

THE IEICE TRANSACTIONS ON INFORMATION AND SYSTEMS (JAPANESE EDITION)

# **EIC** | **電子情報通信学会** **D** | **論文誌**

情報・システム

DOI:10.14923/transinfj.2019JDP7011

早期公開日:2020/01/06

本PDFは、早期公開版である。本論文を引用する場合には、電子情報通信学会和文論文誌投稿のしおり(情報・システムソサイエティ)の「8.早期公開」を参照すること。

**情報・システムソサイエティ**

一般社団法人 **電子情報通信学会**

THE INFORMATION AND SYSTEMS SOCIETY

THE INSTITUTE OF ELECTRONICS, INFORMATION AND COMMUNICATION ENGINEERS

# Augmented Naive Bayes 制約を持つベイジアンネットワーク分類器の 厳密学習

菅原 聖太<sup>†a)</sup> 植野 真臣<sup>†b)</sup>

Exact Learning Bayesian Network Classifier with Augmented Naive Bayes structure constraint

Shouta SUGAHARA<sup>†a)</sup> and Maomi UENO<sup>†b)</sup>

あらまし ベイジアンネットワーク分類器 (Bayesian Network Classifier: BNC) は離散変数を扱う高精度の分類器として知られている。BNC は生成モデルよりも識別モデルの方が分類精度が高いことが先行研究で報告されてきた。しかし、なぜ識別モデルの方が分類精度が高いかという理由については未だ明らかにされていない。本論では、まず識別モデルとしての BNC と厳密に学習した生成モデルとしての BNC の分類精度の比較実験を行なった。その結果、生成モデルはサンプルサイズが大きい時は識別モデルより分類精度が高かった。一方で、サンプルサイズが小さい場合は、生成モデルでは目的変数の親変数が増加してしまい、パラメータ数の指数的増加によって分類精度が識別モデルよりも著しく低下していた。そこで、本論では、目的変数が親変数を持たない Augmented Naive Bayes (ANB) 構造を制約とした、生成モデルとしての BNC の厳密学習手法を提案する。生成モデルとしての ANB の説明変数は ANB 制約のない BNC における目的変数のマルコフブランケットに一致していると考えられる。そこで提案手法では、まず制約なしで BNC を厳密学習する。次に学習された BNC の目的変数のマルコフブランケットのみを説明変数集合とした ANB 構造を制約として BNC を学習する。リポジトリデータセットを用いた実験により、提案手法が従来の識別モデルの BNC よりも分類精度が有意に高いことを示す。

キーワード ベイジアンネットワーク, 分類器, 生成モデル, 機械学習, 確率的グラフィカルモデル

## 1. ま え が き

ベイジアンネットワークは、離散確率変数をノードとし、ノード間の条件付き従属関係を非循環有向グラフ (Directed Acyclic Graph: DAG) で表し、同時確率分布を各ノードの親ノード集合を所与とした条件付き確率パラメータの積に分解する、確率的グラフィカルモデルである。ベイジアンネットワークにおける一つのノードを目的変数とし、その他のノードを説明変数としたベイジアンネットワーク分類器 (Bayesian Network Classifier: BNC) は、離散変数を扱う分類器として知られている [1]。

一般にベイジアンネットワークの DAG 構造はデータから推定する必要がある、この問題をベイジアンネットワークの構造学習と呼ぶ。構造学習では、候補構造から最適な学習スコアを持つ構造を探索するスコアベースアプローチが従来から行われてきた。本論では候補構造に制約を課さずに学習した BNC を General Bayesian Network (GBN) と呼ぶ。一般にスコアベースアプローチでは、漸近一致性を有する、構造の周辺尤度 (Marginal Likelihood: ML) を学習スコアとして用いる。

ML を用いると、全変数の同時確率分布をモデル化する生成モデルとして BNC を学習できる。しかし、Friedman ら [1] は、BNC の構造学習スコアとして、生成モデルではなく、説明変数を所与とした目的変数の条件付き確率分布をモデル化する識別モデルのためのスコアを用いるべきだと主張した。そのような学習スコアとして、説明変数を所与とした目的変数の条件

<sup>†</sup> 電気通信大学大学院情報理工学研究所, 調布市  
Graduate school of Informatics and Engineering, The  
University of Electro-Communications, 1-5-1 Chofugaoka,  
Chofu-shi, 182-8585, Japan

a) E-mail: sugahara@ai.lab.uec.ac.jp

b) E-mail: ueno@ai.lab.uec.ac.jp

付き対数尤度 (Conditional Log Likelihood: CLL) が提案された。しかし、CLL を最大にするパラメータ推定式は閉形式で表せないため、構造の探索に効率的なアルゴリズムを用いることができず、学習時間が膨大になってしまう。これを解決するため、Carvalho ら [2], [3] は構造探索も効率にできるよう CLL を線形近似した approximate CLL (aCLL) を提案した。Grossman ら [4] は CLL をスコアとして、貪欲法の Hill-Climbing アルゴリズム [5] を用いて構造を探索する手法を提案した。これらの近似手法で学習した BNC の方が、ML で学習した BNC よりも分類精度が高いことが報告されている。

しかし、ML 最大化より CLL 最大化の方がなぜ良いかという理由については未だ明らかにされていない。ML は推定構造に対して漸近一致性が保証されており、サンプルサイズが大きい時に ML の分類精度が CLL に劣るのは奇異である。また、BNC の ML は閉形式で表せるため CLL より計算効率がよく、ML を大域的に最大化する構造を探索する厳密学習を効率的に行える。先行研究の比較実験では、ML を局所的に最大化する構造を探索する近似学習を行なっているため、探索精度の悪さが影響したのかもしれない。

そこで、本研究ではまず ML による厳密学習と CLL による近似学習によって得られた BNC の分類精度を比較する。結果として、ML 最大化による BNC は厳密学習することで、大きく精度が向上することがわかった。特にサンプルサイズが大きいときに、最も分類精度が高いことが示された。しかし、厳密学習ではサンプルサイズが小さくなると ML を最大化する BNC の分類精度が低くなり、最も単純な構造をもつ Naive Bayes よりも低い場合もあった。特に、目的変数の親変数が多く子変数が少ないような構造を学習する場合に分類精度が低くなっていることがわかった。その理由は、目的変数の親変数が多いと、パラメータ数が指数的に増えるため、一つのパラメータ学習のためのサンプルサイズが小さくなり、推定精度が悪くなってしまふからである。

この問題を緩和するため、本論では、目的変数が親変数を持たず、説明変数が必ず目的変数の子となる Augmented Naive Bayes (ANB) 構造を制約とした BNC の厳密学習を提案する。ANB はこれまで識別モデルとして扱われてきたため、ML を最大化して学習することはなかった。本論の提案は、識別モデルの学習ではなく、生成モデルとしての GBN の学習に、目

的変数の親変数が増えないように ANB 構造を制約とすることである。生成モデルとしての ANB の説明変数は、生成モデルとしての GBN における目的変数のマルコフブランケットに一致していると考えられる。そこで提案手法では、まず ML を用いて GBN を厳密学習する。次に学習された GBN の目的変数のマルコフブランケットのみを説明変数集合とした ANB 構造を制約として GBN を学習する。

ベンチマークデータによる比較実験で、提案手法の分類精度が従来手法よりも有意に高いことを示した。

## 2. ベイジアンネットワーク

### 2.1 モデル

ベイジアンネットワークは、確率変数をノードとし、ノード間の条件付き従属関係を非循環有向グラフで表し、各ノードの親ノード集合を所与とした条件付き確率で表現される確率的グラフィカルモデルである。今、離散確率変数集合  $\{X_0, X_1, \dots, X_i, \dots, X_n\}$  において、各変数  $X_i$  は  $r_i$  個の状態集合  $\{1, \dots, r_i\}$  から一つの値をとるとし、各変数  $X_i$  が値  $k$  をとるとき、 $X_i = k$  と書く。また、変数  $X_i$  の親変数集合を  $G_i$  とし、ベイジアンネットワークの構造を  $G = (G_0, G_1, \dots, G_n)$  と定義する。さらに、 $\theta_{ijk}$  を  $G_i$  が  $j$  番目のパターンをとったとき ( $G_i = j$  と書く) に  $X_i = k$  となる条件付き確率  $P(X_i = k | G_i = j)$  を示すパラメータとし、 $\Theta_{ij} = \bigcup_{k=1}^{r_i} \{\theta_{ijk}\}$ 、 $\Theta = \bigcup_{i=0}^n \bigcup_{j=1}^{q_i} \{\Theta_{ij}\}$  とする。ここで、 $q_i = \prod_{l: X_l \in G_i} r_l$  である。ベイジアンネットワークでは、次式のように同時確率分布  $P(X_0, X_1, \dots, X_n | G, \Theta)$  を各変数の条件付き確率パラメータの積に分解して表せる。

$$P(X_0, X_1, \dots, X_n | G, \Theta) = \prod_{i=0}^n P(X_i | G_i, \Theta)$$

### 2.2 ベイジアンネットワークのパラメータ推定

今、サンプルが  $N$  個あり、各サンプルは独立で同一な分布に従うとする。 $t$  番目のサンプルを  $\mathbf{d}^t = \langle x_0^t, x_1^t, \dots, x_n^t \rangle$  と表し、学習データを  $D = \langle \mathbf{d}^1, \dots, \mathbf{d}^t, \dots, \mathbf{d}^N \rangle$  と表す。 $D$  が得られた時のベイジアンネットワーク  $\langle G, \Theta \rangle$  の尤度は以下で表される。

$$P(D | G, \Theta) = \prod_{t=1}^N P(x_0^t, x_1^t, \dots, x_n^t | G, \Theta)$$

$$= \prod_{i=0}^n \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} \theta_{ijk}^{N_{ijk}} \quad (1)$$

ここで,  $P(x_0^t, x_1^t, \dots, x_n^t \mid G, \Theta)$  は  $P(X_0 = x_0^t, X_1 = x_1^t, \dots, X_n = x_n^t \mid G, \Theta)$  を表し,  $N_{ijk}$  は  $X_i = k$  かつ  $G_i = j$  となる頻度を表す. 式 (1) の尤度を最大にする  $\theta_{ijk}$  の最尤推定量は以下で表される.

