Circuit (AC)
  - [Sang et al. AAAI-05]
  - [Minato et al., IJCAI-07]:
    ZDD (Zero-suppressed BDD) for Multi-linear Function (MLF)
  - [Ishihata et al., ACML-10]:
    BDD-EM (BO-EM):
    順序符号化 (order encoding) → BDD (binary decision diagram)

- CSI／ゼロ確率の情報を用いて圧縮

命題論理に基づく確率推論：ベイズネットのコンパイル技法 (2 of 3)

- [Chavira & Darwiche, IJCAI-05] の符号化：



変数 A: $\lambda_{a1} \vee \lambda_{a2}$ $\quad \neg\lambda_{a1} \vee \neg\lambda_{a2}$

変数 B: $\lambda_{b1} \vee \lambda_{b2}$ $\quad \neg\lambda_{b1} \vee \neg\lambda_{b2}$

変数 C: $\lambda_{c1} \vee \lambda_{c2} \vee \lambda_{c3}$ $\quad \neg\lambda_{c1} \vee \neg\lambda_{c2}$

$\neg\lambda_{c1} \vee \neg\lambda_{c3}$

$\neg\lambda_{c2} \vee \neg\lambda_{c3}$

at least one

at most one

ゼロ確率，CSI を利用して最適化

A の CPT: $\lambda_{a1} \Leftrightarrow \theta_{a1} \quad \lambda_{a2} \Leftrightarrow \theta_{a2}$

B の CPT: $\lambda_{b1} \Leftrightarrow \theta_{b1} \quad \lambda_{b2} \Leftrightarrow \theta_{b2}$

C の CPT: $\lambda_{a1} \wedge \lambda_{b1} \wedge \lambda_{c1} \Leftrightarrow \theta_{c1|a1b1} \quad \lambda_{a1} \wedge \lambda_{b1} \wedge \lambda_{c2} \Leftrightarrow \theta_{c2|a1b1} \quad \lambda_{a1} \wedge \lambda_{b1} \wedge \lambda_{c3} \Leftrightarrow \theta_{c3|a1b1}$

$\lambda_{a1} \wedge \lambda_{a2} \wedge \lambda_{c1} \Leftrightarrow \theta_{c1|a1b2} \quad \lambda_{a1} \wedge \lambda_{b2} \wedge \lambda_{c2} \Leftrightarrow \theta_{c2|a1b2} \quad \lambda_{a1} \wedge \lambda_{b2} \wedge \lambda_{c3} \Leftrightarrow \theta_{c3|a1b2}$

…

<u>命題変数</u>

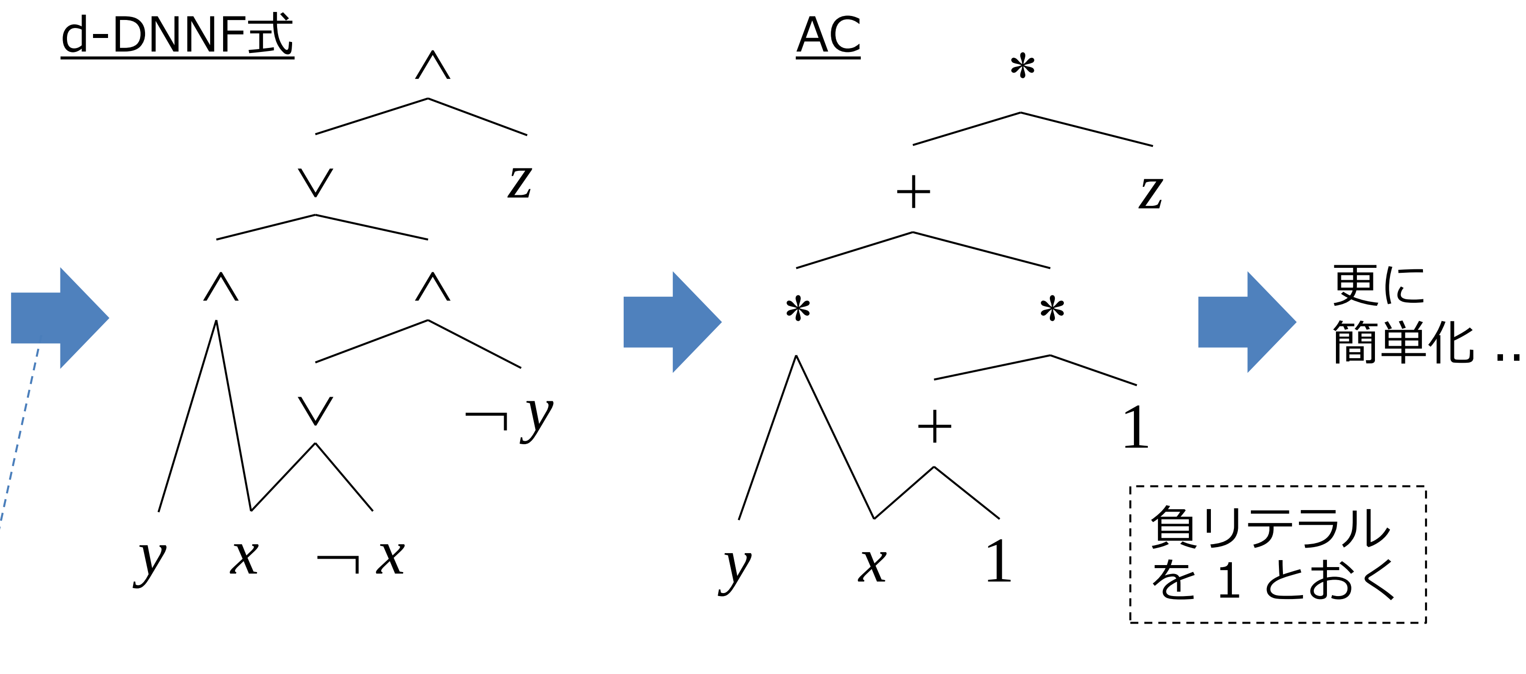
λ_{xi}: 指示変数（X が値 xi をとれば 1）

θ_{xi}: パラメータ変数（X が値 xi をとる確率のマーカ）

命題論理に基づく確率推論：ベイズネットのコンパイル技法 (2 of 3)

- [Chavira & Darwiche, IJCAI-05]
  - d-DNNF 式への変換 + Arithmetic Circuit (AC) への変換

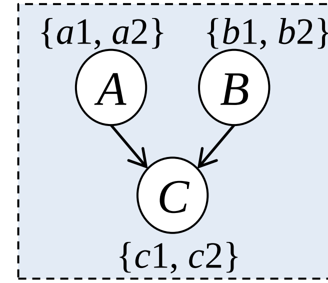
$$y \vee z$$
$$\neg y \vee z$$
$$x \vee \neg y$$
$$x \vee z$$



因数分解：変数 u を一つ選んで（この例では z, y, x の順 + smoothing 処理）

$$factor(\Delta) = u \wedge factor(\Delta \mid u) \vee \neg u \wedge factor(\Delta \mid \neg u)$$

命題論理に基づく確率推論：
ベイズネットのコンパイル技法 (3 of 3)



- [Minato et al., IJCAI-07] の符号化：
  - Multi-linear Function (MLF) 式

$$\begin{aligned}\mathrm{MLF} = {} & \lambda_{a1}\lambda_{b1}\lambda_{c1}\theta_{a1}\theta_{b1}\theta_{c1|a1b1} + \lambda_{a1}\lambda_{b1}\lambda_{c2}\theta_{a1}\theta_{b1}\theta_{c2|a1b1} + \\ & \lambda_{a1}\lambda_{b2}\lambda_{c1}\theta_{a1}\theta_{b2}\theta_{c1|a1b2} + \lambda_{a1}\lambda_{b2}\lambda_{c2}\theta_{a1}\theta_{b2}\theta_{c2|a1b2} + \ldots\end{aligned}$$

  - CSI の利用：値が同じパラメータを共通の命題変数で表現

$$\begin{aligned}\mathrm{MLF} = {} & \lambda_{a1}\lambda_{b1}\lambda_{c1}\theta_{a(0.6)}\theta_{b(0.2)}\theta_{c(0.7)} + \lambda_{a1}\lambda_{b1}\lambda_{c2}\theta_{a(0.6)}\theta_{b(0.2)}\theta_{c(0.3)} + \\ & \lambda_{a1}\lambda_{b2}\lambda_{c1}\theta_{a(0.6)}\theta_{b(0.8)}\theta_{c(0.7)} + \lambda_{a1}\lambda_{b2}\lambda_{c2}\theta_{a(0.6)}\theta_{b(0.8)}\theta_{c(0.3)} + \ldots\end{aligned}$$

  - MLF 式を組み合わせ集合と見なす（要素 = MLF 式の各項）
  - ZDD の積和演算を用いて再帰的に MLF 式を組み上げ

$$\mathrm{MLF}(C = c1) = \lambda_{c1} \cdot \sum_{(a,b)\in\{a1,a2\}\times\{b1,b2\}} \left(\theta_{c(\Pr(c|ab))} \cdot \mathrm{MLF}(A = a) \cdot \mathrm{MLF}(B = b)\right)$$

$$\mathrm{MLF} = \mathrm{MLF}(C) = \textstyle\sum_{c\in\{c1,c2\}} \mathrm{MLF}(C = c)$$

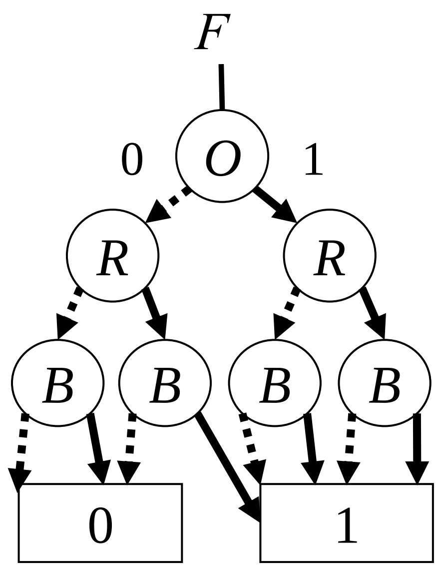
命題論理に基づく確率推論：BDD-EM (1 of 3)

