2E3-6

二値データに対するデータ生成過程の推定

Experimental evaluation of a method to estimate the data generating process of a binary variable causal model

稲積 孝紀*1 Takanori Inazumi　鷲尾 隆*1 Takashi Washio　清水 昌平*1 Shohei Shimizu　鈴木 譲*2 Joe Suzuki　山本 章博*3 Akihiro Yamamoto　河原 吉伸*1 Yoshinobu Kawahara

*1大阪大学 産業科学研究所 The Institute of Scientific and Industrial Research, Osaka University
*2大阪大学 大学院理学研究科 Graduate School of Science, Osaka University
*3京都大学 大学院情報学研究科 Graduate School of Informatics, Kyoto University

Estimation of the data generation process of a given data set is a major objective in studies of statistics and artificial intelligence. Recently, some non-Gaussianity based techniques to estimate a data generation process have been explored, but most of them are limited to continuous data. In this presentation, we review the performance of our new approach [Inazumi 11] to estimate a unique process generating binary data based on some characteristics of external noise through some numerical experiments for both artificial and real world data sets.

1. はじめに

統計学や人工知能の分野では，ベイジアンネットワークにおける因果推論や構造学習に関する手法が盛んに研究されている [Pearl 00, Spirtes 93]．これらの手法の多くは因果関係に非巡回性を仮定することで，観測データから因果構造の候補を推定する．その際，これらは主に観測変数の二次統計量までを用いることで，有向非巡回グラフ (DAG) の候補を絞り込む．しかし，これらの手法はしばしば同じ二値統計量を有するデータを生成する複数の候補を出力し，一意に因果構造を同定できない場合がある．

近年，非巡回モデルの推定にデータの非ガウス性を利用する LiNGAM 法などのいくつかの手法が提案されている [Shimizu 06, Shimizu 11]．これらの手法は観測変数に作用する外乱ノイズの非ガウス性を利用することで，モデルの仮定が正しければ，データの非ガウス性を最もよく説明する因果構造を一意に同定できることが知られている．しかしこれらの手法は連続変数のみを対象としており，対象となるシステムが線形関係によって表されることを必要とする．最近では，この原理の非線形モデルへの拡張 [Hoyer 09] や，順序尺度を持つ離散変数への拡張 [Peters 10] が行われている．しかし，これらの研究は外乱ノイズの加法性を仮定するため，データに含まれる変数が 3 個以上の場合には適用できず，また一般のカテゴリ変数を対象としていない．

一方，コンピュータネットワーク，薬学 [Pearl 00]，バイオインフォマティクス，社会学 [Spirtes 93] などの実世界の分野では，近年確率的二値データの蓄積が進んでおり，さまざまな用途においてこれらの二値データからデータ生成過程を推定したいというニーズが高まっている．しかし我々が知る限り，既存の手法では任意の二値データに対して一意に生成過程を推定することを保証できない．そこで本稿では，与えられた確率的二値データセットのデータ生成過程を一意に推定する我々の新しい手法 [Inazumi 11] を説明し，いくつかの人工データおよび実世界データについてその性能を検証した結果を報告する．

連絡先: 稲積 孝紀，大阪大学 産業科学研究所，567-0047 大阪府茨木市美穂ヶ丘 8-1，inazumi@ar.sanken.osaka-u.ac.jp

2. BExSAM モデルとその推定法

この節では，与えられた二値データに対してデータ生成過程の推定を行う我々の新しい手法 [Inazumi 11] について説明する．まず 2.1 項で，データの生成過程を表すモデルの定義を述べる．次に 2.2 項で，モデルの推定に用いられるモデルの性質を述べる．そして 2.3 項で，モデルの性質に基づく推定アルゴリズムについて述べる．

2.1 モデルの定義

最初に，二値変数間の非巡回因果関係を分析するための新しいモデルの定義について説明する．このモデルは有向非巡回グラフ (DAG) で表される構造を持つ．また，連続値データを対象とする多くのモデルにおいて外乱ノイズは対応する観測変数に加法演算によって作用するが，このモデルでは外乱ノイズは排他的論理和演算によって作用する．