$$\hat{\theta}_{ijk} = \frac{N_{ijk}}{N_{ij}} \quad (2)$$

ここで,  $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$  である. 一般には, ベイジアンネットワークのパラメータ推定量として,  $\theta_{ijk}$  の期待値である Expected A Posteriori (EAP) が用いられる. ベイジアンネットワークの構造  $G$  に対し, 式 (3) のようにパラメータの事前分布にディリクレ分布を仮定すると, 式 (4) の事後分布  $p(\Theta_{ij} \mid D, G)$  が得られ, その事後分布から式 (5) のように EAP を求めることができる.

$$p(\Theta_{ij} \mid G) = \frac{\Gamma(\sum_{k=1}^{r_i} N'_{ijk})}{\prod_{k=1}^{r_i} \Gamma(N'_{ijk})} \prod_{k=1}^{r_i} \theta_{ijk}^{N'_{ijk}-1} \quad (3)$$

$$p(\Theta_{ij} \mid D, G) = \frac{\Gamma(\sum_{k=1}^{r_i} (N'_{ijk} + N_{ijk}))}{\prod_{k=1}^{r_i} \Gamma(N'_{ijk} + N_{ijk})} \prod_{k=1}^{r_i} \theta_{ijk}^{N'_{ijk} + N_{ijk} - 1} \quad (4)$$

$$\begin{aligned} \hat{\theta}_{ijk} &= \int \theta_{ijk} \cdot p(\Theta_{ij} \mid D, G) d\Theta_{ij} \\ &= \frac{N'_{ijk} + N_{ijk}}{N'_{ij} + N_{ij}} \end{aligned} \quad (5)$$

ここで,  $N'_{ijk}$  はディリクレ事前分布のハイパーパラメータを表し,  $N'_{ij} = \sum_{k=1}^{r_i} N'_{ijk}$  である.

### 2.3 ベイジアンネットワークの構造学習

ベイジアンネットワークのパラメータを推定するためには, 最適な構造をデータから推定する必要がある. この問題をベイジアンネットワークの構造学習と呼ぶ. 構造学習では, 候補構造から最適な学習スコアを持つ構造を探索するスコアベースアプローチが従来から行われてきた. 一般に学習スコアとして漸近一致性を有する周辺尤度  $P(D \mid G)$  (Marginal Likelihood: ML) が用いられ, ML を最大にする構造を最適解とする. パラメータの事前分布がディリクレ分布と仮定すると, ML は次のように閉形式で表される.

$$P(D \mid G) = \prod_{i=0}^n \prod_{j=1}^{q_i} \frac{\Gamma(N'_{ij})}{\Gamma(N'_{ij} + N_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(N'_{ijk} + N_{ijk})}{\Gamma(N'_{ijk})} \quad (6)$$

式 (6) の ML は Bayesian Dirichlet (BD) と呼ばれる. Heckerman ら [5] は, 同一の同時確率分布を表す構造は, それらの ML の値も同一でなければならないという尤度等価を導入した. そして, 尤度等価に矛盾しないディリクレ分布の条件として, 以下のハイパーパラメータを提案している.

$$N'_{ijk} = N' P(X_i = k, G_i = j \mid G^h)$$

ここで,  $N'$  は事前知識の重みを示す擬似サンプルである.  $G^h$  はユーザの仮説構造であり, この構造を所与として  $N'$  を  $N'_{ijk}$  に分配する. このスコアは, Bayesian Dirichlet equivalent (BDe) と呼ばれる. さらに,  $N'$  をパラメータ数で除し,  $N'_{ijk} = N' / (r_i \cdot q_i)$  としたスコアを提案している. このスコアは BDe の特殊形とみなすことができ, Bayesian Dirichlet equivalent uniform (BDeu) と呼ばれる. Heckerman ら [5] や Ueno [6] [7] [8] の研究では, 無情報事前分布を用いた BDeu が最も有用であると報告している.

一方,  $-\log ML$  の近似である最小記述長 (Minimum Description Length: MDL) [9] は, ベイジアンネットワークと学習データ  $D$  の同時記述長を表す.

$$\begin{aligned} MDL(D \mid G, \Theta) \\ = \frac{\log N}{2} \sum_{i=0}^n q_i (r_i - 1) - \log P(D \mid G, \Theta) \end{aligned} \quad (7)$$

MDL を用いた学習では, 式 (7) を最小にする構造を最適解とする. 式 (7) の第一項は構造の複雑さに対するペナルティ項である. 式 (7) の第二項は構造のデータへの当てはまりを反映するフィッティング項を表す対数尤度である.

$\log BDeu$  と MDL は, スコアとして次の性質を満たす.

$$Score(G) = \sum_{i=0}^n Score_i(G_i) \quad (8)$$

ここで,  $Score_i(G_i)$  は変数  $X_i$  とその親変数集合  $G_i$  のみに依存する関数であり, ローカルスコアと呼ぶ. 例えば  $\log BDeu$  の変数  $X_i$  と親変数集合  $G_i$  についてのローカルスコア  $Score_i(G_i)$  は以下のように表せる.

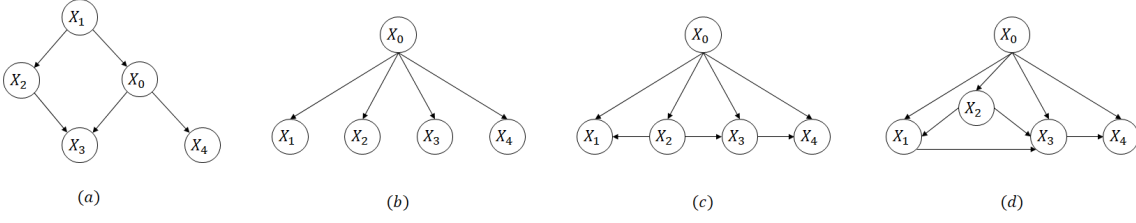


図 1 (a) GBN の例; (b) Naive Bayes; (c) TAN の例; (d) ANB の例  
Fig. 1 (a) Example of GBN; (b) Example of Naive Bayes; (c) Example of TAN; (d) Example of ANB.

$$Score_i(G_i) = \sum_{j=1}^{q_i} \left( \log \frac{\Gamma(N'_{ij})}{\Gamma(N'_{ij} + N_{ij})} + \sum_{k=1}^{r_i} \log \frac{\Gamma(N'_{ijk} + N_{ijk})}{\Gamma(N'_{ijk})} \right) \quad (9)$$

また、式 (8) を満たすスコアを分解可能であると言い、分解可能なスコアを用いると効率的に構造を探索できる [10] [11].

### 3. ベイジアンネットワーク分類器

#### 3.1 ベイジアンネットワーク分類器による分類

ベイジアンネットワークにおける一つのノードを目的変数とし、その他のノードを説明変数としたベイジアンネットワーク分類器 (Bayesian Network Classifier: BNC) は、離散変数を扱う分類器として知られている [1]. 今、 $X_1, \dots, X_n$  を説明変数とし、 $X_0$  を目的変数とした BNC を考える. 説明変数のデータ  $\langle x_1, \dots, x_n \rangle$  が与えられた時、目的変数の推定値  $\hat{c}$  は以下のように得られる.

$$\begin{aligned} \hat{c} &= \arg \max_{c \in \{1, \dots, r_0\}} P(c \mid x_1, \dots, x_n, G, \Theta) \quad (10) \\ &= \arg \max_{c \in \{1, \dots, r_0\}} \frac{P(c, x_1, \dots, x_n \mid G, \Theta)}{P(x_1, \dots, x_n \mid G, \Theta)} \\ &= \arg \max_{c \in \{1, \dots, r_0\}} P(c, x_1, \dots, x_n \mid G, \Theta) \\ &= \arg \max_{c \in \{1, \dots, r_0\}} \prod_{i=0}^n \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} (\theta_{ijk})^{1_{ijk}} \\ &= \arg \max_{c \in \{1, \dots, r_0\}} \left[ \prod_{j=1}^{q_0} \prod_{k=1}^{r_0} (\theta_{0jk})^{1_{0jk}} \right. \\ &\quad \left. \times \prod_{i: X_i \in \mathbf{Ch}} \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} (\theta_{ijk})^{1_{ijk}} \right] \end{aligned}$$

ここで、 $1_{ijk}$  は変数列  $\langle c, x_1, \dots, x_n \rangle$  において  $X_i = k$  かつ  $G_i = j$  の時に 1 をとり、それ以外の時は 0 をとる変数である. また、 $\mathbf{Ch}$  は目的変数の子変数の集合である. 式 (10) の最右辺からわかるように目的変数の分類に影響を及ぼす説明変数は、目的変数の親変数と子変数、および目的変数と子を共有する変数のみである. これらの変数集合をマルコフブランケットと呼ぶ.

#### 3.2 ベイジアンネットワーク分類器

一般に、ベイジアンネットワークの構造学習で探索する候補構造はとりうる全ての構造であり、そのような候補構造に対して BDeu や MDL など最適化して学習される BNC は General Bayesian Network (GBN) と呼ばれる (図 1 の (a)). つまり、分類器として用いられる、制約のない一般的なベイジアンネットワークを GBN と呼ぶ. 大きいネットワークでは GBN の学習に膨大な時間がかかってしまうため、候補構造に制約を入れて学習することが多い. 例えば、GBN の下位構造として、全説明変数が目的変数のみを親に持つと仮定する Naive Bayes [12] (図 1 の (b)) や、全説明変数が目的変数を親に持ち、説明変数間で木構造をとると仮定した Tree-Augmented Naive Bayes (TAN) [1] (図 1 の (c)) などが知られている. Naive Bayes の構造は一意に定まるため、構造学習の必要はない. 尤度を学習スコアとした TAN の学習は多項式時間で学習でき、MDL によって近似的に学習した GBN と同等の分類精度を持つことが数値実験により示されている [1] [13]. また、Naive Bayes や TAN を一般化した、より表現力の高い制約として、全説明変数が目的変数の子変数となり説明変数間関係には仮定をおかない Augmented Naive Bayes (ANB) (図 1 の (d)) [1] が知られている.