- BDD-EM [石畠ら, 09a]
  - BDD (binary decision diagram) 上で動的計画法に基づく効率的な確率・期待値計算を行う EMアルゴリズム
  - 任意の命題論理式で動作する

- BDD: 論理関数をコンパクトに表現する有向非循環グラフ (DAG)

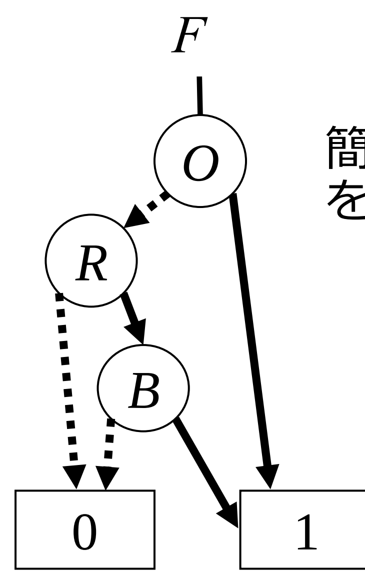
$$F \Leftrightarrow O \vee (R \wedge B)$$

O	R	B	F
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	1
1	0	1	1
1	1	0	1
1	1	1	1



Binary Decision Tree

簡約化規則を適用



Binary Decision Diagram

命題論理に基づく確率推論：BDD-EM (2 of 3)

- パラメータ推定問題
  - O, R, B は真になる確率 $\theta_O, \theta_R, \theta_B$ をもつ
  - 命題 F の真偽を観測 → $\theta_O, \theta_R, \theta_B$ を推定したい
- BDD-EM:
  - $\theta_O, \theta_R, \theta_B$ を初期化
  - E-step, M-step を収束するまで繰り返す

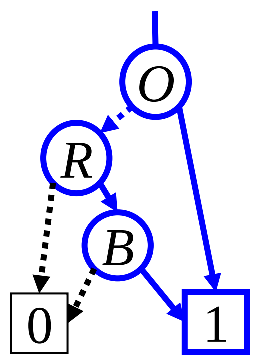
モデル：

$$F \Leftrightarrow O \vee (R \wedge B)$$

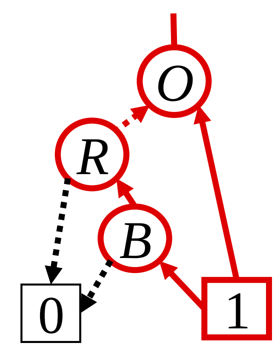
F: 遅刻する
O: 寝坊する
R: 雨が降っている
B: 渋滞している

E-step: BDD上で前向き確率 $\mathrm{F}_\theta[n]$ と後ろ向き確率 $\mathrm{B}_\theta[n]$ を計算し，
条件付き期待値 $\mathrm{E}_\theta[R=1 | F=1] = \mathrm{F}_\theta[R]\theta_R \mathrm{B}_\theta[B] / \mathrm{F}_\theta[1]$ を計算

$$\begin{aligned}
\mathrm{F}_\theta[O] &= 1 \\
\mathrm{F}_\theta[R] &= \theta_{\overline{O}} \\
\mathrm{F}_\theta[B] &= \theta_{\overline{O}}\theta_R \\
\mathrm{F}_\theta[1] &= \theta_O + \theta_{\overline{O}}\theta_R\theta_B
\end{aligned}$$



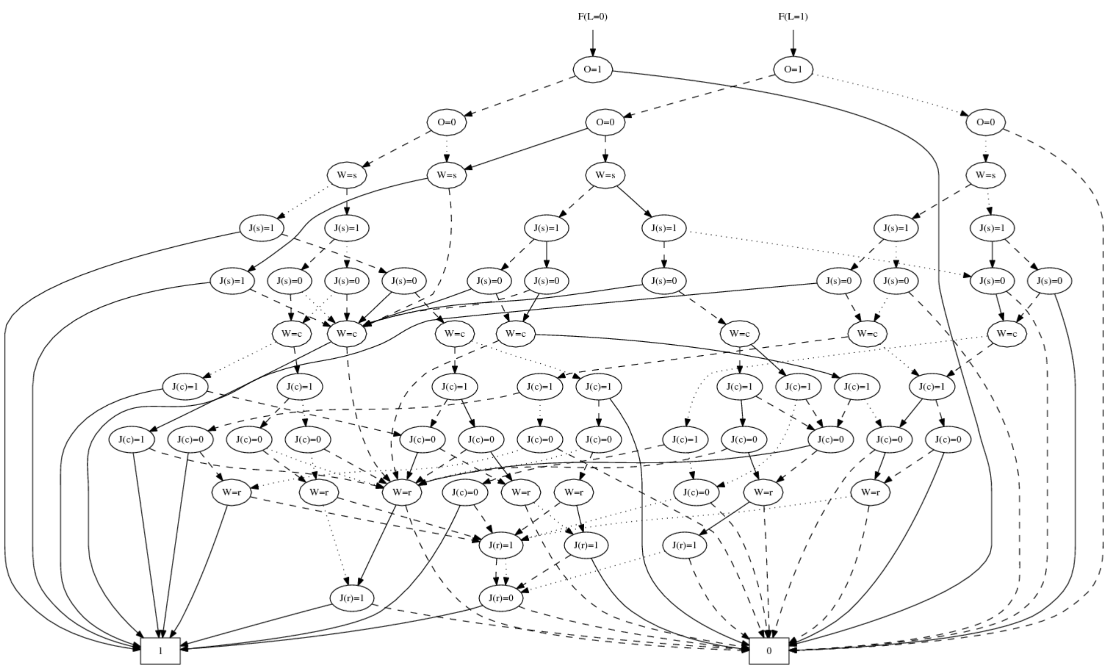
$$\begin{aligned}
\mathrm{B}_\theta[O] &= \theta_O + \theta_{\overline{O}}\theta_R\theta_B \\
\mathrm{B}_\theta[R] &= \theta_R\theta_B \\
\mathrm{B}_\theta[B] &= \theta_B \\
\mathrm{B}_\theta[1] &= 1
\end{aligned}$$



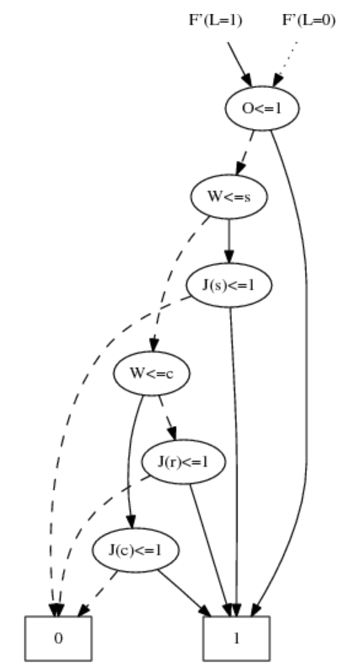
M-step: 条件付き期待値を用いて $\theta_O, \theta_R, \theta_B$ を更新する

命題論理に基づく確率推論：BDD-EM (3 of 3)

- ZDD [Ishihata et al., 08]
- Decomposed BDD [石畠ら, 09a]: 確率文脈自由文法に適用可能
- Shared BDD + 否定枝 [石畠ら, 09b & AMBN-10]
- 順序符号化の導入 [Ishihata et al., ACML-10]: 多値確率変数を取り扱う



直接符号化による命題式から得られた BDD



順序符号化による命題式から得られた BDD

発表内容

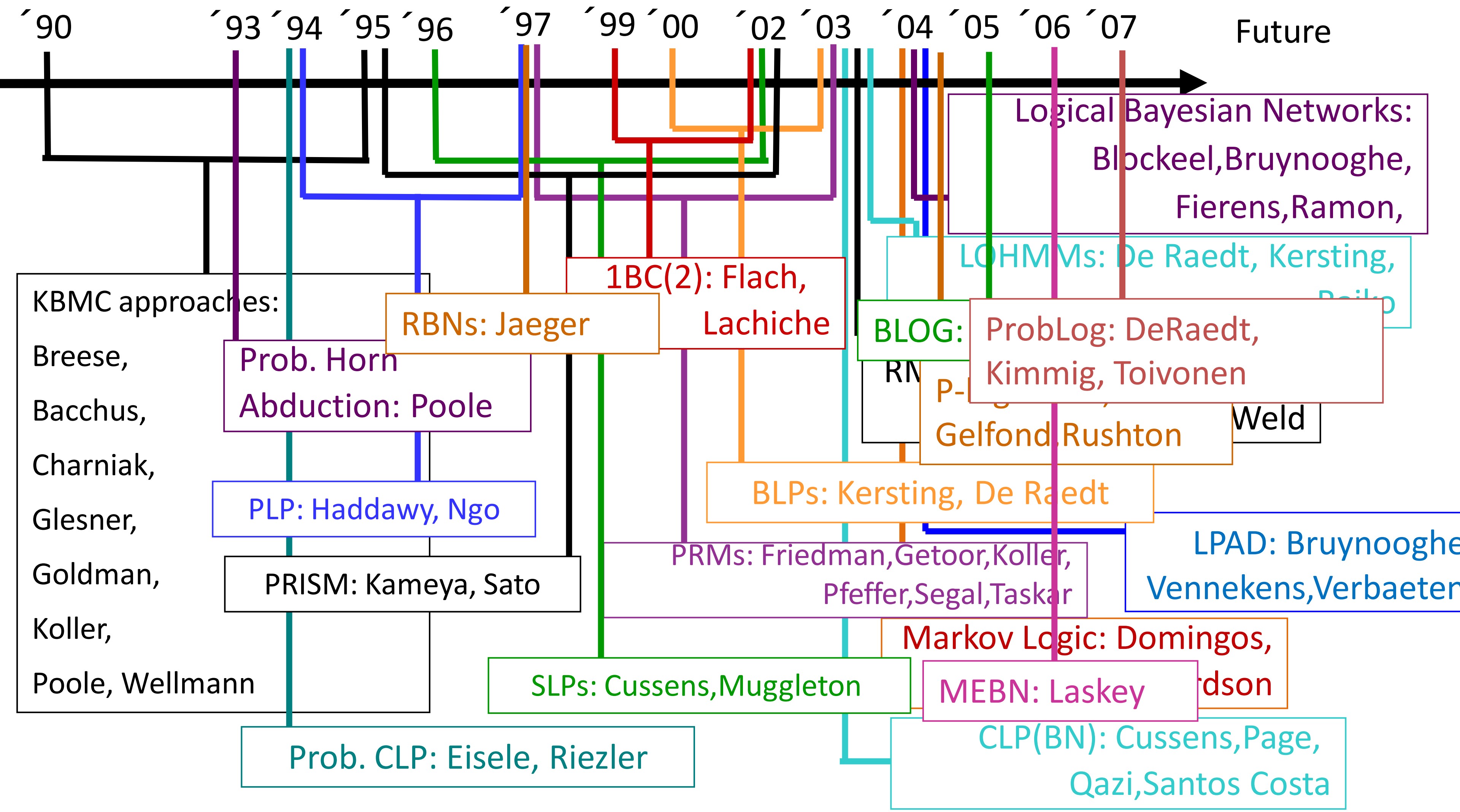
- ✓ はじめに
- ✓ “論理 + 確率” における諸論点
- 論理に基づく確率モデリングのこれまで
  - ✓ 命題論理に基づく確率推論の高速化
  - – 統計的関係学習 (SRL) / 確率論理学習 (PLL)
  - – 確率モデリング言語処理系
    - PRISM
- 論理に基づく確率モデリングのこれから
- まとめ