定義 1. d 個の二値観測変数の集合 $X = \{x_i | x_i \in \{0,1\}, i = 1,\ldots,d\}$ を考える．また，d 個の二値外乱ノイズの集合を $E = \{e_i | e_i \in \{0,1\}, i = 1,\ldots,d\}$ とする．ここで，各 e_i は互いに独立な確率変数とする．このとき，添字 i は以下の式が成り立つような置換 $i : \{1,\ldots,d\} \to \{1,\ldots,d\}$ で表される．

$$x_{i(1)} = f_{i(1)} \oplus e_{i(1)},\ \text{かつ}$$
$$x_{i(k)} = f_{i(k)}(x_{i(1)},\ldots,x_{i(k-1)}) \oplus e_{i(k)}$$
$$\text{for } k = 2,\ldots,d.$$

ここで，$f_{i(1)}$ は $\{0,1\}$ のいずれかを取る定数であり，$f_{i(k)} : \{0,1\}^{k-1} \to \{0,1\}$ $(k \geq 2)$ は決定的ブール関数，$\oplus$ は排他的論理和演算子である．

複数の二値変数間における任意の決定的関係は決定的ブール代数式 [Gregg 98] によって表されるため，$f_{i(k)}$ は一般性を失うことなく任意の決定的二値関係を表すことができる．以降，記述を簡潔にするため定義 1 における $i(k)$ を i，$f_{i(k)}(x_{i(1)},\ldots,x_{i(k-1)})$ を f_i で表す．

さらに外乱ノイズの確率分布は，以下のような偏りを有するベルヌーイ分布とする．

定義 2. 外乱ノイズ e_i について $e_i = 1$ となる確率を $p(e_i = 1)$ と表す．すべての $i = 1, \dots, d$ について，$p(e_i = 1)$ は $0 < p_i < 1$ かつ $p_i \neq 1/2$ を満たす定数 p_i と等しい．

$e_i = 0$ となる確率についても，$p(e_i = 0) = 1 - p_i$ は同じ制約 $0 < p(e_i = 0) < 1$ かつ $p(e_i = 0) \neq 1/2$ を満たす．この外乱ノイズの性質は連続値の場合における非ガウス性に対応し，モデルの推定を行うために原理的に必要となる性質である．定義 1 および 定義 2 に基づくモデルを Binary Exclusive-or Skew Acyclic Model，略して **BExSAM** と呼ぶ．[Inazumi 11] では，このモデルにおいて与えられた二値データセットの分布を最も良く説明する因果構造を一意に推定する手法を提案している．

DAG モデルにおいて，変数 x_j から別の変数 $x_i (i \neq j)$ への有向辺が存在するとき x_j を x_i の **親**，x_i を x_j の **子** と呼ぶ．因果推論の研究 [Pearl 00, Spirtes 93] で広く用いられるように，DAG モデルにおいて親を持たない変数を **外生変数**，それ以外の変数を **内生変数** と呼ぶ．[Inazumi 11] では，さらに特定の内生変数に関して次のような定義を用いる．

定義 3. DAG モデルにおいて子を持たない内生変数を **最終内生変数** と呼ぶ．

例 1. $i(k) = k$ である場合の 4 変数からなる BExSAM モデルの例を次に示す．

$$
\begin{aligned}
x_1 &= e_1, \\
x_2 &= x_1 \oplus e_2, \\
x_3 &= x_1 x_2 \oplus e_3, \\
x_4 &= (x_1 + x_3) \oplus e_4.
\end{aligned}
$$

ここでは記述を簡潔にするため x_1 と x_2 の論理積演算を $x_1 x_2$，x_1 と x_3 の論理和演算を $x_1 + x_3$ と表す．x_1 は外生変数，x_4 は最終内生変数である．

2.2 モデルの性質

次に，最終内生変数とその親となる変数に関する BExSAM モデルの性質について説明する．まず，観測変数の集合 X のうちいくつかの観測変数の値を固定する **統制** と呼ばれる操作を定義する．

