2001 年 ベイジアンネットチュートリアル
Tutorial on Bayesian Networks (BN2001)
Tokyo, Japan, July 29, 2001.

# 不確実性モデリングと AI
# ( Modeling uncertainty in AI )

佐藤 泰介*
東京工業大学大学院情報理工学研究科

**Abstract:**
最初にベイジアンネットワークを概観し、次に AI に於ける不確実性へのアプローチの若干の歴史を述べる. 最後に ( 離散 ) ベイジアンネットワークは成熟しつつあるが、 次の段階の一つとして述語論理レベルへの拡張があり、 現在他分野も巻き込んだ展開が始まっている事を述べる。

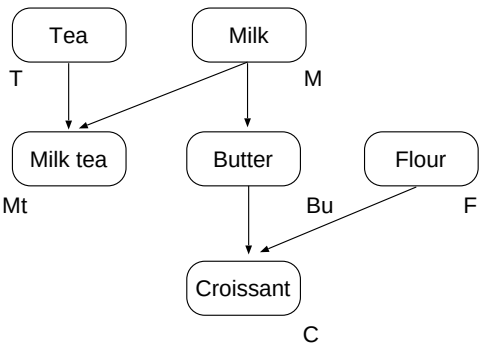
## 1 ベイジアンネットワークとは

実体： 同時確率分布の有向グラフ表現.

$$\begin{aligned} & p(x_1, \ldots, x_n) \\ & \quad = p(x_1)p(x_2 \mid x_1) \cdots p(x_n \mid x_1, \ldots, x_{n-1}) \\ & \quad = \textstyle\prod_i p(x_i \mid \pi_i) \end{aligned}$$

を使う. $\pi_i$ は $x_i$ を子ノードとする親ノード達を表す. 無向グラフとしてループのない単結合 ( singly connected ) ベイジアンネットワーク とループのある複結合 ( multiply connected network ) ベイジアンネットワークに分かれる. 離散分布の場合個々の $p(x_i \mid \pi_i)$ を 条件付き確率テーブル CPT ) で与える ( 以後離散分布を仮定 ).

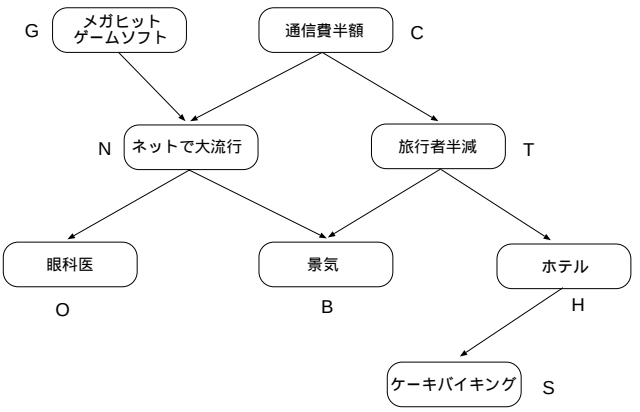
単結合ベイジアンネットワーク例：




P(T,M,Mt,Bu,F,Cr) =
P(T)P(M)P(Mt | T,M)P(Bu | M)P(Cr | Bu,F)

図 1: 単結合ベイジアンネットワークの例

複結合ベイジアンネットワーク例：




P(G, C, N, T, O, B, H, S) =
P(G) P(C) P(N | G,C) P(T | C) P(O | N) P(B | N,T) P(H | T) P(S | H)

図 2: 複結合ベイジアンネットワークの例

使い方：

診断ではベイジアンネットワークにより因果的関係をして、 結果 $e$ から原因 $c$ をベイズの事後確率 $P(c \mid e)$ が最大になるように推定する. これは周辺分布 $p(x, e)$ の計算に帰着する. 予測では例えば ネットワークゲーム ( N ) を観測してケーキバイキング ( S ) を予測する.

作業：

- グラフ構造の推定
- CPT のパラメータの推定
- 周辺分布の計算 ( 一般に NP-hard )

*〒 152-8552 東京都目黒区大岡山 2-12-1、e-mail: sato@mi.cs.titech.ac.jp

計算：

周辺分布 $p(x,e)$ の計算は exponential に時間が掛かる.

厳密計算： 単結合ベイジアンネットワーク用( Pearl によるメッセージパッシング計算は $O(|V|)$ ), 複結合ベイジアンネットワーク( junction tree にコンパイル ), VE ( variable elimination )

近似計算： variatinal method、サンプリング、その他

特色：

相関関係に方向性を入れて因果的知識の表現を行なう. その分通常の無向のグラフィカルモデルより記述が複雑になる場合あり ( 例えば d-separation など ). 計算量を強く意識.