#### 3.3 ベイジアンネットワーク分類器の学習

BDeu や MDL で学習した BNC は、全変数の同時

確率分布をモデル化する生成モデルである。しかし、Friedman ら [1] は、BNC の構造学習には、説明変数を所与とした目的変数の条件付き確率分布をモデル化する識別モデルのためのスコアを用いるべきだと主張した。そのようなスコアとして、以下で表される、説明変数を所与とした目的変数の条件付き対数尤度 (Conditional Log Likelihood: CLL) が提案された。

$$\sum_{t=1}^N \log P(x_0^t | x_1^t, \dots, x_n^t, G, \Theta)$$

しかし、CLL は分解可能ではないため、効率的な構造探索アルゴリズムを用いることができない。そこで、Grossman ら [4] は近似的な構造探索法として、構造に対しエッジを一つ追加、消去、反転のいずれかの操作を行った時に最もスコアが良くなるようなエッジを選びその操作を行うというプロセスを繰り返して構造を更新する Hill-Climbing アルゴリズム [5] を用いた。Hill-Climbing アルゴリズムでは、任意のエッジの追加、消去、反転のどの操作を行ってもスコアが改善されない時に更新を終了し、その時の構造を解とする。

一方、Carvalho ら [3] は、候補構造集合に ANB を仮定し、分解可能となるよう CLL を近似した aCLL (approximate Conditional Log Likelihood) を提案した。ここで、 $t \in \{1, \dots, N\}$ ,  $c \in \{1, \dots, r_0\}$  に対して、 $J_{t,c} = P(c, x_1^t, \dots, x_n^t | G, \Theta)$  とすると、CLL は次のように表せる。

$$\sum_{t=1}^N \log P(x_0^t | x_1^t, \dots, x_n^t, G, \Theta) = \sum_{t=1}^N f(J_{t,1}, \dots, J_{t,r_0})$$

ただし、

$$f(J_{t,1}, \dots, J_{t,r_0}) = \log J_{t,x_0^t} - \log \left( \sum_{c=1}^{r_0} J_{t,c} \right)$$

である。今、 $(J_{t,1}, \dots, J_{t,r_0})$  が対称ディリクレ分布に従うことを仮定すると、 $f$  を以下の  $\hat{f}$  で近似できる。

$$\hat{f}(J_{t,1}, \dots, J_{t,r_0}) = \log J_{t,x_0^t} + \sum_{c=1}^{r_0} \beta \log J_{t,c} + \gamma$$

ここで、 $\beta$  と  $\gamma$  は、2 点  $A = -\log(\sum_{c=1}^{r_0} J_{t,c})$  と  $B = \sum_{c=1}^{r_0} \log J_{t,c}$  をランダムに多数発生させ、 $A = \beta B + \gamma$  を満たすように推定するパラメータである。Carvalho らはこの近似を用いて以下の aCLL スコアを提案した。

$$\begin{aligned} aCLL(D | G) &= \sum_{t=1}^N \hat{f}(J_{t,1}, \dots, J_{t,r_0}) \\ &= \sum_{t=1}^N \left( \log J_{t,x_0^t} + \sum_{c=1}^{r_0} \beta \log J_{t,c} + \gamma \right) \\ &\propto \sum_{i=1}^n \sum_{j=1}^{q_i^*} \sum_{k=1}^{r_i} \sum_{c=1}^{r_0} \left( N_{ijck} + \beta \sum_{c'=1}^{r_0} N_{ijc'k} \right) \log(\theta_{ijck}) \end{aligned}$$

ここで、 $N_{ijck}$  は  $X_i = k$  かつ  $G_i \setminus \{X_0\} = j$  かつ  $X_0 = c$  となる頻度を表し、 $\theta_{ijck}$  は  $G_i \setminus \{X_0\} = j$  かつ  $X_0 = c$  の時に  $X_i = k$  となる条件付き確率パラメータを表す。また、 $q_i^* = \prod_{l: X_l \in G_i \setminus \{X_0\}} r_l$  である。ハイパーパラメータ  $N'_{ijck} > 0$  に対して、aCLL を最大にするパラメータは次のように推定できる。

$$\hat{\theta}_{ijck} = \frac{N_{ij+ck}}{N_{ij+c}}$$

ここで、

$$N_{ij+ck} = \begin{cases} N_{ijck} + \beta \sum_{c'=1}^{r_0} N_{ijc'k} & (\text{if } N_{ijck} + \beta \sum_{c'=1}^{r_0} N_{ijc'k} \geq N'_{ijck}) \\ N'_{ijck} & (\text{if } N_{ijck} + \beta \sum_{c'=1}^{r_0} N_{ijc'k} < N'_{ijck}), \end{cases}$$

$$N_{ij+c} = \sum_{k=1}^{r_i} N_{ijc+k}$$

である。最適な  $\beta$  と  $\gamma$  に対して、aCLL は CLL の最小分散不偏推定量である。これらの近似手法で学習した BNC の方が、ML (BDeu や MDL など) によって近似的に学習した BNC よりも分類精度が高いことが報告されてきた。

しかし、ML 最大化より CLL 最大化の方がなぜ良いかという理由についてはこれまで明らかにされていない。ML は推定構造に対して漸近一致性が保証されており、サンプルサイズが大きい時に ML の分類精度が CLL に劣るのは奇異である。また、BNC の ML は分解可能であるため、ML による厳密学習は CLL による厳密学習とは異なり、現実的な時間で学習できる。先行研究の比較実験では、ML による近似学習を行なっているため、探索精度の悪さが影響したのかもしれない。

そこで、次章では ML によって厳密に学習した BNC と CLL によって近似的に学習した BNC の分類精度を比較する。

#### 4. GBN の厳密学習と識別モデルの比較実験

本章では、ML によって厳密に学習した BNC と CLL によって近似的に学習した BNC の分類精度を比較する。ML として用いるスコアは BDeu とし、擬似サンプル（ハイパーパラメータ）の値は Ueno [7], [8] の提案に従い 1.0 とした。この実験では次の 6 つの手法を比較する。

- GBN-BDeu: BDeu を用いて厳密学習した GBN
- Naive Bayes
- GBN-CMDL (Grossman ら [4]): MDL のフィッティング項を CLL に置き換えた Conditional MDL (CMDL) を用いて近似学習した GBN
- BNC2P (Grossman ら [4]): 各変数が最大 2 つまでしか親を持たない構造を候補として、CLL を用いて近似学習した BNC
- TAN-aCLL (Carvalho ら [3]): aCLL を用いて厳密学習した TAN
- gGBN-BDeu: BDeu を用いて近似学習した GBN

近似学習の構造探索法としては Hill-Climbing を用い、厳密学習の構造探索手法としては動的計画法 [10] を用いた。UCI レポジトリデータベース [14] の 43 個のベンチマークデータセットを用い、各データセットに含まれる連続量はいずれも中央値を境に 2 値に離散化し、欠損値を含むサンプルはデータセットから除去した。いずれの手法においても、構造学習後の BNC のパラメータは全て EAP で推定した。

各手法、各データセットに対して、10 分割交差検証によるテストデータの平均一致率を求め、分類精度として表 1 に示した。ただし、表 1 に記載されている ANB-BDeu については 6 章で議論する。各データセットに対し、各手法の中で最も高い分類精度を太字で示している。また、分類精度と学習した GBN-BDeu の構造との関係を見るために、10 分割交差検証における GBN-BDeu の学習構造の目的変数の親変数数の平均を表 2 の "Parents" に示した。表 2 の "Children" は、10 分割交差検証における GBN-BDeu の学習構造の目的変数の子変数数の平均を示している。表 2 の "Sparse data" は  $N_{0j} = 0$  となる空データをとる  $j$  のパターン数の平均を示している。ただし、表 2 に記載されている MBsize, Missing variables, Extra variables,

Max parents については 6 章で言及する。

表 1 を見ると、GBN-BDeu はサンプルサイズの大きいデータセット 11 番, 12 番, 22 番, 32 番, 40 番について、比較手法 6 つの中で分類精度が最も高い。この結果から、ML によって学習した BNC が必ずしも CLL によって学習した BNC より分類精度が低いとは限らないことがわかる。しかし、サンプルサイズの小さいデータセットである 3 番, 9 番, 31 番では GBN-BDeu の分類精度が著しく低い。これらのデータセットでは、目的変数の子変数が少なく、親変数が多いことが表 2 の "Children", "Parents" からわかる。また、これらのデータセットでは "Sparse data" の値が大きいことがわかる。このように、目的変数の親変数が多いと、親変数集合  $G_0$  の状態数  $q_0$  が指数的に増えるため、各状態  $G_0 = j$  に対して  $N_{0j}$  が小さくなる。したがって、目的変数のパラメータ推定値 (式 (5)) の精度が悪くなってしまう。さらに、式 (10) より、目的変数の子変数が少ないほど目的変数のパラメータが分類に大きく影響を及ぼす。以上から、目的変数の親変数が多く子変数が少ないことが GBN-BDeu の分類精度の著しい低下の原因だと考えられる。次章でこの問題を解決する手法とそのアルゴリズムを提案する。

#### 5. Augmented Naive Bayes 制約を持つ BNC の厳密学習

4 章では、サンプルサイズの小さいデータセットにおいて GBN-BDeu が目的変数の親変数が多く子変数が少ない構造を学習する場合があります、それが分類精度の著しい低下を招くことを示した。GBN-BDeu は目的変数のマルコフブランケットとそれ以外の変数の構造を同時に学習するが、分類には目的変数のマルコフブランケットのみが影響する。したがって、GBN-BDeu の目的変数のマルコフブランケットを説明変数集合として、上記の問題を緩和するような制約下で構造を学習すれば、分類精度を向上できる可能性がある。目的変数の親変数数と子変数数を制約する手法として、目的変数が親変数を持たず、全説明変数を子変数として持つ Augmented Naive Bayes (ANB) が知られている。本論では、GBN-BDeu の目的変数についてのマルコフブランケットのみを説明変数集合とし、ANB 構造を制約として BDeu により厳密学習する BNC (以降、このモデルを Markov blanket Augmented Naive Bayes (MANB) と呼ぶ) を提案する。ANB はこれまで識別モデルとして扱われてきたため、BDeu を最

表 1 GBN-BDeu, ANB-BDeu と従来手法の分類精度（太字は最大の分類精度）

Table 1 Classification accuracies of GBN-BDeu, ANB-BDeu, and traditional methods (bold text means the highest accuracy).