定義 4. 観測変数の集合 X から x_i を除いた部分集合 $X_i = X \setminus \{x_i\}$ を考える．部分集合 X_i に含まれる各変数 x_h の値をそれぞれ v_h としたとき，集合 $\{v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_d\}$ を X_i の値集合 V_i と呼ぶ．この X_i と V_i の対応付けを V_i による X_i の **統制** と呼び，$X_i = V_i$ で表す．同じように，$X_{ij} = X \setminus \{x_i, x_j\}$ を考える．V_{ij} による X_{ij} の **統制** を同様に定義し，$X_{ij} = V_{ij}$ で表す．

最終内生変数の条件付き確率分布に関する重要な定理を以下に掲げる．

定理 1. 次の二つの条件は同値である．

1. 変数 $x_i \in X$ は最終内生変数である．
2. 変数集合 $X_i = X \setminus \{x_i\}$ のすべての統制 $X_i = V_i \in \{0,1\}^{d-1}$ に亘って，$p(x_i = 1 | X_i = V_i) = q_i$ あるいは $1 - q_i$ となるような共通の定数 q_i が存在する（したがって，$p(x_i = 0 | X_i = V_i) = 1 - q_i$ あるいは q_i である）．

次に，最終内生変数の親変数の推定に関する重要な命題を以下に掲げる．

命題 1. 変数 $x_i \in X$ を最終内生変数とする．他の変数 $x_j \in X$ $(j \neq i)$ について，$X_{ij} = X \setminus \{x_i, x_j\}$ の統制を $X_{ij} = V_{ij}$ とする．このとき，$V_{ij} \in \{0,1\}^{d-2}$ において

1. x_j が x_i の親である． $\Leftrightarrow$
   少なくとも一つの V_{ij} について，$p(x_i = v_i | x_j = v_j, X_{ij} = V_{ij}) = p(x_i = \bar{v}_i | x_j = \bar{v}_j, X_{ij} = V_{ij})$ が成り立つ．
2. x_j が x_i の親でない．$\Leftrightarrow$
   任意の V_{ij} について，$p(x_i = v_i | x_j = v_j, X_{ij} = V_{ij}) = p(x_i = v_i | x_j = \bar{v}_j, X_{ij} = V_{ij})$ が成り立つ．

ここで，$v_* \in \{0,1\}$，$\bar{v}_* = v_* \oplus 1$ である．

2.3 モデルの推定アルゴリズム

二値データセットに対してデータ生成過程の推定を行うアルゴリズムについて説明する．このアルゴリズムは，与えられた二値データセット $D = \{V^{(h)} | h = 1, \dots, n\}$ の分布を最もよく説明するデータ生成過程を一意に推定する．ここで，各 $V^{(h)}$ は観測変数 X に対応する値を持つ d 次元ベクトルである．データセット D は BExSAM モデルによって生成されていると仮定する．すなわち，$V^{(h)}$ のすべての要素 $v_i^{(h)}$ は定義 1 の式 $x_i = f_i \oplus e_i$ によって生成され，e_i の分布は互いに独立かつ定義 2 の制約に従うものとする．この仮定より，サンプルサイズ n が 2^d よりも十分に大きければ，D はあらゆるベクトル $V^{(h)}$ を含み，あらゆる統制における条件付き確率を推定することができるであろう．第 3 節の数値実験で示されるように，D に含まれる観測変数の数が中程度までであれば D のサイズは現実的である．

推定アルゴリズムの全体の流れを以下に述べる．まず，データセット D の度数表 FT を計算する．これ以降アルゴリズムは度数表 FT のみを用いるため，データセットとして FT が提供されている場合はアルゴリズムを直接 FT に適用することができる．次に，最終内生変数 $x_{i(k)}$ を推定し，さらに発見した $x_{i(k)}$ の親変数集合 $P_{i(k)}$ を推定する．そして，発見した $x_{i(k)}$ をモデルから取り除き，新しい入力モデルとする．この操作は度数表 FT を $x_{i(k)}$ で周辺化することによって行われる．以降，新しい入力モデルに対して最終内生変数とその親変数集合を見つける手順を繰り返すことで，BExSAM モデルの全ての構造 $[\{x_{i(k)}, P_{i(k)}\} | k = 1, \dots, d]$ を得る．この反復的な変数削減は，DirectLiNGAM アルゴリズム [Shimizu 11] の構造と似ている．ただし，DirectLiNGAM アルゴリズムが外生変数，すなわち因果的順序が最上位となる変数から取り除くのに対し，このアルゴリズムは最終内生変数，すなわち因果的順序が最下位となる変数から取り除く．