確率モデルの述語論理化 (1 of 2)

- プレート記法
- 統計的関係学習 (statistical relational learning, SRL)
  - ベイジアンネットワーク研究の周辺から生まれ主に北米で発展
- 確率論理学習 (probabilistic logic learning, PLL)
  - 帰納論理プログラミング研究の周辺から生まれ主に欧州で発展

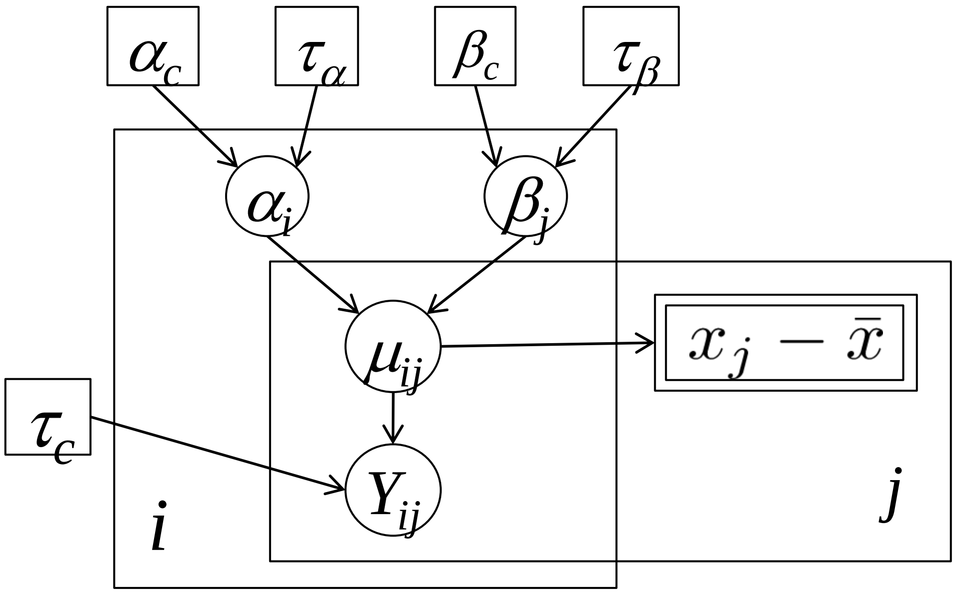
確率モデルの述語論理化 (2 of 2)

(from PLL Turotial in ICML 2004 by K.Kersting, modified by T. Sato)



プレート記法

- グラフィカルモデルのマクロ記述法 [Buntine 94, Gilks 94]
  - プレートにより繰り返しを表現



$$Y_{ij} \sim \mathrm{Normal}(\mu_{ij}, \tau_c)$$

$$\mu_{ij} = \alpha_i + \beta_i(x_j - \bar{x})$$

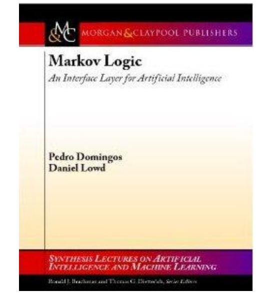
$$\alpha_i \sim \mathrm{Normal}(\alpha_c, \tau_\alpha)$$

$$\beta_j \sim \mathrm{Normal}(\beta_c, \tau_\beta)$$

統計的関係学習：KBMC

- KBMC: Knowledge-based Model Construction
  - 知識ベースからベイズネット（グラフィカルモデル）を構築
- 初期の試み：
  Breese (1992), Goldman & Charniak (1993), Ngo & Haddawy (1997), Koller & Pfeffer (1997)
- 発展形：
  - PRMs (Probabilistic Relational Models) [Friedman et al. 99]
    ➔ 関係データベース + ベイズネット
  - BLPs (Bayesian Logic Programs) [Kersting & De Raedt, 01],
    RBNs (Relational Bayesian networks) [Jaeger, 97]
    ➔ 論理プログラム + ベイズネット
  - DAPER (Directed Acyclic Probabilistic Entity-Relationship) [Heckerman et al. 04]
    ➔ ER モデル + ベイズネット
  - RMNs (Relational Markov Networks) [Taskar et al. 02]
    ➔ SQLクエリ + マルコフネットワーク
  - ...

統計的関係学習：MLN (1 of 6)



ISBN: 9781598296921

- Markov Logic Network [Richardson & Domingos, MLJ-06]
  - ➔ （任意の）一階述語論理式 + マルコフネットワーク
- 関係情報を属性 (feature) として扱える
- 述語論理式と重みの対の集合 $L = \{(F_i, w_i)\}_{i=1,2,\ldots,|L|}$
- 定数の有限集合 C

→ マルコフネットワーク $M_{L,C}$ を形成

Given:
$C = \{Anna, Bob\}$

F_i	w_i
$\forall x\ Smokes(x) \Rightarrow Cancer(x)$	1.5
$\forall x \forall y\ Friends(x,y) \wedge Smokes(x) \Rightarrow Smokes(y)$	1.1

F_i	w_i
$Smokes(Anna) \Rightarrow Cancer(Anna)$ $Smokes(Bob) \Rightarrow Cancer(Bob)$	1.5
$Friends(Anna, Anna) \wedge Smokes(Anna) \Rightarrow Smokes(Anna)$ $Friends(Anna, Bob) \wedge Smokes(Anna) \Rightarrow Smokes(Bob)$ $Friends(Bob, Anna) \wedge Smokes(Bob) \Rightarrow Smokes(Anna)$ $Friends(Bob, Bob) \wedge Smokes(Bob) \Rightarrow Smokes(Bob)$	1.1

統計的関係学習：MLN (2 of 6)

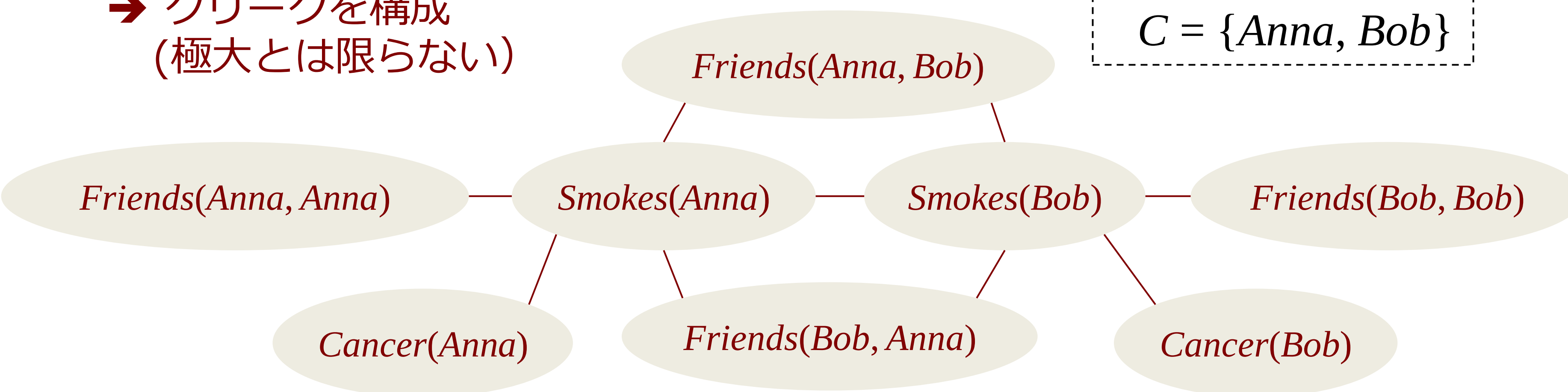
- 述語論理式と重みの対の集合 $L = \{(F_i, w_i)\}_{i=1,2,\ldots,|L|}$
- 定数の有限集合 C

→ マルコフネットワーク $M_{L,C}$

F_i	w_i
$Smokes(Anna) \Rightarrow Cancer(Anna)$ $Smokes(Bob) \Rightarrow Cancer(Bob)$	1.5
$Friends(Anna, Anna) \wedge Smokes(Anna) \Rightarrow Smokes(Anna)$ $Friends(Anna, Bob) \wedge Smokes(Anna) \Rightarrow Smokes(Bob)$ $Friends(Bob, Anna) \wedge Smokes(Bob) \Rightarrow Smokes(Anna)$ $Friends(Bob, Bob) \wedge Smokes(Bob) \Rightarrow Smokes(Bob)$	1.1

- ➢ 基礎アトム = ノード
- ➢ 同じ節に含まれる基礎アトム
  ➔ クリークを構成
  (極大とは限らない)

