

# Augmented Naive Bayes による大規模ベイジアンネットワーク分類器学習

菊谷 成慎<sup>†a)</sup>      菅原 聖太<sup>†b)</sup>      名取 和樹<sup>†c)</sup>      植野 真臣<sup>†d)</sup>

## Learning Huge Bayesian Network Classifier with Augmented Naive Bayes

Naruchika KIKUYA<sup>†a)</sup>, Shouta SUGAHARA<sup>†b)</sup>, Kazuki NATORI<sup>†c)</sup>, and Maomi UENO<sup>†d)</sup>

あらまし ベイジアンネットワーク分類器 (Bayesian Network Classifier: BNC) は確率モデルによる分類器である。BNC の構造学習はこれまで候補構造から近似的に識別モデルの学習スコアが最大となる構造を探索する手法が用いられてきた。近年、菅原ら (2020) はベイジアンネットワークの学習スコアにより厳密学習した BNC は分類精度が低いとは限らないと報告している。更に彼らは、構造に Augmented Naive Bayes (ANB) を仮定し生成モデルとして BNC を厳密学習する手法を提案した。これによりデータが少ない場合も分類精度の高い BNC を学習できることを示した。しかし、厳密学習は変数の増加に対して計算量が指数的に増加するため、菅原ら (2020) の手法は数十変数の ANB 学習が限界である。そこで、本論文では大規模変数をもつ BNC を学習できる手法を提案する。因果モデルの研究分野では、条件付き独立性検定 (CI テスト) とエッジの方向付けによる計算効率の高い構造学習法が提案されており、制約ベースアプローチと呼ぶ。名取らは、CI テストに Bayes factor を用いることで真の構造への漸近一致性を有しつつ 1000 変数以上の構造学習を実現し、本田らはその手法に推移性を組み込むことで 3500 変数の構造学習を実現した。そこで本論文では、本田らの手法を用いて従来より大規模な ANB を学習できる手法を提案し、提案手法が ANB 構造について漸近的にパラメータ数を最小にして真の同時確率分布に収束することを示す。実験により、大規模構造学習において提案手法が有用であることを示す。

キーワード ベイジアンネットワーク, 確率的グラフィカルモデル, 機械学習, 分類器, 制約ベースアプローチ

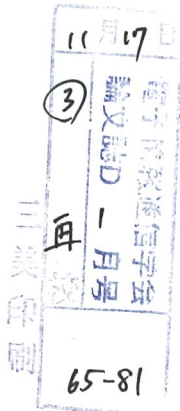
### 1. まえがき

ベイジアンネットワークは、離散確率変数をノードで表しノード間の依存関係を非循環有向グラフ (Directed Acyclic Graph: DAG) で表現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は、確率構造に DAG を仮定することにより、同時確率分布を条件付き確率の積に分解する。ベイジアンネットワークの構造は一般にデータから推定する必要があり、これをベイジアンネットワークの構造学習と呼ぶ。ベイジアンネットワークの構造学習法として、真の構造への漸近

一致性を有する学習スコアを用いて、候補構造から最適なスコアをもつ構造を探索する厳密解探索アプローチが従来から用いられている。また、一般に学習スコアとして周辺ゆ度が用いられている [1]。

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 [2]。BNC では説明変数を所与とした目的変数の条件付き確率を表現する識別モデルのほうが、全変数の同時確率を表現する生成モデルよりも分類精度が高いことが報告されている [1], [3]~[5]。しかし近年、Sugahara ら [6] は、データが十分大きいとき、生成モデルのほうが分類精度が高いことを示した。一方で、データが少なく目的変数の親変数が多い構造をとる場合、パラメータ学習に用いるデータがスパースになり、精度が著しく低下してしまうと指摘した。この問題を解決するために、彼らは目的変数が親をもたず、全ての

<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: kikuya@ai.lab.uec.ac.jp  
b) E-mail: sugahara@ai.lab.uec.ac.jp  
c) E-mail: natori@ai.lab.uec.ac.jp  
d) E-mail: ueno@ai.lab.uec.ac.jp  
DOI:10.14923/transinfj.2020JDP7008



説明変数の子にもつ構造を仮定した Augmented Naive Bayes (ANB) [2] を厳密学習する手法を提案した。これにより、データが少ない場合でも高い分類精度を得られることを示した。更に、菅原ら [7] は ANB 構造を厳密学習する効率的なアルゴリズムを提案した。しかし、彼らは動的計画法を用いており、数十変数の構造学習が限界であるため大規模変数には適用できない。

一方、因果モデル分野では、真の構造への漸近一致性はもたないが、より計算効率の高い構造学習法が提案されている。この手法は制約ベースアプローチと呼ばれ、完全無向グラフに、二変数間の条件付き独立性検定 (Conditional Independence test: CI テスト) を適用して学習される無向グラフに対し、オリエンテーションルール [8] によるエッジの方向付けを行うことで構造を学習する。制約ベースアプローチの研究では、PC アルゴリズム [9]、TPDA アルゴリズム [10]、MMHC アルゴリズム [11]、RAI アルゴリズム [12] が提案されている。名取らは従来の CI テストがもたない、独立性に対する漸近一致性を有する Bayes factor を RAI アルゴリズムの CI テストに用いることで、1000 変数以上の大規模構造学習を実現している [13]~[15]。更に、本田らは RAI アルゴリズムにベイジアンネットワークの推移性を組み込むことで CI テスト数と計算時間を削減し、3500 変数の大規模構造学習を実現している [16]。

そこで本論文では、本田ら [16] の手法を ANB 構造を仮定した BNC の学習に拡張することで、従来よりも大規模な BNC を学習できる手法を提案する。また、提案手法が ANB についてパラメータ数を最小にして真の同時確率分布に収束することを証明する。更に、ランダムネットワークを用いたシミュレーション実験とベンチマークデータセットを用いた数値実験により、提案手法が厳密探索アプローチよりも大規模なネットワークを学習し、大規模ネットワークにおいて従来手法よりも高い分類精度をもつことを示す。また、BNC 以外の分類器と提案手法を比較することで、提案手法の特徴を分析する。

## 2. ベイジアンネットワーク

ベイジアンネットワークは、確率変数をノードとし、ノード間の依存関係を非循環有向グラフ (Directed Acyclic Graph: DAG) と各ノードの条件付き確率で表現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の構造は確率変数間の因果関係を表して

いるため、高い説明可能性をもつ。例えば、図 1 は肺がんに関するベイジアンネットワークである。図 1 の構造から、公害病や喫煙は肺がんの原因となり、肺がんを患うと呼吸困難を発症したり X 線検査で異常が見つかるという因果関係を表現している。今、離散確率変数集合  $\mathbf{X} = \{X_0, X_1, \dots, X_n\}$  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  $k$  をとる ( $X_i = k$  と書く) とする。このとき、ベイジアンネットワークの構造  $G$  において、各変数  $X_i$  の親変数集合を  $\Pi_i$  としたときの同時確率分布  $P(X_0, X_1, \dots, X_n | G, \Theta)$  は以下のように表現できる。

$$P(X_0, X_1, \dots, X_n | G, \Theta) = \prod_{i=0}^n P(X_i | \Pi_i, G). \quad (1)$$

ここで、 $\theta_{ijk}$  を  $\Pi_i$  が  $j$  番目のパターンをとるとき ( $\Pi_i = j$  と書く) に  $X_i = k$  となる条件付き確率  $P(X_i = k | \Pi_i = j, G)$  を示すパラメータとし、パラメータ集合  $\Theta = \{\theta_{ijk}\}, (i = 0, \dots, n; j = 1, \dots, q_i; k = 1, \dots, r_i)$  とする。また、 $q_i = \prod_{l: X_l \in \Pi_i} r_l$  である。

変数  $X$  から変数  $Y$  までの隣接している変数の連なりを二変数  $X, Y$  を結ぶ道と言う。道上で、三変数  $A, B, C$  が  $A \rightarrow C \rightarrow B, A \leftarrow C \rightarrow B, A \rightarrow C \leftarrow B$  と結合することをそれぞれ、変数  $C$  で逐次結合、分岐結合、合流結合すると言う。ベイジアンネットワーク構造は二変数を結ぶ道をブロックすることで条件付き独立性を表現する。ブロックを以下で定義する。

[定義 2.1] 二変数  $X, Y$  を結ぶ道  $p$  が以下のいずれかの条件を満たすとき、道  $p$  は変数集合  $\mathbf{Z}$  でブロックされる。

- (1) 道  $p$  が変数  $Z \in \mathbf{Z}$  で逐次結合か分岐結合をする。
- (2) 道  $p$  が変数  $Z \notin \mathbf{Z}$  で合流結合をし、かつ  $Z$  の子孫が  $\mathbf{Z}$  に属さない。

二変数  $X, Y$  を結ぶ全ての道が変数集合  $\mathbf{Z}$  でブロックされるとき、ネットワーク構造  $G$  において  $X, Y$  は  $\mathbf{Z}$  を

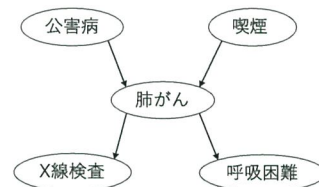


図 1 肺がんのベイジアンネットワーク  
Fig. 1 Lung cancer Bayesian network.