最終内生変数の推定アルゴリズムについて説明する．定理 1 の条件 2 より，変数 x_i が最終内生変数のとき確率 $p(x_i = 1 | X_i = V_i)$ は統制に応じて q_i と $1 - q_i$ のいずれかの値を取る．一方，エントロピー $H = -q_i \log q_i - (1 - q_i) \log(1 - q_i)$ は q_i と $1 - q_i$ の交換に対して不変であるため，この値は統制によらず一定となる．この性質に注目し，各統制 $X_i = V_i$ について確率 $p(x_i = 1 | X_i = V_i)$ に対する条件付きエントロピー $H(x_i | X_i = V_i)$ の値を計算し，これらの値の分散を独立性を表すスコア $S_B(x_i)$ とする．

$$S_B(x_i) = E[(H(x_i | X_i = V_i) - E[H(x_i | X_i = V_i)])^2]$$

ここで，期待値 E は各 $H(x_i | X_i = V_i)$ を度数表 FT における $X_i = V_i$ の度数で重み付けて計算される．変数 x_i の条件付きエントロピー $H(x_i | X_i = V_i)$ が統制 $X_i = V_i$ に対して完全に独立であるとき，スコア $S_B(x_i)$ は最小値ゼロを取る．これよ

り，最終内生変数の推定はすべての x_i についてスコア $S_B(x_i)$ を計算し，最小の $S_B(x_i)$ を取る x_i を選ぶことによって行われる．

次に，親変数の推定アルゴリズムについて説明する．変数 x_i を最終内生変数としたとき，別の変数 $x_j (j \neq i)$ について，度数 $FT(x_i = 1, x_j = v_j, X_{ij} = V_{ij})$ は二項分布 $B(FT(x_j = v_j, X_{ij} = V_{ij}), p(x_i = 1|x_j = v_j, X_{ij} = V_{ij}))$ に従う．これより $\delta p(X_{ij} = V_{ij}) := p(x_i = 1|x_j = v_j, X_{ij} = V_{ij}) - p(x_i = 1|x_j = \bar{v}_j, X_{ij} = V_{ij})$ の分散 $\sigma^2(X_{ij} = V_{ij})$ を容易に導くことができる．そしてこの分散より，命題 1 の 2 の条件式 $p(x_i = 1|x_j = v_j, X_{ij} = V_{ij}) = p(x_i = 1|x_j = \bar{v}_j, X_{ij} = V_{ij})$ を帰無仮説とした場合の P 値を計算することができる．この帰無仮説は，統制 $X_{ij} = V_{ij}$ のもとで $x_i = 1$ となる確率が $x_j = v_j$ に対して独立であることを表している．親変数の推定はすべての統制 $V_{ij} \in \{0,1\}^{d-2}$ についてこの帰無仮説を検定することで行われる．これは同時に複数の帰無仮説を検定するため，多重検定の問題がある．そこで，[Inazumi 11] では [Benjamini 95] の方法を用いて False Discovery Rate (FDR) を制御する．この検定で用いる有意水準 α はこのアルゴリズム全体における唯一のパラメータである．

推定アルゴリズム全体の計算量について説明する．変数が d 個のときサンプルサイズ $n > 2^d$ との仮定および度数表 FT のサイズが 2^d であることより，このアルゴリズムの空間計算量は $O(n)$ となる．また，時間計算量は $O(d^2 n)$ となる．従来手法のうち最も効率のよい手法の一つである DirectLiNGAM アルゴリズム [Shimizu 11] の時間計算量が $O(d^3 n)$ であることを考えると，このアルゴリズムの時間計算量は従来のアルゴリズムとの比較において優れているといえる．

3. 評価実験

この節では，第 2 節で説明した手法を人工データと実世界データの両方に適用した結果 [Inazumi 11] について報告する．まず 3.1 項で，手法を人工データに適用した結果を述べる．次に 3.2 項で，手法を実世界データに適用した結果を述べる．

