人　工　知　能　96－2
（１９９４．１０．４）

回帰分析に基づいた帰納学習アルゴリズム

月本 洋　下郡 信宏　森田 千絵
(株) 東芝 研究開発センター

本論文では回帰分析に基づく帰納学習アルゴリズムを提示し、C4.5 と比較して良好な結果を得たことを述べる。アルゴリズムの概要は、データ（連続属性、離散属性）を２値化し、回帰分析を行ない線形関数（$\{0,1\}^n \rightarrow \{0,1\}$）を得、この線形関数をブール関数で近似することである。線形関数をブール関数で近似するのは準最尤法であり、回帰分析は最小２乗法であるから、本アルゴリズムはパターン学習の原理で記号学習のアルゴリズムが得られることを意味する。

An Inductive Learning Algorithm Based on Regression Analysis

Hiroshi Tsukimoto　Nobuhiro Shimogori　Chie Morita
Research & Development Center, Toshiba Corporation

This paper presents an inductive learning algorithm based on regression analysis and shows the algorithm works better than C4.5. The algorithm consists of preprocessing continuous data and discrete data, obtaining linear functions by multiple regression analysis, and approximating the functions by Boolean functions. Since approximating linear functions by Boolean functions is a pseudo maximum likelihood method and regression is the least square method, the algorithm means that an algorithm for symbol learning can be obtained by the principles of pattern learning.

1 はじめに

従来、回帰分析等の多変量解析のいわゆるパターン学習のアルゴリズムと人工知能の分類木等の帰納学習等の記号学習のアルゴリズムは別物であった。本論文では代表的なパターン学習アルゴリズムである回帰分析で得られた線形関数をブール関数で近似することによって（近似するだけで）記号学習アルゴリズムが得られることを示す。

アルゴリズムの概要は以下の通りである。

1. 連続属性、離散属性を2値化する。
2. 回帰分析を行ない線形関数（$\{0,1\}^n \rightarrow \{0,1\}$）を得る。
3. この線形関数をブール関数で近似する。

線形関数をブール関数で近似するのは無差別原理を仮定すれば準最尤法であることが言える。回帰分析は最小2乗法であるから、原理的には本アルゴリズムは最小2乗法と準最尤法から構成されることになる。すなわちパターン学習の原理で記号学習のアルゴリズムが得られることになる。このアルゴリズムの予測率はC4.5と比較して良好であることを実験的に確認した。

2節で線形関数（$\{0,1\}^n \rightarrow \{0,1\}$）をブール関数で近似する方法について述べる。3節ではこの近似方法が無差別原理を仮定すれば準最尤法であることを述べる。4節では2節で提示したアルゴリズムの効率化について述べる。5節では誤差の理論的分析について述べる。そして最後に6節では実験について述べる。

2 線形関数のブール関数による近似

$\{0,1\}^n \rightarrow \{0,1\}$ の線形関数をブール関数で近似するには、線形関数がブール代数の原子（論理の最小項）で張られる線形空間（ユークリッド空間）に属すことに注目すれば良い。

2.1 ブール代数の原子で張られる線形空間（ユークリッド空間）

n 変数のブール代数の原子は以下の通りである。

$\phi_i = \prod_{j=1}^{n} e(x_j)$ $(i = 1 \sim 2^n, j = 1 \sim n)$ ここで $e(x_j) = \overline{x_j}$ または x_j である。例えば2変数の場合にはブール代数の原子は $xy, x\overline{y}, \overline{x}y, \overline{xy}$ である。任意のブール関数はこの原子の線形結合で表現される。すなわち $\sum e_i a_i$。ここで a_i は原子であり、e_i は係数（1か0）である。ここで係数を $\{1,0\}$ から実数に拡張すれば線形空間になる。即ちその線形関数は $\sum r_i a_i$ となる。但し r_i は実数である。また関数間に自然に2乗距離が入れられる。即ち二つの関数を $\sum r_i a_i, \sum s_i a_i$ とすればその距離は $(\sum(r_i - s_i)^2)^{1/2}$ である。

2.2 線形関数のブール代数の原子による展開

線形関数を

$p_1x_1 + ... + p_nx_n$

とし、それの原子による展開式を

$a_1x_1 \cdot\cdot x_n + ... + a_{2^n}\overline{x_1} \cdot\cdot \overline{x_n}$

とすると、

$$p_1x_1 + ... + p_nx_n = a_1x_1 \cdot \cdot x_n + ... + a_{2^n}\overline{x_1} \cdot \cdot \overline{x_n}$$