関連分野：

隠れマルコフ、連鎖解析、カルマンフィルター、ターボコード、統計的自然言語処理 [8] など

応用製品など：

PATHFINDER：80 年代のシステム、リンパ節 ( lymph-node ) の病気の 60 以上の症状と 100 以上のテスト結果を判断、 マイクロソフト社の Office 95 の Answer Wizard、Office Assistant Office 97、トラブルシューター、NASA の Vista システム ( スペースシャトルの推進システムのモニタ、保全 ), RICHO のコピー機故障診断など

参考情報：

学術雑誌：

Artificial intelligence、Machine leanring、International Journal of Approximate Reasoning など

国際会議：

UAI(Uncertainty in AI)、IJCAI、AAAI、Int'l Workshop on AI and Statistics など

教科書：

[5]( ISBN 1-85728-332-5 ), [2]( ISBN 0-387-94858-9 ) など

電子的テキスト ：

`http://www.cs.engr.uky.edu/~dekhtyar/dblab/resources.html#bnets` など

団体：

Association for Uncertainty in Artificial Intelligence ( `http://auai.org/` )

## 2 不確定性の表現

- 1960 年代、 fuzzy logic, Dempster-Shafer theory
- 1970 年代、 CF(ceratinty factor)
- 80 年代に入り、 確率の重要性が訴えられ始めた (e.g. In Defenece of Probability (Cheeseman, IJCAI'85) ).
- 1985 年 第 1 回 UAI( Uncertainty in Artificial Intelligence ) が開催され、ベイジアンネットワークのコミュニティが成立.
- 1986 年、 単結合 ( singly connected ) ベイジアンネットワークに対するメーセージ伝搬法提案が Pearl により提案された [10]. ⇒ 変数具体化によるルー プ解体 (NP-hard 問題) により多重結合 ( multiply connected network ) ベイジアンネットワークも 単結合ベイジアンネットワークの集まりに分解して計算出来る ( $\sum_c P(x \mid c_1, \ldots, c_n, e) = \sum_{c_1,\ldots,c_n} P(x \mid e, c_1, \ldots, c_n) P(c_1, \ldots, c_n \mid e)$ ).
- 1988 年、ベイジアンネットワークの教科書 "Probabilistic Reasoning in Intelligent Systems" Pearl が出版される. 同年、multiply connected ベイジアンネットワークに対する clustering 法 が Lauritzen と Spegelhalter により提案される [7].
- 1990 年、ベイジアンネットワークの周辺分布計算が NP-hard である事が Cooper により証明される [3]. 同年、junction tree アルゴリズムが Jensen により提案される( ⇒ 商用ソフト Hugin へと展開 ).
- 以後順調に発展して、マイクロソフト社の Office などの製品に組み込まれようになり、今年は 第 17 回 UAI ( UAI-2001 ) が シアトルで開催予定.

# 3 ベイジアンネットワークと論理的知識表現

ベイジアンネットワークの表現力の向上を目指して、論理的表現と組み合わせる.

## 3.1 ベイジアンネットワークの限界

**(i)** 関係の表現：

論理記述のレベルが命題論理レベルなので、関係が表現出来ず、関係の具体例しか表せない. 例えば遺伝で言うと、父親から 遺伝子 gf もらう確率が $p$ であり、母親から 遺伝子 gm をもらう確率が $q$ ならば、子の遺伝型が $\langle \mathtt{gf}, \mathtt{gm} \rangle$ である確率は $pq$ であるというような一般的関係が表現出来ない.

**(ii)** 無限個の確率変数：

ダイナミックベイジアンネットワークは固定した大きさのネットワークの繰り返しを表現出来るので HMM が表現できた. しかし無限個の確率変数を必要とするより一般的確率過程の表現がうまく出来ない. 例えば（HMM の一般化であるところの）統計的自然言語処理の基本である確率文脈自由文法など.

## 3.2 論理式と確率を混ぜる際の問題点

**(i)** 確率の一貫性：

"確率" という言葉を使う限り制約がある.

$$P(\neg A) = 1 - P(A) = 0.5$$
$$P(A \wedge B) = \{minP(A), P(B)\}$$

と約束すると

$$P(A \wedge \neg A) = min\{P(A), 1 - P(A)\} = 0.5$$

になる.

コルモゴロフの公理（$0 \leq P(A) \leq 1$、$P(\neg A) = 1 - P(A)$、$P(A \vee B) = P(A) + P(B)$（$A, B$ が素））を満たすようにどのように論理式に確率を割り当てるかは自明ではない.

断片的に条件確率の集合が与えられたとしても、それらが、一つの確率分布から生じている保証はない.

**(ii)** 確率空間の構成：

現在知られている唯一の一貫して確率を論理式に割り当てる方法は、解釈の世界（可能多世界）$\Omega = \{\omega\}$ を固定し、$P(\phi) = P(\{\omega \mid \omega \models \phi\})$ とするものである. 逆も成立する事が 60 年代から知られている. この場合、使う言語の関数記号、述語記号が有限でなければ、$\Omega$ の濃度が可算を越えてしまうので、従来の、可算個の点 $\omega_1, \omega_2, \ldots$ に確率 $P(\omega_1), P(\omega_2), \ldots$、$\sum P(\omega_i) = 1$ を割り当てる（分かり易い）手法が使えない.

## 3.3 表現力の向上を目指して

Prolog の節（definite clause）を使ったベイジアンネットワークの表現の研究が [1] から始まった.