所与として条件付き独立であり、 $I_G(X, Y|Z)$ と表す。また、真の同時確率分布において  $X$  と  $Y$  が  $Z$  を所与として条件付き独立であるとき、 $I_P(X, Y|Z)$ と書くとする。それに対し、真の同時確率分布において  $X$  と  $Y$  が  $Z$  を所与として従属であるとき、 $D_P(X, Y|Z) \equiv \neg I_P(X, Y|Z)$ と書くとする。

ベイジアンネットワークは全変数の同時確率分布を DAG 構造  $G$  とパラメータ集合  $\Theta$  で表現する。しかし、DAG は同時確率分布における全ての条件付き独立性を表現できるわけではない。そのため、ベイジアンネットワークは DAG で表現できる条件付き独立性のみを扱うこととして、以下の I-map を導入する。

[定義 2.2] ベイジアンネットワーク構造  $G$  が表現する全ての条件付き独立性が真の条件付き独立性に含まれているとき、 $G$  をインディペンデントマップ (independent map: I-map) という。

$$\forall X, Y \in \mathbf{V}, Z \subseteq \mathbf{V} \setminus \{X, Y\}, I_G(X, Y|Z) \Rightarrow I_P(X, Y|Z).$$

I-map が表現する同時確率分布は漸近的に真の分布に確率収束する。また、真のベイジアンネットワークとその I-map との間には以下の関係が成り立つ。

[定理 2.1] 真の構造を  $G^* = (\mathbf{V}, \mathbf{E}^*)$  とおく。ある構造  $G = (\mathbf{V}, \mathbf{E})$  が以下の三つの条件を満たすとき、 $G$  は I-map である。ここで、 $E_{XY}$  は  $XY$  間のエッジを表す。

(1)  $\forall X, Y \in \mathbf{V}$  について、 $E_{XY} \in \mathbf{E}^*$  ならば  $E_{XY} \in \mathbf{E}$ 。

(2)  $\forall X, Y, Z \in \mathbf{V}$  について、 $G^*$  において合流結合  $X \rightarrow Z \leftarrow Y$  するならば、 $E_{XY} \in \mathbf{E}$  または  $G$  において合流結合  $X \rightarrow Z \leftarrow Y$  する。

(3)  $\forall X, Y, Z \in \mathbf{V}$  について、 $G^*$  において逐次結合  $X \rightarrow Z \rightarrow Y$  または分岐結合  $X \leftarrow Z \rightarrow Y$  するならば、 $E_{XY} \in \mathbf{E}$  または、 $G$  において逐次結合  $X \rightarrow Z \rightarrow Y$  または分岐結合  $X \leftarrow Z \rightarrow Y$  する。

証明は Chickering [17] を参照してほしい。

## 2.1 ベイジアンネットワークのパラメータ推定

パラメータの事前分布としてディリクレ分布  $P(\Theta)$  を仮定する。変数集合  $\mathbf{X}$  に対する  $N$  個のデータを  $\mathbf{D} = \{D_1, \dots, D_N\}$  とすると、事後分布  $P(\Theta | D, G)$  が得られる。

$$P(\Theta) = \prod_{i=0}^n \prod_{j=1}^{q_i} \frac{\Gamma(\sum_{k=1}^{r_i} \alpha_{ijk})}{\prod_{k=1}^{r_i} \Gamma(\alpha_{ijk})} \prod_{k=1}^{r_i} \theta_{ijk}^{\alpha_{ijk}-1}. \quad (2)$$

$$P(\Theta | D, G) \quad (3)$$

$$= \prod_{i=0}^n \prod_{j=1}^{q_i} \frac{\Gamma(\sum_{k=1}^{r_i} (\alpha_{ijk} + N_{ijk}))}{\prod_{k=1}^{r_i} \Gamma(\alpha_{ijk} + N_{ijk})} \prod_{k=1}^{r_i} \theta_{ijk}^{\alpha_{ijk} + N_{ijk} - 1}.$$

ここで、 $N_{ijk}$  は、 $X_i$  の親変数集合  $\Pi_i$  が  $j$  番目のパターンを取ったときの  $X_i = k$  となるデータの頻度を表し、 $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$  を表す。また、 $\alpha = \{\alpha_{ijk}\}, (i = 0, \dots, n; j = 1, \dots, q_i; k = 1, \dots, r_i)$  はディリクレ事前分布のハイパーパラメータであり、 $\alpha_{ij} = \sum_{k=1}^{r_i} \alpha_{ijk}$  である。ベイジアンネットワークでは、パラメータ推定値として、期待事後確率推定値 (Expected a Posteriori: EAP) が最もよく用いられる。EAP は式 (3) に期待値をとることで以下を得る。

$$\hat{\theta}_{ijk} = \frac{\alpha_{ijk} + N_{ijk}}{\alpha_{ij} + N_{ij}}. \quad (4)$$

## 2.2 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。これをベイジアンネットワークの構造学習という。構造学習法として、候補構造から最適なスコアをもつ構造を探索する厳密解探索アプローチが従来から用いられている。一般に学習スコアとして周辺ゆう度  $P(\mathbf{D} | G)$  が用いられる。周辺ゆう度は、式 (3) の事後分布からパラメータ推定値を周辺化することで閉形式で表せる。

$$P(\mathbf{D} | G) = \prod_{i=0}^n \prod_{j=1}^{q_i} \frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij} + N_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(\alpha_{ijk} + N_{ijk})}{\Gamma(\alpha_{ijk})}. \quad (5)$$

ハイパーパラメータを  $\alpha_{ijk} = 1$  (事前分布が一様分布) とした学習結果は、Bayesian Information Criterion (BIC) [18]、Minimum Description Length (MDL) [19] の結果に漸近的に一致することが知られている [20]。近年では、 $\alpha_{ijk} = \alpha / (r_i q_i)$  とした Bayesian Dirichlet equivalent uniform (BDeu) が最も用いられる [1], [21]。ここで、 $\alpha$  は Equivalent Sample Size (ESS) と呼ばれる事前知識の重みを示す疑似サンプルである。

BDeu による厳密解探索アプローチは以下で定義される漸近一致性を有する。

[定義 2.3] データ数  $N \rightarrow \infty$  のとき、ある構造学習手法で推定された構造がパラメータ数最小の I-map に概収束するならば、その手法はベイジアンネットワーク構造についての漸近一致性を有するという。

ただし、この手法は NP 困難であり、変数に対して計



算量が指数的に増加してしまう。厳密解を効率的に探索するために、動的計画法 [22]~[26], A\* 探索 [27], 整数計画法 [28] などが提案されている。しかし, 最先端手法 [28] でさえ 60 変数程度の構造学習が限界であり, 大規模構造を学習できない。

### 3. ベイジアンネットワーク分類器

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, それ以外のノードを説明変数とすることで, ベイジアンネットワークを分類器として扱うことができる。分類器としてのベイジアンネットワークをベイジアンネットワーク分類器 (Bayesian Network Classifier: BNC) と呼ぶ [2]。今,  $X_0$  を目的変数とし,  $X_1, \dots, X_n$  を説明変数とした BNC を考える。説明変数のデータ  $\langle x_1, \dots, x_n \rangle$  が与えられたとき, 目的変数の推定値  $\hat{c}$  は以下のように得られる。

$$\begin{aligned} \hat{c} &= \arg \max_{c \in \{1, \dots, r_0\}} P(c | x_1, \dots, x_n, G, \Theta) \quad (6) \\ &= \arg \max_{c \in \{1, \dots, r_0\}} \prod_{i=0}^n \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} (\theta_{ijk})^{1_{ijk}} \\ &= \arg \max_{c \in \{1, \dots, r_0\}} \left[ \prod_{j=1}^{q_0} \prod_{k=1}^{r_0} (\theta_{0jk})^{1_{0jk}} \right. \\ &\quad \left. \times \prod_{i: X_i \in \mathbf{Ch}} \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} (\theta_{ijk})^{1_{ijk}} \right]. \end{aligned}$$

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  $\Pi_i = j$  のときに 1 をとり, それ以外の場合は 0 をとる変数である。ここで,  $\mathbf{Ch}$  は目的変数の子集合である。式 (6) より, 目的変数  $X_0$  の親集合, 子集合,  $X_0$  と子を共有している変数集合の和集合のみで推定値  $\hat{c}$  が求まる。また, この集合を  $X_0$  のマルコフブランケット (Markov Blanket) と呼ぶ。BNC の構造学習は通常, ベイジアンネットワークと同様にデータから学習する。一般的なベイジアンネットワークを用いた BNC を General Bayesian Network (GBN) と呼ぶ (図 2(a))。BDeu など学習された GBN は目的変数が子変数をもたず, 親変数を多くもつ構造をとることがある。このような構造では, 式 (6) が目的変数の条件付きパラメータ  $\theta_{0jk}$  のみに依存する。更に, 目的変数の親変数集合のとりうる値のパターン数  $q_0$  が大きくなり, パラメータの推定に用いられるデータ数  $N_{0jk}$  が小さくなる。したがって,  $\theta_{0jk}$  のパラメータの分類精度が低

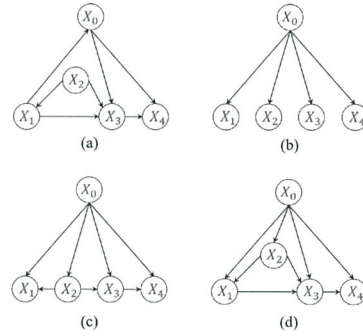


図 2 (a) GBN の例; (b) Naive Bayes の例; (c) TAN の例; (d) ANB の例

Fig. 2 (a) Example of GBN; (b) Example of Naive Bayes; (c) Example of TAN; (d) Example of ANB.

下するため, 結果として分類精度が著しく低下することが多い [29], [30]。この問題を緩和する BNC として, 目的変数が全ての説明変数の子にもち, 説明変数間が独立であると仮定する Naive Bayes [31] (図 2(b)), 目的変数が全ての説明変数の子にもち, 説明変数間で木構造をとると仮定した Tree-Augmented Naive Bayes (TAN) [2] (図 2(c)) が提案されている。ゆう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ることが知られている [2], [32]。また, Naive Bayes や TAN を一般化した, より表現力の高いモデルとして, 目的変数が全ての説明変数の子にもつことを仮定する Augmented Naive Bayes (ANB) [2] (図 2(d)) が知られている。一般のベイジアンネットワークと同様に, ANB 構造から説明変数間の因果関係を解釈できる。

周辺ゆう度を最大にする GBN は全変数の同時確率を表現する生成モデルであるが, 説明変数を所与とした目的変数の条件付き確率を表現した識別モデルのほうが, 漸近的な分類精度が高いことが報告されている [4], [5]。しかし先行研究では識別モデルとしての GBN がベイジアンネットワーク構造についての漸近一致性をもつ生成モデルとしての GBN よりも分類精度が高くなる根拠は理論的に示されていない。また, 分類精度の比較実験において, 生成モデルとしての BNC は厳密解探索アプローチを用いて学習できるにもかかわらず近似学習を用いている。そこで, 菅原ら [6], [7] は周辺ゆう度で厳密学習した GBN は, データが十分なとき識別モデルの GBN よりも分類精度が高いことを示した。しかし, データが少ないとき生成モデルの GBN では目的変数の親変数数が過多になる場合があり, このとき分類精度が著しく低下してしまうことを

実験で示した。この問題を解決するため、彼らは目的変数が全ての変数の親変数になる ANB を生成モデルとして厳密学習する手法を提案した。結果は、提案手法はデータが少ない場合でも安定した分類精度を得ることができ、識別モデルの BNC より有意に分類精度が高いことを示した。しかし、厳密解探索アプローチは数十変数の学習が限界であり、変数数が多い場合に適用することができない。そこで本論文では大規模変数をもつ BNC を学習できる手法を提案する。