となる。a_i は次の通りである。

$a_1 = p_1 + ... + p_n,$
$a_2 = p_1 + ... + p_{n-1},$
...
$a_{2^{n-1}} = p_1,$
$a_{2^{n-1}+1} = p_2 + p_3 + ... + p_{n-1} + p_n,$
...
$a_{2^n-1} = p_n,$
$a_{2^n} = 0$

各係数 a_i はそれに対応する原子の値のみを 1 にし、他の原子の値を 0 にするような変数の値の組合わせを代入することによって簡単に求められる。例えば a_1 ならばこれに対応する原子 $x_1 \cdot \cdot x_n$ の値を 1 にする変数の値の組合わせは $x_1 = x_2 = \cdot\cdot = x_n = 1$ である。これを代入すると右辺は a_1 となり、左辺は $p_1 + ... + p_n$ となり、$a_1 = p_1 + ... + p_n$ が得られる。同様に $x_1 = x_2 = \cdot\cdot = x_{n-1} = 1, x_n = 0$ を代入すると $a_2 = p_1 + ... + p_{n-1}$ となる。他の a_i も同様にして計算できる。

2.3 線形関数のブール関数による近似

今までの議論に基づき、線形関数の原子による展開（ベクトル表現）を f としブール関数のベクトル表現を g とすると線形関数をブール関数で近似することは f を g で近似する事になる。$\mathbf{f}(= (f_i))$ とし、$\mathbf{g}(= (g_i))$ とすると、最小にしたい量は $\sum(f_i - g_i)^2$ となる。各項 $(f_i - g_i)^2$ は独立なので各項を最小にすることが $\sum(f_i - g_i)^2$ を最小にすることと同じになる。g_i は 1, 0 のいずれかなので、$(f_i - g_i)^2$ を最小にする g_i は $f_i \geq 0.5$ ならば $g_i = 1$ その他で $g_i = 0$ となる。例えば線形関数を $z = 0.6x - 1.1y + 0.3$ としよう。これを直交展開すると、$z = -0.2xy + 0.9x\overline{y} - 0.8\overline{x}y + 0.3\overline{x}\overline{y}$ となり、ブール関数で近似すると $x\overline{y}$ となる。

3 準最尤法

前節の近似法が準最尤法であることを証明するためにまず論理と確率の対応関係を調べる [月本 92]。

3.1 情報量の導入

下式の H を情報量として定義する。

$$H(f) = n - log_2(N(f))^2$$

ただし $N(f)$ は関数 f の（2 乗）ノルムである [月本 90]。この情報量は無差別原理を仮定すると確率の情報量 $I(= n - H_e)$ と等しいことが証明できる。ただし $H_e = -\sum_1^{2^n}(p_i log_2 p_i)$ で n は変数の数である。

3.2 論理ベクトルと確率ベクトルの基本的対応

情報量 H が 0 の命題はトートロジである。例えば $x \vee \overline{x}$。$x \vee \overline{x}$ のベクトル表示 (論理ベクトル) は (1,1) である。一方、情報量 0 の確率分布は (1/2, 1/2) である。この情報量とは I である。ここで、命

題で最も情報量の小さいものと事象で最も情報量の小さいものを対応させると、論理ベクトルの (1,1) と確率ベクトルの $(1/2, 1/2)$ が対応することになる。この考えは無差別原理と同等のものである [Keynes 21]。同様に考えると 2 変数では論理ベクトル $(1,1,1,1)$ と確率ベクトル $(1/4,1/4,1/4,1/4)$ が対応することになる。そして一般に n 変数のトートロジの場合には以下の対応が存在する事になる。

$$\begin{array}{ll} 2 & \text{変数}\ (1,1,1,1) \Leftrightarrow (1/4,1/4,1/4,1/4) \\ n & \text{変数}\ (1,...,1) \Leftrightarrow (1/(2^n),...,1/(2^n)) \end{array}$$

同様にして以下の対応が得られる。

$$\begin{array}{ll} x & \text{1変数}\ (1,0) \Leftrightarrow (1,0) \\ x & \text{2変数}\ (1,1,0,0) \Leftrightarrow (0.5,0.5,0,0) \end{array}$$