3.1 人工データ

2.1 項で定義した BExSAM モデルに従い人工的に生成したデータセットを用いて行った数値実験の結果を示す．ここで，各 $f_{i(k)}$ は以下の Algebraic Normal Form (ANF) によって表すことができる．

$$\begin{aligned} f_{i(k)} = a_0^k \oplus & \\ & a_1^k x_{i(1)} \oplus a_2^k x_{i(2)} \oplus \cdots \oplus a_d^k x_{i(d)} \oplus \\ & a_{1,2}^k x_{i(1)} x_{i(2)} \oplus \cdots \oplus a_{k-1,k}^k x_{i(k-1)} x_{i(k)} \oplus \\ & \cdots \oplus \\ & a_{1,2,\ldots,k}^k x_{i(1)} x_{i(2)} \ldots x_{i(k)} \oplus e_{i(k)}. \end{aligned}$$

各係数 a_*^k は $\{0,1\}$ のいずれかを取る定数である．データセットの生成手順を以下に述べる．まず，パラメータ d, n, $p(a_*^k = 1)$ のもとで，順列 $i : \{1,\ldots,d\} \to \{1,\ldots,d\}$ をランダムに決定し，さらにすべての係数 $a_*^k (k = 1,\ldots,d)$ をランダムに生成した．任意のブール関数 $f_{i(k)}$ は適切な係数 a_*^k をもつ ANF で表せるため [Carlet 19]，ランダムに生成した a_*^k の集合は一般的な BExSAM モデルを生成する．次に，パラメータ $p(e_{i(k)} = 1)$ $(k = 1,\ldots,d)$ のもとで，外乱ノイズ $e_{i(1)}^{(h)}, \ldots, e_{i(d)}^{(h)}$ $(h = 1,\ldots,n)$ をランダムに生成した．そして，生成した BExSAM モデルに従いデータセット $D = \{V^{(h)}|h = 1,\ldots,n\}$ を生成した．他の制約ベースの手法 [Spirtes 93] 同様，[Inazumi 11] の手法は 2.3 項で述べたように X のすべての値の組み合わせのパターンを含む完全なデータセットを原理的に必要とする．このため，生成されたデータセット D が不完全であった場合はこれを除外し，完全なデータセットのみを評価の対象とした．

表 1　n および d の各組み合わせにおける性能．ER_o は因果的順序の推定精度，ER_s は因果構造全体の推定精度を表し，CT は計算時間（ミリ秒）である．

$n \backslash d$		2	4	6	8
100	ER_o	0.04	0.28	0.42	-
	ER_s	0.053	0.25	0.37	-
	CT	0.96	2.6	7.1	-
500	ER_o	0.004	0.10	0.29	0.18
	ER_s	0.011	0.13	0.29	0.23
	CT	0.90	2.7	7.9	27.0
1000	ER_o	0.001	0.042	0.26	0.40
	ER_s	0.01	0.064	0.27	0.39
	CT	0.89	2.8	8.3	32.0
5000	ER_o	0.00	0.0033	0.12	0.27
	ER_s	0.0078	0.012	0.13	0.27
	CT	1.10	3.4	9.9	35.0
10000	ER_o	0.00	0.0013	0.075	0.24
	ER_s	0.0082	0.0078	0.081	0.24
	CT	1.2	4.2	11.0	40.0

評価には以下に述べる 3 つの性能指標を用いた．生成した BExSAM モデルの真の DAG 構造を表す隣接行列 B を考える．次に，2.3 項のアルゴリズムによって推定された変数の因果的順序 $[x_{i(k)}, k = 1,\ldots,d]$ に従い，隣接行列 B の行と列を同時に並び換える．一つ目の指標は，この並び換えられた B の厳密な上三角部分に含まれる非ゼロの要素の割合として定義されるエラー率 $0 \leq ER_o \leq 1$ である．真の B は厳密な下三角行列であるため，エラー率 ER_o は因果的順序の推定精度を表す．二つ目の指標は，B のすべての要素について真の B と推定された $\hat{B}$ の間の不一致要素の割合によって定義されるエラー率 $0 \leq ER_s \leq 1$ である．$\hat{B}$ は 2.3 項のアルゴリズムの出力 $[\{x_{i(k)}, P_{i(k)}\}|k = 1,\ldots,d]$ から得られる．ER_s は因果的順序を含む因果構造全体の推定の精度を表す．三つ目の指標は，2.3 項のアルゴリズムの計算時間 CT（ミリ秒）である．