Given:
$C = \{Anna, Bob\}$



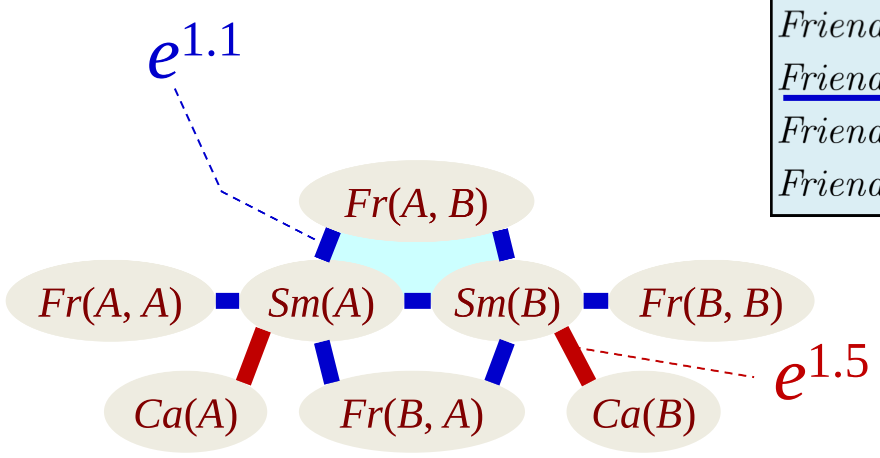
統計的関係学習：MLN (3 of 6)

- マルコフネットワーク $M_{L,C}$ の表す分布

$$P(X=x)=\frac{1}{Z}\exp\left(\sum_i w_i n_i(x)\right)$$
$$=\frac{1}{Z}\prod_i \underbrace{\phi_i(x_{\{i\}})}_{=e^{w_i}}{}^{n_i(x)}$$

x: 可能世界 = ノードへの値割り当て
$n_i(x)$: F_i に変数代入した式のうち，可能世界 x で成り立つものの数
Z: 正規化定数（分割関数）

F_i	w_i
$Smokes(Anna) \Rightarrow Cancer(Anna)$ $Smokes(Bob) \Rightarrow Cancer(Bob)$	1.5
$Friends(Anna, Anna) \wedge Smokes(Anna) \Rightarrow Smokes(Anna)$ $Friends(Anna, Bob) \wedge Smokes(Anna) \Rightarrow Smokes(Bob)$ $Friends(Bob, Anna) \wedge Smokes(Bob) \Rightarrow Smokes(Anna)$ $Friends(Bob, Bob) \wedge Smokes(Bob) \Rightarrow Smokes(Bob)$	1.1



統計的関係学習：MLN (4 of 6)

- MLN の利点：汎用性（任意の一階述語論理式 + マルコフネット）
- 正規化定数 Z があるため，一般には確率推論が困難

$$P(X=x)=\frac{1}{Z}\exp\left(\sum_i w_i n_i(x)\right)$$

- 確率推論：
  - MPE 推論（質問変数に対する最尤な値割り当てを見つける）
    - MaxWalkSAT を使用
  - 条件付き確率計算
    - MCMC (Slice sampling) → MC-SAT (SampleSAT)
  - 改良・発展形
    - LazySAT, Lazy-MC-SAT：必要に応じてマルコフネットワークに展開
    - CPI (Cutting Plane Inference) [Riedel, UAI-08]：整数線形計画法に基づく
- 重み学習
  - Markov blanket に基づく疑似尤度
  - 判別タスクの場合は L-BFGS などの準ニュートン法を使う

統計的関係学習：MLN (5 of 6)

- 応用 [Domingos & Lowd, 09]
  - 社会ネットワーク解析（リンク予測）
  - Collective classification
  - Entity resolution
  - 情報抽出
  - 共参照解析
  - Robot mapping
  - テキストからの意味ネットワーク抽出
  - ...

ロジスティック回帰式

$$\log\left(\frac{P(C=1 \mid F=f)}{P(C=0 \mid F=f)}\right) = a + \sum_i b_i f_i$$

F_i	w_i
$PageClass(p, +c)$	$w_1(c)$
$Has(p,+w) \Rightarrow PageClass(p, +c)$ ※	$w_2(w, c)$
$PageClass(p, c_1) \wedge (c_1 \neq c_2) \Rightarrow \neg PageClass(p, c_2)$	∞
$\exists c\ PageClass(p, c)$	∞
$Linked(u_1, u_2) \wedge PageClass(u_1, +c_1) \wedge PageClass(u_2, +c_2)$	$w_3(c_1, c_2)$

Collective feature

※ $Has(p,+w) \wedge PageClass(p, +c)$?

統計的関係学習：MLN (6 of 6)

- 直観的理理解は？

$$P(X=x)=\frac{1}{Z}\exp\left(\sum_i w_i n_i(x)\right)$$
$$=\frac{1}{Z}\prod_i \underbrace{\phi_i(x_{\{i\}})}_{=e^{w_i}}{}^{n_i(x)}$$

- トートロジーを知識ベース L に追加すると確率分布が変わる
- 規則の前提部が偽になる割り当て x に対し確率 $P(x)$ が大きくなる

トートロジー

F_i	w_i
$Smokes(Anna) \Rightarrow Cancer(Anna)$	1.5
$Smokes(Bob) \Rightarrow Cancer(Bob)$	
$Friends(Anna, Anna) \wedge Smokes(Anna) \Rightarrow Smokes(Anna)$	1.1
$Friends(Anna, Bob) \wedge Smokes(Anna) \Rightarrow Smokes(Bob)$	
$Friends(Bob, Anna) \wedge Smokes(Bob) \Rightarrow Smokes(Anna)$	
$Friends(Bob, Bob) \wedge Smokes(Bob) \Rightarrow Smokes(Bob)$	

$$\begin{cases} \mathrm{F} & Fr(A, B)=\mathrm{T},\ Sm(A)=\mathrm{T},\ Sm(B)=\mathrm{F} \\ \mathrm{T} & \text{otherwise} \end{cases}$$

$Friends(Anna, Bob) \wedge Smokes(Anna) \wedge Smokes(Bob)$?

確率論理学習

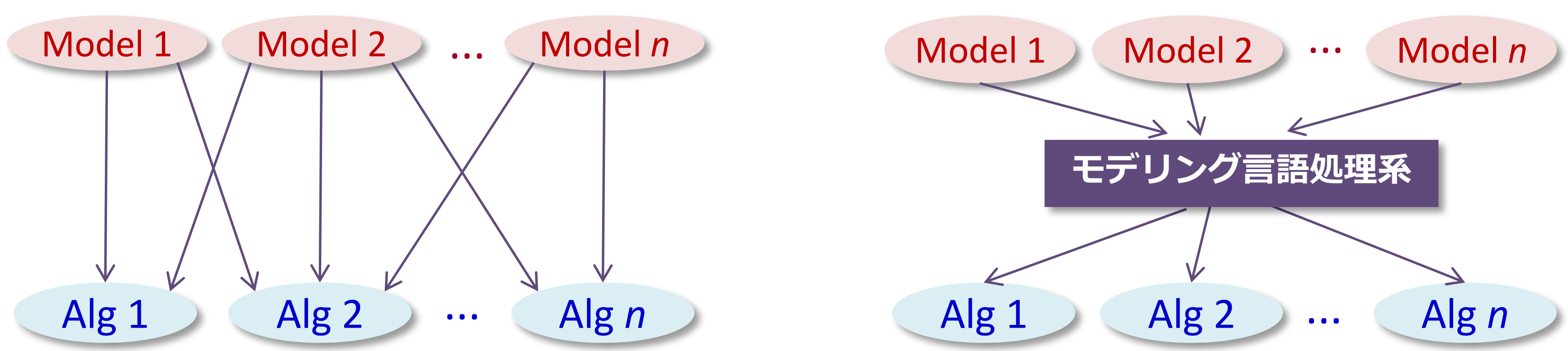
- 帰納論理プログラミング (inductive logic programming) 分野から発展
- 代表的手法：
  - SLPs (Stochastic Logic Programs) [Muggleton, 96]
    - 確率的文脈自由文法 の一般化として考案
  - PHA (Probabilistic Horn Abduction) [Poole, AIJ-93]
    - ホーン節に限定したアブダクションの確率化
    - abducible と呼ばれる，原因事象を表現する命題に確率を付与
  - PRISM [Sato, ICLP-95][Sato et al., IJCAI-97]
    - PHA を厳密に形式化（特に無限世界の扱い） → 分布意味論
    - EM アルゴリズムを考案
  - ProbLog [De Raedt et al., IJCAI-07]
    - BDD (binary decision diagram) に基づく確率計算を提案

発表内容

- ✓ はじめに
- ✓ "論理 + 確率" における諸論点
- 論理に基づく確率モデリングのこれまで
  - ✓ 命題論理に基づく確率推論の高速化
  - ✓ 統計的関係学習 (SRL) / 確率論理学習 (PLL)
  - – 確率モデリング言語処理系
    - PRISM
- 論理に基づく確率モデリングのこれから
- まとめ

確率モデリング言語とその処理系

- 確率モデリング言語：確率モデルを（コンパクトに）記述する表現言語
- 確率モデリング言語処理系：
  - 確率推論および関連するデータ処理を行う処理系
  - ベイズネット処理系も含まれる
- 新たに考案したモデルに対して確率推論アルゴリズムを別途導出する必要はない
- ソフトウェア工学的な視点 (how) が新たに加わる
  - 開発ループ（試行錯誤），プロトタイピング，再利用

