記号的統計モデリングの世界を探る

佐藤泰介(東工大)

記号的統計モデリングとは

文やグラフなど記号的な対象の離散的統計モデリングを指す。

- ベイジアンネット（BN）、隠れマルコフモデル（HMM）、確率文脈自由文法（PCFG）などは命題論理レベルの記号的統計モデルであり、バイオインフォマティクスやロボティクス、意志決定など幅広く使われている。
- 形式言語、確率、学習の3要素を持つ。
- 90年代後半から述語論理化の流れが生じ、ヨーロッパ（PLL）やアメリカ（SRL）で発達し、論理と確率の統合へと歩み始めた。

  "Norvig said he was encouraged by recent research in AI, such as work on "probabilistic first-order logic" being at Stanford University and UC Berkeley" (News Blog – CNT, Sept. 2007)

- 本講演で、統合への歩みを紹介する。

論理と確率

論理:事物の必然性

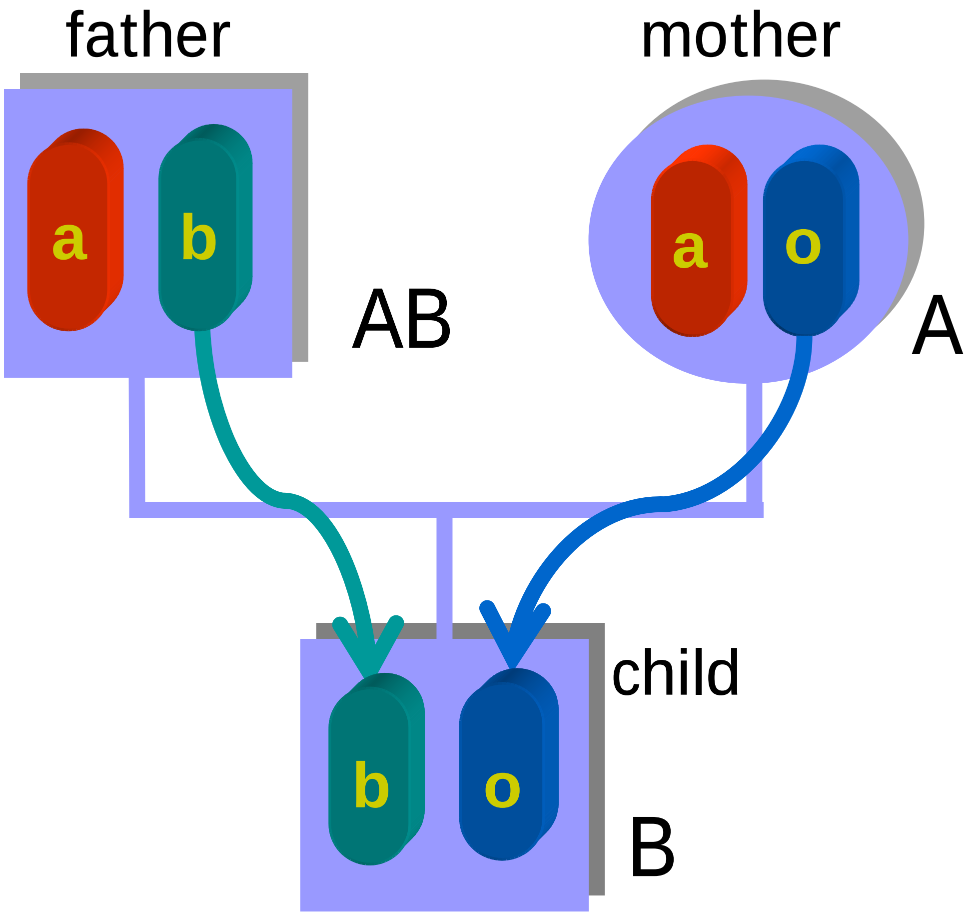
- Default ＝nothing is connected
- 表現＝論理式
- 知識＝関係を表す
- 推論＝演繹(LP),帰納(ILP),発想(ALP)
- 関心事＝健全性、完全性、計算量

確率:事物の偶然性

- Default ＝everything is connected
- 表現＝同時分布
- 知識＝独立性を表す
- 学習＝パラメータ、構造
- 関心事＝統計的推定、予測、計算量

現実:両者の絡み合い
(記号的統計モデリング)

記号的統計モデリングの例：血液型の遺伝



PRISMプログラム

```
btype(X):- pg_table(X,[Gf,Gm]), gtype(Gf,Gm).
pg_table(X,Gtype):-
   ((X=a;X=b), (GT=[X,o] ; GT=[o,X] ; GT=[X,X])
   ; X=o, GT=[o,o]
   ; X=ab, (GT=[a,b] ; GT=[b,a])).
gtype(Gf,Gm):- msw(abo,Gf), msw(abo,Gm).
```

(確率的選択)

P(msw(abo,a)) = $\theta_{(abo,a)}$ = 0.3,…（パラメータ）

➔ P(btype(a)) = 0.4

（パラメータから計算された確率）

初期の取り組み

- Fenstad’67（logic）
  - $c(\alpha \vee \beta) + c(\alpha \,\&\, \beta) = c(\alpha) + c(\beta)$, $c(\sim\alpha) = 1 - c(\alpha)$
  - $c(\alpha) = c(\beta)$ if $\vdash \alpha \Leftrightarrow \beta$, $c(\alpha) = 1$ if $\vdash \alpha$

    ならば $c(\alpha) = \int \mu_M(\alpha)\, d\lambda(M)$ (可能世界上の確率測度による期待値）
- Gaifman’82（logic）
  - $e_1^w,\ldots,e_n^w$ を世界 w における事実の観測列とせよ