[Inazumi 11] では，すべての $k = 1,\ldots,d$ についてパラメータ $0 \leq p(a_*^k = 1) \leq 1$ と $p(e_{i(k)} = 1) \in \{0.1, 0.2, 0.3, 0.4, 0.6, 0.7, 0.8, 0.9\}$ をランダムに選択し，変数の数 $d = 2, 4, 6, 8$, サンプルサイズ $n = 100, 500, 1000, 5000, 10000$ のすべての組み合わせに対し 1000 回の試行を行った．このパラメータ $p(e_{i(k)} = 1)$ の選択は，外乱ノイズ $e_{i(k)}$ の分布が定義 2 の制約を満たすことを保証する．また，パラメータ $p(a_*^k = 1)$ は生成された BExSAM モデルにおける変数の組み合わせの密度を表す．1000 回の試行における三つの指標の平均値を表 1 に示す．$n = 100$ かつ $d = 8$ の場合は n が D の完全性の点において不十分であったため，各指標を得ることができなかった．$n \geq 1000$ かつ $d \leq 4$ の場合，エラー率 ER_o，ER_s はどちらも 0.05 より小さいかほぼ同じであった．この結果は変数が 4 個以下であれば，1000 程度のデータサイズで我々の手法がうまく働くことを示している．一

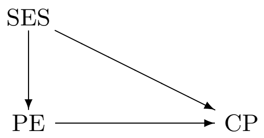

図 1　CP, PE, SES の間に発見された因果関係.

方, $d \geq 6$ の場合は 10000 以上の多くのデータを用意する必要がある. また, CT は主に d の増加に合わせて大きくなる. これは 2.3 項で説明した時間計算量と一致しているが, $d = 8$ かつ $n = 10000$ の場合でも必要な計算時間 CT は十分に小さく, 実用上の問題とはならないといえる.

3.2　実世界データ

[Inazumi 11] では二つの実世界データに対して手法の適用を行った. 一つ目はネバダ州での核実験での放射性落下物による放射線被爆の高低 ($EX = 1/0$) と, ユタ州南部の子供の白血病による生死 ($LE = 1/0$) の関係のデータセットである [Finkelstein 90, Pearl 00]. このデータセットは二つの二値変数しか含まないので, 標準的な制約ベース [Spirtes 93], スコアベース [Chickering 02] の手法は一意に因果構造を推定することができない. 一方, [Inazumi 11] の手法は有意水準 $\alpha = 0.05$ の検定のもとでは EX と LE の間に因果関係を検出できなかったが, $\alpha = 0.15$ の検定のもとでは因果の向き $EX \rightarrow LE$ を検出することができた. この結果は直感と一致しており, この二つの変数の間に弱い因果関係がある可能性が示唆された.

二つ目はウィスコンシン州の高校生 10318 人の大学進学意思に関するデータセットである [Sewell 68, Spirtes 93]. [Spirtes 93] では制約ベースの手法を用いて Y 構造 [Mani 06] を持つ 5 変数間の因果構造を発見することを目的としているが, 我々は社会的要素を表す 3 つの変数, すなわち性別 ($SEX = 0/1$) と知能指数の高低 ($IQ = 0, \ldots, 3$) を除いた大学進学意思の有無 ($CP = 0/1$), 親の働き掛けの高低 ($PE = 0/1$), 社会的ステータスの高低 ($SES = 0, \ldots, 3$) の 3 つの変数間の因果構造に注目した. 標準的な制約ベースの手法は同値なクラスに属する複数の因果構造の候補を出力することが知られている. ここではサンプルのサイズを維持しつつ個々のサンプルの特性を調整するために, 比較的高い IQ を持つ男子高校生 3756 人 ($SEX = 0$, $IQ \neq 0$) を選択した. このとき, $SES = 0$ のサンプルでは $p(CP = 0) = 0.21$ であるのに対し, $SES \neq 0$ のサンプルでは $p(CP = 0) = 0.51$ である. すなわち SES が最も低いサンプルは大学進学意思を持つ確率が特に低い. この観察に基づき, 元々 $SES = 0$ である場合は 0, $SES \neq 0$ の場合は 1 となるように SES を二値化した. 以上より得た 3 個の二値変数 CP, PE, SES からなるデータセットに対し有意水準 $\alpha = 0.05$ で [Inazumi 11] の手法を適用した結果, 図 1 に示すような因果構造を一意に得ることができた. この結果は, 親の社会的ステータスが親の働き掛けおよび高校生の大学進学意思に影響し, さらに親の働き掛けが大学進学意思に影響するということを示している. これは直感と一致し, 加えて [Sewell 68] で述べられている仮説とも一致する.