#### 4. 制約ベースアプローチ

因果モデル分野では、計算量を大幅に削減できる制約ベースアプローチと呼ばれる構造学習法が提案されている。本章では制約ベースアプローチの概要と Bayes factor を用いた RAI アルゴリズムについて紹介する。

##### 4.1 Bayes factor を用いた RAI アルゴリズム

今、真のベイジアンネットワークを次のように仮定する。

[仮定 1]  $G^* = (\mathbf{V}, \mathbf{E}^*)$  を真のベイジアンネットワーク構造とし、 $\mathbf{V}$ ,  $\mathbf{E}^*$  をそれぞれ、 $G^*$  に含まれる変数集合、エッジ集合とする。このとき、 $G^*$  は以下を満たす。

$$\forall X, \forall Y \in \mathbf{V}, \forall Z \subseteq \mathbf{V} \setminus \{X, Y\},$$

$$I_P(X, Y | Z) \Leftrightarrow E_{XY} \notin \mathbf{E}^*.$$

次に、ベイジアンネットワークの統計的識別可能性についての性質であるマルコフ等価性を定義する。

[定義 4.1]  $G_1, G_2$  を変数集合  $V$  からなる二つの DAG とする。  $\forall X, \forall Y \in \mathbf{V}, \forall Z \subseteq \mathbf{V}$  に対して以下が成り立つとき、 $G_1$  と  $G_2$  はマルコフ等価である。

$$I_{G_1}(X, Y | Z) \Leftrightarrow I_{G_2}(X, Y | Z).$$

一つのマルコフ等価なクラスを表すグラフは無向エッジを使って表現されるのが一般的であり、そのような部分グラフを含む DAG は PDAG (Partially DAG) と呼ばれる。制約ベースアプローチでは仮定 1 を満たす真のベイジアンネットワーク構造が存在するとき、その構造とマルコフ等価なクラスの PDAG を推定することを目的とする。

このアプローチの基本的なアルゴリズムは以下のとおりである。

- (1) 完全無向グラフを生成する。

- (2) (1) で生成された完全無向グラフに対し条件付き独立性検定 (Conditional Independence test: CI テスト) によりエッジを削除する。

- (3) (2) で得られた無向グラフに対してオリエンテーションルール [8] を用いて方向付けを行う。

一般に、制約ベースアプローチの学習精度は CI テストの精度に依存し、学習速度は実行する CI テスト数に依存する。制約ベースアプローチとして、PC アルゴリズム [9], TPDA アルゴリズム [10], MMHC アルゴリズム [11], RAI アルゴリズム [12] が提案されてきた。しかし、これらのアルゴリズムでは  $\chi^2$  検定,  $G^2$  検定, 条件付き相互情報量などを CI テストとして用いており、以下で定義される漸近一致性をもたない。

[定義 4.2] データ数  $N \rightarrow \infty$  のとき、任意の変数間について、CI テストが確率 1 で真の条件付き独立性を判定するとき、その CI テストは独立性に対する漸近一致性を有するという。

一方で、Steck らは、二変数間が従属・独立モデルの周辺ゆう度の比による Bayes factor を用いた CI テストを提案した [33]。例として、ベイジアンネットワークの二変数  $X, Y$  において  $X$  と  $Y$  間について各変数の共通親変数集合を  $Z$  としたときの従属モデルを  $G_1$ 、独立モデルを  $G_2$  とし、それぞれ図 3 の (a), (b) に示す。このときの Bayes factor を  $BF(X, Y | Z)$  とすると、対数 Bayes factor は次のように表せる。

$$\log BF(X, Y | Z) = \log \frac{P(\mathbf{D} | G_1, \alpha)}{P(\mathbf{D} | G_2, \alpha)}. \quad (7)$$

ここで、 $P(\mathbf{D} | G_1, \alpha)$ ,  $P(\mathbf{D} | G_2, \alpha)$  は式 (5) の BDeu を用いる。Bayes factor を用いた CI テストでは、対数 Bayes factor が 0 以上か否かで図 3 の (a), (b) のどちらを選択するか判断する。しかし、Steck らはこれを理論分析に用いるだけでベイジアンネットワーク学習に用いていない。一方で、名取らは Bayes factor を用いた CI テストが独立性に対する漸近一致性をもつことを示した [13]~[15]。

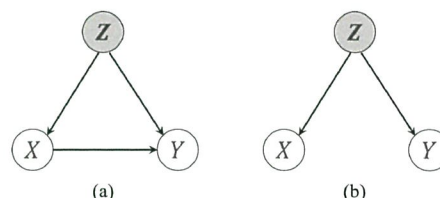


図 3 (a) 従属モデル  $G_1$ ; (b) 独立モデル  $G_2$   
Fig. 3 (a) Dependent model; (b) independent model.

[定理 4.1] データ数  $N \rightarrow \infty$  のとき,

(1) 真の構造が  $\mathbf{Z}$  を所与として  $X$  と  $Y$  が条件付き独立でないとき, 確率 1 で  $\log BF(X, Y | \mathbf{Z}) > 0$ .

(2) 真の構造が  $\mathbf{Z}$  を所与として  $X$  と  $Y$  が条件付き独立のとき, 確率 1 で  $\log BF(X, Y | \mathbf{Z}) < 0$ .

証明は名取ら [15] を参照してほしい. 更に, Bayes factor を用いた CI テストを RAI アルゴリズムに組み込んだ手法を提案し, ペイジアンネットワーク構造についての漸近一致性を有しつつ 1000 変数を超える大規模構造学習を実現した. 証明は名取ら [15] を参照してほしい.

#### 4.2 推移性を用いた RAI アルゴリズム

前述のように, 制約ベースの学習速度は CI テストの回数に依存する. 一般に CI テストの回数は学習の早期にエッジを削除するほど削減できる. そのために以下の推移性を用いる手法が提案されている.

[定理 4.2]  $G = (\mathbf{V}, \mathbf{E})$  を DAG とし,  $X, Y \in \mathbf{V}$  で,  $Y$  は  $X$  の非子孫とする. このとき,  $A \in \mathbf{V} \setminus (\{X, Y\} \cup \mathbf{Pa}(X, G) \cup \mathbf{W})$  とすると, 以下が成り立つ.

$$I_G(X, Y | \mathbf{Pa}(X, G)) \quad (8)$$

$$\rightarrow I_G(X, A | \mathbf{Pa}(X, G)) \text{ or } I_G(A, Y | \mathbf{Pa}(X, G)).$$

ここで,  $\mathbf{W}$  は  $X$  の子孫であり,  $X$  と  $Y$  が合流結合する変数とその子孫からなる変数集合を表し,  $\mathbf{Pa}(X, G)$  は  $X$  の  $G$  における親変数集合を表す. 証明は本田ら [16] を参照してほしい. 定理 4.2 より, ある二変数の条件付き独立性からその各変数と他変数との条件付き独立性の少なくとも一つを保証できる. これを利用することで, 二変数の条件付き独立性から少なくとも一つの条件付き独立が保証される二組のエッジを列挙できる. 列挙されたエッジの CI テストを優先して行うことで, より学習の早期にエッジを削除できる. 本田らは以上の推移性によるエッジの削除法を Bayes factor を用いた RAI アルゴリズムに組み込むことで CI テストの回数を削減し, 3500 変数の大規模構造学習を実現した.

本論文では, 本田ら [16] の手法を用いることで, 従来よりも大規模な BNC の構造学習の実現する.

## 5. 提案手法

ANB は識別モデルとして扱われていたが, 菅原らは BNC の探索空間において, 構造に ANB を仮定し, 生成モデルとして BNC を厳密学習する手法を提案した. 更に, 学習した ANB 構造を用いることで高い分類精度を得られることを示した [6], [7]. 制約ベースアプロー

チの構造学習法においても初期構造や CI テストの実行範囲に制約を設けることで ANB 構造を学習できる. 本章ではまず, 本田らの手法に基づいて ANB 構造を学習するアルゴリズムを説明する. 更に, 提案手法が以下で定義される漸近一致性を有することを示す.

[定義 5.1] データ数  $N \rightarrow \infty$  のとき, ある構造学習手法で推定された構造が I-map となる ANB (I-map ANB と呼ぶ) の中でパラメータ数最小の構造に概収束するならば, その手法は ANB 構造についての漸近一致性を有するという.

### 5.1 RAI アルゴリズムを用いた ANB 構造学習

まず本田らの手法 [16] の基本的な動作を紹介する. 今, グラフを  $G = (\mathbf{V}, \mathbf{E})$  と表し,  $\mathbf{V}, \mathbf{E}$  はそれぞれ  $G$  に含まれる変数集合, エッジ集合を表す. ここで,  $G$  は有向エッジと無向エッジを併せもつとする. また,  $G_{ex} = (\mathbf{V}_{ex}, \mathbf{E}_{ex})$  を RAI アルゴリズムによって分割された部分グラフとする.

(1) 完全無向グラフ  $G_{uc}$  とデータ  $\mathbf{D}$  を入力する.

(2) 各次数の CI テストにおいて  $\log BF(X, Y | \mathbf{Z}) < 0$  となり  $X, Y$  が条件付き独立と判定されるとき,  $XY$  間のエッジを削除する. また, 1 次以降の CI テストによるエッジの削除の直後, 推移性に基づくエッジの削除を行う.

(3) (2) で得られたグラフに対してオリエンテーションルールを適用して方向付けを行う.

(4) 方向付けの結果から部分グラフ  $G_{ex}$  に分割する.

(5) 各部分グラフで再帰的に RAI を呼び出す.

ここで, Bayes factor を用いた CI テストは漸近的に真の条件付き独立性を判定し, 本田らの手法は漸近的に真の構造を学習する [16].

以上の本田らの手法を用いて ANB 構造を学習する方法を説明する. ANB は目的変数が親をもたず, 全ての説明変数の子にもつ構造である. 上記のアルゴリズムにおいて手順 (1) の初期グラフと, 手順 (2) の CI テストに変更を加えることで RAI アルゴリズムを用いた ANB 学習を実現する.

