Parameter learning for Bayesian networks on Shared Binary Decision Diagrams

東工大 石畠正和

自己紹介

2

- 名前　　石畠正和
- 所属　　東京工業大学 佐藤泰介研究室 D1
- 興味
  - 命題論理を用いた確率モデリング
  - BDDを用いた確率学習

2011/5/2　　ERATO合宿

論理と確率

3

Logic
Deterministic rules
Expressive power

Probability
Uncertainty
Modeling power

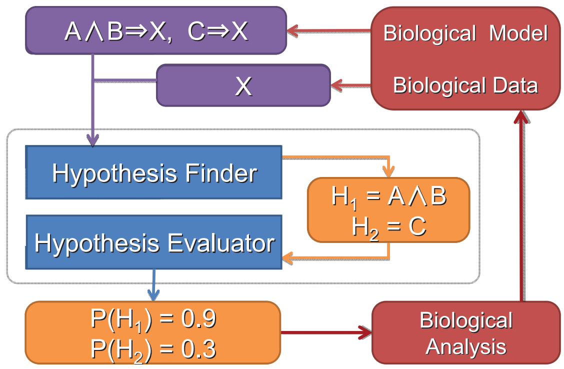
LBPMs
Logic-based Probabilistic Models

- *PHA* [Poole 93]
- *PRISM* [Sato+ 97]
- *ProbLog* [Deraedt+ 07]
- *MLNs* [Richardson+ 06]　　etc...

2011/5/2

Ex) Hypothesis generator using LBPMs

4



Inoue et al., "Evaluating abductive hypotheses using an EM algorithm on BDDs. ", IJCAI09

2011/5/2

佐藤研的アプローチ

5

- 命題変数集合上に確率分布を導入
  - 命題変数 → 確率変数
  - 命題論理式 → 確率事象
- うまく命題変数と論理式を定めれば任意の離散確率変数の同時分布を表現可能
- 今日は論理を使ってBayesian networkを表現

2011/5/2　　ERATO合宿

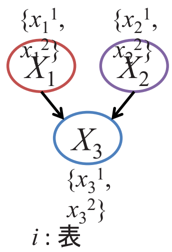
Bayesian networks

6

pa_1^j	$P(x_1^1 \mid$	$P(x_1^2 \mid$
{}	θ_{111}	θ_{112}

pa_2^j	$P(x_2^1 \mid$	$P(x_2^2 \mid$
{}	θ_{211}	θ_{212}

pa_3^j	$P(x_3^1 \mid$	$P(x_3^2 \mid$
$(x_1^1,$	θ_{311}	θ_{312}
$(x_1^1,$	θ_{321}	θ_{322}
$(x_1^2,$	θ_{331}	θ_{332}
(x_1^2, x_2^2)	θ_{341}	θ_{342}

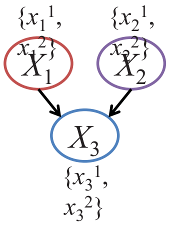


i : 表
j : 行
k : 列

2011/5/2　　ERATO合宿

Bayesian networks

X_1	X_2	X_3	$P(.)$
x_1^1	x_2^1	x_3^1	$\theta_{111}\ \theta_{211}\ \theta_{311}$
x_1^1	x_2^1	x_3^2	$\theta_{111}\ \theta_{211}\ \theta_{312}$
x_1^1	x_2^2	x_3^1	$\theta_{111}\ \theta_{212}\ \theta_{321}$
x_1^1	x_2^2	x_3^2	$\theta_{111}\ \theta_{212}\ \theta_{322}$
x_1^2	x_2^1	x_3^1	$\theta_{112}\ \theta_{211}\ \theta_{331}$…
.	.	.	



$$P(X_1=x_1^1) = \theta_{111}\ \theta_{211}\ \theta_{311} + \theta_{111}\ \theta_{211}\ \theta_{312} + \theta_{111}\ \theta_{212}\ \theta_{321} + \theta_{111}$$

論理とBN

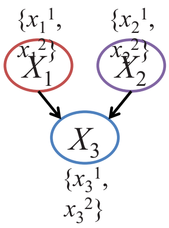
- 命題論理を用いて任意の離散BNを表現可能
- 表現可能なだけですか？
  - **Determinism** (0/1確率)
  - **Context specific independence** (値単位の独立性)