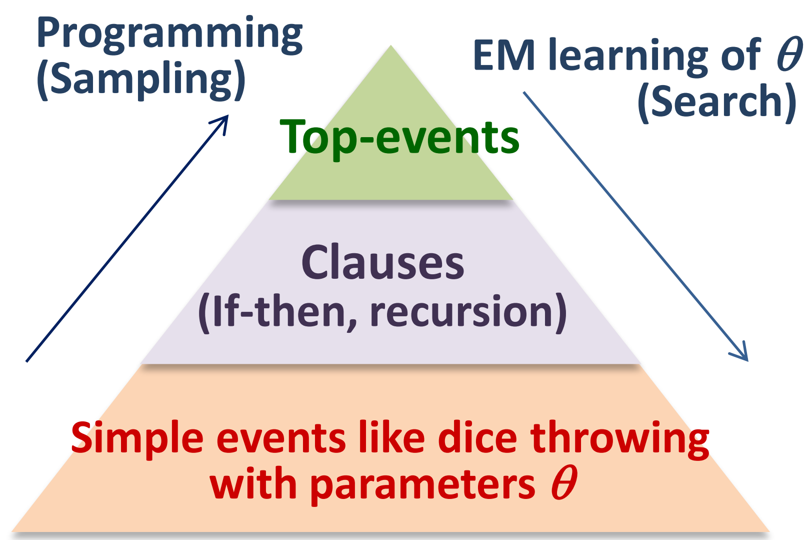


PRISM 処理系：概要

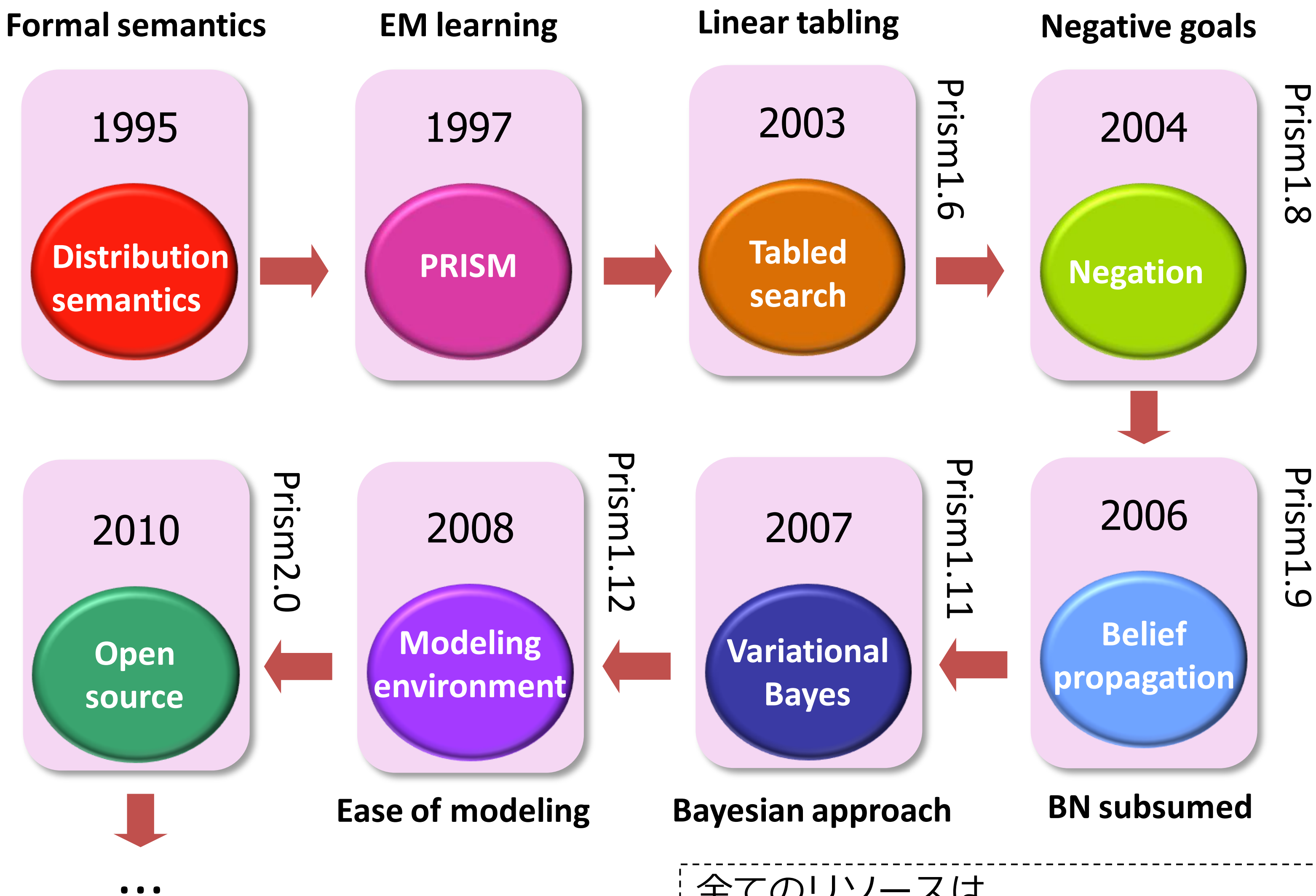
- 分布意味論 [Sato ICLP-95] に基づく
- 離散的な生成モデル（HMM, 確率文脈自由文法, ベイズネット含む）を記述可能
- 多様な推論機構を提供する

- B-Prolog (by N.-F. Zhou at CUNY) 上に実装
  – 多くの推論機構の内部は C 言語
  – テーブル探索機能を多用

- プログラミング処理系である
  – データ：Prolog 項（リスト，木構造）
  – すべての推論結果は Prolog 項として受け取ることができる



PRISM 処理系：開発の歴史



全てのリソースは
http://sato-www.cs.titech.ac.jp/prism/

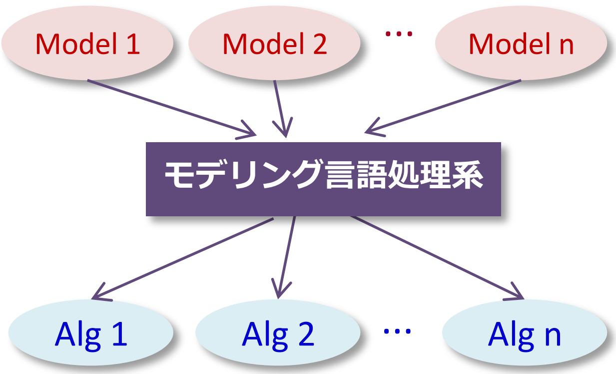
PRISM 処理系：なぜ論理型言語なのか?

- 高級プログラミング言語としての基本要素を備える
  - 単一化（パターンマッチング）機構を提供
  - リスト操作が容易
  - メモリ管理が不要 (GC)

- 確率モデルの宣言的記述が可能

- 深さ優先探索に基づく組み込みの非決定性計算機構
  - 場合の数え上げ：確率計算の基本
  - 一部の処理系ではテーブル探索を提供 ➔ 動的計画法を実現
  - amb オペレータ in LISP（一般的？）

PRISM 処理系：多様な推論機構



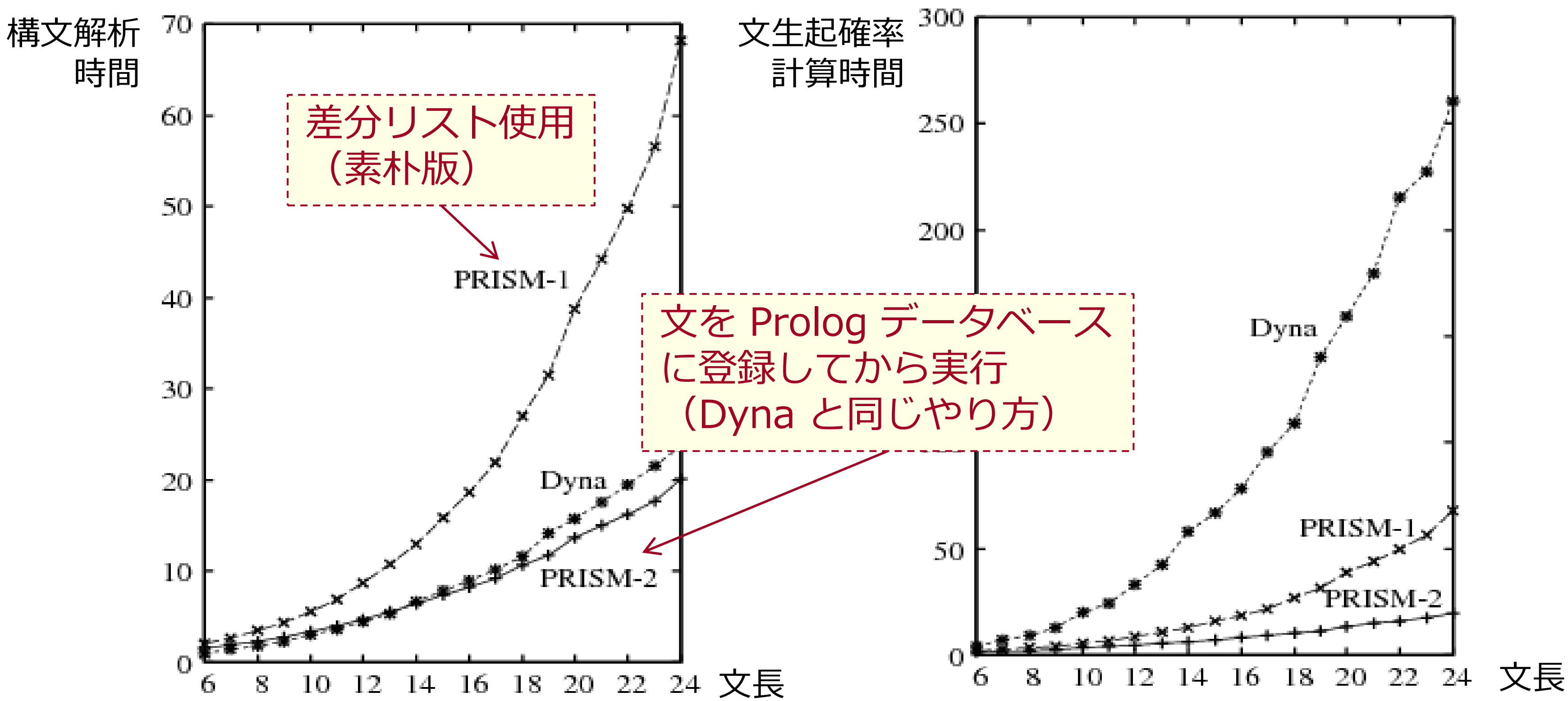
- PRISM が提供する確率推論ルーチン
  1. サンプリング
  2. 式確率計算
  3. (Top-n) Viterbi 計算
  4. Hindsight 計算
  5. EM 学習: 最尤推定, MAP 推定
  6. 並列 EM 学習 [Izumi et al., 06]
  7. 変分ベイズ学習 [Sato et al., 09]
  8. モデルスコア: BIC, Cheeseman-Stutz 近似，変分自由エネルギー
  9. Deterministic Annealing EM (DAEM)
  10. 疑似乱数（一様，正規），統計量（平均，分散, etc.）の計算ユーティリティ