まず手順 (1) において, 入力する初期グラフを全変数集合からなる完全無向グラフではなく, 説明変数集合からなる完全無向グラフとする. 次に手順 (2) において, ANB 構造では説明変数は必ず目的変数を親にもつので, CI テストを実施する際に目的変数  $X_0$  を所与とするように制約を加える. Algorithm1 に RAI アルゴリズムにおける Bayes factor を用いた CI テストの詳細



**Algorithm 1** CI test for ANB

```

1: function CI_test( $n_z, G_s, G_{ex}, X_0, D$ )
    $n_z$ : CI テストの次数
    $G_s = (V_s, E_s)$ : 入力グラフ
    $G_{ex}$ : 分割されたグラフの集合
    $G_{all} = (V, E)$ : CI テストと方向付けによって得られる出力グラフ
    $X_0$ : 目的変数
   // CI テストによるエッジの削除
2: for  $G_{ex} = (V_{ex}, E_{ex}) \in G_{ex}$  do
3:   for  $X \in V_s, Y \in V_{ex}$  do
4:     for  $Z \subseteq Pa_P(X, G_s) \cup Pa(X, G_{ex}) \setminus \{Y\}$  do
5:       if  $|Z| = n_z - 1$  かつ  $\log BF(X, Y | Z \cup \{X_0\}) < 0$  then
6:          $E_{all} \leftarrow E_{all} \setminus \{E_{XY}\}$ 
            $\triangleright E_{XY}$ :  $XY$  間のエッジ
           // 推移性によるエッジの削除
7:         TRANSITIVE_CUT( $G_s, G_{all}, X, Y, Z, X_0, D$ )
8:       end if
9:     end for
10:   end for
11: end for
12: for  $X \in V_s, Y \in V_s$  do
13:   for  $Z \subseteq Pa_P(X, G_s) \cup Pa(X, G_{ex}) \setminus \{Y\}$  do
14:     if  $|Z| = n_z - 1$  かつ  $\log BF(X, Y | Z \cup \{X_0\}) < 0$  then
15:        $E_{all} \leftarrow E_{all} \setminus \{E_{XY}\}, E_s \leftarrow E_s \setminus \{E_{XY}\}$ 
           // 推移性によるエッジの削除
16:       TRANSITIVE_CUT( $G_s, X, Y, Z, X_0, D$ )
17:     end if
18:   end for
19: end for
20: return ( $G_s, G_{all}$ )
21: end function

```

細を示す。ここで、入力グラフを  $G = (V_s, E_s)$  と表す。 $\text{Adj}(X, G)$  はグラフ  $G$  における変数  $X$  の隣接変数集合を表し、 $\text{Ch}(X, G)$  はグラフ  $G$  における変数  $X$  の子変数集合を表す。このとき、 $\text{Pa}_P(X, G)$  は  $\text{Adj}(X, G) \setminus \text{Ch}(X, G)$  を表し、 $\text{Pa}(X, G)$  はグラフ  $G$  に存在する変数  $X$  の親変数集合を表す。また、 $\text{Pa}(X, G)$  はグラフ集合  $\mathbf{G}$  において  $\bigcup_{G \in \mathbf{G}} \text{Pa}(X, G)$  を表す。関数  $\text{CI\_test}$  は入力グラフの変数集合  $V_s$  に対し、各次数の CI テストにおいて  $\log BF(X, Y | Z) < 0$  となり  $X, Y$  が条件付き独立と判定されるとき、 $XY$  間のエッジ  $E_{XY}$  を削除する。ANB 学習では目的変数と説明変数は必ず接続されているため、CI テストの実施する際に目的変数  $X_0$  を条件部に必ず含める。つまり、 $Z \cup \{X_0\}$  を所与とした CI テストを実施する (5 行目と 14 行目)。また、CI テストによるエッジ削除の直後 (7 行目と 16 行目) に関数  $\text{TRANSITIVE\_CUT}$  を呼び出す。ここでは推移性によるエッジの削除を行う。Algorithm2 に詳細を示す。関数  $\text{TRANSITIVE\_CUT}$  では CI テストで推定した条件付き独立性から推移性により検出された変数集合  $\mathbf{A}$  に対して CI テストを行うが、ここでも目的変数  $X_0$  を

**Algorithm 2** Edge cutting with transitivity for ANB

```

1: function TRANSITIVE_CUT( $G_s, G, X, Y, Z, X_0, D$ )
    $G_s = (V_s, E_s)$ : 自律的部分構造
    $G = (V, E)$ : 全体グラフ
    $X, Y, Z$ :  $I_P(X \perp Y | Z)$  となる二変数  $X, Y$  と変数集合  $Z$ 
    $X_0$ : 目的変数
2:  $\mathbf{A} \leftarrow \text{Adj}(X, G) \cap \text{Adj}(Y, G) \setminus$ 
   ( $\text{Ch}(X, G) \cap \text{Ch}(Y, G) \cup Z \cup \{X_0\}$ )
3: for  $A \in \mathbf{A}$  do
4:   if  $\log BF(X, A | Z \cup \{X_0\}) < 0$  then
5:      $E_s \leftarrow E_s \setminus \{E_{AY}\}, E \leftarrow E \setminus \{E_{AY}\}$ 
6:   else
7:      $E \leftarrow E \setminus \{E_{XA}\}$ 
            $\triangleright E_{XA}$ :  $XA$  間のエッジ
8:   end if
9:   if  $\log BF(A, Y | Z \cup \{X_0\}) < 0$  then
10:    if  $A \in V_s$  かつ  $Y \in V_s$  then
11:       $E_s \leftarrow E_s \setminus \{E_{AY}\}, E \leftarrow E \setminus \{E_{AY}\}$ 
12:    else
13:       $E \leftarrow E \setminus \{E_{AY}\}$ 
14:    end if
15:   end if
16: end for
17: return ( $G_s, G$ )
18: end function

```

条件部に必ず含める。これらの変更を加えることで説明変数集合から構成される PDAG が出力される。最後に、目的変数から全ての説明変数へエッジを接続することで ANB 構造をもつ PDAG が出力される。

**5.2 パラメータ数最小の I-map ANB**

本節では、提案手法が ANB 構造についての漸近一致性をもつことを示す。まず、以下の三つの補題を示す。ここで、 $\mathbf{V} = \{X_0, X_1, \dots, X_n\}$  を変数集合とし、 $X_0$  を目的変数、 $X_1, \dots, X_n$  を説明変数とする。 $G^* = (\mathbf{V}, \mathbf{E}^*)$ 、 $G_{ANB} = (\mathbf{V}, \mathbf{E}_{ANB})$  をそれぞれ、真のベイジアンネットワーク構造、提案アルゴリズムによって学習した ANB 構造とする。

[補題 5.1] データ数  $N \rightarrow \infty$  のとき、 $\forall X, Y \in \mathbf{V}$  について、 $E_{XY} \in \mathbf{E}^*$  ならば  $E_{XY} \in \mathbf{E}_{ANB}$ 。

[証明]  $E_{XY} \in \mathbf{E}^*$  のとき、 $\forall Z \subseteq \mathbf{V} \setminus \{X, Y\}$ 、 $D_P(X, Y | Z)$  である。これより、 $\forall Z \subseteq \mathbf{V} \setminus \{X, Y, X_0\}$ 、 $D_P(X, Y | Z \cup \{X_0\})$  が成り立つ。定理 4.1 より、提案アルゴリズムで行われる CI テストは  $XY$  間の独立性を検出しないため、エッジ  $E_{XY}$  が削除されず、 $E_{XY} \in \mathbf{E}_{ANB}$ 。□

[補題 5.2] データ数  $N \rightarrow \infty$  のとき、 $\forall X, Y, Z \in \mathbf{V}$  について、 $G^*$  において合流結合  $X \rightarrow Z \leftarrow Y$  となるならば、 $E_{XY} \in \mathbf{E}_{ANB}$  または  $G_{ANB}$  において合流結合  $X \rightarrow Z \leftarrow Y$  となる。

[証明]  $Z$  が目的変数の場合とそれ以外の場合に分けられる。

(1)  $Z = X_0$  のとき :

$G^*$  において合流結合  $X \rightarrow X_0 \leftarrow Y$  となるため,  $\forall Z \subseteq \mathbf{V} \setminus \{X, Y, X_0\}$ ,  $D_P(X, Y | Z \cup \{X_0\})$  である. 定理 4.1 より, 提案アルゴリズムで行われる CI テストは  $XY$  間の独立性を検出しないため, エッジ  $E_{XY}$  が削除されず,  $E_{XY} \in \mathbf{E}_{ANB}$ .

(2)  $Z \neq X_0$  のとき :

$X$  または  $Y$  が目的変数の場合とそれ以外の場合に分けられる.

(a)  $X = X_0$  または  $Y = X_0$  のとき :

ANB の仮定より目的変数は常に説明変数の親であるため,  $E_{XY} \in \mathbf{E}_{ANB}$ .

(b)  $X \neq X_0$  かつ  $Y \neq X_0$  のとき :

以下の背反な事象で場合分けする.

i.  $\exists Z \subseteq \mathbf{V} \setminus \{X, Y, Z, X_0\}$ ,  $I_P(X, Y | Z \cup \{X_0\})$

のとき :

定理 4.1 より, 提案アルゴリズムは CI テストによって独立性を検出してエッジ  $E_{XY}$  を削除する.  $G^*$  において合流結合  $X \rightarrow Z \leftarrow Y$  となるため,  $Z$  を所与としたとき  $X$  と  $Y$  は従属であるので, このとき検出した独立性の条件部に  $Z$  が含まれない. したがって, オリエンテーションルールによって  $G_{ANB}$  において合流結合  $X \rightarrow Z \leftarrow Y$  として方向付けされる.

ii.  $\forall Z \subseteq \mathbf{V} \setminus \{X, Y, Z, X_0\}$ ,  $D_P(X, Y | Z \cup \{X_0\})$  (9)

のとき :

$G^*$  において合流結合  $X \rightarrow Z \leftarrow Y$  となるため,  $Z$  を所与としたとき  $X$  と  $Y$  は従属であるので,  $\forall Z \subseteq \mathbf{V} \setminus \{X, Y, Z, X_0\}$ ,  $D_P(X, Y | Z \cup \{Z, X_0\})$  が成り立つ. これと式 (9) より,  $\forall Z \subseteq \mathbf{V} \setminus \{X, Y, X_0\}$ ,  $D_P(X, Y | Z \cup \{X_0\})$  が成り立つ. 定理 4.1 より, 提案アルゴリズムで行われる CI テストは  $XY$  間の独立性を検出しないため, エッジ  $E_{XY}$  が削除されず,  $E_{XY} \in \mathbf{E}_{ANB}$ .