| No.     | Dataset                      | Variables | Sample size | Classes | Naive-Bayes   | GBN-CMDL      | BNC2P         | TAN-aCLL      | gGBN-BDeu     | GBN-BDeu      | ANB-BDeu      |
|---------|------------------------------|-----------|-------------|---------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| 1       | Balance Scale                | 5         | 625         | 3       | <b>0.9152</b> | 0.3333        | 0.8560        | 0.8656        | <b>0.9152</b> | <b>0.9152</b> | <b>0.9152</b> |
| 2       | banknote authentication      | 5         | 1372        | 2       | 0.8433        | <b>0.8819</b> | 0.8797        | 0.8761        | <b>0.8819</b> | 0.8812        | 0.8812        |
| 3       | Hayes-Roth                   | 5         | 132         | 3       | <b>0.8182</b> | 0.6136        | 0.6894        | 0.6742        | 0.7525        | 0.6136        | <b>0.8182</b> |
| 4       | iris                         | 5         | 150         | 3       | 0.7133        | 0.7800        | 0.8200        | 0.8200        | 0.8133        | <b>0.8267</b> | 0.8200        |
| 5       | lenses                       | 5         | 24          | 3       | 0.7500        | <b>0.8333</b> | 0.6667        | 0.7083        | <b>0.8333</b> | <b>0.8333</b> | 0.7500        |
| 6       | Car Evaluation               | 7         | 1728        | 4       | 0.8571        | <b>0.9497</b> | 0.9416        | 0.9433        | 0.9416        | 0.9416        | 0.9427        |
| 7       | liver                        | 7         | 345         | 2       | 0.6319        | 0.6145        | 0.6290        | <b>0.6609</b> | 0.6029        | 0.6087        | 0.6348        |
| 8       | MONK's Problems              | 7         | 432         | 2       | 0.7500        | <b>1.0000</b> | <b>1.0000</b> | <b>1.0000</b> | 0.8449        | <b>1.0000</b> | <b>1.0000</b> |
| 9       | mux6                         | 7         | 64          | 2       | 0.5469        | 0.3750        | <b>0.5625</b> | 0.4688        | 0.4063        | 0.4531        | 0.5469        |
| 10      | led7                         | 8         | 3200        | 10      | 0.7294        | 0.7366        | <b>0.7375</b> | 0.7350        | 0.7297        | 0.7294        | 0.7294        |
| 11      | HTRU2                        | 9         | 17898       | 2       | 0.7031        | 0.7096        | 0.7070        | 0.7018        | 0.7188        | <b>0.7305</b> | 0.7188        |
| 12      | Nursery                      | 9         | 12960       | 3       | 0.6782        | <b>0.7126</b> | 0.6092        | 0.5862        | <b>0.7126</b> | <b>0.7126</b> | 0.6782        |
| 13      | pima                         | 9         | 768         | 9       | 0.8966        | 0.9086        | 0.9118        | 0.9130        | 0.9092        | 0.9112        | <b>0.9141</b> |
| 14      | post                         | 9         | 87          | 5       | 0.9033        | 0.5823        | <b>0.9442</b> | 0.9177        | 0.9291        | 0.9340        | 0.9181        |
| 15      | Breast Cancer                | 10        | 277         | 2       | <b>0.9751</b> | 0.8917        | 0.9473        | 0.9488        | 0.7058        | <b>0.9751</b> | <b>0.9751</b> |
| 16      | Breast Cancer Wisconsin      | 10        | 683         | 2       | <b>0.7401</b> | 0.6209        | 0.6823        | 0.7184        | 0.7094        | 0.7184        | 0.7040        |
| 17      | Contraceptive Method Choice  | 10        | 1473        | 3       | 0.4671        | 0.4501        | <b>0.4745</b> | 0.4705        | 0.4440        | 0.4542        | 0.4650        |
| 18      | glass                        | 10        | 214         | 6       | 0.5561        | 0.5654        | 0.5794        | 0.6308        | 0.4626        | 0.5701        | <b>0.6449</b> |
| 19      | shuttle-small                | 10        | 5800        | 6       | 0.9384        | 0.9660        | 0.9703        | 0.9583        | 0.9683        | 0.9693        | <b>0.9716</b> |
| 20      | threeOf9                     | 10        | 512         | 2       | 0.8164        | <b>0.9434</b> | 0.8691        | 0.8828        | 0.8652        | 0.8887        | 0.8730        |
| 21      | Tic-Tac-Toe                  | 10        | 958         | 2       | 0.6921        | <b>0.8841</b> | 0.7338        | 0.7203        | 0.6754        | 0.8340        | 0.8497        |
| 22      | MAGIC Gamma Telescope        | 11        | 19020       | 2       | 0.7482        | 0.7849        | 0.7806        | 0.7631        | 0.7844        | 0.7873        | <b>0.7874</b> |
| 23      | Solar Flare                  | 11        | 1389        | 9       | 0.7811        | 0.8265        | 0.8315        | 0.8229        | <b>0.8431</b> | <b>0.8431</b> | 0.8229        |
| 24      | heart                        | 14        | 270         | 2       | <b>0.8259</b> | 0.8185        | 0.8037        | 0.8148        | 0.8222        | <b>0.8259</b> | 0.8185        |
| 25      | wine                         | 14        | 178         | 3       | 0.9270        | <b>0.9438</b> | 0.9157        | 0.9326        | 0.9045        | 0.9270        | 0.9270        |
| 26      | cleve                        | 14        | 296         | 2       | <b>0.8412</b> | 0.8209        | 0.8007        | 0.8378        | 0.7973        | 0.7973        | 0.8277        |
| 27      | australian                   | 15        | 690         | 2       | 0.8290        | 0.8312        | 0.8348        | 0.8464        | 0.8420        | <b>0.8536</b> | 0.8246        |
| 28      | crx                          | 15        | 653         | 2       | 0.8377        | 0.8346        | 0.8208        | 0.8560        | <b>0.8622</b> | 0.8591        | 0.8515        |
| 29      | EEG                          | 15        | 14980       | 2       | 0.5778        | 0.6787        | 0.6374        | 0.6125        | 0.6732        | 0.6814        | <b>0.6864</b> |
| 30      | Congressional Voting Records | 17        | 232         | 2       | 0.9095        | 0.9698        | 0.9612        | 0.9181        | <b>0.9741</b> | 0.9655        | 0.9483        |
| 31      | zoo                          | 17        | 101         | 5       | 0.9802        | 0.9109        | 0.9505        | <b>1.0000</b> | 0.9505        | 0.9307        | 0.9505        |
| 32      | pendigits                    | 17        | 10992       | 10      | 0.8032        | 0.9062        | 0.8719        | 0.8700        | 0.9253        | <b>0.9290</b> | 0.9279        |
| 33      | letter                       | 17        | 20000       | 26      | 0.4466        | 0.5796        | 0.5132        | 0.5093        | 0.5761        | 0.5761        | <b>0.5935</b> |
| 34      | ClimateModel                 | 19        | 540         | 2       | 0.9222        | <b>0.9407</b> | 0.9241        | 0.9333        | 0.9370        | 0.9000        | 0.8426        |
| 35      | Image Segmentation           | 19        | 2310        | 7       | 0.7290        | 0.7918        | 0.7991        | 0.7407        | 0.8026        | 0.8156        | <b>0.8225</b> |
| 36      | lymphography                 | 19        | 148         | 4       | <b>0.8446</b> | 0.7939        | 0.7973        | 0.8311        | 0.7905        | 0.7500        | 0.7770        |
| 37      | vehicle                      | 19        | 846         | 4       | 0.4350        | 0.5910        | 0.5910        | 0.5816        | 0.5461        | 0.5768        | <b>0.6253</b> |
| 38      | hepatitis                    | 20        | 80          | 2       | 0.8500        | 0.7375        | <b>0.8875</b> | 0.8750        | 0.8500        | 0.5875        | 0.6250        |
| 39      | german                       | 21        | 1000        | 2       | 0.7430        | 0.6110        | 0.7340        | <b>0.7470</b> | 0.7140        | 0.7210        | 0.7380        |
| 40      | bank                         | 21        | 30488       | 2       | 0.8544        | 0.8618        | 0.8928        | 0.8618        | 0.8952        | <b>0.8956</b> | 0.8950        |
| 41      | waveform-21                  | 22        | 5000        | 3       | 0.7886        | 0.7862        | 0.7754        | 0.7896        | 0.7698        | 0.7846        | <b>0.7966</b> |
| 42      | Mushroom                     | 22        | 5644        | 2       | 0.9957        | <b>1.0000</b> | <b>1.0000</b> | 0.9995        | <b>1.0000</b> | 0.9949        | <b>1.0000</b> |
| 43      | spect                        | 23        | 263         | 2       | 0.7940        | 0.7940        | 0.7903        | 0.8090        | 0.7603        | 0.7378        | <b>0.8240</b> |
| average |                              |           |             |         | 0.7764        | 0.7721        | 0.7936        | 0.7943        | 0.7867        | 0.7963        | <b>0.8061</b> |
| p-value |                              |           |             |         | 0.0031        | 0.0414        | 0.0067        | 0.0561        | 0.0629        | 0.2263        | -             |

大化して学習することはなかった。本論の提案は、識別モデルの学習ではなく、生成モデルとしての GBN の学習に、目的変数の親変数が増えないように ANB 構造に制約することである。

MANB の厳密学習では、GBN の目的変数についてのマルコフブランケット  $M = \{X_1, \dots, X_m\}$ , ( $m = |M|$ ) と目的変数  $X_0$  によって構成される全ての ANB 構造の中で、スコアを最大にする構造を探索する。ANB では目的変数が親変数を持たないため、全ての ANB 構造の目的変数についてのローカルスコア

は等しい。このため、厳密解を得るには、構造全体のスコアから目的変数のローカルスコアを除いた  $Score_{ANB}(G) = Score(G) - Score_0(\phi)$  を最大化する ANB 構造を探索すればよい。しかし、ANB の全構造数は説明変数数  $m$  の増加に対して超指数関数的に増加するため、構造を全探索すると計算ステップ数が  $\mathcal{O}(m!2^{\binom{m}{2}})$  と膨大になってしまう。そこで本論では全構造を列挙せずに、変数間の半順序を推定することにより構造の計算ステップ数を大幅に削減するアルゴリズムを提案する。以下では提案アルゴリズムで用