ここで、m 個の要素が 1 であるような論理ベクトル $\mathbf{f} = (f_i)(i = 1 \sim n)$ とそれに対応する確率ベクトル $\mathbf{p} = (p_i)(i = 1 \sim n)$ 考える。この場合上記の議論に従えば以下の対応が成立することになる。

f_i が 1 ならば p_i は $1/m$

f_i が 0 ならば p_i は 0

これは f は p と方向が同じだがノルムが違うベクトルであることを意味する。以降この関係を非古典論理にも拡張して使用する。

3.3 対応式

3.1 より $H = I$ であり、これより、

$N = 2^{H_e/2}$

となる。又、3.2 よりベクトルの向きが同じなため、対応式は以下のようになる。

$\mathbf{f} = (2^{H_e/2}/|\mathbf{p}|)\mathbf{p}$

ここで、$H_e = -\sum_1^{2^n}(p_i log_2 p_i)$, $|\mathbf{p}| = (\sum_1^{2^n} p_i^2)^{1/2}$, $\mathbf{p} = (p_1, ..., p_{2^n})$

3.4 準最尤法の証明

前節の近似法が無差別原理を仮定すれば準最尤法であることを示す。

いま線形関数 f とブール関数 g が十分近いと仮定する。即ち

$f_i \simeq g_i, (i = 1, ..., n)$

とする。この時前節の近似は

$\mathbf{f} - \mathbf{g} \to min$

であり、今 $f_i \simeq g_i, (i = 1, ..., n)$ であるから上式は

$||\mathbf{f}| - |\mathbf{g}|| \to min$

となる。$|\mathbf{f}|, |\mathbf{g}|$ を前節の議論に基づいて $2^{H_e/2}, 2^{H'_e/2}$, とおくと上式は

$|2^{H_e/2} - 2^{H'_e/2}| \to min$

となり、

$|H_e - H'_e| \to min$

となる。$H'_e > H_e$ の場合を考えると

$H'_e - H_e \to min$

となり、$H_e = -\sum_1^{2^n}(p_i log_2 p_i), H'_e = -\sum_1^{2^n}(q_i log_2 q_i)$ とおけば、

$\sum_1^{2^n}(p_i log_2 p_i) - \sum_1^{2^n}(q_i log_2 q_i) \to min$

$f_i \simeq g_i, (i = 1, ..., n)$ を仮定しているから $p_i \simeq q_i, i = 1, ..., n$ であり、対数関数が定義域 [0,1] で急降下関数であることを考えると式は

$\sum_1^{2^n}(p_i log_2 p_i) - \sum_1^{2^n}(p_i log_2 q_i) \to min$

と変形できる。これは K-L 情報量であり、この近似法は最尤法となる [坂元 83]。従って $f_i \simeq g_i, (i = 1, ..., n)$ を仮定すれば最尤法になるので、前節の近似法は準最尤法と言える。

4　効率的なアルゴリズム

2節で提示した近似アルゴリズムは n 変数の時は 2^n 個の係数を調べる必要があり、計算量的に指数オーダーであり、現実的ではない。本節では効率的な近似アルゴリズムを提示する。[月本 93]

4.1　近似後のブール関数に x_i 等が存在する条件

ここで x_1 に注目する。ブール関数に近似された後で x_1 が存在することは

$x_1 x_2 \cdot \cdot x_n$,

$x_1 x_2 \cdot \cdot \overline{x_n}$,

...

$x_1 \overline{x_2} \cdot \cdot \overline{x_n}$.

の全ての原子が存在することである。即ち各原子の係数, $a_1, a_2, ..., a_{2^{n-1}}$ が全て 0.5 以上のことである。即ち $min\{a_i\} \geq 0.5 (i = 1 \leq i \leq 2^{n-1})$ （以下 $mina_i$ と略記）である。

ところで $a_i (1 \leq i \leq 2^{n-1})$ は

$a_1 = p_1 + ... + p_n$,

$a_2 = p_1 + ... + p_{n-1}$,

...