  を効率よく扱える！！

Determinism

CPT for X_3

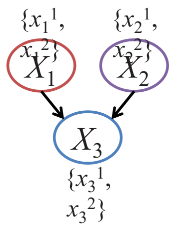
pa_3^j	$P(x_3^1 \mid$	$P(x_3^2 \mid$
$(x_1^1,$	0.2	0.8
$(x_1^1,$	0.2	0.8
$(x_1^2,$	0	1
(x_1^2, x_2^2)	0.6	0.4



Context specific independence

CPT for X_3

pa_3^j	$P(x_3^1 \mid$	$P(x_3^2 \mid$
$(x_1^1,$	0.2	0.8
$(x_1^1,$	0.2	0.8
$(x_1^2,$	0	1
(x_1^2, x_2^2)	0.6	0.4



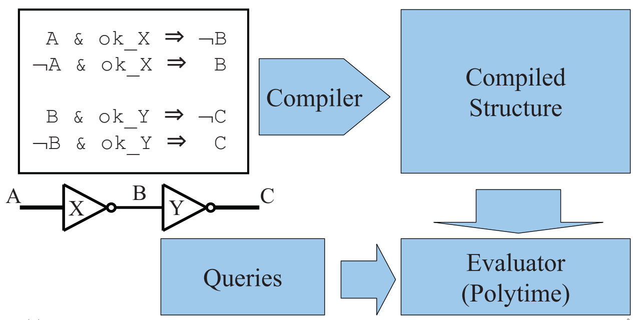
$$X_2 \perp\!\!\!\perp X_3 \mid X_1 = x_1^1$$

２つのアプローチ

1. Darwiche らのアプローチ
   - 昨日の浜口先生、川原さんのアプローチもこちらに近い
2. 佐藤研究室のアプローチ

Darwiche らのアプローチ

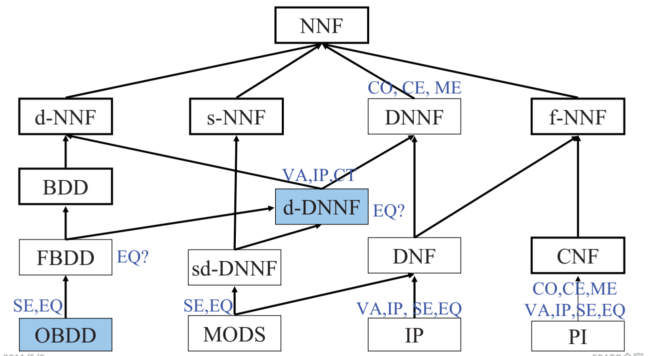
Knowledge Compilation



DAGs for knowledge representation

13

NNF Subsets

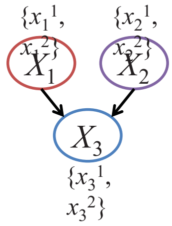


2011/5/2 ERATO合宿

Darwiche 的な encoding

14

X_1	X_2	X_3	$P(.)$
x_1^1	x_2^1	x_3^1	$\theta_{111}\ \theta_{211}\ \theta_{311}$
x_1^1	x_2^1	x_3^2	$\theta_{111}\ \theta_{211}\ \theta_{312}$
x_1^1	x_2^2	x_3^1	$\theta_{111}\ \theta_{212}\ \theta_{321}$
x_1^1	x_2^2	x_3^2	$\theta_{111}\ \theta_{212}\ \theta_{322}$
x_1^2	x_2^1	x_3^1	$\theta_{112}\ \theta_{211}\ \theta_{331}$…
.	.	.	



$$P(X_1=x_1^1) = \theta_{111}\ \theta_{211}\ \theta_{311} + \theta_{111}\ \theta_{211}\ \theta_{312} + \theta_{111}\ \theta_{212}\ \theta_{321} + \theta_{111}$$

2011/5/2 ERATO合宿

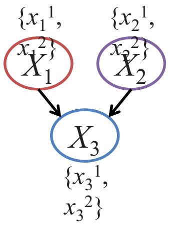
Darwiche 的な encoding

15

Indicator variables

X_1	X_2	X_3	
x_1^1	x_2^1	x_3^1	$\lambda_{11}\ \lambda_{21}\ \lambda_{31}\ \theta_{111}\ \theta_{211}\ \theta_{311}$
x_1^1	x_2^1	x_3^2	$\lambda_{11}\ \lambda_{21}\ \lambda_{32}\ \theta_{111}\ \theta_{211}\ \theta_{312}$
x_1^1	x_2^2	x_3^1	$\lambda_{11}\ \lambda_{22}\ \lambda_{31}\ \theta_{111}\ \theta_{212}\ \theta_{321}$
x_1^1	x_2^2	x_3^2	$\lambda_{11}\ \lambda_{22}\ \lambda_{32}\ \theta_{111}\ \theta_{212}\ \theta_{322}$
x_1^2	x_2^1	x_3^1	… θ_{112}
			…