既存の確率推論アルゴリズムと同オーダの計算量

いったんモデルを記述すれば，ユーザはこれらの機能を一通り試すことができる

PRISM 処理系：性能

- Penn Treebank における構文解析 (Viterbi)・文生起確率計算 [亀谷ら, 07]
  - 比較対象: Dyna [Eisner et al., HLT/EMNLP-05]
    - ボトムアップチャートパーザ（優先度付きキュー使用）に基づく動的計画法アルゴリズム記述言語およびその処理系
    - アルゴリズム記述を C++ プログラムに翻訳して実行



PRISM 処理系：応用 (1 of 4)

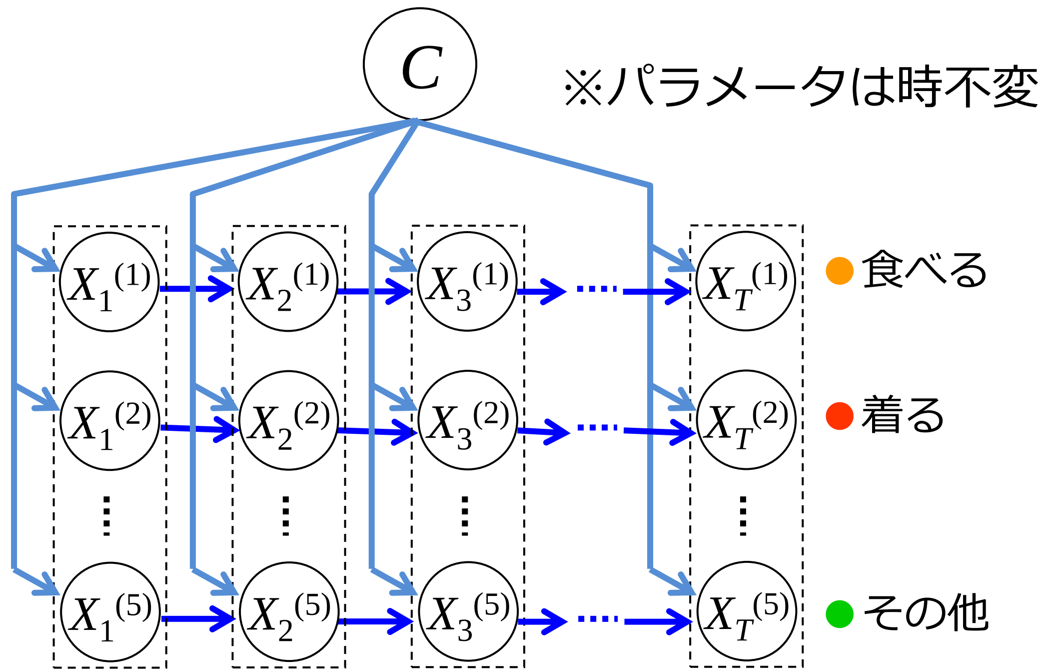
- 音楽の解析・生成モデル [Sneyers et al., PADL-06]
  - 階層的な HMM
- RNA の2次構造解析モデル [Christiansen et al., ICLP-09]
  - 確率文脈自由文法 + コドン構造 + ...
- 確率的プランニング [Jiménez et al., ICAPS-06 WS]

PRISM 処理系：応用 (2 of 4)

- 商店街来訪者の回遊行動解析 [韓, 06][新福, 08]
  - 対象：名古屋市大須地区（門前町 + 電気街 + ファッション街）
  - 回遊行動アンケート（2003年9月の週末, 404人）
    - 回遊行動は通行路の系列（不定長）として記録
    - 前処理：通行路を店舗種別で抽象化
    - 混合マルコフ連鎖 [Cadez et al., KDD-00] に基づくクラスタ分析

食べる ●	1
着る ●	1
買う ●	0
家電 ●	1
その他 ●	0

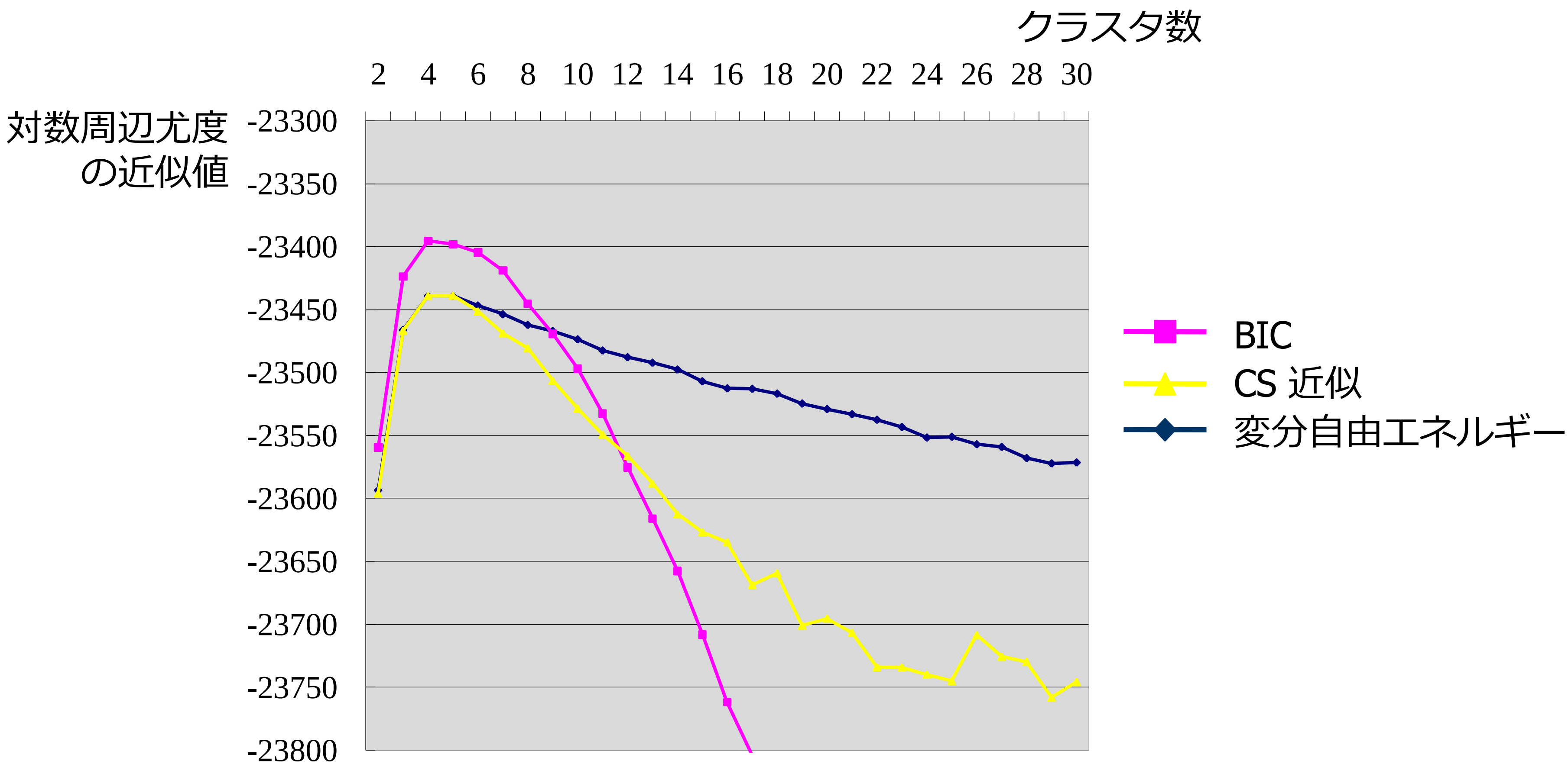
店舗種別の
5次元ベクトル



PRISM ではモデル作成
の試行錯誤にかかる手間
が少ない（ことを実感）

PRISM 処理系：応用 (3 of 4)

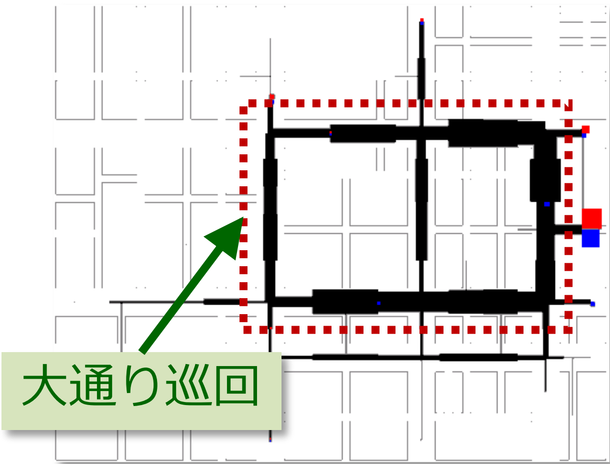
- 商店街来訪者の回遊行動解析 [韓, 06][新福, 08]
  - ベイズスコアに基づくクラスタ数の選択



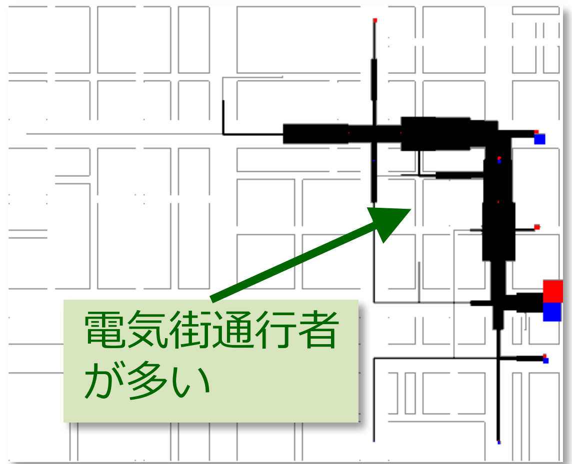
PRISM 処理系：応用 (4 of 4)