$a_{2^{n-1}} = p_1$,

であり、$a_i (1 \leq i \leq 2^{n-1})$ は全て p_1 を含んでいる。従ってもし p_i が全て非負ならば $a_{2^{n-1}} = p_1$ が最小になる。なぜならばその他の a_i は他の p_i を含んでいるから $a_{2^{n-1}} = p_1$ より大きくなるからである。一般には p_i は全て非負ではない。その場合には負である p_i を全て含んでいる a_i が最小になる。即ち $mina_i = p_1 + \sum_{1 \leq j \leq n, j \neq 1, p_j \leq 0} p_j$ となる。そしてこのような a_i は必ず $a_i (1 \leq i \leq 2^{n-1})$ の中に存在する。なぜならば $a_i (1 \leq i \leq 2^{n-1})$ は $p_1 + (p_j (2 \leq j \leq n)$ の任意の組合せ) であるからである。以上の議論より

$mina_i = p_1 + \sum_{1 \leq j \leq n, j \neq 1, p_j \leq 0} p_j$

となる。従って、近似後のブール関数に x_1 が存在する条件は $p_1 + \sum_{1 \leq j \leq n, j \neq 1, p_j \leq 0} p_j \geq 0.5$ となる。回帰式 $p_1 x_1 + ... + p_n x_n$ は x_i に関して対称であるから以上の議論は他の x_i についても成立する。即ち

$p_i + \sum_{1 \leq j \leq n, j \neq i, p_j \leq 0} p_j \geq 0.5$

ならば x_i が近似後のブール関数に存在する。同様に考えると $\overline{x_i}$ が近似後のブール関数に存在する条件は

$\sum_{1 \leq j \leq n, j \neq i, p_j \leq 0} p_j \geq 0.5$

となる [高島 94]。今までの議論は最も低次の項（$x_i, \overline{x}_j$）についての議論であったが同様の議論が高次の項（$x_i x_j$ 等）の存在についても可能であり、その結果一般に $x_{i_1} \cdot \cdot x_{i_k} \overline{x}_{i_{k+1}} \cdot \cdot \overline{x}_{i_l}$ が近似後のブール関数に存在する条件は

$\sum_{i_1}^{i_k} p_{i_j} + \sum_{1 \leq j \leq n, j \neq i_1, .., i_l, p_j \leq 0} p_j \geq 0.5$

になる。

4.2 DNF 式の生成

上記の判定条件を用いて低次の項から生成して行き、その項を論理和で接続して最終的なブール関数を得る。低次の項で存在が確認された変数に関してはその項より高次の項での存在の判定の必要がない。簡単な例で言えば x が存在することが分かれば xy, xz 等は存在する（$x = x \vee xy \vee xz$）ので xy, xz の存在の判定は必要ないのである。従ってこのアルゴリズムは簡単化を同時に行なっている。このブール関数同定アルゴリズムの例を以下に挙げる。$y = 0.65x_1 + 0.23x_2 + 0.15x_3 + 0.20x_4 + 0.02x_5$ を回帰分析の回帰式とする。まず x_i に関しては x_1 の存在条件 p_1 のみが 0.5 以上なので x_1 が存在することになる。つぎに x_1 を除外した $x_i x_j$ に関してはどの p_i と p_j の和も 0.5 を越えないので $x_i x_j$ と言う項は存在しないことになる。そして次に x_1 を除外した $x_i x_j x_k$ に関しては $x_2 x_3 x_4$ の存在条件 $p_1 + p_2 + p_3$ が 0.5 以上なので $x_2 x_3 x_4$ が存在することになる。x_1, x_2, x_3, x_4 を除外すると x_5 が残るが、これ一つでは更に高次の項は作れないのでここでアルゴリズムは終了する。従って存在するのは x_1 と $x_2 x_3 x_4$ となり、その論理和 $x_1 \vee x_2 x_3 x_4$ が求めるブール関数となる。

5 誤差について

5.1 k 次までで近似する場合の誤差

まず回帰分析の計算量は回帰関数の変数（＝項）の多項式である。これはブール関数の変数の数 n の多項式である。従って回帰分析の計算量はブール関数の変数の数 n の多項式であると言える。回帰式からブール関数を求めるアルゴリズムの計算量は簡単な計算より多項式であることが分かる。但し高次の項まで同定しようとすると計算量が増加して行き最高次まで同定するとその計算量は指数関数になる。従って計算量を多項式にするには項の生成をある次数で止めねばならない。