Multi liner function



$$F \equiv \lambda_{11}\ \lambda_{21}\ \lambda_{31}\ \theta_{111}\ \theta_{211}\ \theta_{311} + \lambda_{11}\ \lambda_{21}\ \lambda_{32}\ \theta_{111}\ \theta_{211}\ \theta_{312} + \lambda_{11}\ \lambda_{22}\ \lambda_{31}\ \theta_{111}\ \theta_{212}\ \theta_{321} + \lambda_{11}\ \lambda_{22}\ \lambda_{32}\ \theta_{111}\ \theta_{212}\ \theta_{322}$$

2011/5/2 ERATO合宿

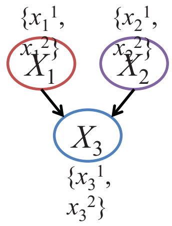
Determinism & CSI

16

CPT for X_3

pa_3^j	$P(x_3^1\|$	$P(x_3^2\|$
$(x_1^1,$	$\theta_{311} =$	$\theta_{312} =$
$(x_1^1,$	$\theta_{321} =$	$\theta_{322} =$
$(x_1^2,$	$\theta_{331} = 0$	$\theta_{332} = 1$
$(x_1^2,$	$\theta_{341} =$	$\theta_{342} =$

$x_2^2)$ 0.6 0.4



$$F \equiv \lambda_{11}\ \lambda_{21}\ \lambda_{31}\ \theta_{111}\ \theta_{211}\ \theta_{311} + \lambda_{11}\ \lambda_{21}\ \lambda_{32}\ \theta_{111}\ \theta_{211}\ \theta_{312} + \lambda_{11}\ \lambda_{22}\ \lambda_{31}\ \theta_{111}\ \theta_{212}\ \theta_{321} + \lambda_{11}\ \lambda_{22}\ \lambda_{32}\ \theta_{111}\ \theta_{212}\ \theta_{322} + \lambda_{12}\ \lambda_{21}\ \lambda_{31}\ \theta_{112}\ \theta_{211}\ \theta_{331} + \lambda_{12}\ \lambda_{21}\ \lambda_{32}$$

2011/5/2 ERATO合宿

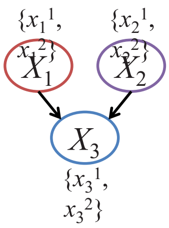
Determinism & CSI

17

CPT for X_3

pa_3^j	$P(x_3^1\|$	$P(x_3^2\|$
$(x_1^1,$	$\theta_{311} =$	$\theta_{312} =$
$(x_1^1,$	$\theta_{321} =$	$\theta_{322} =$
$(x_1^2,$	$\theta_{331} = 0$	$\theta_{332} = 1$
$(x_1^2,$	$\theta_{341} =$	$\theta_{342} =$

$x_2^2)$ 0.6 0.4

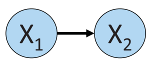


$$F \equiv \lambda_{11}\ \lambda_{21}\ \lambda_{31}\ \theta_{111}\ \theta_{211}\ \theta_{311} + \lambda_{11}\ \lambda_{21}\ \lambda_{32}\ \theta_{111}\ \theta_{211}\ \theta_{312} + \lambda_{11}\ \lambda_{22}\ \lambda_{31}\ \theta_{111}\ \theta_{212}\ \theta_{311} + \lambda_{11}\ \lambda_{22}\ \lambda_{32}\ \theta_{111}\ \theta_{212}\ \theta_{312} + \lambda_{12}\ \lambda_{21}\ \lambda_{32}\ \theta_{112}\ \theta_{211}$$

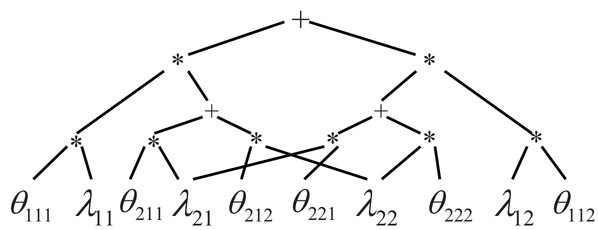
2011/5/2 ERATO合宿

18

$$F=\lambda_{11}\lambda_{21}\theta_{111}\theta_{211}+\lambda_{11}\lambda_{21}\theta_{11}\theta_{212} +\lambda_{12}\lambda_{11}\theta_{112}\theta_{221}+\lambda_{12}\lambda_{22}\theta_{12}\theta_{222}$$