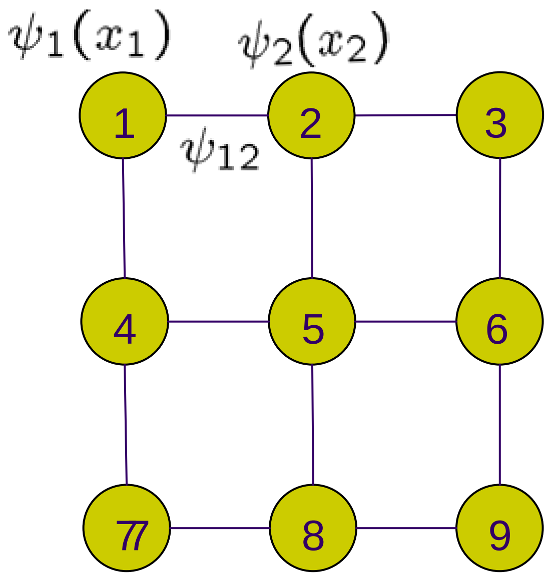
    $\lim_n P(\psi \mid e_1^w,\ldots,e_n^w) \rightarrow \psi^w$ (a.s.)
- Nilsson’86（AI）
  - P(A)+P(A&B) ≦ 0.3, 不等式の推論
- Halpern’90（UAI）
  - Type I（個体上の分布）, Type II（可能世界の分布）
- Breese’92（UAI）
  - KBMC（knowledge based model construction）
  - KB ➔ ベイズネットを生成
- Poole’93（AI）
  - Probabilistic Horn Abduction, ベイズネットも表現できる

90年代以降:BNとILPの接近

- BNの動機:楽してBNを書くため、(論理)変数を使う
  - KBMC(knowledge based model construction)
    - 定節により記述されたKBからBNを合成(Breeze'92)
  - PRM(probabilistic relatioinal model)
    - 関係データベースとBNの融合(Friedman,Getoor'99)
  - BLP(Bayesian logic program)
    - A|B,C でP(A|B,C)を表しBNを定義(Kersting,De Raedt'00)
  - MEBN(multi-entity BN)
    - 論理式と制約で定義された局所BNを集めて無限BNを定義(Laskey'06)
- LPの動機:LPに確率を融合して表現力を上げたい
  - ICL(independent choice logic)
    - BNをLPで表現・計算、宣言的意味論なし(Poole'97)
  - SLP(stochastic logic programming)
    - PCFGとLPのアナロジー、証明木の分布を定義(Muggleton'96,Cussens'01)
  - PRISM(programming in statistical modeling)
    - BN,PCFG,Prolog を包含、可能世界意味論(Sato'95,Kameya'00)

グラフィカルモデル（MRF）の発展

Ising モデル



- ポテンシャル関数f_αの積$\Pi_\alpha f_\alpha(x_\alpha)$の形の分布を表す($\alpha$はグラフのクリーク）
- log-linear モデルとも言う

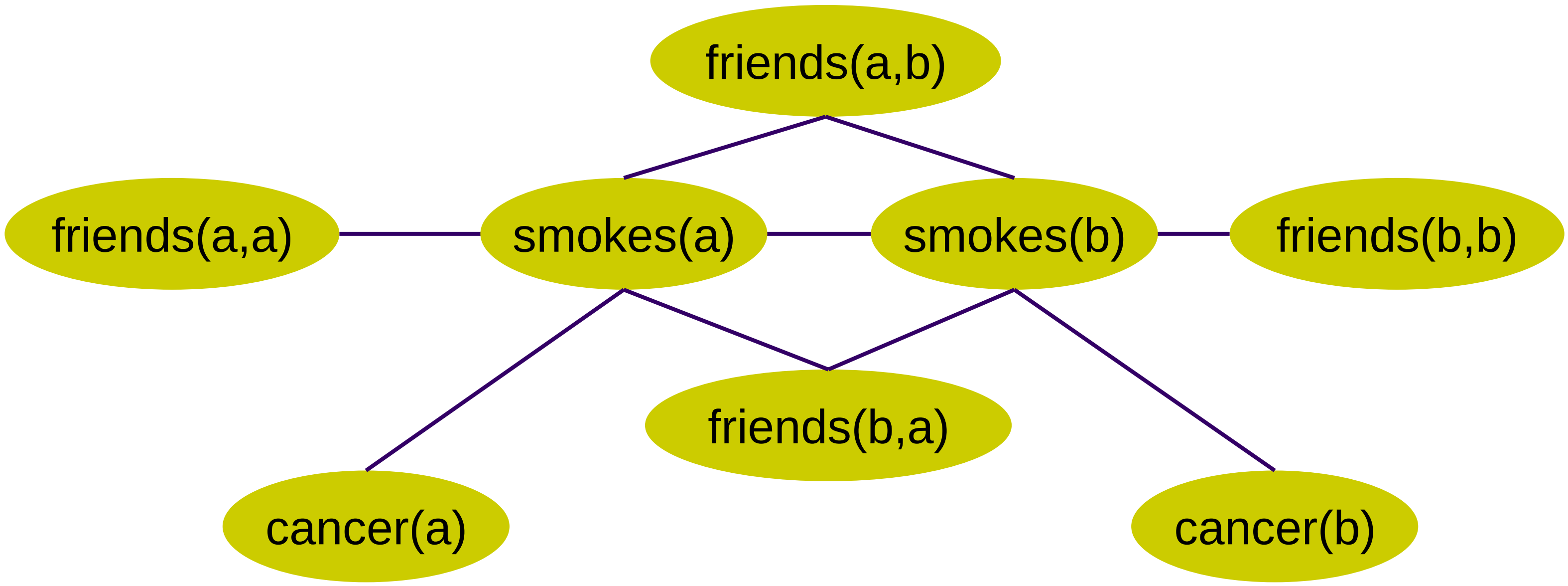
  $Z^{-1} \exp(\Sigma_i w_i F_i(x))$ （F_i は素性）
- 確率計算はMCMCを使う
- 無限グラフ（Gibbs 測度）では相転移が起こる
- F_i を節、x をHerbrand 解釈にしたものがMLN(Markov logic network, Domingos'05)