(1), (2) より,  $\forall X, Y, Z \in \mathbf{V}$  について,  $G^*$  において合流結合  $X \rightarrow Z \leftarrow Y$  となるならば,  $E_{XY} \in \mathbf{E}_{ANB}$  または  $G_{ANB}$  において合流結合  $X \rightarrow Z \leftarrow Y$  となる.  $\square$

[補題 5.3] データ数  $N \rightarrow \infty$  のとき,  $\forall X, Y, Z \in \mathbf{V}$  について,  $G^*$  において逐次結合  $X \rightarrow Z \rightarrow Y$  または分岐結合  $X \leftarrow Z \rightarrow Y$  となるならば,  $E_{XY} \in \mathbf{E}_{ANB}$  または  $G_{ANB}$  において逐次結合  $X \rightarrow Z \rightarrow Y$  または分岐結合  $X \leftarrow Z \rightarrow Y$  となる.

[証明]  $Z$  が目的変数の場合とそれ以外の場合に分けられる.

(1)  $Z = X_0$  のとき :

$G^*$  において逐次結合  $X \rightarrow X_0 \rightarrow Y$  または分岐結合  $X \leftarrow X_0 \rightarrow Y$  となるため,  $\exists Z \subseteq \mathbf{V} \setminus \{X, Y, X_0\}$ ,  $I_P(X, Y | Z \cup \{X_0\})$  が成り立つ. 定理 4.1 より, 提案アルゴリズムは  $N \rightarrow \infty$  のとき CI テストにより確率 1 で独立性を検出してエッジ  $E_{XY}$  を削除する. ANB の仮定より, 目的変数は常に説明変数の親であるため,  $G_{ANB}$  において分岐結合  $X \leftarrow X_0 \rightarrow Y$  となる.

(2)  $Z \neq X_0$  のとき :

$X$  または  $Y$  が目的変数の場合とそれ以外の場合に分けられる.

(a)  $X = X_0$  または  $Y = X_0$  のとき :

ANB の仮定より目的変数は常に説明変数の親であるため,  $E_{XY} \in \mathbf{E}_{ANB}$ .

(b)  $X \neq X_0$  かつ  $Y \neq X_0$  のとき :

以下の背反な事象で場合分けする.

i.  $\exists Z \subseteq \mathbf{V} \setminus \{X, Y, Z, X_0\}$ ,  $I_P(X, Y | Z \cup \{Z, X_0\})$  の

とき :

定理 4.1 より, 提案アルゴリズムは  $N \rightarrow \infty$  のとき CI テストにより確率 1 で独立性を検出してエッジ  $E_{XY}$  を削除する.  $G^*$  において逐次結合  $X \rightarrow Z \rightarrow Y$  または分岐結合  $X \leftarrow Z \rightarrow Y$  となるため,  $Z$  を所与としないとき  $X$  と  $Y$  は従属であるので, このとき検出した独立性の条件部に  $Z$  が含まれる. したがって, オリエンテーションルールにより,  $G_{ANB}$  は逐次結合  $X \rightarrow Z \rightarrow Y$  または分岐結合  $X \leftarrow Z \rightarrow Y$  として方向付けされる.

ii.  $\forall Z \subseteq \mathbf{V} \setminus \{X, Y, Z, X_0\}$ ,  $D_P(X, Y | Z \cup \{Z, X_0\})$  (10) のとき :

$G^*$  において逐次結合  $X \rightarrow Z \rightarrow Y$  または分岐結合  $X \leftarrow Z \rightarrow Y$  となるため,  $Z$  を所与としないとき  $X$  と  $Y$  は従属であるので,  $\forall Z \subseteq \mathbf{V} \setminus \{X, Y, Z, X_0\}$ ,  $D_P(X, Y | Z \cup \{X_0\})$ . これと式 (10) より,  $\forall Z \subseteq \mathbf{V} \setminus \{X, Y, X_0\}$ ,  $D_P(X, Y | Z \cup \{X_0\})$  が成り立つ. 定理 4.1 より, 提案アルゴリズムで行われる CI テストは  $XY$  間の独立性を検出しないため, エッジ  $E_{XY}$  が削除されず,  $E_{XY} \in \mathbf{E}_{ANB}$ .

(1), (2) より,  $\forall X, Y, Z \in \mathbf{V}$  について,  $E_{XY} \in \mathbf{E}_{ANB}$  または  $G_{ANB}$  において逐次結合  $X \rightarrow Z \rightarrow Y$  または分岐結合  $X \leftarrow Z \rightarrow Y$  となる.  $\square$

以上の三つの補題から, 以下を示す

[定理 5.1]  $N \rightarrow \infty$  のとき, 提案アルゴリズムが学習する構造  $G_{ANB}$  は I-map ANB に概収束する.

[証明] 補題 5.1, 5.2, 5.3 より,  $G_{ANB}$  は  $N \rightarrow \infty$  のとき確率 1 で定理 2.1 の三つの条件をそれぞれ満たす.



よって、提案アルゴリズムが学習する構造  $G_{ANB}$  は I-map ANB に概収束する。□

更に、以下の定理が示すように、 $G_{ANB}$  は漸近的にパラメータ数最小の I-map ANB に概収束する。

[定理 5.2] 提案アルゴリズムは ANB 構造についての漸近一貫性をもつ。

[証明] 定理 5.1 より、 $N \rightarrow \infty$  のとき  $G_{ANB}$  は I-map ANB に概収束する。定理 4.1 より、 $N \rightarrow \infty$  のとき提案アルゴリズムの CI テストは確率 1 で真の条件付き独立性を検出するため、 $X_0$  を所与としたとき真に条件付き独立となる変数間のエッジは全て削除される。したがって、 $G_{ANB}$  は  $N \rightarrow \infty$  のとき I-map ANB の中でエッジ数が最小となる。すなわち、 $G_{ANB}$  はパラメータ数が最小の I-map ANB に概収束する。□

したがって、提案アルゴリズムは厳密学習した ANB [6], [7] の分類精度と同程度の精度をより短い計算時間で得られると期待できる。

一方で、ANB 構造を仮定すると、目的変数から説明変数へのエッジを強制的に引くため、GBN よりもパラメータ数が増加してしまう。このとき推定構造が表現する同時確率分布の真値への収束は GBN よりも理論的に遅くなるはずである。しかし、3. で述べたように GBN は目的変数の親変数数が多い場合、目的変数の事前分布パラメータが指数的に増え、推定精度が不安定になることが知られている [6], [7]。ANB 構造を仮定することでパラメータ数は冗長になるが、目的変数の親変数数は抑制されて事前分布をロバストに推定でき、分類精度の改善が期待できる。

## 6. 評価実験

本章では、提案手法の有効性を示すために以下の評価実験を行う。まず、シミュレーション実験により提案手法が厳密解探索アプローチよりも大規模なネットワークを学習できることを示す。次に、実データを用いて提案手法と他の BNC 学習手法の分類精度を比較する。

本章では以下の六つの手法を比較する。

- Naive Bayes
- TAN: 対数ゆがみを最適化する TAN を学習
- exact-GBN: BDeu スコアで GBN を厳密学習
- exact-ANB: BDeu スコアで ANB を厳密学習
- RAI-GBN: 本田らの手法を用いて GBN を学習
- RAI-ANB: 提案アルゴリズムを用いて ANB を学習

本論では、RAI-ANB を提案手法とする。TAN は Friedman ら [2] の手法を用いて学習した。exact-GBN と exact-ANB は学習スコアを BDeu とし、動的計画法 [25] を用いて厳密に学習した。RAI-GBN と RAI-ANB は本田ら [16] の手法を用いて学習した。BDeu スコアと Bayes factor の疑似サンプル (ハイパーパラメータ) の値は、Ueno [20], [34] の提案に従い事後分散を最大にする 1.0 とした。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。各手法の計算環境を表 1 に示す。

### 6.1 ランダムネットワークを用いた評価実験

本節では、提案手法が厳密解探索アプローチよりも大規模ネットワークを学習できることを示すためにランダムネットワークを用いたシミュレーション実験を行う。ランダムネットワークの生成には BNGenerator [35], [36] を使用する。BNGenerator はマルコフ連鎖モンテカルロ法により一様分布からネットワークをランダムに生成する。本節では、変数数を {5, 10, 20, 50, 100, 200, 500, 1000} とし、各ネットワークの最大次数を 5 と指定してネットワークを生成する。また、生成されたネットワークからデータ数 10000 のデータを発生させ、各手法について各ネットワークごとに構造学習を行い計算時間を求める。ただし、6 時間の制限時間を設け、超過する場合は学習を打ち切った。

結果を図 4 に示す。横軸が変数数 (個)、縦軸が計算時間 (秒) を表している。図 4 より、構造学習を必要としない Naive Bayes を除き、変数数が増加するにつれて計算時間が増加していることがわかる。Naive Bayes の次に TAN の計算時間が短い。これは TAN は多項式時間で計算できるためである [2], [32]。次に、厳密

表 1 計算環境  
Table 1 Computational environment.

Naive Bayes	
CPU	2.10GHz 8-Cores Intel XEON
System Memory	128GB
OS	OS ubuntu 16.04.4 lts
ソフトウェア	Python
TAN, exact-GBN, exact-ANB	
CPU	2.10GHz 8-Cores Intel XEON
System Memory	128GB
OS	OS ubuntu 16.04.4 lts
ソフトウェア	JAVA
RAI-GBN, RAI-ANB	
CPU	2.10GHz 8-Cores Intel XEON
System Memory	128GB
OS	OS ubuntu 16.04.4 lts
ソフトウェア	MATLAB

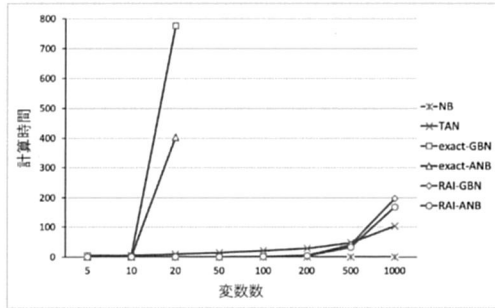


図4 変数数と計算時間の関係

Fig. 4 Relationship between the number of variables and computation time.