Factoring

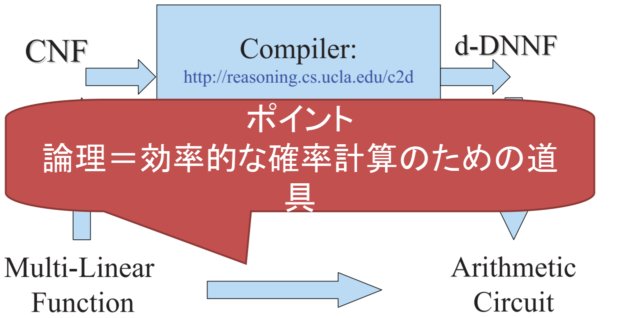


2011/5/2 ERATO合宿

Darwiche 的な encoding

19

Reduction to Logic



佐藤研究室のアプローチ

20

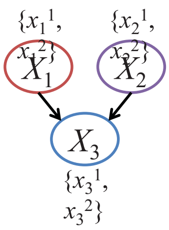
- 命題変数集合上の確率分布を導入
  - Rule based generative models [Sato+ 95]
  - Constraint based probabilistic models [Sato+ 10]

- 命題変数 → 確率変数
- 命題論理式 → 確率事象
- 論理推論 + 確率推論

佐藤研的な encoding

21

X_1	X_2	X_3	$P(.)$
x_1^1	x_2^1	x_3^1	$\theta_{111}\ \theta_{211}\ \theta_{311}$
x_1^1	x_2^1	x_3^2	$\theta_{111}\ \theta_{211}\ \theta_{312}$
x_1^1	x_2^2	x_3^1	$\theta_{111}\ \theta_{212}\ \theta_{321}$
x_1^1	x_2^2	x_3^2	$\theta_{111}\ \theta_{212}\ \theta_{322}$
x_1^2	x_2^1	x_3^1	$\theta_{112}\ \theta_{211}\ \theta_{331}$…
.	.	.	

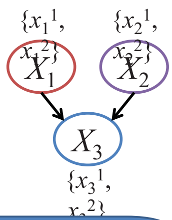


$$P(X_1=x_1^1) = \theta_{111}\ \theta_{211}\ \theta_{311} + \theta_{111}\ \theta_{211}\ \theta_{312} + \theta_{111}\ \theta_{212}\ \theta_{321} + \theta_{111}$$

佐藤研的な encoding

22

X_1	X_2	X_3	説明
x_1^1	x_2^1	x_3^1	$A_{111}A_{211}A_{311}$
x_1^1	x_2^1	x_3^2	$A_{111}A_{211}A_{312}$
x_1^1	x_2^2	x_3^1	$A_{111}A_{212}A_{321}$
x_1^1	x_2^2	x_3^2	$A_{111}A_{212}A_{322}$
x_1^2	x_2^1	x_3^1	$A_{112}A_{211}A_{331}$
.	.	.	…

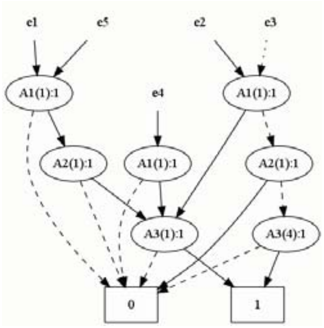


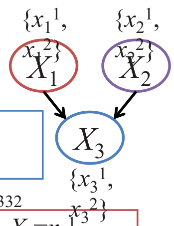
確率的命題変数

$$“X_1=x_1^1” = A_{111}A_{211}A_{311} \lor A_{111}A_{211}A_{312} \lor A_{111}A_{212}A_{321} \lor A_{111}A_{212}A_{322}$$