$$
\begin{aligned}
p(x) &= Z^{-1}\psi_1(x_1)\psi_{12}(x_1,x_2)\psi_2(x_2)\cdots \\
&= Z^{-1} \exp -(\sigma_1 x_1 + \sigma_{12} x_1 x_2 + \sigma_2 x_2 + \ldots)
\end{aligned}
$$

Markov Logic Network

1.5: $\forall x \ \mathrm{smokes}(x) \rightarrow \mathrm{cancer}(x)$

1.1: $\forall x,y \ \mathrm{friends}(x,y) \rightarrow (\mathrm{smokes}(x) \leftrightarrow \mathrm{smokes}(y))$



- 世界は a,bの二人から成る
- グラウンドアトムは0,1の値をとる

グラフィカルモデル（BN）の発展 -- going to NewYork

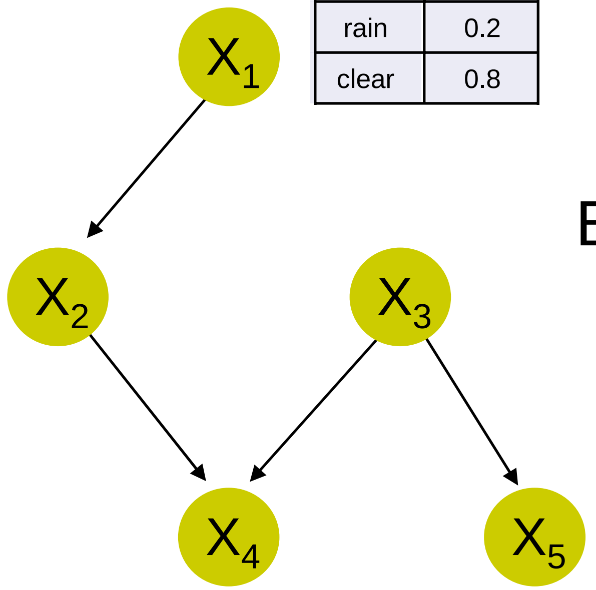
X_1: Tokyo weather
$x_1 \in \{$ rain, clear $\}$
X_2: departure delay
$x_2 \in \{$ yes, no $\}$
X_3: NY weather
$x_3 \in \{$ rain, clear $\}$
X_4: arrival delay
$x_4 \in \{$ yes, no $\}$
X_5: game-cancelled
$x_5 \in \{$ yes, no $\}$

CPT

	$P(x_2 \mid x_1)$	$P(x_2 \mid x_1)$
x_2 / x_1	rain	clear
yes	0.3	0.1
no	0.7	0.9

CPT

x_1	$P(x_1)$
rain	0.2
clear	0.8

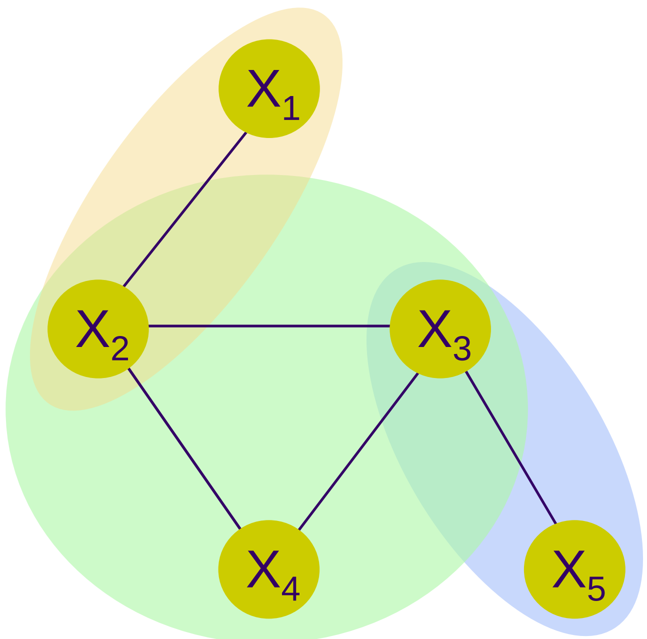


BN_{NY}

$$p(X_1=x_1, X_2=x_2, X_3=x_3, X_4=x_4, X_5=x_5) = p(x_1)p(x_2 \mid x_1)p(x_4 \mid x_2, x_3)p(x_3)p(x_5 \mid x_3)$$

ポテンシャル関数の積へ

BN$_{NY}$のモラルグラフ



$$p(X_1=x_1,X_2=x_2,X_3=x_3,X_4=x_4,X_5=x_5)$$
$$= p(x_1)p(x_2\,|x_1)p(x_4\,|x_2,x_3)p(x_3)p(x_5\,|x_3)$$
$$= \phi_2(x_1,x_2)\,\phi_4(x_4,x_3,x_1)\,\phi_5(x_5,x_3)$$