表 2 GBN-BDeu と MANB-BDeu の学習構造の詳細  
Table 2 Statistics summary of GBN-BDeu and MANB-BDeu

| No. | Dataset                      | Variables | Classes | Sample size | Parents | Children | Sparse data | MBsize | Missing variables | Extra variables | Max parents |
|-----|------------------------------|-----------|---------|-------------|---------|----------|-------------|--------|-------------------|-----------------|-------------|
| 1   | Balance Scale                | 5         | 3       | 625         | 0.4     | 3.6      | 0.0         | 4.0    | 0.0               | 0.0             | 1.0         |
| 2   | banknote authentication      | 5         | 2       | 1372        | 0.0     | 2.0      | 0.0         | 4.0    | 0.0               | 0.0             | 4.0         |
| 3   | Hayes-Roth                   | 5         | 3       | 132         | 3.0     | 0.0      | 17.2        | 3.0    | 0.0               | 0.0             | 1.0         |
| 4   | iris                         | 5         | 3       | 150         | 1.8     | 1.2      | 0.0         | 3.0    | 0.0               | 0.0             | 2.0         |
| 5   | lenses                       | 5         | 3       | 24          | 1.1     | 1.0      | 0.0         | 2.0    | 0.0               | 0.1             | 1.1         |
| 6   | Car Evaluation               | 7         | 4       | 1728        | 2.0     | 3.0      | 0.0         | 5.0    | 0.0               | 0.0             | 2.0         |
| 7   | liver                        | 7         | 2       | 345         | 0.0     | 1.9      | 0.0         | 3.4    | 1.6               | 0.0             | 2.0         |
| 8   | MONK's Problems              | 7         | 2       | 432         | 3.0     | 0.0      | 0.0         | 3.0    | 0.0               | 0.0             | 3.0         |
| 9   | mux6                         | 7         | 2       | 64          | 5.8     | 0.0      | 5.2         | 5.8    | 0.2               | 0.0             | 1.0         |
| 10  | led7                         | 8         | 10      | 3200        | 0.9     | 6.1      | 0.0         | 7.0    | 0.0               | 0.0             | 1.0         |
| 11  | HTRU2                        | 9         | 2       | 17898       | 1.8     | 4.2      | 0.0         | 7.0    | 0.0               | 0.0             | 3.0         |
| 12  | Nursery                      | 9         | 5       | 12960       | 4.0     | 3.0      | 0.0         | 7.0    | 0.0               | 0.0             | 3.0         |
| 13  | pima                         | 9         | 2       | 768         | 1.4     | 1.7      | 0.0         | 4.0    | 0.0               | 0.2             | 2.0         |
| 14  | post                         | 9         | 3       | 87          | 0.0     | 0.0      | 0.0         | 0.0    | 0.0               | 0.0             | 0.0         |
| 15  | Breast Cancer                | 10        | 2       | 277         | 0.9     | 8.0      | 0.0         | 8.0    | 0.0               | 0.9             | 1.0         |
| 16  | Breast Cancer Wisconsin      | 10        | 2       | 683         | 0.7     | 0.3      | 0.0         | 1.0    | 0.0               | 0.0             | 1.0         |
| 17  | Contraceptive Method Choice  | 10        | 3       | 1473        | 0.7     | 0.8      | 0.0         | 1.5    | 0.5               | 0.2             | 1.2         |
| 18  | glass                        | 10        | 6       | 214         | 0.6     | 3.1      | 0.0         | 4.2    | 0.8               | 0.1             | 1.6         |
| 19  | shuttle-small                | 10        | 6       | 5800        | 2.0     | 4.0      | 0.0         | 7.0    | 0.0               | 0.0             | 5.0         |
| 20  | threeOf9                     | 10        | 2       | 512         | 5.0     | 2.1      | 0.0         | 7.6    | 1.4               | 0.0             | 2.4         |
| 21  | Tic-Tac-Toe                  | 10        | 2       | 958         | 1.2     | 2.2      | 0.0         | 5.1    | 0.9               | 0.2             | 3.0         |
| 22  | MAGIC Gamma Telescope        | 11        | 2       | 19020       | 0.0     | 6.1      | 0.0         | 8.0    | 0.0               | 0.0             | 4.0         |
| 23  | Solar Flare                  | 11        | 9       | 1389        | 0.8     | 0.2      | 0.0         | 0.1    | 0.9               | 0.9             | 1.0         |
| 24  | heart                        | 14        | 2       | 270         | 1.8     | 4.2      | 0.0         | 6.0    | 1.0               | 0.3             | 1.5         |
| 25  | wine                         | 14        | 3       | 178         | 1.7     | 5.3      | 0.0         | 7.0    | 0.0               | 1.1             | 2.1         |
| 26  | cleve                        | 14        | 2       | 296         | 1.8     | 4.5      | 0.0         | 6.3    | 0.7               | 0.3             | 2.0         |
| 27  | Australian                   | 15        | 2       | 690         | 1.4     | 2.8      | 0.0         | 4.2    | 0.8               | 0.3             | 2.2         |
| 28  | crx                          | 15        | 2       | 653         | 1.3     | 2.8      | 0.0         | 2.9    | 1.1               | 1.3             | 2.0         |
| 29  | EEG                          | 15        | 2       | 14980       | 0.4     | 8.2      | 0.0         | 12.8   | 0.2               | 0.0             | 5.0         |
| 30  | Congressional Voting Records | 17        | 2       | 232         | 1.3     | 2.6      | 0.1         | 5.2    | 1.8               | 1.0             | 2.8         |
| 31  | zoo                          | 17        | 5       | 101         | 4.3     | 1.6      | 20.3        | 6.9    | 3.1               | 0.5             | 3.5         |
| 32  | pendigits                    | 17        | 10      | 10992       | 2.6     | 13.4     | 0.1         | 16.0   | 0.0               | 0.0             | 5.6         |
| 33  | letter                       | 17        | 26      | 20000       | 2.9     | 9.1      | 0.0         | 13.0   | 0.0               | 0.0             | 5.0         |
| 34  | ClimateModel                 | 19        | 2       | 540         | 1.8     | 4.4      | 0.0         | 15.9   | 1.1               | 0.7             | 14.0        |
| 35  | Image Segmentation           | 19        | 7       | 2310        | 0.7     | 10.4     | 0.0         | 12.7   | 0.3               | 0.5             | 4.1         |
| 36  | lymphography                 | 19        | 4       | 148         | 1.6     | 5.9      | 0.2         | 9.0    | 1.0               | 4.1             | 8.0         |
| 37  | vehicle                      | 19        | 4       | 846         | 1.1     | 5.1      | 0.1         | 9.0    | 2.0               | 1.1             | 3.6         |
| 38  | hepatitis                    | 20        | 2       | 80          | 1.3     | 6.1      | 0.4         | 13.1   | 1.9               | 2.9             | 10.7        |
| 39  | German                       | 21        | 2       | 1000        | 1.1     | 2.8      | 0.0         | 3.9    | 2.1               | 0.2             | 1.2         |
| 40  | bank                         | 21        | 2       | 30488       | 4.1     | 2.0      | 32.5        | 9.9    | 0.1               | 0.0             | 5.0         |
| 41  | waveform-21                  | 22        | 3       | 5000        | 3.8     | 10.1     | 0.0         | 13.9   | 0.1               | 0.6             | 4.0         |
| 42  | Mushroom                     | 22        | 2       | 5644        | 1.3     | 3.3      | 9.0         | 6.1    | 12.9              | 0.3             | 5.2         |
| 43  | spect                        | 23        | 2       | 263         | 2.0     | 3.4      | 0.0         | 6.4    | 2.6               | 1.3             | 2.5         |

いる表記や用語を定義する。  $M \cup \{X_0\}$  の部分集合  $W$  で構成される ANB 構造の中で、  $Score_{ANB}$  を最大にする構造を  $G^*(W)$  とする。 子変数を持たない変数をシンクと呼び、  $G^*(W)$  におけるシンクを  $X_s^*(W)$  と定義する。  $W$  のべき集合に対して、  $X_0$  を含む部分集合を  $\Pi(W)$  と定義する。  $X_i \in M$  の親変数集合  $G_i \subseteq M \cup \{X_0\} \setminus \{X_i\}$ ,  $(X_0 \in G_i)$  について、  $\Pi(G_i)$  を  $X_i$  の親変数候補集合とよぶ。 次式のように、  $\Pi(G_i)$  の中でローカルスコアを最大にする親変数集合を最適親変数集合とよぶ。

$$g_i^*(\Pi(G_i)) = \arg \max_{g \in \Pi(G_i)} Score_i(g)$$

提案アルゴリズムは以下の 5 つのプロセスからなる。

(1) GBN-BDeu を学習し、目的変数  $X_0$  のマルコフブランケット  $M$  を取り出す。 GBN-BDeu の学習法は Silander ら [10] が提案しており、それをそのまま用いる。 以降の学習プロセスは、Silander らの GBN-BDeu の学習アルゴリズムをそのまま適用したものではなく、  $M \cup \{X_0\}$  についての学習構造が ANB の制約を満たすように Silander らのアルゴリズムを改良したものである。

(2) 各説明変数  $X_i \in M$  の親変数集合  $G_i \subseteq M \cup \{X_0\} \setminus \{X_i\}$ ,  $(X_0 \in G_i)$  の各組合せについて logBDeu のローカルスコア (式 (9)) を計算する。 全てのローカルスコアの計算ステップ数は  $m2^{m-1}$  である。

(3) 各説明変数  $X_i \in M$  の親変数集合  $G_i \subseteq M \cup \{X_0\} \setminus \{X_i\}$ , ( $X_0 \in G_i$ ) の各組合せについて  $g_i^*(\Pi(G_i))$  を計算する. 全ての最適親変数集合の計算ステップ数は  $m2^{m-1}$  である.

(4) 全ての変数集合  $W \subseteq M \cup \{X_0\}$ , ( $X_0 \in W$ ) に対して,  $X_s^*(W)$  を計算する. 全てのシンクの計算ステップ数は  $2^m$  である.

(5) プロセス (3) と (4) の結果を用いて  $G^*(M \cup \{X_0\})$  を計算する.