黒実線の太さ=通行率
■ 入場通行率
■ 退場通行率

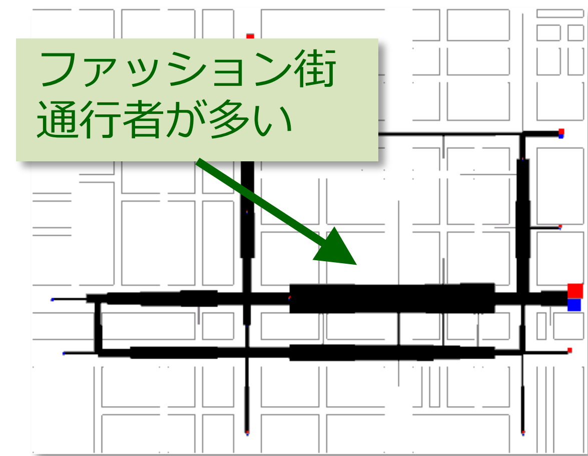
- 商店街来訪者の回遊行動解析 [韓, 06][新福, 08]



クラスタA（52人）

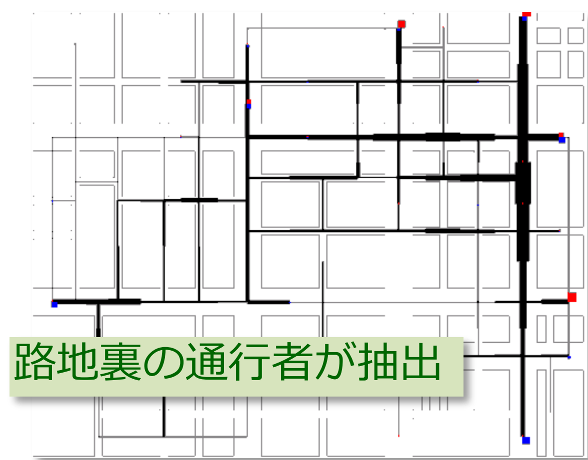


クラスタB（55人）



クラスタC（120人）

クラスタD（140人）



クラスタE（37人）

データ全体（404人）

発表内容

- ✓ はじめに
- ✓ “論理 + 確率” における諸論点
- ✓ 論理に基づく確率モデリングのこれまで
  - ✓ 命題論理に基づく確率推論の高速化
  - ✓ 統計的関係学習 (SRL) / 確率論理学習 (PLL)
  - ✓ 確率モデリング言語処理系
    - ✓PRISM
- 論理に基づく確率モデリングのこれから
- まとめ

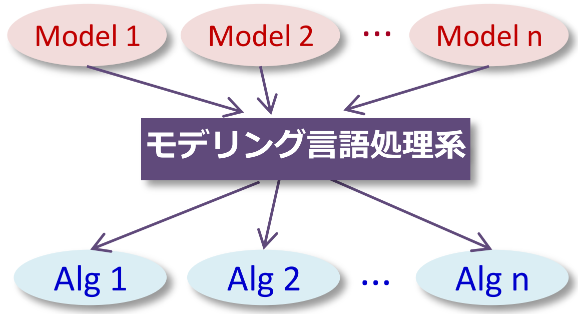
命題論理に基づく確率推論の深化

- 符号化，コンパイル手法の探求
- 他の decision diagram の利用
  - AADD (Affine Algebraic Decision Diagram) の利用
    [Sanner & McAllester, IJCAI-05]

確率モデリング言語処理系の発展

- MLN, ProbLog, PRISM それぞれの発展
- 関数型確率モデリング言語処理系
  - "Probabilistic programming"
  - 代表手法:
    FACTORIE [McCallum et al., NIPS-08 WS], IBAL [Pfeffer, 07], Church [Goodman et al. UAI-08], Figaro [Pfeffer, 09]
  - MCMC に基づく確率推論機構
    - Church では HDP (Hierachical Dirichlet Process) を容易に記述可能

今後の PRISM 処理系



- BDD-EM アルゴリズムの実装
  - PRISMプログラムに排反性を要求しない

- MCMC 推論の導入
  - [Johnson et al., NAACL/HLT-07] の PCFG 用 MCMC を一般化
  - PRISM が提供する変分ベイズ, Viterbi, 式確率計算の組み込みルーチンを内部ルーチンとして利用

- Viterbi training の導入
  - EM アルゴリズムのような繰り返しアルゴリズム
    - aka. Classification EM, Sparse EM, ... ($\supset$ K-means)
  - Viterbi アルゴリズムで得られた最尤説明式上の統計量でパラメータ更新
  - 音声認識では古くから，統計的 NLP でも最近試されている [Spitkovsky et al., CoNLL-10]
  - 使用メモリを大幅に節約できる，PRISMプログラムに排反性を要求しない

制約に基づく確率モデリング (1 of 3)

- CBPMs (constraint-based probabilistic models) [Sato et al., MLJ, to appear]
  - 同時分布 $P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$
  - CBPM: P_c (“X_1=x_1”, “X_2=x_2”, …, “X_n=x_n” | KB)
    - P_c において命題確率変数 {“X_1=x_1”, “X_2=x_2”, …, “X_n=x_n”} は独立
    - “X_i = x_i” は真 ⇔ $X_i = x_i$
    - KB: {“X_1=x_1”, …, “X_n=x_n”}に対する論理的制約
    - $P_c(\cdot \mid KB)$ が所望の分布 $P(\cdot)$ となるよう KB でうまく制約づける
  - CBPM によりマルコフ確率場，ベイズネット，確率文脈自由文法を表現可能
  - 循環する依存関係もモデル化可能
- EMC (EM with constraints) アルゴリズム
  - FAM (Failure-Adjusted Maximization) アルゴリズム [Cussens, MLJ-01] を一般化
  - BDD 上で動作する

制約に基づく確率モデリング (2 of 3)

- CBPM で $P(X = a, Y = b)$ を表現する例
  - KB を定義

$$KB = XOR(X) \wedge XOR(Y) \wedge EQU$$

$$XOR(Z) = \left(\bigvee_{c} \text{“}Z = c\text{”} \right) \wedge \bigwedge_{c_1 \neq c_2} \neg (\text{“}Z = c_1\text{”} \wedge \text{“}Z = c_2\text{”})$$

$$EQU = \bigwedge_{a,b} (\text{“}X = a\text{”} \wedge \text{“}Y = b\text{”} \Leftrightarrow \theta_{ab})$$

  - $P_c(\cdot)$ の下では命題確率変数は独立

$$P_c(\text{“}X = a\text{”}, \text{“}X = b\text{”}, \theta_{ab}) = P_c(\text{“}X = a\text{”}) P_c(\text{“}Y = b\text{”}) P_c(\theta_{ab})$$

$$P_c(\text{“}X = a\text{”}) = \frac{1}{2} \quad \text{for } \forall a$$

$$P_c(\theta_{ab}) = \frac{P(X = a, X = b)}{1 + P(X = a, X = b)}$$

制約に基づく確率モデリング (3 of 3)

$$
\begin{aligned}
KB \quad &\Leftrightarrow \quad \bigvee_{a_1,b_1} \Big\{ \Big(\text{“}X = a_1\text{”} \wedge \bigwedge_{a_2 \neq a_1} \text{“}X = a_2\text{”} \Big) \\
&\qquad \wedge \Big(\text{“}Y = b_1\text{”} \wedge \bigwedge_{b_2 \neq b_1} \text{“}Y = b_2\text{”} \Big) \wedge \Big(\theta_{a_1 b_1} \wedge \bigwedge_{(a_2,b_2) \neq (a_1,b_1)} \neg \theta_{a_2 b_2} \Big) \Big\} \\
P_c(KB) \quad &= \quad K \sum_{a,b} \left(\frac{P_c(\text{“}X = a\text{”})}{P_c(\neg \text{“}X = a\text{”})} \right) \left(\frac{P_c(\text{“}Y = b\text{”})}{P_c(\neg \text{“}Y = b\text{”})} \right) \left(\frac{P_c(\theta_{ab})}{P_c(\neg \theta_{ab})} \right)
\end{aligned}
$$

このとき，$P_c(\text{“}X = a\text{”}) = \ldots = 1/2$ であるので

$$
\begin{aligned}
P_c(\text{“}X = a\text{”}, \text{“}Y = b\text{”} \mid KB) \quad &= \quad \frac{\dfrac{P_c(\theta_{ab})}{P_c(\neg \theta_{ab})}}{\displaystyle\sum_{a',b'} \frac{P_c(\theta_{a'b'})}{P_c(\neg \theta_{a'b'})}} \\
&= \quad \frac{P(X = a, Y = b)}{\sum_{a',b'} P(X = a', Y = b')} \\
&= \quad P(X = a, Y = b)
\end{aligned}
$$

発表内容

- ✓ はじめに
- ✓ "論理 + 確率" における諸論点
- ✓ 論理に基づく確率モデリングのこれまで
  - ✓ 命題論理に基づく確率推論の高速化
  - ✓ 統計的関係学習 (SRL) / 確率論理学習 (PLL)
  - ✓ 確率モデリング言語処理系
    - ✓ PRISM
- ✓ 論理に基づく確率モデリングのこれから
- まとめ

まとめ

- 論理に基づく確率モデリング手法を概観した
  - 命題論理に基づくモデリング
    - ベイジアンネットのコンパイル技法, BDD-EM
  - 述語論理に基づくモデリング
    - 統計的関係学習 (SRL) / 確率論理学習 (PLL) – Markov Logic Network 等
  - 確率モデリング処理系 (PRISM)
  - 制約に基づく確率モデル