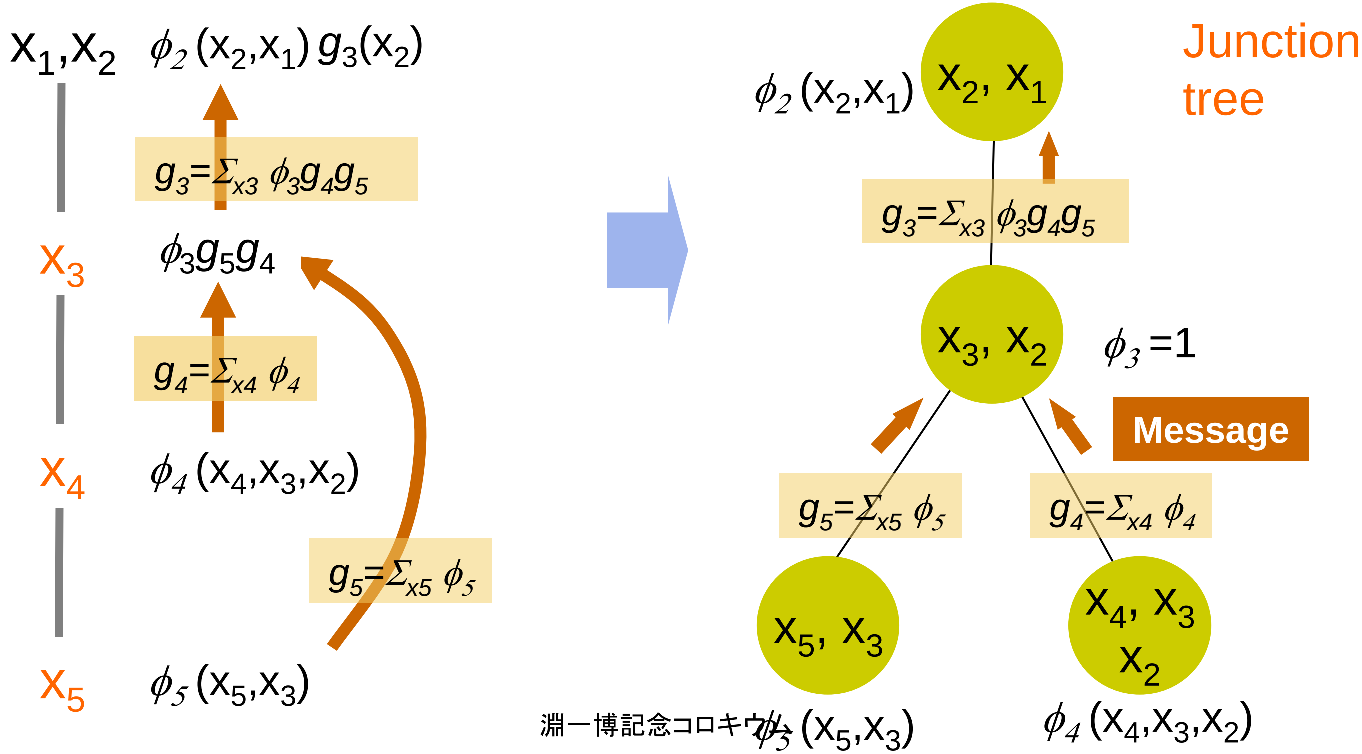
$\phi_2\,(x_2,x_1)=P(x_1)P(x_2|x_1)$

$\phi_4\,(x_4,x_3,x_2)=P(x_4|x_2,x_3)$

$\phi_5\,(x_5,x_3)=P(x_5|x_3)P(x_3)$

積和計算とBP(信念伝播)

$$p(x_1,x_2) = \Sigma_{x5,x4,x3}\, p(x_1,x_2,x_3,x_4,x_5)$$
$$= \Sigma_{x5,x4,x3,}\, \phi_2(x_2,x_1)\ \phi_4(x_4,x_3,x_2)\ \phi_5(x_5,x_3)$$
$$= \phi_2(x_2,x_1)\ \Sigma_{x3}(\Sigma_{x4}\phi_4(x_4,x_3,x_2)\ \Sigma_{x5}\phi_5(x_5,x_3))$$



統計的自然言語処理の発展 -- 確率文脈自由文法(PCFG)

PCFG = CFG + 確率

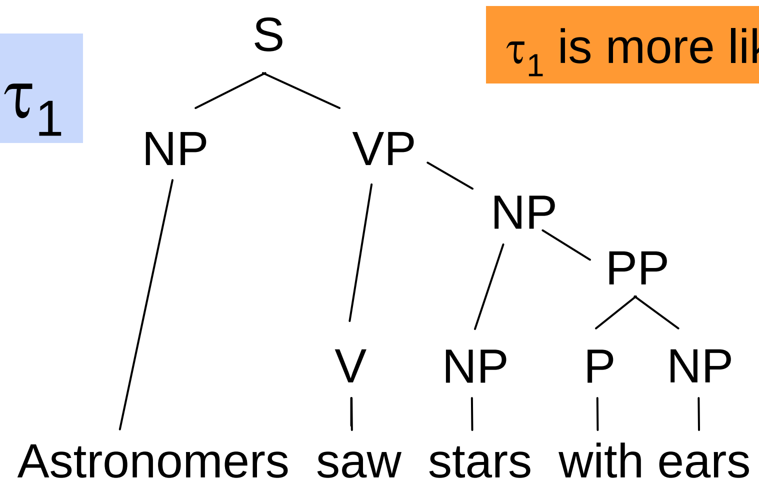
S	$\rightarrow$	NP VP (1.0)
NP	$\rightarrow$	NP PP (0.2) | ears (0.1) | stars (0.2)
		telescopes (0.3) | astronomers (0.2)
PP	$\rightarrow$	P NP (1.0)
VP	$\rightarrow$	VP PP (0.4) | V NP (0.6)
V	$\rightarrow$	see (0.5) | saw (0.5)
P	$\rightarrow$	in (0.3) | at (0.4) | with (0.3)