解探索アプローチである exact-GBN と exact-ANB の両方において変数数 50 より大きいネットワークは時間内に学習が終了しなかった。一方、制約ベース手法である RAI-GBN と RAI-ANB は 1000 変数のネットワークも構造学習することができた。提案手法である RAI-ANB は RAI-GBN よりも短い時間で構造学習を完了している。これは、RAI-GBN が全変数間の独立性を判定するのに対し、RAI-ANB は説明変数間の独立性のみを判定すればよいからである。以上の結果より、提案手法は厳密解探索アプローチよりも大規模なネットワークを学習できることを示した。また、提案手法が RAI-GBN よりも計算時間が短いことを示した。

## 6.2 実データを用いた評価実験

本節では、提案手法の有意性を示すために、実データによる分類精度と計算時間の比較を行う。まず、提案手法と厳密解探索アプローチを比較するため小規模ネットワークについて分類精度の比較を行う。次に、厳密学習不可能な大規模ネットワークについて分類精度の比較を行う。本実験では、UCI リポジトリ [37] に登録されているデータセットを用いた。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 値に離散化した。各手法、各データセットに対して、10 分割交差検証を用いて平均分類精度を求め、また、提案手法の有意性を示すために Hommel 法による多重検定 [38] を行い p 値を求め、各手法、各データセットに対して、構造学習にかかる平均計算時間を求めている。

### 6.2.1 小規模ネットワークにおける分類精度

本節では、厳密解探索アプローチを適用可能なデータセットを用いて各手法で学習した BNC の分類精度を比較する。実験には変数数 5 から 23 をもつ 43 個のデータセットを用いた。表 2 に各データセットに対

する各手法の分類精度を示している。表 2 の下部に示されている“分類精度”において“平均”は全データセットに対する各手法の平均分類精度であり、“p 値”は多重検定により得られた p 値を示している。各データセットに対し、各手法において最も高い分類精度を太字で示している。また、“計算時間”において“平均”は全データセットに対する各手法の構造学習にかかる平均計算時間であり、“標準誤差”は各手法の計算時間の標準誤差を示している。

結果として、提案手法は平均分類精度が 1 番高かった。また、提案手法は、Naive Bayes, TAN, RAI-GBN 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。RAI-GBN と RAI-ANB に対しては有意差を示すことができなかった。しかし、これらの分類精度が低かったデータセット 31 番, 36 番, 38 番において提案手法の分類精度が高いことがわかる。

次に、各手法を比較するために四つの指標を導入する。表 3 は各データセットに対する各手法の Max parents の平均を示している。“Max parents”は 10 分割交差検証法における各手法で学習した構造内の変数をもつ最大親変数数の平均を示している。Max parents の値が大きいほど構造がより複雑であることを表す [39]。表 4 は各データセットに対する各手法の目的変数の推定に関わるエッジ数の平均を表している。値が大きいほど目的変数のマルコフブランケットが複雑な構造を学習している。

まず、Naive Bayes と TAN は説明変数のとる親変数数を制限しているため Max Parents は 1 と 2 で固定されている。そのためデータ数が少ない場合でもパラメータの推定精度が安定するが、複雑な分布を表現できないため分類精度はあまり高くない。Naive Bayes は構造学習をする必要がないので計算時間は 0.0 である。また、TAN は多項式時間で計算できるため他の手法と比べて計算時間が短い [2], [32]。

exact-ANB と比較すると、有意差を示すことができなかったが、提案手法は厳密学習した ANB と同程度の分類精度が得られることがわかった。ただし、exact-ANB は提案手法よりも計算時間が長い。また、表 4 より、全てのデータセットにおいて提案手法の目的変数の推定に関わるエッジ数が exact-ANB よりも小さいことから、提案手法は exact-ANB よりもスパースな構造を学習する傾向がある。データセット 31 番, 36 番, 38 番において、exact-ANB よりも提案手法の分類精度が高い。表 2 よりこれらのデータ数は少なく、表 3, 表 4

表2 小規模ネットワークにおける各手法の分類精度  
Table 2 Accuracies of respective classifiers for small networks.

	データセット	変数数	データ数	目的変数の状態数	Naive Bayes	TAN	exact-GBN	exact-ANB	RAI-GBN	RAI-ANB
1	Balance	5	625	3	<b>0.9168</b>	0.8640	<b>0.9168</b>	<b>0.9168</b>	0.6352	<b>0.9168</b>
2	banknote	5	1372	2	0.8433	<b>0.8819</b>	0.8812	0.8812	0.8776	0.8812
3	Hayes-Roth	5	132	3	<b>0.8484</b>	0.6808	0.5610	<b>0.8484</b>	0.5132	<b>0.8484</b>
4	iris	5	150	3	0.7133	<b>0.8267</b>	<b>0.8267</b>	0.8200	0.8133	<b>0.8267</b>
5	lenses	5	24	3	0.6833	0.6833	0.6833	0.6833	<b>0.8500</b>	0.6833
6	Car	7	1728	4	0.8583	0.9381	0.9415	<b>0.9421</b>	0.8137	0.9410
7	liver	7	345	2	0.6461	0.6402	0.6200	0.6402	0.5830	<b>0.6634</b>
8	monk1	7	432	2	0.7500	<b>1.0000</b>	<b>1.0000</b>	<b>1.0000</b>	0.7500	<b>1.000</b>
9	mux6	7	64	2	0.5762	<b>0.6357</b>	0.5500	0.5762	0.3262	0.5333
10	led7	8	3200	10	0.7294	<b>0.7306</b>	0.7294	0.7294	0.7056	0.7294
11	HTRU2	9	17898	2	0.8966	<b>0.9141</b>	0.9112	<b>0.9141</b>	0.9027	0.9084
12	Nursery	9	12960	3	0.9033	0.9250	0.9340	0.9181	0.8921	<b>0.9356</b>
13	pima	9	768	9	0.7031	0.7175	<b>0.7279</b>	0.7175	0.7122	0.7032
14	post	9	87	5	0.6764	0.5986	<b>0.7125</b>	0.6764	<b>0.7125</b>	0.6764
15	Breast Cancer	10	277	2	<b>0.7443</b>	0.7187	0.7295	0.7119	0.6972	0.7192
16	Breast Cancer Wisconsin	10	683	2	<b>0.9752</b>	0.9649	<b>0.9752</b>	<b>0.9752</b>	0.9254	<b>0.9752</b>
17	cmc	10	1473	3	0.4657	0.4704	0.4548	0.4677	0.4358	<b>0.4752</b>
18	glass	10	214	6	0.5524	0.5431	0.5656	<b>0.6361</b>	0.5890	0.6087
19	shuttle-small	10	5800	6	0.9384	0.9566	0.9693	<b>0.9716</b>	0.9659	0.9707
20	threeOf9	10	512	2	0.8144	0.8477	<b>0.8865</b>	0.8673	0.7071	0.8399
21	TicTac	10	958	2	0.6919	0.7567	0.8319	<b>0.8549</b>	0.6992	0.7546
22	magic	11	19020	2	0.7482	0.7768	0.7873	<b>0.7874</b>	0.7801	0.7700
23	Flare	11	1389	9	0.7804	0.7948	<b>0.8431</b>	0.8229	<b>0.8431</b>	0.8236
24	heart	14	270	2	<b>0.8259</b>	<b>0.8259</b>	<b>0.8259</b>	0.8185	0.7815	<b>0.8370</b>
25	wine	14	178	3	<b>0.9330</b>	0.9275	0.9327	0.9216	0.8938	<b>0.9330</b>
26	cleve	14	296	2	<b>0.8410</b>	0.8338	0.7900	0.8344	0.7798	0.8308
27	australian	15	690	2	0.8290	0.8348	0.8536	0.8246	<b>0.8551</b>	0.8377
28	crx	15	653	2	0.8393	0.8531	0.8592	0.8531	<b>0.8639</b>	0.8531
29	EEG	15	14980	2	0.5778	0.6305	0.6814	<b>0.6864</b>	0.6411	0.6697
30	Congressional	17	232	2	0.9092	0.9478	<b>0.9652</b>	0.9478	<b>0.9652</b>	0.9522
31	zoo	17	101	5	<b>0.9800</b>	0.9600	0.9400	0.9600	0.9000	<b>0.9800</b>
32	pendigits	17	10992	10	0.8032	0.8504	<b>0.9289</b>	0.9278	0.8790	0.9085
33	letter	17	20000	26	0.4466	0.4868	0.5761	<b>0.5935</b>	0.5448	0.5609
34	ClimateModel	19	540	2	0.9222	<b>0.9315</b>	0.9000	0.8426	0.8963	0.9222
35	ImageSegmentation	19	2310	7	0.7290	0.7515	0.8156	<b>0.8225</b>	0.7758	0.8039
36	lymphography	19	148	4	<b>0.8386</b>	0.7648	0.7586	0.8186	0.7033	<b>0.8386</b>
37	vehicle	19	846	4	0.4339	0.5722	0.5732	<b>0.6241</b>	0.5543	0.5793
38	hepatitis	20	80	2	<b>0.8625</b>	0.8375	0.5875	0.6250	0.7375	<b>0.8625</b>
39	german	21	1000	2	<b>0.7430</b>	0.7310	0.7210	0.7380	0.6830	0.7390
40	bank	21	30488	2	0.8544	0.8774	<b>0.8956</b>	0.8949	0.8939	0.8907
41	waveform-21	22	5000	3	0.7886	0.7896	0.7846	<b>0.7966</b>	0.7336	0.7826
42	Mushroom	22	5644	2	0.9957	<b>1.0000</b>	0.9949	<b>1.0000</b>	<b>1.0000</b>	<b>1.0000</b>
43	spect	23	263	2	0.8013	0.8128	0.7450	<b>0.8164</b>	0.7946	0.8051
	分類精度	平均			0.7770	0.7927	0.7985	0.8071	0.7583	<b>0.8086</b>
		p 値			0.00004	0.00126	0.46812	0.46812	0.00004	-
	計算時間 (s)	平均			0.00	2.58	1790.93	500.76	11.94	1.57
		標準誤差			0.00	0.16	895.76	252.69	9.55	0.6.2

より exact-ANB において Max parents と目的変数の推定に関わるエッジ数は提案手法より小さい。データ数が少なく、複雑な構造を学習したため、厳密解探索アプローチでは目的変数の分類精度が低かったと考えられる。一方、提案手法は厳密解探索アプローチよりスパースな構造を学習する傾向にあり、データ数が少なくてもそれほど分類精度を低下させなかったと考えら

れる。

RAI-GBN と RAI-ANB を比較すると、構造に ANB を仮定することで厳密解探索アプローチと同様に分類精度を改善している。RAI-GBN は比較手法のなかで最も分類精度が低い。これは CI テストの精度に起因する。Bayes factor を用いた CI テストはデータ数が小さい場合に誤ったエッジを削除してしまう可能性がある



表3 小規模ネットワークにおける各手法の Max parents  
Table 3 The number of Max parents of respective classifiers for small networks.