プロセス (4) は以下の性質に基づいている.  $G^*(W)$  はシンク  $X_s^*(W)$  を必ず持っており,  $G^*(W)$  において  $X_s^*(W)$  は最適な親変数集合  $g_s^*(\Pi(W \setminus \{X_s^*(W)\}))$  を持っていなければならない. さらに,  $G^*(W)$  において残りの  $W \setminus \{X_s^*(W)\}$  で構成される構造も最適でなければならないため, 次式が成り立つ.

$$X_s^*(W) = \arg \max_{X_i \in W \setminus \{X_0\}} \{Score_i(g_i^*(\Pi(W \setminus \{X_i\}))) + Score_{ANB}(G^*(W \setminus \{X_i\}))\} \quad (11)$$

式 (11) によって  $G^*(M \cup \{X_0\})$  をシンクとそれ以外の変数からなる構造に分解できる. この分解によりシンクを順に削除して新しいシンクを求めることができる. これを再帰的に適用して順に得られるシンク  $X_s^*(W)$  とその最適親変数集合  $g_s^*(\Pi(W \setminus \{X_s^*(W)\}))$  の組をそれぞれ  $(X_{s_1}, g_{s_1}^*), \dots, (X_{s_i}, g_{s_i}^*) \dots, (X_{s_m}, g_{s_m}^*)$  とすると,  $G^*(M \cup \{X_0\})$  における変数  $X_{s_i}$  の親変数集合は  $g_{s_i}^*$  である. 以上により  $G^*(M \cup \{X_0\})$  を求めることができる. 構造探索を行うプロセス (3) と (4) における最適親変数集合とシンクの計算ステップ数は合わせて  $\mathcal{O}(m2^m)$  であり, 全ての構造パターンの計算ステップ数  $\mathcal{O}(m!2^{\binom{m}{2}})$  より小さい. このように提案アルゴリズムはシンクの順序付けによって, 全ての構造パターンを計算する場合より, 大幅に計算ステップ数を削減できる.

次に, 最も計算時間のかかるプロセス (3) の最適親変数集合の計算と, プロセス (4) のシンクの計算の具体的なアルゴリズムを説明する.  $g_i^*(\Pi(G_i))$  は  $G_i$  そのものか,  $G_i$  から一つだけ説明変数を取り除いた親変数候補集合  $\Pi(G_i \setminus \{X\})$ , ( $X \in G_i \setminus \{X_0\}$ ) についての最適親変数集合のうち, ローカルスコアが最大のものである [10]. したがって, 次式が得られる.

$$Score_i(g_i^*(\Pi(G_i))) = \max\{Score_i(G_i), Score1(G_i)\}, \quad (12)$$

ここで,

$$Score1(G_i) = \max_{X \in G_i \setminus \{X_0\}} Score_i(g_i^*(\Pi(G_i \setminus \{X\})))$$

である. 式 (12) によって  $X_i \in M$  の全ての親変数候補集合に対して最適親変数集合を再帰的に求めることができる. しかし, 再帰的な計算では, 既に計算済みの最適親変数集合を重複して計算してしまう. 例えば,  $g_1^*(\Pi(\{X_0, X_2, X_3, X_4\}))$  を計算する場合,  $g_1^*(\Pi(\{X_0, X_2, X_3\}))$ ,  $g_1^*(\Pi(\{X_0, X_2, X_4\}))$ ,  $g_1^*(\Pi(\{X_0, X_3, X_4\}))$  を再帰的に計算する必要がある. しかし,  $g_1^*(\Pi(\{X_0, X_2, X_3\}))$ ,  $g_1^*(\Pi(\{X_0, X_2, X_4\}))$  の計算時に, それぞれ  $g_1^*(\Pi(\{X_0, X_2\}))$  を重複して計算してしまう. したがって, 式 (12) を再帰的に計算する場合は, 既に計算済みかどうかの確認が必要となり, 計算時間が増加してしまう. そこで, 再帰的な計算は行わず, Silander ら [10] と同様に, 重複計算がないことを保証できる計算順序として  $G_i$  のビット列表記に関する辞書式順序で式 (12) を適用する. ここで, 変数集合  $W \subseteq M \cup \{X_0\}$  のビット列表記は以下の長さ  $m+1$  のビット列で定義される.

$$(b_m, \dots, b_i, \dots, b_1, b_0), \quad b_i = \begin{cases} 1 & (\text{if } X_i \in W) \\ 0 & (\text{otherwise}) \end{cases}$$

$G_i$  のビット列表記に関する辞書式順序は,  $G_i \succ G_i \setminus \{X\}$ , ( $X \in G_i \setminus \{X_0\}$ ) と定義され, この順序で式 (12) を適用すると, 各計算に必要な最適親変数集合は必ず既知である. このため新たに最適親変数集合を計算する必要がなく, 重複計算は行われぬ. 例えば,  $X_1$  の親変数集合  $\{X_0, X_2, X_3, X_4\}$  (ビット列表記は 11101) について,  $g_1^*(\Pi(11101))$  を計算する場合を考える. この計算を行うには,  $g_1^*(\Pi(01101))$ ,  $g_1^*(\Pi(10101))$ ,  $g_1^*(\Pi(11001))$  を求める必要がある. 辞書式順序では 11101  $\succ$  01101, 10101, 11001 であるから, それら 3 つは既に計算済みである. よって重複計算を行わずに目的の  $\Pi(11101)$  に対する最適親変数集合を計算できる. 式 (11) によるシンクの計算でも同様に, 再帰的な計算では重複によって計算時間が増加してしまう. 例えば, 変数集合 11111 に対してシンクを計算する場合, 01111, 10111, 11011, 11101 に対するシンクを再帰的に計算する必要がある. しかし, 01111, 10111 に対するシンクの計算時に, それぞれ 00111 に対するシンクを重複して計算してしまう. そこで, 最適親変数集合の計算と同様

に、変数集合のビット列表記に関する辞書式順序で式 (11) を適用して、この問題を回避する。辞書式順序では 11111  $\succ$  01111, 10111, 11011, 11101 であるから、11111 に対するシンクの計算時に 01111, 10111, 11011, 11101 に対するシンクは既に計算済みである。よって重複計算を行わずに目的の 11111 に対するシンクを計算できる。

変数選択を行わない場合、提案手法におけるローカルスコア、最適親変数集合、シンクの各計算ステップ数は、全説明変数数を  $n$  とするとそれぞれ  $n2^{n-1}$ ,  $n2^{n-1}$ ,  $2^n$  である。Silander らの動的計画法 [10] におけるそれら 3 つの計算ステップ数はそれぞれ  $(n+1)2^n$ ,  $(n+1)2^n$ ,  $2^{n+1}$  である。したがって、変数選択を行わない場合、提案手法におけるそれら 3 つの各計算ステップ数は Silander らの動的計画法における各計算ステップ数の約半分に減少する。動的計画法の計算時間はこれらの計算ステップ数の減少に対して線形的に減少するため、計算時間も約半分に短縮される。

次章で、提案手法で学習した BNC と、CLL を用いて学習した BNC の分類精度を比較する。

## 6. 評価実験

本章では、提案手法と CLL をベースとした先行研究の手法の分類精度を比較するため、リポソトリデータを用いた評価実験を行う。まず、BDeu によって生成モデルとして厳密に学習した ANB (ANB-BDeu) と、4 章の実験に用いた手法の分類精度を比較する。ANB-BDeu は 5 章で紹介したアルゴリズムにおけるマルコフブランケット  $M$  を全説明変数集合に置き換えたものを用いて学習できる。ANB-BDeu とその他の手法との有意性を示すため、分類精度の多重検定手法として標準的に用いられる Hommel の多重検定 [15], [16] を行った。検定の  $p$  値を表 1 の最下部に示した。また、表 2 の "MBSIZE" は、10 分割交差検証における GBN-BDeu の学習構造の目的変数のマルコフブランケットの要素数の平均を示している。

結果として、ANB-BDeu は Naive Bayes, GBN-CMDL, BNC2P よりも有意水準 5% のもとで有意に分類精度が高かった。さらに、ANB-BDeu は GBN-BDeu の分類精度が著しく悪かったデータセット 3 番, 9 番, 31 番において、分類精度を改善していることがわかる。しかし、データセット 5 番と 14 番では ANB-BDeu は GBN-BDeu の分類精度を大きく下回っている。これらのデータセットでは、目的変数

のマルコフブランケットの要素数が少ないことが表 2 の "MBSIZE" からわかる。したがって、マルコフブランケットによる変数選択は ANB-BDeu の分類精度を改善すると期待できる。

次に、表 1 に示される手法と BDeu を用いて厳密学習した MANB (MANB-BDeu) の分類精度を比較する。公正に比較するため、MANB-BDeu 以外の手法にも GBN-BDeu の目的変数のマルコフブランケットによる変数選択を行った。変数選択をした各手法の名前は、元々の手法名の最初に "M" を付けて表す。表 3 に、各データセットに対する変数選択をした各手法の分類精度を示す。また、表 3 の最下部には MANB-BDeu とその他の手法の分類精度に対する Hommel の多重検定の  $p$  値を示している。結果として、MANB-BDeu は全比較手法に対して有意水準 5% のもとで有意に分類精度が高かった。

表 2 の "Max parents" は 10 分割交差検証における MANB-BDeu の学習構造内の変数が持つ最大親変数数の平均を示している。"Max parents" の値が大きいほど MANB-BDeu の構造がより複雑であることを表す。表 2, 表 3 より "Max parents" の値が大きいデータセット 36 番と 38 番では、MANB-BDeu は MNaive Bayes の分類精度を下回っていることがわかる。4 章で述べたように、変数の親変数が多くなるとその変数のパラメータの推定精度は不安定になってしまう。MNaive Bayes は "Max parents" の値がわずかに 1 でありパラメータの推定精度が安定するため、サンプルサイズが小さすぎると MNaive Bayes は MANB-BDeu よりも分類精度が高くなると考えられる。しかし、MNaive Bayes は説明変数間の相関を考慮できないため、サンプルサイズの大きいデータセット 8 番や 29 番では他のどの手法よりも分類精度が著しく悪くなっている。

MGBN-CMDL はデータセット 1 番, 3 番, 9 番, 14 番, 15 番などの多くのデータセットで他手法よりも分類精度が著しく悪くなっている。これは、Grossman ら [4] の実験結果と一致している。この理由として、CMDL スコアのペナルティ項とフィッティング項の整合性が取れていないことが考えられる。

MBNC2P と MTAN-aCLL は、"Max Parents" が 2 であるから、MNaive Bayes と同様にサンプルサイズの小さい 38 番のデータセットでは MANB-BDeu よりも分類精度が高い。しかし、データセット 29 番では、MANB-BDeu の "Max parents" の値が 2 より大き