S $\Rightarrow^{*}$ Astronomers saw stars with ears

尤度による曖昧性解消

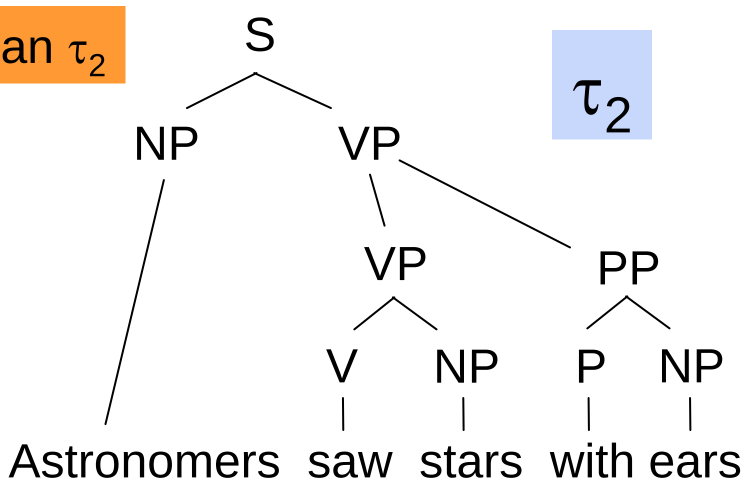
S	$\rightarrow$	NP VP (1.0)
NP	$\rightarrow$	NP PP (0.2) | ears (0.1) | stars (0.2)
		telescopes (0.3) | astronomers (0.2)
PP	$\rightarrow$	P NP (1.0)
VP	$\rightarrow$	VP PP (0.4) | V NP (0.6)
V	$\rightarrow$	see (0.5) | saw (0.5)
P	$\rightarrow$	in (0.3) | at (0.4) | with (0.3)

τ_1



τ_1 is more likely than τ_2

τ_2



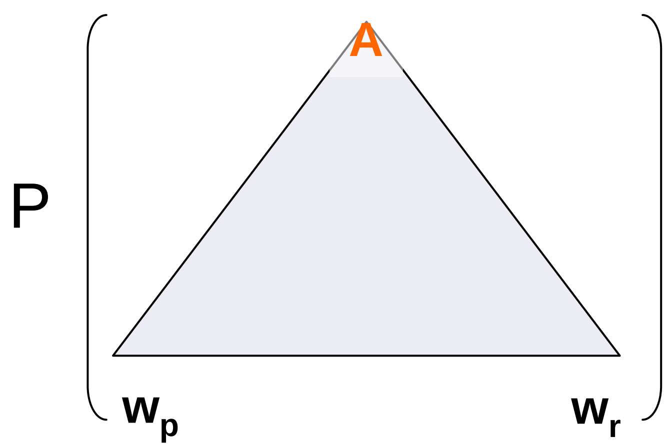
$P(\tau_1)=1.0*0.2*0.6*0.5*0.4*0.2*1.0*0.3*0.1$
$=0.00192$

$P(\tau_2)=1.0*0.2*0.4*0.6*0.5*0.2*1.0*0.3*0.1$
$=0.00144$

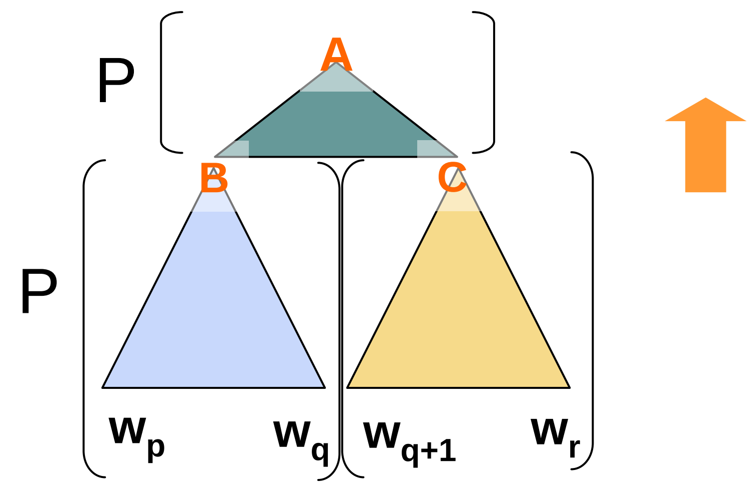
内側確率の計算

$$
\begin{aligned}
P(w_{1:N}) &= P(S \overset{*}{\Rightarrow} w_{1:N}) \\
&= \beta(S, 1, N) \\
\beta(A, p, r) &= P(A \overset{*}{\Rightarrow} w_{p:r}) \\
&= \begin{cases} \sum_{A \to BC,\, p \le q \le r} \\ \quad \theta_{A \to BC} \cdot \beta(B, p, q) \cdot \beta(C, q+1, r) \\ \qquad \text{if } r > p + 1 \\ P(A \to w_{p:p}) \text{ if } r = p \end{cases}
\end{aligned}
$$