参考文献

- [Boutilier et al. 96] C. Boutilier, N. Friedman, M. Goldszmidt and D. Koller (1996). Context-specific independence in Bayesian networks, Proc. of UAI-96, pp.115-123.
- [Breese, 92] J. Breese (1992). Construction of belief and decision networks. Computational Intelligence, Vol.8, 425-456.
- [Buntine 94] W. L. Buntine (1994). Operations for learning with graphical models. J. of Artificial Intelligence Research, Vol.2, pp.159-225.
- [Cadez et al., 00] I. V. Cadez, S. Gaffney and P. Smyth (2000). A general probabilistic framework for clustering individuals and objects. Proc. of KDD-00, pp.140-149.
- [Chavira & Darwiche, 05] M. Chavira and A. Darwiche (2005). Compiling {Bayesian} networks with local structure. Proc. of IJCAI-05, pp.1306-1312.
- [Christiansen et al., 09] H. Christiansen and O. Torp Lassen (2009). Preprocessing for optimization of probabilistic-logic models for sequence analysis. Proc. of the 25th Int'l. Conf. on Logic Programming (ICLP-09), pp.70-83.
- [Cussens, 01] J. Cussens (2001). Parameter estimation in stochastic logic programs. Machine Learning, Vol.44, pp.245-271.
- [De Raedt et al., 07] L. De Raedt, A. Kimmig and H. Toivonen (2007). ProbLog: A probabilistic Prolog and its application in link discovery. Proc. of IJCAI-07, pp.2462-2467.
- [Domingos & Lowd, 09] P. Domingos and D. Lowd (2009). Markov Logic: An Interface Layer for Artificial Intelligence. Morgan and Claypool Publishers.
- [Eisner et al., 05] J. Eisner, E. Goldlust and N. A. Smith (2005). Compiling comp ling: Weighted dynamic programming and the Dyna language. Proc. of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP-05), pp.281-290.

参考文献

- [Friedman et al. 99] N. Friedman, L. Getoor, D. Koller and A. Pfeffer (1999). Learning probabilistic relational models. Proc. of IJCAI-99, pp.1300-1309.
- [Gilks 94] W. R. Gilks, A. Thomas and D. J. Spiegelhalter (1994). A language and program for complex Bayesian modeling. The Statistician, Vol.43, pp.169-117.
- [Goldman & Charniak, 93] R. P. Goldman and E. Charniak (1993). Language for construction of belief networks. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.15, pp.196-208.
- [Goodman et al., 08] N. Goodman, V. Mansinghka, D. Roy, K. Bonawitz and D. Tarlow (2008). Church: a language for generative models. Proc. of UAI-08, pp.220-229.
- [Heckerman et al, 04] D. Heckerman and C. Meek and D. Koller (2004). Probabilistic Models for Relational Data. Microsoft Research Technical Report, MSR-TR-2004-30.
- [Ishihata et al., 08] M. Ishihata, Y. Kameya, T. Sato and S. Minato (2008). Propositionalizing the EM algorithm by BDDs. Technical Report TR08-0004, Dept. of Computer Science, Tokyo Institute of Technology.
- [石畠ら, 09a] 石畠正和, 亀谷由隆, 佐藤泰介, 湊真一 (2009). BDD上の命題化計算に基づくEMアルゴリズム. 人工知能学会論文誌, Vol.25, No.3, pp.475-484.
- [石畠ら, 09b] 石畠正和, 亀谷由隆, 佐藤泰介, 湊真一 (2009). 否定枝を含む shared BDD 上で動作する EM アルゴリズム. 人工知能学会第9回データマイニングと統計数理研究会 (SIG-DMSM) 予稿集, SIG-DMSM-A803, pp.11-19.
- [Ishihata et al., ACML-10] M. Ishihata, Y. Kameya, T. Sato and S. Minato (2010). An EM algorithm on BDDs with order encoding for logic-based probabilistic models. Proc. of ACML-2010, pp.161--176.

参考文献

- [Ishihata et al., AMBN-10] M. Ishihata, T. Sato and S. Minato (2010). Parameter learning for Bayesian networks on Shared Binary Decision Diagrams. Proc. of the 1st Int'l Workshop on Advanced Methodologies for Bayesian Networks (AMBN-2010).
- [Izumi et al., 06] Y. Izumi, Y. Kameya and T. Sato (2006). Parallel EM learning for symbolic-statistical models. Proc. of the Int'l Workshop on Data-Mining and Statistical Science (DMSS-2006), pp.133--140.
- [Jaeger, 97] M. Jaeger (1997). Relational Bayesian networks. Proc. of UAI-97, pp.266-273.
- [Jimenez et al., 06] S. Jiménez, F. Fernández and D. Borrajo (2006). Inducing non-deterministic actions behavior to plan robustly in probabilistic domains. Proc. of the ICAPS-06 Workshop on Planning under Uncertainty and Execution Control for Autonomous Systems.
- [Johnson et al., 07] M. Johnson, T. Griffiths and S. Goldwater (2007). Bayesian inference for PCFGs via Markov chain Monte Carlo. Proc. of NAACL/HLT-07, pp.139-146.
- [韓, 06] 韓方傑 (2006). 回遊行動データに基づく商店街来訪者のクラスタリング. 学士論文，東京工業大学.
- [Kersting & De Raedt, 01] K. Kersting and L. De Raedt (2001). Bayesian Logic Programs. Technical Report 151, University of Freiburg, 2001.
- [Koller & Pfeffer, 97] D. Koller and A. Pfeffer (1997). Learning probabilities for noisy first-order rules. Proc. of IJCAI-97, pp.1316-1321.
- [Kramer et al. 02] S. Kramer, N. Lavrač and P. Flach (2002). Propositionalization approaches to relational data mining. In Relational Data Mining, pp.262-291, Springer, 2002.

参考文献

- [McCallum et al., 08] A. McCallum, K. Rohanemanesh, M. Wick, K. Schultz, S. Singh (2008). FACTORIE: Efficient probabilistic programming for relational factor graphs via imperative declarations of structure, inference and learning. Proc. of the NIPS-08 Workshop on Probabilistic Programming.
- [Milch, 06] B. Milch (2006). Probabilistic Models with Unknown Objects. Ph.D. Thesis, Computer Science Division, University of California, Berkeley.
- [Minato et al., 07] S. Minato, K. Satoh and T. Sato (2007). Compiling Bayesian networks by symbolic probability calculation based on zero-suppressed BDDs. Proc. of IJCAI-07, pp.2550-2555.
- [Muggleton, 96] S. H. Muggleton (1996). Stochastic logic programs. In Advances in Inductive Logic Programming, pp.254-264. IOS Press.
- [Ngo & Haddawy, 97] L. Ngo and P. Haddawy (1997). Answering queries from context-senstive probabilistic knowledge bases. Theoretical Computer Science, Vol.171, pp.147-177.
- [Pfeffer, 07] A. Pfeffer (2007). The Design and Implementation of IBAL: A General-Purpose Probabilistic Language. In Statistical Relational Learning, The MIT Press.
- [Pfeffer, 09] A. Pfeffer (2009). Figaro: An Object-Oriented Probabilistic Programming Language. Charles River Analytics Technical Report.
- [Poole, 93] D. Poole (1993). Probabilistic Horn abduction and Bayesian networks. Artificial Intelligence, Vol.64, pp.81-129.
- [Pynadath & Wellman, 98] D. V. Pynadath and M. P. Wellman (1998). Generalized queries on probabilistic context-free grammars. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.20, pp.65-77.

参考文献

- [Richardson & Domingos, 06] M. Richardson and P. Domingos (2006). Markov logic networks. Machine Learning, Vol.62, pp.107-136.
- [Riedel, 08] S. Riedel (2008). Improving the accuracy and efficency of MAP inferece for Markov logic. Proc. of UAI-08, pp.468-475.
- [Sang et al. 05] T. Sang, P. Beame and H. A. Kautz (2005). Performing Bayesian Inference by Weighted Model Counting. Proc. of AAAI 2005, pp.475-482.
- [Sanner & McAllester, 05] S. Sanner and D. McAllester (2005). Affine Algebraic Decision Diagrams (AADDs) and their application to structured probabilistic inference. Prof. of IJCAI-05.
- [Sato et al., to appear] T. Sato, M. Ishihata and K. Inoue. Constraint-based probabilistic modeling for statistical abduction. Machine Learning, to appear.
- [Sato, 95] T. Sato (1995). A statistical learning method for logic programs with distribution semantics. Proc. of the 12th Int'l Conf. on Logic Programming (ICLP95), pp.715-729.
- [Sato et al., 97] T. Sato and Y. Kameya (1997). PRISM: A symbolic-statistical modeling language. Proc. of IJCAI97, pp.1330-1335.
- [Sato et al., 01] T. Sato and Y. Kameya (2001). Parameter learning of logic programs for symbolic-statistical modeling. J. of Artificial Intelligence Research, Vol.15, pp.391-454.
- [Sato et al., 08] T. Sato and Y. Kameya (2008). New advances in logic-based probabilistic modeling by PRISM. In Probabilistic Inductive Logic Programming, LNCS 4911, Springer, pp.118-155.
- [Sato et al., 09] T. Sato, Y. Kameya and K. Kurihara (2009). Variational Bayes via propositionalized probability computation in PRISM. Annals of Mathematics and Artificial Intelligence, Vol.54, No.1-3, pp.135-158.

参考文献

- [新福, 08] 新福 崇史 (2008). 変分ベイズ法を用いたクラスタリング手法の研究, 修士論文，東京工業大学.
- [Sneyers et al., 06] J. Sneyers, J. Vennekens and D. De Schreye (2006). Probabilistic-logical modeling of music. Proc. of the 8th Intl. Sympo. on Practical Aspects of Declaration Languages (PADL-2006), pp.60-72.
- [Spitkovsky et al., 10] V. I. Spitkovsky, H. Alshawi, D. Jurafsky and C. D. Manning (2010). Viterbi Training Improves Unsupervised Dependency Parsing. In Proc. of the 14th Conf. on Computational Natural Language Learning (CoNLL-2010).
- [Taskar et al., 02] B. Taskar, P. Abbeel and D. Koller (2002). Discriminative probabilistic models for relational data. Proc. of UAI-02, pp.485-492.
- [Wrobel, 02] S. Wrobel (2002). Inductive logic programming for knowledge discovery in databases. In Relational Data Mining, pp.74-104, Springer, 2002.