ある次数までの論理式で近似する場合の誤差評価に関して [Linial 93] の結果が有効である。Linial は一様分布の仮定のもとで以下の式を提示した。

$\sum_{|S|>k} \hat{f}(S)^2 \leq 2M2^{-k^{1/2}/20}$,

ただし f はブール関数である。S は項（もしくは項の集合）、$|S|$ は項（もしくは項の集合）の次数であり、k は整数、$\hat{f}(S)$ は f の S 次フーリエ変換であり、M は回路のサイズである。

上式は高次の項は非常に少ないパワーしかなく、低次の項に情報が集中していることを意味している。即ちある次数までの項の近似で十分良い近似になっていることになる。即ち若干の誤差を付加するだけで計算量を多項式に落せることを意味する。

5.2 サンプル計算量

このアルゴリズムのサンプル計算量を簡単に説明する。仮説空間を線形関数空間にした場合の正解と経験的な仮説間の誤差は経験的な仮説と最適な仮説の誤差と最適な仮説と正解の間の誤差の和となる。

まず経験的な仮説と最適な仮説の誤差は疑似次元（pseudo dimension[Haussler 92]）で評価されることを述べる。この疑似次元は VC 次元を拡張したものである。VC 次元は値域が $\{0,1\}$ の関数空間に対して定義されている [Blumer 89]。Haussler は VC 次元を、値域が実数の関数空間に拡張して、疑似次元を定義した [Haussler 92]。Haussler の定理は次の通りである。

定理 [Haussler 92]

d を仮説空間の疑似次元とする。h_t と h_e を各々最適な仮説と経験的な仮説とする。m をサンプル数とする。$|\bullet|$ をノルムとする。ϵ と δ を実数とする。線形関数の値域が [0,1] ならば, 任意の確率分布に対して以下の式が成立する。

Probability$(|t-h| \geq \epsilon) \leq \delta$ if $m \geq \frac{a}{\epsilon^2}(2dlog_2\frac{b}{\epsilon} + log_2\frac{c}{\delta})$,

但し a, b, c は係数である。

n 変数の場合、線形関数空間の疑似次元は n の多項式である [Dudley 78]。それゆえ、サンプル計算量は n, $\frac{1}{\epsilon}$ and $\frac{1}{\delta}$ の多項式になる。

5.3 ブール関数を多重線形関数で近似する場合の誤差

正解と最適な仮説との誤差は [Linial 93] によって評価される。

定理 [Linial 93]

f をブール関数とする。一様分布の場合、$\| f-g \| < \epsilon$ を満たす k- 多重線形関数 g が存在する。但し k は高々 $O(log(n/\epsilon)^2)$ である。

この式はブール関数が低次元の多重線形関数で良く近似されることを示している。誤差解析に関して簡単に述べたが、さらに詳細な解析は将来の課題である。

6 実験

6.1 事例の前処理

事例の前処理は、欠損値補填、連続属性の離散化、離散属性の 2 値化、変数の削減である。

6.1.1 欠損値補填

欠損値の補填は次のように行なう。まず、欠損値を持つ事例のクラスと同じクラスである事例の集合を作り、欠損している属性の値に注目する。このとき、

離散属性 ： クラスが同じ事例の集合の中で最頻の値
連続属性 ： クラスが同じ事例の集合の中の平均値

によって欠損値を補充することとする。

表 1: 例

事例	A	B	クラス
E1	a_1	?	1
E2	a_1	b_2	2
E3	a_2	b_1	1
E4	a_1	b_1	1
E5	a_3	b_2	2
E6	a_2	b_1	2
E7	a_3	b_2	1

6.1.2 連続属性の離散化

連続属性の離散化は C4.5 のアルゴリズムを流用した。これはアルゴリズムの性能を C4.5 と比較するため、（非本質的な）連続属性の（離散化）での性能を同じにするためである。

6.1.3 離散属性の表現

上述の連続属性の離散化によって属性はすべて離散になった。離散値をとる属性に対しては、ダミー変数によって $\{0, 1\}$ の 2 値で表現する。ダミー変数とは次のように値を定める。

$$x_{ij} = \begin{cases} 1 & \text{属性 } i \text{ について値 } j \text{ を持つとき} \\ 0 & \text{その他のとき} \end{cases}$$

例えば、属性 A の属性値が $\{a_1, a_2, a_3\}$ のとき、属性値「a_2」は「0, 1, 0」と表される。 このとき、各属性の属性値の数をそれぞれ n_i とすると、ダミー変数の数は $\sum n_i$ になる。