**KBMC：** Knowledge-Based Model Constructuion. Breeze [1], Koller [6] らのアプローチ.

Breeze は CPT を記述するのにアトム（原子論理式）CPT を組織するのに Prolog プログラム（定節（definite clause）の集合）を使った. Koller は節に確率を直接振り

$$genotype(P, G) \overset{0.5}{\leftarrow} parent(P, Q), genenotype(Q, G)$$

（親 $Q$ が遺伝子 $G$ を持つと子ども $P$ はそれを 0.5 の確率で受け継ぐ）のような表現を使った. しかし、意味論も確率空間もなかったので、論理式の確率変数としての解釈は困難. 単にベイジアンンネットワークを記述するマクロ言語としての役割に留まった.

**SLP：** Stochastic Logic Progrmming, Muggleton [9], Cussens [4] らのアプローチ.

確率文脈自由文法 (PCFG) の拡張である. 非終端記号 $A$ が $N$ 個の規則 $\{A \rightarrow \alpha_i \mid 1 \leq i \leq N\}$ を持つ時、確率 $p_i$ が各規則 $A \rightarrow \alpha_i$ $(1 \leq i \leq N)$ に割り当てられる（$\sum_{i=1}^{N} p_i = 1$）. Muggleton らは Koller と同じように規則を変数を含む節に拡張した. 確率空間の議論がなくまた論理式の証明に確率を割り当てたので、$P(A \wedge A) \neq P(A)$ となる可能性があり、意味論的に問題が多い.

**PRISM：** Poole [11], Sato [12] らのアプローチ.

最初から単一の確率空間を構成して、その上で論理式の確率を議論する立場をとる. Poole は Prolog プログラムを適切に拡張することによりベイジアンネットワークを書き下せる事を証明した [11]. Sato [12] は 論理プログラムの意味論の確率的拡張を与え、パラメータ学習のための EM アルゴリズムを導いた.

**PRISM [13] プログラム例：** PRISM は定義された確率的意味論と EM 学習ルーチンを持つ統計的記号モデリング言語. 離散分布の表現手段としてプログラムを使う事を可能にした.

図 3 は ABO 血液型の遺伝の記述を与えている. `X,Gf,Gm` などは論理変数、 `btype(a)` などは確率変数である. 存在限量子 $\exists$ が 確率和 $\sum$ に対応する. `msw(abo,P,G)` は 名前が `abo` のさいころ ( 確率的スイッチ ) で `P` 回目に振ると値 $\in \{\mathtt{a},\mathtt{b},\mathtt{o}\}$ を `G` に返す.

```
btype(X):-
   pg_table(X,Gf,Gm),genotype(Gf,Gm).

   pg_table(a,a,a). pg_table(a,a,o).
   pg_table(a,o,a). pg_table(b,b,b).
   pg_table(b,b,o). pg_table(b,o,b).
   pg_table(o,o,o). pg_table(ab,a,b).
   pg_table(ab,b,a).

genotype(Gf,Gm):-
   gene(father,Gf),gene(mother,Gm).

gene(P,G):-msw(abo,P,G).
```

図 3: ABO blood type program $DB_{\mathtt{abo}}$

# 参考文献


[1] Breese, J. S., "Construction of belief and decision networks," Computational Intelligence, vol. 8, No. 4, pp. 624-647, 1992.

[2] Castillo,E.C., Gutierrez,J.M. and Had,A.S., "Expert Systems and Probabilistic Network Models," Springer-Verlag, 1997.

[3] Cooper,G.F., "The Computational Complexity of Probabilitic Inferences Using Bayesian Belief Networks," Artificial Intelligence vol.42, pp. 393-405, 1990.

[4] Cussens, J., "Loglinear models for first-order probabilistic reasoning," Proc. of UAI'99, pp. 126-133, 1999.

[5] Jensen,F.V., "An Introduction to Bayesian Networks," UCL Press 1996.

[6] Koller, D. and Pfeffer, A., "Learning probabilities for noisy first-order rules," Proc. of IJCAI'97, pp. 1316-1321, 1997.

[7] Lauritzen,S.L. and Spiegelhalter,D.J., "Local Computations with Probabilities on Graphical Structures and Their Applications to Expert Systems," J. of the Royal Statistical Socity, B, vol. 50, pp. 157-224.

[8] Manning, C. D. and Schütze, H., "Foundations of Statistical Natural Language Processing," The MIT Press,1999.

[9] Muggleton, S., "Stochastic logic programs," in *Advances in Inductive Logic Programming*,ed. de Raedt, L., OSP Press, pp. 254-264, 1996.

[10] Pearl,J.,"Fusion, Propagation and Structuring in Belief Networks," Artificial Intelligence vol.29, pp. 357-369, 1986.

[11] Poole, D., "Probabilistic Horn abduction and Bayesian networks," Artificial Intelligence, vol. 64, No. 1,pp. 81-129, 1993.

[12] Sato, T., "A statistical learning method for logic programs with distribution semantics," Proc. of ICLP'95, pp. 715-729, 1995.

[13] Sato, T. and Kameya, Y., "PRISM: a language for symbolic-statistical modeling," Proc. of IJCAI'97, pp. 1330-1335, 1997.