- Bottom-up 計算
- 計算の複雑さ O(N^3)
- ダイナミックプログラミング

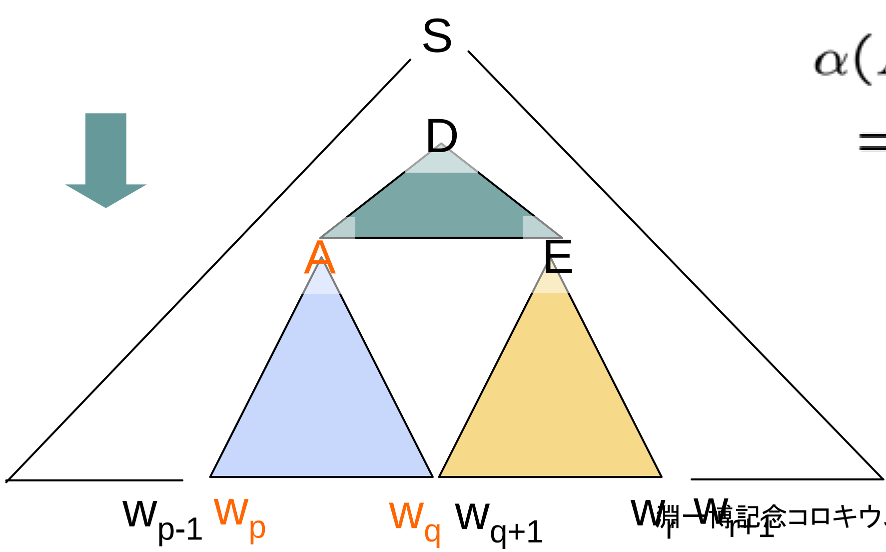


$= \Sigma_{B,C,q}$



外側確率の計算

$$
\begin{aligned}
&\alpha(A,p,q)\\
&= P(S \overset{*}{\Rightarrow} w_{1:p-1} A w_{q+1:N})\\
&= \begin{cases} 1 \text{ if } p=1, q=N \\ \sum_{D\to AE, p\le q\le r} \alpha(D,p-1,r+1)\cdot\theta_{D\to AE}\cdot\beta(E,q+1,r) \\ \quad +\sum_{D\to EA, p\le q\le r} \alpha(D,p-1,r+1)\cdot\theta_{D\to E}\cdot\beta(E,p,q) \\ \qquad \text{if o.w.} \end{cases}
\end{aligned}
$$



$$
\begin{aligned}
&\alpha(A,p,q)\cdot\beta(A,p,q)\\
&= S \overset{*}{\Rightarrow} w_{1:p-1} A w_{q+1} \overset{*}{\Rightarrow} w_{1:N}
\end{aligned}
$$

- Top-down 計算
- 計算の複雑さ O(N^3)
- ダイナミックプログラミング

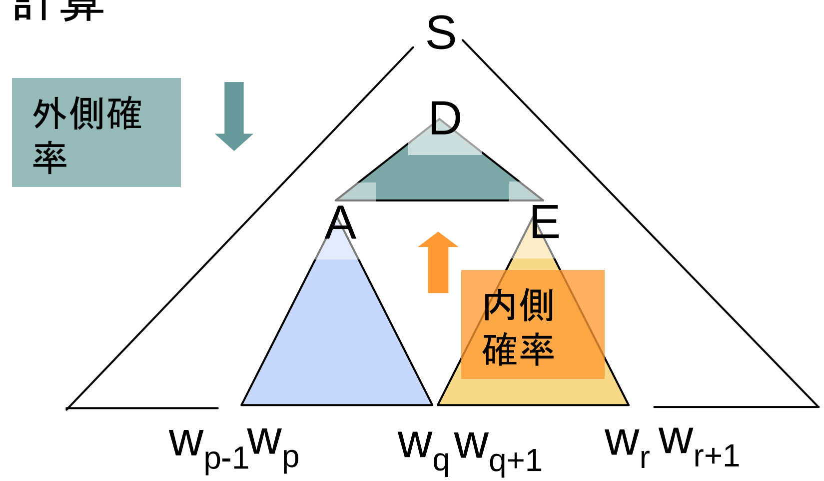
命題化計算による統合

- 命題化計算：“X=x”を命題変数として表現、計算する
  - 命題化計算は値毎の計算なので、値の間の依存関係を利用できる
  - BNでPCFGを計算しようとすると計算量は exponential
- 無限個の命題変数を用意してPCFGの命題化計算を行うと$O(N^3)$で確率計算ができる -- PRISM（Sato’97）
  - “wd(i,i)”, “ NP(i,j)”が命題変数
- MRFは命題化により計算できる – CFD(Macllester’04)
- BNも命題化により効率良く計算できる -- ACE(Chavira’05)
  - “A=a | B=b,C=c”が命題変数、確率P(A=a | B=b, C=c)を持つ
- BNのBPも命題化計算で実行できる – PRISM(Sato’07)
  - Junction treeを定節で表現、cpt(X_A, parent(Y_B,Z_c)) がアトム
  - P(cpt(a, parent(b,c))) = P(A=a | B=b, C=c)
- Quantifier Elimination(QE, 変数除去)
  - （命題化計算による）sum-product計算はQEの一種
  - resolution, linear programming, 制約解消もQEの一種