表 3 変数選択を用いた場合の各手法の分類精度  
Table 3 Classification accuracies of respective classifiers achieved using feature selection for 43 datasets

| No.     | Dataset                      | Variables | Sample size | Classes | MNaive-Bayes  | MGBN-CMDL     | MBNC-2P       | MTAN-aCLL     | MgGBN-BDeu    | GBN-BDeu      | MANB-BDeu     |
|---------|------------------------------|-----------|-------------|---------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| 1       | Balance Scale                | 5         | 625         | 3       | <b>0.9152</b> | 0.3333        | 0.8560        | 0.8656        | <b>0.9152</b> | <b>0.9152</b> | <b>0.9152</b> |
| 2       | banknote authentication      | 5         | 1372        | 2       | 0.8433        | <b>0.8819</b> | 0.8783        | 0.8761        | 0.8812        | 0.8812        | 0.8812        |
| 3       | Hayes-Roth                   | 5         | 132         | 3       | <b>0.8333</b> | 0.6136        | 0.7197        | 0.7879        | 0.7980        | 0.6136        | <b>0.8333</b> |
| 4       | iris                         | 5         | 150         | 3       | <b>0.8267</b> | 0.7800        | 0.8200        | 0.8200        | 0.8200        | <b>0.8267</b> | <b>0.8267</b> |
| 5       | lenses                       | 5         | 24          | 3       | 0.8333        | 0.8333        | 0.8333        | 0.8333        | <b>0.8750</b> | 0.8333        | 0.8333        |
| 6       | Car Evaluation               | 7         | 1728        | 4       | 0.8559        | 0.9242        | 0.9375        | 0.9363        | <b>0.9416</b> | <b>0.9416</b> | <b>0.9416</b> |
| 7       | liver                        | 7         | 345         | 2       | <b>0.6348</b> | <b>0.6348</b> | 0.6000        | 0.5942        | 0.6000        | 0.6087        | 0.5855        |
| 8       | MONK's Problems              | 7         | 432         | 2       | 0.7500        | <b>1.0000</b> | <b>1.0000</b> | <b>1.0000</b> | 0.8194        | <b>1.0000</b> | <b>1.0000</b> |
| 9       | mux6                         | 7         | 64          | 2       | 0.5469        | 0.3750        | <b>0.6250</b> | 0.4688        | 0.3906        | 0.4531        | 0.5469        |
| 10      | led7                         | 8         | 3200        | 10      | 0.7294        | 0.7363        | <b>0.7375</b> | 0.7350        | 0.7303        | 0.7294        | 0.7294        |
| 11      | HTRU2                        | 9         | 17898       | 2       | 0.7083        | 0.7057        | 0.7044        | 0.7070        | <b>0.7305</b> | <b>0.7305</b> | 0.7227        |
| 12      | Nursery                      | 9         | 12960       | 3       | <b>0.7126</b> | <b>0.7126</b> | <b>0.7126</b> | <b>0.7126</b> | <b>0.7126</b> | <b>0.7126</b> | <b>0.7126</b> |
| 13      | pima                         | 9         | 768         | 9       | 0.9102        | 0.9046        | 0.9076        | <b>0.9141</b> | 0.9083        | 0.9112        | <b>0.9141</b> |
| 14      | post                         | 9         | 87          | 5       | 0.8996        | 0.8775        | 0.9322        | 0.9103        | 0.9258        | <b>0.9340</b> | 0.9174        |
| 15      | Breast Cancer                | 10        | 277         | 2       | <b>0.9751</b> | 0.8909        | 0.9663        | 0.9458        | 0.9429        | <b>0.9751</b> | <b>0.9751</b> |
| 16      | Breast Cancer Wisconsin      | 10        | 683         | 2       | <b>0.7184</b> | <b>0.7184</b> | <b>0.7184</b> | <b>0.7184</b> | <b>0.7184</b> | <b>0.7184</b> | 0.7166        |
| 17      | Contraceptive Method Choice  | 10        | 1473        | 3       | 0.4549        | 0.4542        | <b>0.4555</b> | 0.4535        | 0.4501        | 0.4542        | 0.4549        |
| 18      | glass                        | 10        | 214         | 6       | <b>0.5841</b> | 0.5514        | 0.5467        | <b>0.5841</b> | 0.5047        | 0.5701        | 0.5654        |
| 19      | shuttle-small                | 10        | 5800        | 6       | 0.9360        | 0.9645        | 0.9666        | 0.9605        | 0.9690        | <b>0.9693</b> | <b>0.9693</b> |
| 20      | threeOF9                     | 10        | 512         | 2       | 0.8145        | 0.8750        | 0.8750        | 0.8809        | 0.8652        | <b>0.8887</b> | 0.8711        |
| 21      | Tic-Tac-Toe                  | 10        | 958         | 2       | 0.7182        | <b>0.8476</b> | 0.7244        | 0.7213        | 0.7359        | 0.8340        | <b>0.8476</b> |
| 22      | MAGIC Gamma Telescope        | 11        | 19020       | 2       | 0.7520        | 0.7841        | 0.7807        | 0.7699        | 0.7875        | 0.7873        | <b>0.7880</b> |
| 23      | Solar Flare                  | 11        | 1389        | 9       | <b>0.8431</b> | <b>0.8431</b> | <b>0.8431</b> | <b>0.8431</b> | <b>0.8431</b> | <b>0.8431</b> | <b>0.8431</b> |
| 24      | heart                        | 14        | 270         | 2       | 0.8222        | 0.8185        | 0.8148        | 0.8259        | 0.7889        | 0.8259        | <b>0.8296</b> |
| 25      | wine                         | 14        | 178         | 3       | <b>0.9607</b> | 0.9494        | 0.9438        | 0.9494        | 0.9326        | 0.9270        | 0.9326        |
| 26      | cleve                        | 14        | 296         | 2       | <b>0.8176</b> | <b>0.8176</b> | 0.7804        | 0.8108        | 0.7905        | 0.7973        | 0.8108        |
| 27      | australian                   | 15        | 690         | 2       | 0.8536        | <b>0.8580</b> | 0.8493        | 0.8522        | 0.8507        | 0.8536        | 0.8507        |
| 28      | crx                          | 15        | 653         | 2       | <b>0.8622</b> | 0.8545        | 0.8545        | <b>0.8622</b> | 0.8576        | 0.8591        | <b>0.8622</b> |
| 29      | EEG                          | 15        | 14980       | 2       | 0.5774        | 0.6790        | 0.6389        | 0.6111        | 0.6670        | 0.6814        | <b>0.6935</b> |
| 30      | Congressional Voting Records | 17        | 232         | 2       | 0.9353        | <b>0.9698</b> | 0.9655        | 0.9397        | 0.9655        | 0.9655        | 0.9569        |
| 31      | zoo                          | 17        | 101         | 5       | 0.9406        | 0.9406        | 0.9307        | 0.9307        | <b>0.9505</b> | 0.9307        | <b>0.9505</b> |
| 32      | pendigits                    | 17        | 10992       | 10      | 0.8032        | 0.9062        | 0.8719        | 0.8700        | 0.9253        | 0.9290        | <b>0.9297</b> |
| 33      | letter                       | 17        | 20000       | 26      | 0.4536        | <b>0.5796</b> | 0.5068        | 0.5036        | 0.5636        | 0.5761        | 0.5779        |
| 34      | ClimateModel                 | 19        | 540         | 2       | 0.9259        | <b>0.9407</b> | 0.9222        | 0.9352        | 0.9370        | 0.9000        | 0.8667        |
| 35      | Image Segmentation           | 19        | 2310        | 7       | 0.7662        | 0.7848        | 0.7918        | 0.7922        | 0.8022        | 0.8156        | <b>0.8203</b> |
| 36      | lymphography                 | 19        | 148         | 4       | <b>0.8176</b> | 0.7027        | 0.7770        | 0.8041        | 0.7770        | 0.7500        | 0.8108        |
| 37      | vehicle                      | 19        | 846         | 4       | 0.4634        | 0.5816        | 0.5721        | 0.5922        | 0.5437        | 0.5768        | <b>0.6028</b> |
| 38      | hepatitis                    | 20        | 80          | 2       | <b>0.8750</b> | 0.8500        | 0.8625        | 0.8500        | 0.8625        | 0.5875        | 0.6625        |
| 39      | german                       | 21        | 1000        | 2       | 0.7210        | 0.7250        | <b>0.7350</b> | 0.7230        | 0.7230        | 0.7210        | 0.7240        |
| 40      | bank                         | 21        | 30488       | 2       | 0.8680        | 0.8955        | 0.8924        | 0.8777        | 0.8954        | 0.8956        | <b>0.8966</b> |
| 41      | waveform-21                  | 22        | 5000        | 3       | 0.7852        | 0.7912        | 0.7806        | 0.7814        | 0.7626        | 0.7846        | <b>0.7920</b> |
| 42      | Mushroom                     | 22        | 5644        | 2       | 0.9970        | 0.9991        | 0.9991        | 0.9972        | <b>1.0000</b> | 0.9949        | <b>1.0000</b> |
| 43      | spect                        | 23        | 263         | 2       | <b>0.7865</b> | 0.7303        | 0.7416        | 0.7715        | 0.7715        | 0.7378        | 0.7603        |
| average |                              |           |             |         | 0.7867        | 0.7801        | 0.7993        | 0.7981        | 0.7961        | 0.7963        | <b>0.8074</b> |
| p-value |                              |           |             |         | 0.0089        | 0.0054        | 0.0104        | 0.0057        | 0.0188        | 0.0301        | -             |

く、MBNC2P と MTAN-aCLL よりも分類精度が高くなっている。このように、サンプルサイズが大きくなると、より複雑な確率分布を表現できる MANB-BDeu の方が分類精度が高くなる。

MANB-BDeu はサンプルサイズが大きいデータセット 22 番, 29 番, 32 番, 33 番, 40 番で MgGBN-BDeu より分類精度が高い。この理由として、MgGBN-BDeu は BDeu を用いて近似学習しているが、MANB-BDeu は BDeu を用いて厳密学習しているため、サンプルサイズが大きくなると MANB-BDeu の方がデータの確率分布を正確に表現できることが考えられる。

また MANB-BDeu は、GBN-BDeu の学習構造の目的変数の子変数が少なく、親変数が多かったデータセット 3 番, 9 番, 31 番において、GBN-BDeu の分類精度を改善している。この理由は、5 章で述べた分類精度低下の問題を MANB-BDeu で緩和できるからである。