6.1.4 変数削減

次に、ダミー変数によって表現されたデータに対し、変数の削減を行なう。これはダミー変数の導入による線形従属性を消去するためである。各離散属性に対応するダミー変数の組 (属性値に相当する) から一つずつ変数を削除する。削除される変数はクラスとの負の相関が最も高いものである。

6.1.5 例

例えば表 1のデータに対して上記の処理を行なうと表 2のようになる。クラスもダミー変数で表現してクラス毎にデータを持つ。

6.2 実験

本アルゴリズムの実験を行ない、機械学習アルゴリズムの一つである C4.5 [Quinlan 93] と比較した。回帰分析はクラス毎に行なう。従って回帰式はクラスの数だけ存在する。予測の時は値が最も大きい回帰式のクラスとする。表 3 が実験結果であるが、本アルゴリズムの方が C4.5 より優れていることがわかる。なお表中の $(2, m, c-2)$ は属性が 2 値、多値、連続でクラスが 2 値であることを意味する。

表 2: 事例表現の全ての処理を終了

表 2-a: クラス 1 に対して

	A		B	
事例	a_1	a_2	b_1	クラス 1
E1	1	0	1	1
E2	1	0	0	0
E3	0	1	1	1
E4	1	0	1	1
E5	0	0	0	0
E6	0	1	1	0
E7	0	0	0	1

表 2-b: クラス 2 に対して

	A		B	
事例	a_1	a_2	b_1	クラス 1
E1	0	0	0	0
E2	0	0	1	1
E3	1	0	0	0
E4	0	0	0	0
E5	0	1	1	1
E6	1	0	0	1
E7	0	1	1	0

表 3: 実験結果

	本アルゴリズム		C4.5	
データ名	回帰式	命題	決定木	ルール
mushrooms(2,m—2)	99.3%	99.3%	98.7%	98.7%
voting-records(2—2)	97.0%	97.0%	93.0%	93.5%
zoo(2,m—m)	93.1%	84.9%	86.9%	87.8%
iris(c—m)	94.1%	94.1%	93.6%	93.6%
heart-disease(2,m,c—2)	78.1%	70.2%	66.9%	71.6%

7 おわりに

本論文では回帰分析に基づく帰納学習アルゴリズムを提示し、C4.5 と比較して良好な結果を得たことを述べた。アルゴリズムの概要は、データ（連続属性、離散属性）を２値化し、回帰分析を行ない線形関数（$\{0,1\}^n \to \{0,1\}$）を得、この線形関数をブール関数で近似することであった。実験の結果、代表的なアルゴリズムである C4.5 より優れていることが分かった。また線形関数をブール関数で近似するのは準最尤法であり、回帰分析は最小２乗法であるから、本アルゴリズムはパターン学習の原理で記号学習のアルゴリズムが得られることを意味する。変数選択等の改良は今後の課題である。

参考文献

[Blumer 89] A. Blumer, A. Ehrenfeucht, D. Haussler and M. K. Warmuth: Learnability and the Vapnik-Chervonenkis dimension *J. ACM*, Vol.36, No.4, pp.929-965, 1989.

[Haussler 92] D. Haussler: Decision theoretic generalization of the PAC Model for neural net and other learning applications , *Information and computation*, pp.78-150, 1992.

[Keynes 21] J.M. Keynes: *A Treatise on Probability*, Macmillan, London, 1921.

[Linial 93] N. Linial, Y. Mansour and N. Nisan: Constant depth circuits, Fourier Transform, and Learnability, *Journal of the ACM*, Vol.40, No.3, pp.607-620, 1993.

[Quinlan 93] J.R. Quinlan: *C4.5: Programs for machine learning*, Morgan Kaufmann Pub., 1993.

[坂元 83] 坂元 慶行, 石黒真木夫, 北川 源四郎: 情報量統計学, 共立出版, 1983.

[高島 94] 高島 文次郎: 私信

[月本 90] 月本 洋: 命題論理の幾何的モデル, 情報処理学会論文誌, Vol.31, pp.783-791, 1990.

[月本 92] 月本 洋: 確率データからの帰納学習, 人工知能学会誌, Vol.7, No.5, pp.870-876, 1992.

[月本 93] 月本 洋: ブール関数の多項式時間学習アルゴリズム, 情報処理学会研究報告 93-AI-87, Vol.93, No.17, pp.105-114, 1993.