データセット	Naive Bayes	TAN	exact-GBN	exact-ANB	RAI-GBN	RAI-ANB
1 Balance	1	2	1	1	1	1
2 banknote	1	2	4	4	3.9	4
3 Hayes-Roth	1	2	3	1	1.8	1
4 iris	1	2	2	2	2.7	2
5 lenses	1	2	1.3	1.1	2	1.1
6 Car	1	2	2	2	3	2
7 liver	1	2	2	2	1.5	2
8 monk1	1	2	3	3	1	2
9 mux6	1	2	5.8	1	0.7	1
10 led7	1	2	1	1	1.8	1
11 HTRU2	1	2	3	4	5	4
12 Nursery	1	2	4	3	3	3
13 pima	1	2	2	2	2.1	2.3
14 post	1	2	0.2	1	0.3	1
15 Breast Cancer	1	2	1	2	1	2
16 Breast Cancer Wisconsin	1	2	1	1	1.1	1
17 cmc	1	2	2	2.5	2	2.1
18 glass	1	2	2.9	3	2	2.3
19 shuttle-small	1	2	5	5	5	3.7
20 threeOf9	1	2	5	2.7	4.4	2
21 TicTac	1	2	3	3	1.7	2
22 magic	1	2	4	4	4	5
23 Flare	1	2	2	3	1.6	3
24 heart	1	2	2	2	2	2
25 wine	1	2	2.2	2.1	3.2	2.1
26 cleve	1	2	2	2	2	2
27 australian	1	2	2.4	2.9	2	2.3
28 crx	1	2	3	2.2	1.6	2
29 EEG	1	2	5	5	4.9	5.3
30 Congressional	1	2	3.5	4	2.3	3
31 zoo	1	2	4.9	4.9	3.7	3
32 pendigits	1	2	5.5	5.6	8	5.9
33 letter	1	2	6	5	7.6	5.3
34 ClimateModel	1	2	14	14.1	3.1	1
35 ImageSegmentation	1	2	4.1	4	6	5
36 lymphography	1	2	8.7	9.9	2.1	2.3
37 vehicle	1	2	4.2	4.1	3.5	3.6
38 hepatitis	1	2	10.4	11.4	2.4	2.9
39 german	1	2	2	3	2	3
40 bank	1	2	5	6	5.4	5.4
41 waveform-21	1	2	4	4	5	3.7
42 Mushroom	1	2	2.4	7.6	4.8	4.8
43 spect	1	2	2.7	3	2.6	3.2

表4 小規模ネットワークにおける各手法の目的変数の推定に関わるエッジ数

Table 4 The number of edges used to estimate the class variable of respective classifiers for small networks.

データセット	Naive Bayes	TAN	exact-GBN	exact-ANB	RAI-GBN	RAI-ANB
1 Balance	4	7	4	4	1	4
2 banknote	4	7	7	10	6.8	9.8
3 Hayes-Roth	4	7	3	4	1.7	4
4 iris	4	7	4.1	7	3.1	5.2
5 lenses	4	7	2.1	4.1	2	4
6 Car	6	11	7	9	3	6
7 liver	6	11	3.8	10.6	1.1	9
8 monk1	6	11	3	8	1	6
9 mux6	6	11	5.8	6	0.4	6
10 led7	7	13	7	7	5.1	7
11 HTRU2	8	15	12.5	20	5	18
12 Nursery	8	15	8	13	3	8
13 pima	8	15	4.2	15	2.2	12.4
14 post	8	15	0	8	0	8
15 Breast Cancer	9	17	1	13	0.1	11.1
16 Breast Cancer Wisconsin	9	17	8.9	9	1	9
17 cmc	9	17	1.7	16.1	1	13.2
18 glass	9	17	4.3	15.5	3.1	13
19 shuttle-small	9	17	15	23.8	4.4	22
20 threeOf9	9	17	9.7	13.4	4.5	9
21 TicTac	9	17	7.4	18.9	1	12.3
22 magic	10	19	20.4	30	13	26.8
23 Flare	10	19	1	18.9	0.9	17.4
24 heart	13	25	6.6	18.4	2	17.1
25 wine	13	25	9.5	19	3.2	16
26 cleve	13	25	7.5	18.3	2	16.3
27 australian	14	27	6.2	24.1	3.6	20.2
28 crx	14	29	5.3	23.9	2.5	21.2
29 EEG	14	27	34.2	57.5	11.8	48.6
30 Congressional	16	31	7.1	37.1	3.4	27.7
31 zoo	16	31	9.4	36.9	3.9	24.9
32 pendigits	16	31	63.4	66.5	9.2	60.5
33 letter	16	31	41.4	57.9	17.7	51.1
34 ClimateModel	18	35	32.1	69.7	3	18
35 ImageSegmentation	18	35	31.5	48	5.8	38.5
36 lymphography	18	35	16.6	36.7	1.8	23.6
37 vehicle	18	35	14.3	50.8	6.3	41.1
38 hepatitis	19	37	31.6	78.1	1.4	28.2
39 german	20	39	4.1	33.3	1	28.8
40 bank	20	39	13.1	63.9	5.8	47.6
41 waveform-21	21	41	39.8	60.3	5.9	42.9
42 Mushroom	21	41	6.7	83	17.1	61.3
43 spect	22	43	9.3	49.2	2.1	44.8

る。これにより、目的変数のマルコフブランケットが小さいネットワークを学習することがある。実際、表4をみると他の手法よりも目的変数の推定に関わるエッジ数が少なく、9番、14番、15番では目的変数と説明変数がほとんど独立な構造を学習していることがわかる。それに対し、提案手法は、全ての説明変数を子にもつためマルコフブランケットの数は常に説明変数数と同じであり、全ての説明変数を用いて推論することで分類精度を向上させている。また、提案手法は説明

変数間のCIテストのみを行うため、全変数間のCIテストを行うRAI-GBNよりも計算時間が短くなる。

### 6.2.2 大規模ネットワークにおける分類精度

本節では、厳密解探索アプローチで扱えない大規模なデータセットを用いた評価実験について述べる。変数数36~1300である15個のデータセットを用いて、従来手法で学習したBNCと提案手法で学習したBNCの分類精度を比較する。適用できないので厳密解探索

表5 大規模ネットワークにおける各手法の分類精度  
Table 5 Accuracies of respective classifiers for huge networks.

	データセット	変数数	データ数	目的変数の状態数	Naive Bayes	TAN	RAI-GBN	RAI-ANB
1	kr-vs-kp	37	3196	2	0.8774	0.9240	0.9402	<b>0.9524</b>
2	Connect-4	43	67557	3	0.7213	0.7643	0.7467	<b>0.7938</b>
3	Flowmeters D	44	180	4	0.8389	0.8389	0.8389	<b>0.8500</b>
4	movement libras	91	360	15	0.5028	<b>0.5389</b>	0.2278	0.5333
5	dota2	117	102944	2	<b>0.5981</b>	0.5810	0.5564	0.5957
6	Musk1	167	478	2	0.6517	0.7566	0.5756	<b>0.7986</b>
7	Musk2	167	6598	2	0.7445	0.8406	0.9047	<b>0.9615</b>
8	Epileptic Seizure	179	11500	5	0.2344	0.3650	0.1187	<b>0.3808</b>
9	mfeat-fac	219	2000	10	0.3520	0.4590	0.3310	<b>0.4650</b>
10	semeion	257	1600	10	0.8550	0.8719	0.3521	<b>0.8776</b>
11	madelon	501	2000	2	<b>0.5905</b>	0.5270	0.5740	<b>0.5905</b>
12	HART	563	10929	12	0.7967	0.8685	0.8456	<b>0.8805</b>
13	HAR	563	10929	6	0.7633	0.8797	0.8657	<b>0.8987</b>
14	Parkinson's Disease	755	756	2	0.7182	0.7898	0.7419	<b>0.7964</b>
15	MNIST	785	70000	10	0.8258	0.8911	0.9482	<b>0.9493</b>
16	MicroMass	1301	360	10	0.9472	0.9472	0.8756	<b>0.9556</b>
	分類精度	平均			0.6886	0.7402	0.6527	<b>0.7675</b>
		p 値			0.0080	0.0060	0.0022	-
	計算時間 (s)	平均			0.0	545.7	3647.2	565.7
		標準誤差			0.0	434.6	1966.1	220.9

アプローチは比較対象から除外した。6.2.1 と同様に表5に各データセットにおける各手法ごとの分類精度の平均と p 値, 計算時間の平均と標準誤差を示している。また, 前節の分析同様に表6と表7は“Max parents”と目的変数の推定に関わるエッジ数を示している。

表5より, 提案手法の分類精度が最も高かった。また, 提案手法は, Naive Bayes, TAN, RAI-GBN 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。更に, 提案手法は動的計画法によって学習することができない大規模ネットワークをもつ BNC を学習することができた。また, 小規模ネットワークに対する計算時間の結果同様に, 提案手法の計算時間は RAI-GBN の計算時間よりも短く, Naive Bayes と TAN の計算時間より長い。この理由は前節で述べたとおりである。

次に各手法との比較を行う。Naive Bayes と TAN はほとんどのデータセットに対して提案手法の分類精度よりも低かった。これは, 表6からわかるように Naive Bayes と TAN は Max parents が制限されているためである。変数数が増えるほど影響し合う可能性がある変数が増えるため, Max parents が小さいことが分類精度の低下を招く恐れがある。しかし, データセット4番は Naive Bayes における分類精度が最も高く, 5番は TAN における分類精度が最も高い。これは, 変数間の相関が小さいためであると考えられる。表7をみる

表6 大規模ネットワークにおける各手法の Max parents  
Table 6 The number of Max parents of respective classifiers for huge networks.

	データセット	Naive Bayes	TAN	RAI-GBN	RAI-ANB
1	kr-vs-kp	1	2	6	6.2
2	Connect-4	1	2	5.1	5.5
3	Flowmeters D	1	2	3.6	4
4	movement libras	1	2	2.5	3.1
5	dota2	1	2	3.4	4
6	Musk1	1	2	4.7	5
7	Musk2	1	2	9.3	10.3
8	Epileptic Seizure	1	2	3	3
9	mfeat-fac	1	2	5.1	5.1
10	semeion	1	2	5	4
11	madelon	1	2	3.5	4.4
12	HAPT	1	2	5.8	5
13	HAR	1	2	6.5	6.8
14	Parkinson's Disease	1	2	5	6
15	mnist	1	2	8.6	8.5
16	MicroMass	1	2	6.8	4.4

と RAI-GBN のデータセット4, 5番について, 目的変数の推定に関わるエッジ数が2.4と14.3であり非常に小さいことがわかる。これは, 目的変数のマルコフブランケットが小さいことを示し, ほとんどの説明変数が目的変数の推定に関わっていない。また, RAI-ANB における目的変数の推定に関わるエッジ数をみると, 4番において TAN と近い値をとっており, 5番において TAN よりも小さく Naive Bayes に近い値をとっている。

表7 大規模ネットワークにおける各手法の目的変数の推定に関わるエッジ数

Table 7 The number of edges used to estimate the class variable of respective classifiers for huge networks.