4.　まとめ

本稿では, 与えられた確率的二値データセットのデータ生成過程を一意に推定する我々の新しい手法 [Inazumi 11] を説明し, いくつかの人工データおよび実世界データについてその性能を検証した結果を報告した. この手法の計算複雑性は小さく, この手法が必要とするパラメータは 1 つのみである. そして評価実験の結果を通して, この手法が人工データと実世界データの両方について望ましい性能を示すことを述べた.

参考文献

[Benjamini 95] Benjamini, Y. and Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. Royal Statistical Society: Series B*, Vol. 57, pp. 289–300 (1995)

[Carlet 19] Carlet, C. and Guillot, P.: A new representation of Boolean functions, *Proc. of 13th AAECC, Lecture Notes in Computer Science*, pp. 94–103 (1719)

[Chickering 02] Chickering, D.: Optimal structure identification with greedy search, *Journal of Machine Learning Research*, Vol. 3, pp. 507–554 (2002)

[Finkelstein 90] Finkelstein, M. and Levin, B.: *Statistics for lawyers*, Springer-Verlag (1990)

[Gregg 98] Gregg, J.: *Ones and zeros: understanding Boolean algebra, digital circuits, and the logic of sets*, chapter 5, pp. 101–123, Hoboken, NJ. John Wiley & Sons (1998)

[Hoyer 09] Hoyer, P., Janzing, D., Mooij, J., Peters, J., and Schölkopf, B.: Nonlinear causal discovery with additive noise models, *Proc. of NIPS2008: In Advances in Neural Information Processing Systems*, Vol. 21, pp. 689–696 (2009)

[Inazumi 11] Inazumi, T., Washio, T., Shimizu, S., Suzuki, J., Yamamoto, A., and Kawahara, Y.: Discovering causal structures in binary exclusive-or skew acyclic models (2011), 投稿中.

[Mani 06] Mani, S., Cooper, G., and Spirtes, P.: A theoretical study of Y structures for causal discovery, in *Proc. of the 22nd Conference in Uncertainty in Artificial Intelligence*, pp. 314–323 (2006)

[Pearl 00] Pearl, J.: *Causality: models, reasoning, and inference*, Cambridge University Press (2000)

[Peters 10] Peters, J., Janzing, D., and Schölkopf, B.: Identifying cause and effect on discrete data using additive noise models, in *Proc. of AISTATS2010: the 13th International Conference on Artificial Intelligence and Statistics (JMLR Workshop and Conference Proceedings)*, Vol. 9, pp. 597–604 (2010)

[Sewell 68] Sewell, W. and Shah, V.: Social class, parental encouragement, and educational aspirations, *American Journal of Sociology*, Vol. 73, pp. 559–572 (1968)

[Shimizu 06] Shimizu, S., Hoyer, P. O., Hyvärinen, A., and Kerminen, A.: A linear non-gaussian acyclic model for causal discovery, *Journal of Machine Learning Research*, Vol. 7, pp. 2003–2030 (2006)

[Shimizu 11] Shimizu, S., Inazumi, T., Sogawa, Y., Hyvarinen, A., Kawahara, Y., Washio, T., Hoyer, P. O., and Bollen, K.: DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model, *Journal of Machine Learning Research*, Vol. 12, pp. 1225–1248 (2011)

[Spirtes 93] Spirtes, P., Glymour, C., and Scheines, R.: *Causation, prediction, and search*, Springer-Verlag (1993), (2nd ed. MIT Press 2000)