最後に、MANB-BDeu と ANB-BDeu を比較する。二つの違いは GBN-BDeu の目的変数のマルコフブランケットによる変数選択をしているか否かである。表 2 の "Missing variables" は、目的変数と説明変数の間で真に相関があったのに変数選択で除去してしまっ

た変数数の 10 分割交差検証における平均を示している。"Extra variables"は、目的変数と説明変数の間で真に無相関であったのに変数選択で除去しなかった変数数の 10 分割交差検証における平均を示している。実際には目的変数と説明変数の真の相関関係を知ることではできないので、交差検証で分割せずに全データを用いて学習した GBN-BDeu の目的変数のマルコフブランケットを、目的変数と真に相関のある説明変数集合とした。

結果として、MANB-BDeu はマルコフブランケットのサイズが小さいデータセット 5 番や 25 番では ANB-BDeu の分類精度を大きく上回っていることが、表 1 と表 2 の"MBsize", 表 3 からわかる。このように、目的変数と無相関な変数を除去することは分類精度を向上させることがわかる。"Extra variables"はほとんどのデータセットで少ない傾向であった。一方、"Missing variables"が多いデータセット 7 番, 39 番, 43 番では MANB-BDeu は ANB-BDeu の分類精度を大きく下回っている。このように、目的変数と相関のある変数を除去してしまうと、分類精度を低下させてしまう。表 2 より、データセット 12 番, 22 番, 40 番のようにサンプルサイズが大きい時は"Missing variables"が少なくなっている。この理由は、サンプルサイズが大きくなるとマルコフブランケットの変数選択の精度が上がるためだと考えられる。この結果から、MANB-BDeu はサンプルサイズが大きい時に、ANB-BDeu よりも分類精度が高くなると考えられる。

以上の結果は以下のようにまとめられる。

- ML によって厳密に学習した生成モデルの BNC は CLL によって近似的に学習した識別モデルの BNC より分類精度が必ずしも低いとは限らず、サンプルサイズが大きい時はむしろ生成モデルの方が分類精度が高い。
- サンプルサイズが小さいときに生成モデルの BNC では目的変数の親変数数が増え、子変数が減ると分類精度が下がるので、目的変数のマルコフブランケット内で、目的変数が説明変数を必ず子として持つ ANB 構造を制約として ML を最大化する生成モデルの厳密学習法を提案し、CLL を含む従来手法の分類精度を有意に改善することができた。

## 7. む す び

本論では、最初に ML によって厳密に学習した生成モデルの BNC と CLL によって近似的に学習した識

別モデルの BNC の分類精度を比較した。その結果、ML を最大化する生成モデルが必ずしも CLL を最大化する識別モデルより分類精度が低いとは限らないことがわかった。しかし、ML は目的変数の親変数が多く子変数が少ないような構造を学習することがあり、その場合は CLL を最大化する BNC より著しく分類精度が低くなっていることがわかった。その理由は、目的変数の親変数が多いと目的変数のパラメータ推定精度が低下し、目的変数の子変数が少ないと推定精度の低い目的変数のパラメータが分類に大きく影響するからである。次に本論では、この問題を緩和するため、ML で厳密学習した GBN の目的変数についてのマルコフブランケットのみを説明変数集合とし、ANB 構造を制約として ML により厳密学習する BNC を提案した。リポジトリデータセットを用いた実験を行ったところ、提案手法は ML を最大化する生成モデルの GBN の問題点を改善し、分類精度の著しい低下を防ぐことができた。さらに、提案手法は既存の CLL を最大化する識別モデルの BNC よりも分類精度が有意に高いことを示した。サンプルサイズが大きい時は提案手法のマルコフブランケットの変数選択精度が高まり、分類精度も高くなることがわかった。以上から、提案手法はサンプルサイズが小さい時でも分類精度が安定し、サンプルサイズが大きい時も高い分類精度を示す。

本論の重要な知見として、BNC は生成モデルよりも識別モデルの方が分類精度が良いと報告されてきたが、その理由は目的変数の親変数パラメータ数が増えた時、一つのパラメータ学習のためのデータが過疎となって学習精度が下がり、分類精度に影響していたことがわかった。そのため、識別モデルのためのスコアを必ずしも使う必要がなく、GBN の目的変数についてのマルコフブランケットに ANB の制約を入れた生成モデル学習で高い分類精度が実現できることが示唆された。

本研究はサンプルサイズが小さい時の GBN-BDeu の分類精度を改善することが目的の一つであった。近い研究として、Isozaki ら [17] [18] [19] は、少数データで有効なパラメータ学習法と学習スコアを提案している。本研究の提案手法において BDeu スコアの代わりにこれらのパラメータ学習法とスコアを用いれば、サンプルサイズの小さいデータセットの分類精度のさらなる向上が期待できる。これは今後の課題とする。

謝辞 本論の一部は Sugahara ら [20] によって発表

されている。

## 文 献

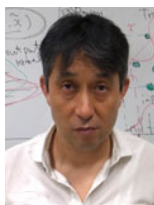
- [1] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian Network Classifiers," *Machine Learning*, vol.29, no.2, pp.131–163, 1997.
- [2] A.M. Carvalho, T. Roos, A.L. Oliveira, and P. Myllymäki, "Discriminative Learning of Bayesian Networks via Factorized Conditional Log-Likelihood," *Journal of Machine Learning Research*, vol.12, pp.2181–2210, 2011.
- [3] A.M. Carvalho, P. Adão, and P. Mateus, "Efficient Approximation of the Conditional Relative Entropy with Applications to Discriminative Learning of Bayesian Network Classifiers," *Entropy*, vol.15, no.7, pp.2716–2735, 2013.
- [4] D. Grossman and P. Domingos, "Learning Bayesian Network classifiers by maximizing conditional likelihood," *Proceedings, Twenty-First International Conference on Machine Learning, ICML 2004*, pp.361–368, 2004.
- [5] D. Heckerman, D. Geiger, and D.M. Chickering, "Learning Bayesian Networks: The Combination of Knowledge and Statistical Data," *Machine Learning*, vol.20, no.3, pp.197–243, 1995.
- [6] M. Ueno, "Learning likelihood-equivalence bayesian networks using an empirical bayesian approach," *Behaviormetrika*, vol.35, no.2, pp.115–135, 2007.
- [7] M. Ueno, "Learning Networks Determined by the Ratio of Prior and Data," *Proceedings of Uncertainty in Artificial Intelligence*, pp.598–605, 2010.
- [8] M. Ueno, "Robust learning Bayesian networks for prior belief," *Proceedings of Uncertainty in Artificial Intelligence*, pp.689–707, 2011.
- [9] J. Rissanen, *Stochastic Complexity in Statistical Inquiry Theory*, World Scientific Publishing Co., Inc., 1989.
- [10] T. Silander and P. Myllymäki, "A Simple Approach for finding the Globally Optimal Bayesian Network Structure," *Proceedings of Uncertainty in Artificial Intelligence*, pp.445–452, 2006.
- [11] M. Barlett and J. Cussens, "Advances in Bayesian Network Learning Using Integer Programming," *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*, pp.182–191, 2013.
- [12] M. Minsky, "Steps toward Artificial Intelligence," *Proceedings of the IRE*, vol.49, pp.8–30, 1961.
- [13] M.G. Madden, "On the classification performance of TAN and general Bayesian networks," *Knowledge-Based Systems*, pp.489–495, 2009.
- [14] M. Lichman, "UCI machine learning repository," 2013. <http://archive.ics.uci.edu/ml>
- [15] G. Hommel, "A stagewise rejective multiple test procedure based on a modified bonferroni test," *Biometrika*, pp.383–386, 1988.
- [16] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol.7, pp.1–30, 2006.
- [17] T. Isozaki, N. Kato, and M. Ueno, "Minimum free energies with "data temperature" for parameter learning of bayesian networks," 2008 20th IEEE International Conference on Tools with Artificial Intelligence, vol.1, pp.371–378, 2008.
- [18] T. Isozaki, N. Kato, and M. Ueno, "data temperature" in minimum free energies for parameter learning of bayesian networks," *International Journal on Artificial Intelligence Tools*, vol.18, pp.653–671, 2009.
- [19] T. Isozaki and M. Ueno, "Minimum free energy principle for constraint-based learning bayesian networks," *Machine Learning and Knowledge Discovery in Databases*, pp.612–627, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
- [20] S. Sugahara, M. Uto, and M. Ueno, "Exact learning augmented naive bayes classifier," *Proceedings of the Ninth International Conference on Probabilistic Graphical Models*, vol.72, pp.439–450, *Proceedings of Machine Learning Research*, PMLR, 2018.

(平成 xx 年 xx 月 xx 日受付)



菅原聖太

2018 年電気通信大学情報理工学部卒。同年、同大学院情報理工学研究科情報・ネットワーク工学専攻博士前期課程入学、現在に至る。



植野真臣 (正員)

1992 年神戸大学大学院教育学研究科修了, 1994 年東京工業大学大学院総合理工学研究科修了。博士(工学)。東京工業大学, 千葉大学, 長岡技術科学大学を経て 2006 年より電気通信大学助教授, 2013 年より教授, 現在に至る。

**Abstract** Bayesian network classifier (BNC) is known as a highly accurate classifier for discrete variables. Earlier reports have described that classification accuracies of BNCs as discriminative models were higher than those as generative models. However, the reports stated no reason why discriminative models outperformed generative models. The present study conducted experiments to compare the classification accuracies of BNCs as discriminative models and ones as generative models. Results demonstrate that the classification accuracies of BNCs as generative models are higher than those as discriminative models when the sample size is large. However, the results also show that the classification accuracies of BNCs as generative models are much worse than those as discriminative models because the class variable has numerous parents and the number of parameters increase exponentially, when the sample size is small. To resolve the difficulty, this study proposes an exact learning the BNC as a generative model with Augmented Naive Bayes (ANB) structure constraint in which the class variable has no parents. The set of feature variables of the ANB as a generative model is the Markov blanket for the class variable of the unrestricted BNC. Therefore, the proposed method first learns exactly the unrestricted BNC. Then, the proposed method learns the BNC with the ANB structure constraint in the Markov blanket for the class variable of the unrestricted BNC. Some comparison experiments demonstrate that the proposed method outperforms the other methods to learn traditional discriminative models.

**Key words** Bayesian networks, classifiers, generative models, machine learnings, probabilistic graphical models.