佐藤研的な encoding

23





$X_2 \perp\!\!\!\perp X_3 \mid X_1=x_1^1$

$\theta_{331} = 0,\ \theta_{332}$

$e_1 \equiv \{X_1=x_1^1, X_2=x_2^1, X_3=x_3^1\}$

$e_2 \equiv \{X_3=x_3^1\}$

$e_3 \equiv \{X_3=x_3^2\}$

$e_4 \equiv \{X_1=x_1^1, X_3=x_3^1\}$

$e_5 \equiv \{X_2=x_2^1, X_3=x_3^1\}$

Darwiche との違い

24

- Darwiche
  - 計算式を簡略化するために論理を使う
  - 1つのBNを1つの d-DNNF で表現
  - 確率推論に有利

- 佐藤研
  - 知識や事象の表現に論理を使う
  - 複数の観測事象を1つの SBDD で表現
  - 確率学習・論理推論に有利

実験1 BNのパラメータ学習

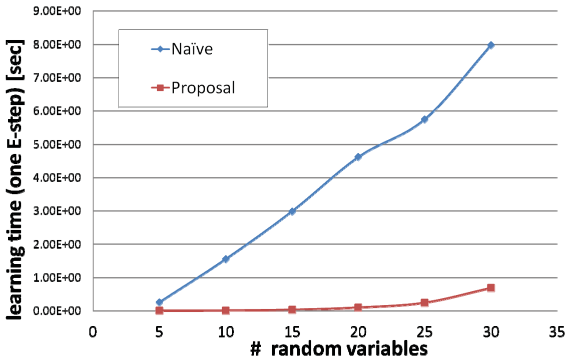
25

- ランダムに生成したBNのパラメータをEMアルゴリズムで学習
- Naïve approach
  - 条件付き期待値を条件付き確率の和として計算
  - 確率計算には Darwiche の方法を使用
- Proposal
  - 観測を表現するSBDD上の動的計画法に基づいて期待値を計算
  - 複数の観測の共通部分を共有化できる

2011/5/2 ERATO合宿

Result 1. No local structures

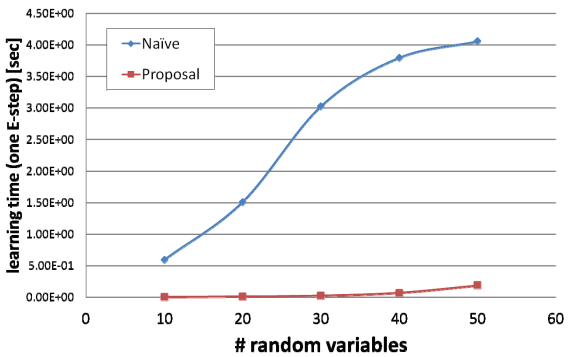
26



2011/5/2 AMBN2010

Result 2. Assume local structures

27



2011/5/2 AMBN2010

実験2 論理回路の故障診断

- 故障素子は確率的に 0/1 に縮退する
- 観測として回路の入出力値が与えられる
- 観測から故障素子の場所を診断したい

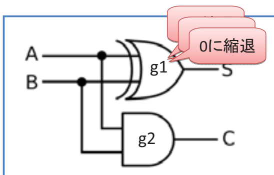


図 半加算器

観測

$O^{(1)}$: ([A,B],[S,C]) = ([0,0],[1,0])

$O^{(2)}$: ([A,B],[S,C]) = ([0,1],[0,1])

$O^{(3)}$: ([A,B],[S,C]) = ([1,1],[0,0])

診断

素子 g1 が故障している

実験2 論理回路の故障診断

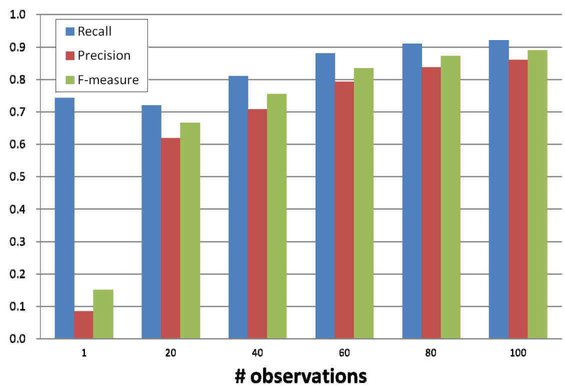
- 観測命題: $obs([A,B], [S,C])$
- 基本命題: $state(G, St)$
  - state(G, ok) -> G は正常
  - state(G, $stk0$) -> G の出力は 0 に縮退
  - state(G, $stk1$) -> G の出力は 1 に縮退
- 例) obs([1,1],[0,0]) ⇔ $h_1 \vee h_2$

  $h_1 = \mathrm{state(g1, stk0)} \wedge \mathrm{state(g2, stk0)}$

  $h_2 = \mathrm{state(g1, stk0)} \wedge \mathrm{state(g2, ok)}$

論理推論により可能な仮説を導出

実験結果



まとめ

31

- 論理と確率を統合するアプローチを紹介
- 任意の離散確率分布は命題変数上の確率分布で表現できる
- 論理推論に確率を導入することで仮説評価を統計的に行うことが可能

2011/5/2 ERATO合宿