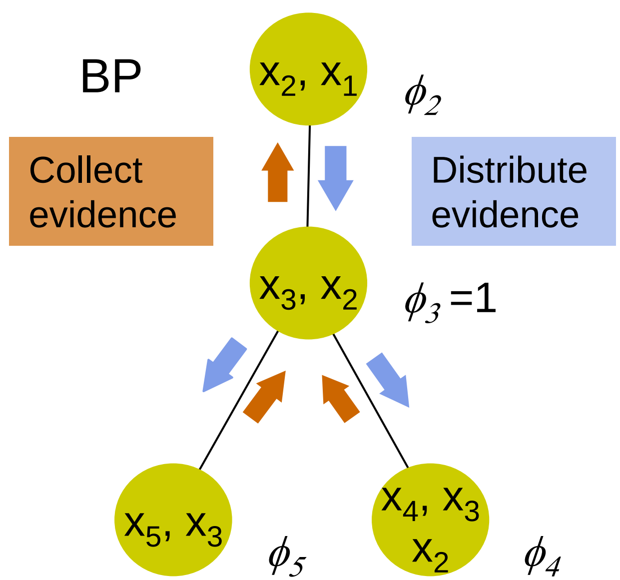
内側外側確率(IO)計算と信念伝播(BP)

- 似ている
  - Tree の上の2段階計算
  - 積和計算
- 違う
  - IO計算(PCFG)は再帰あり、BP(BN)は再帰なし

IO 計算

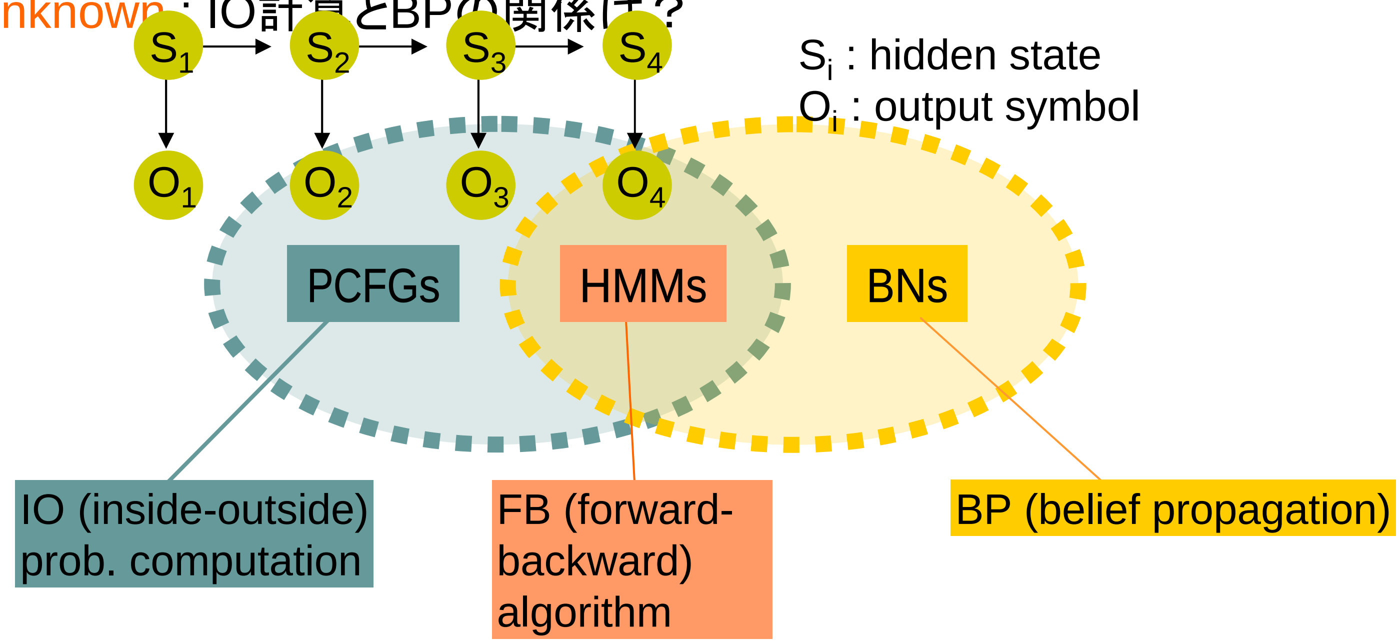


BP



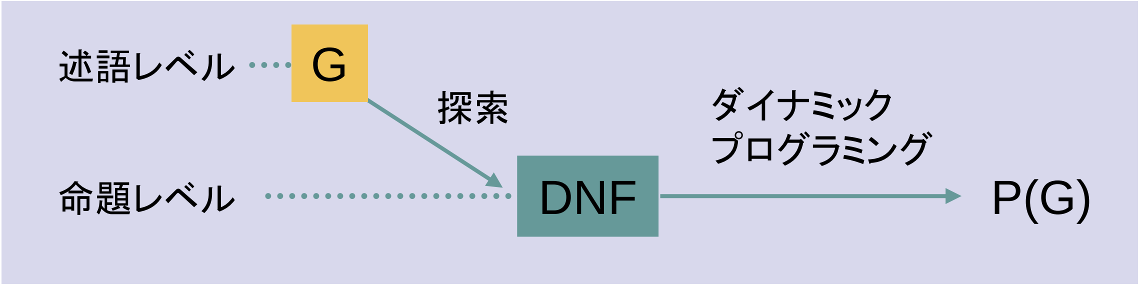
ベイズネットと確率文脈自由文法

- 別々のコミュニティで発達
- 両者は隠れマルコフモデル（HMM）を含む
  - FB (forward-backward) アルゴリズムにより確率計算を行う
  - FB = BP + HMM [Smyth et al.’97]
  - Unknown : IO計算とBPの関係は？



PRISMにおける統合(I)

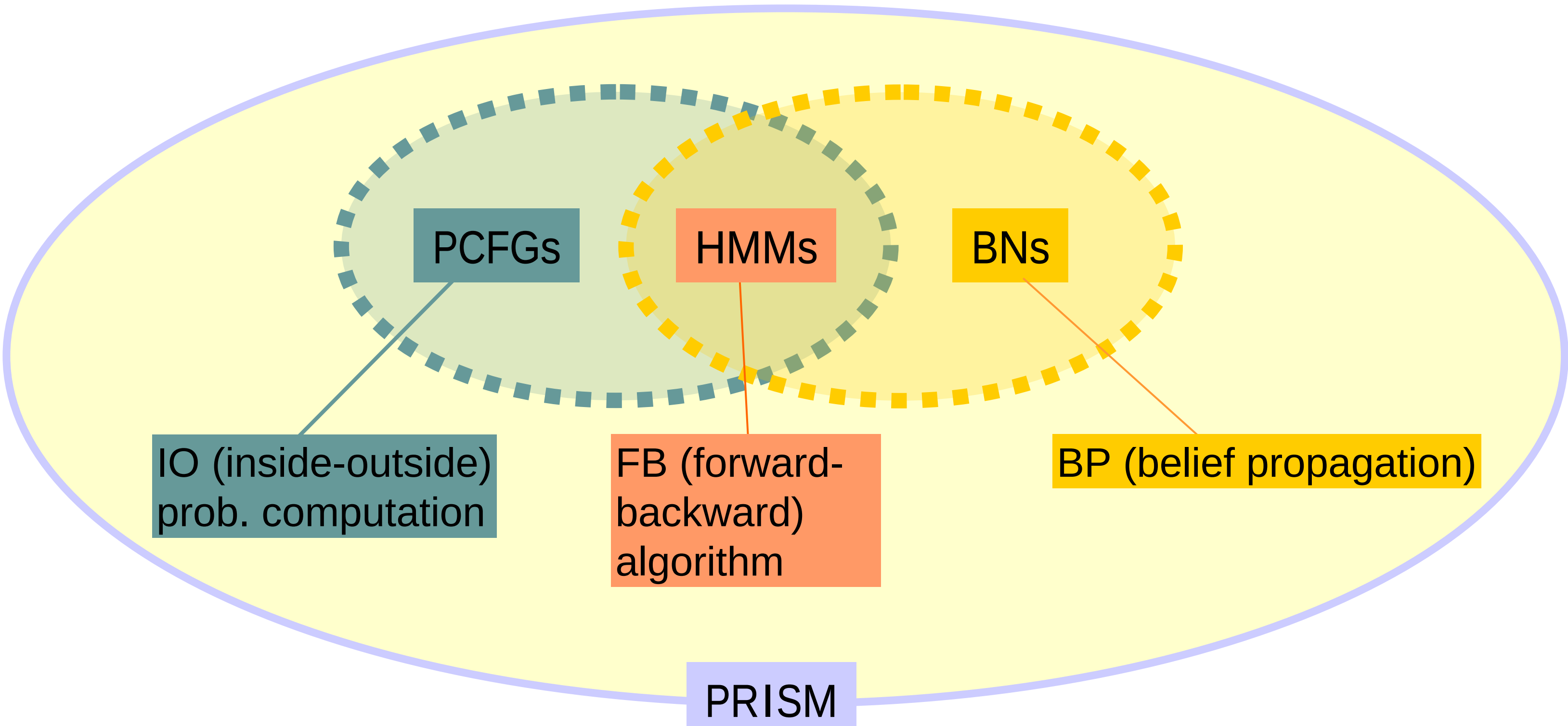
- 意味論:
  - 最小不動点意味論をHerbrand解釈上の確率空間に拡張したもの
  - 可算個のground atomsの同時分布(確率測度)
- 計算:
  - 論理計算はSLD + tabling
  - 確率計算は 一般化されたIO計算による命題化計算



- 学習:
  - 汎用のEMアルゴリズムとVB-EMアルゴリズム
  - Baum-Welchi, IO など包含

PRISMにおける統合(II)

- PCFG,HMM,BNを述語レベルで表現し、個別のアルゴリズムと同じ計算量で確率計算し、EM学習を行う事が出来る



生成的モデルと判別的モデル

- 生成的モデル（generative model）
  - 同時分布P(x,y)を定義、但しx は未知（構文木、クラス）、yは観測値
  - BN, HMM(regular), PCFG, PRISM (Turing machine)は生成的モデルを扱う
  - P(x | y) を知りたい場合が多い
- 判別的モデル（discriminative model）
  - P(x | y) を（log-linear モデルにより）直接的に定義
  - 代表はCRF（conditional random field）で、しばしばSVMをしのぐ判別性能を発揮する
- 生成的モデルと判別的モデルを統合する述語論理・確率モデルはまだない
  - 有限MRFは条件付生成的モデルである事が分かっている
  - MLNはMRFで表現される判別モデルP(MRF | RDB)を節集合で定義
  - PRMはBNを使った判別モデルP(BN | RDB)と考えられる

混沌の世界

- ２０００年頃からILPの中からBNに関心を寄せるグループが現れ、PLL(probabilistic logic learning)の標語の元、様々なLP+確率の枠組みが提案された。
- 同じ頃、BNの中から関係に注目するグループが現れBN+関係システムを追求し始め、SRL(statistical relational learning)という標語を掲げた。
- 両者は合同で論理、確率、学習の統合を追及しているが、まだ混沌としている。
- 本講演では両者のそれぞれの統合の一端を紹介したが、以上の流れとは全く別の論理推論＋確率の区間の流れについては触れられなかった。
- 不確定性は確率だけでなく無知や複雑さによっても引き起こされる。確率の先に不確定性があり、論理と不確定性の統合はまだ先が見えない。

多くのシステム – alphabet soup

(Probabilistic Logic Learning Turotial in ICML 2004 by K.Kersting)

[names in alphabetical order]