	データセット	Naive Bayes	TAN	RAI-GBN	RAI-ANB
1	kr-vs-kp	36	71	46.1	116.2
2	Connect-4	42	83	43.1	115
3	Flowmeters D	43	85	13.1	83.2
4	movement libras	90	179	2.4	187.4
5	dota2	116	231	14.3	209.9
6	Musk1	166	331	1.1	513.1
7	Musk2	166	331	61.6	1026.9
8	Epileptic Seizure	178	355	0	387
9	mfeat-fac	216	431	19.5	561.3
10	semeion	256	511	15.3	753.3
11	madelon	500	999	3	533.5
12	HAPT	561	1121	164.7	1524.6
13	HAR	561	1121	204.9	1625.8
14	Parkinson's Disease	753	1505	2.5	1973.2
15	mnist	784	1567	2218.0	3253.2
16	MicroMass	1300	2599	169.8	1940.4

る。これは、RAI-ANB が Naive Bayes や TAN に近いスパースな構造を推定しているといえる。よって、これらのデータセットは ANB 候補空間において Naive Bayes や TAN に近い構造が真の構造であったためこれらの分類精度が高かったと考えられる。

RAI-GBN と RAI-ANB を比較すると全てのデータセットにおいて提案手法の分類精度のほうが高い。表7をみるとデータセット4, 6, 8, 11, 14 番について、RAI-GBN の目的変数の推定に関わるエッジ数が非常に少なく、CI テストにより過少学習している可能性がある。一方、提案手法は ANB 構造を仮定しているため全ての説明変数を目的変数の推定に用いることで分類精度を改善していると考えられる。

## 7. 他の分類器との比較

前章では、提案手法が大規模ネットワークにおいて他の BNC よりも高い分類精度をもつことを示した。本章では、BNC 以外の分類器と精度の比較を行い提案手法を分析する。実験に用いるデータセットは前章と同じものを用いる。

本章では一般的な分類器として以下の3種類の手法を用いる。

- RF: ランダムフォレスト
- SVM: サポートベクターマシン
- MLP: 深層ニューラルネットワーク (多層パーセプトロン)

これらの手法はいずれも、scikit-learn (<https://scikit-learn.org/stable/>) を使用して学習する。MLP は隠れ層数の候補を {10, 100, 1000}, 学習率の候補を {0.0001, 0.001, 0.01} としてグリッドサーチを用いてチューニングを行う。提案手法である RAI-ANB は構造学習と分類精度の推定において ESS の候補を {1, 5, 10, 20} としてグリッドサーチを用いてチューニングを行う。各手法、各データセットに対して、10 分割交差検証を用いて平均分類精度を求める。ただし、比較手法は欠損値を含む入力データに対応できないため、それらはデータセットから取り除いている。また、提案手法の有意性を示すために Hommel 法による多重検定 [38] を行い p 値を求めた。表8に各手法の分類精度と p 値を示す。また、各データセットに対し各手法において最も高い分類精度を太字で示している。

表8より、提案手法の分類精度は RF と SVM より高く、MLP の分類精度より低い。また、提案手法は全ての比較手法に対して有意差を示すことができなかった。提案手法は同時確率分布を表現する生成モデルであるのに対し、MLP は特徴量を入力とし目的変数の値を出力とする関数型の識別モデルである。先行研究 [40] では、分類問題では生成モデルであるベイジアンネットワークよりも関数型の識別モデルであるニューラルネットワークのほうが表現力が高く、精度が高いことが報告されている。結果として、提案したベイジアンネットワーク分類器が他の分類器より精度で優れているとはいえなかった。しかし、提案手法が生成モデルである利点もある。提案手法は確率モデルであり同時確率分布そのものを推定できる利点がある。例えば、MLP は入力データの欠損値の扱いが難しいのに対し、提案手法は欠損変数を周辺化した確率分布を容易に求められる。そのため、欠損値を含む入力データから分類する場合、MLP より高い精度を示せる可能性が高い。また、提案手法は目的変数の確率を漸近的に求めることができるため、意思決定問題で期待効用 (損失) 関数を容易に計算でき、意思決定システムの構築が容易である。更に、提案手法は説明変数同士の因果関係を示すことができるので、説明可能性に優れている。

BNC はモデル平均を行うことによって分類精度が大きく改善することが知られている [10]。近年では、アンサンブル学習を組み合わせることで分類精度が改善することが報告されている [41]。これらの手法を用いることで、提案手法の分類精度の改善が期待できる。



表 8 各手法の分類精度  
Table 8 Accuracies of respective classifiers.

	データセット	変数数	データ数	目的変数 の状態数	RF	SVM	MLP	RAI- ANB
1	Balance	5	625	3	0.8289	0.9008	0.8565	<b>0.9167</b>
2	banknote	5	1372	2	0.8812	0.8819	<b>0.9665</b>	0.8812
3	Hayes-Roth	5	132	3	0.8022	0.8560	<b>0.8970</b>	0.8110
4	iris	5	150	3	0.8267	0.8133	<b>0.8819</b>	0.8267
5	lenses	5	24	3	<b>0.7333</b>	0.6500	0.7180	0.6333
6	Car	7	1728	4	0.9641	0.9688	<b>0.9707</b>	0.9398
7	liver	7	345	2	0.6146	0.6403	<b>0.9919</b>	0.6811
8	monk1	7	432	2	0.9560	0.9166	0.8106	<b>1.0000</b>
9	mux6	7	64	2	0.7976	0.5786	<b>0.9241</b>	0.8714
10	led7	8	3200	10	0.7288	<b>0.7369</b>	0.4922	0.7319
11	HTRU2	9	17898	2	0.9141	0.9112	<b>0.9482</b>	0.9084
12	Nursery	9	12960	3	<b>0.9891</b>	0.9860	0.8621	0.9398
13	pima	9	768	9	0.6875	0.6967	<b>0.7302</b>	0.7292
14	post	9	87	5	0.5736	0.7125	<b>0.8394</b>	0.6681
15	Breast Cancer	10	277	2	0.6534	0.7299	<b>0.7510</b>	0.7251
16	Breast Cancer Wisconsin	10	683	2	0.9664	0.9650	0.6325	<b>0.9766</b>
17	cmc	10	1473	3	0.4508	0.4895	<b>0.7571</b>	0.4807
18	glass	10	214	6	0.6214	0.5983	<b>0.8074</b>	0.6173
19	shuttle-small	10	5800	6	0.9721	0.9662	0.8500	<b>0.9722</b>
20	threeOf9	10	512	2	<b>0.9824</b>	0.8925	0.9115	0.9337
21	TicTac	10	958	2	<b>0.9133</b>	0.8831	0.8299	0.7798
22	magic	11	19020	2	0.7809	0.7807	<b>0.8267</b>	0.7803
23	Flare	11	1389	9	0.8179	<b>0.8431</b>	0.7313	0.8222
24	heart	14	270	2	0.7926	<b>0.8407</b>	0.7167	0.8222
25	wine	14	178	3	0.8935	0.9105	0.6561	<b>0.9265</b>
26	cleve	14	296	2	<b>0.8339</b>	0.8238	0.6518	0.8307
27	australian	15	690	2	0.8449	<b>0.8609</b>	0.8457	0.8464
28	crx	15	653	2	0.8638	<b>0.8653</b>	0.7877	0.8514
29	EEG	15	14980	2	0.7275	0.6955	<b>1.0000</b>	0.6834
30	Congressional	17	232	2	0.9524	0.9696	<b>1.0000</b>	0.9440
31	zoo	17	101	5	<b>0.9700</b>	0.9300	0.7881	<b>0.9700</b>
32	pendigits	17	10992	10	0.9406	0.9285	<b>0.9998</b>	0.9176
33	letter	17	20000	26	0.6471	0.5877	<b>0.9408</b>	0.5723
34	ClimateModel	19	540	2	0.9204	0.9148	0.7253	<b>0.9241</b>
35	ImageSegmentation	19	2310	7	<b>0.8229</b>	0.8134	0.6764	0.8139
36	lymphography	19	148	4	0.7910	0.7767	<b>0.9705</b>	0.8719
37	vehicle	19	846	4	0.6289	0.6348	<b>0.8170</b>	0.6015
38	hepatitis	20	80	2	0.8000	0.8375	<b>1.0000</b>	0.8750
39	german	21	1000	2	0.7330	0.7390	<b>1.0000</b>	0.7460
40	bank	21	30488	2	0.8803	0.8918	0.6205	<b>0.8940</b>
41	waveform-21	22	5000	3	0.7790	0.8132	<b>0.8168</b>	0.7922
42	Mushroom	22	5644	2	<b>1.0000</b>	<b>1.0000</b>	0.9203	<b>1.0000</b>
43	spect	23	263	2	0.8128	0.8021	<b>0.9709</b>	0.8207
44	kr-vs-kp	37	3196	2	0.9831	0.9384	<b>0.9950</b>	0.9524
45	Connect-4	43	67557	3	0.7901	0.7393	<b>0.8097</b>	0.7938
46	Flowmeters D	44	180	4	<b>0.8778</b>	0.8722	0.8833	0.8500
47	movement libras	91	360	15	0.6750	0.5361	<b>0.7306</b>	0.5333
48	dota2	117	102944	2	0.5314	0.5508	0.5425	<b>0.5957</b>
49	Musk1	167	476	2	0.7921	0.7604	<b>0.8380</b>	0.7986
50	Musk2	167	6598	2	0.9548	0.9406	<b>0.9767</b>	0.9615
51	Epileptic Seizure	179	11500	5	0.4207	0.4618	<b>0.4692</b>	0.3808
52	mfeat-fac	219	2000	10	0.4630	0.4245	<b>0.4675</b>	0.4650
53	semeion	257	1600	10	0.8789	<b>0.9360</b>	0.9272	0.8776
54	madelon	501	2000	2	0.5585	0.5860	0.5605	<b>0.5905</b>
55	HAPT	563	10929	12	0.8756	0.8866	<b>0.9222</b>	0.8805
56	HAR	563	10299	6	0.8880	0.9068	<b>0.9316</b>	0.8987
57	Parkinson's Disease	755	756	2	0.8389	0.4061	<b>0.9316</b>	0.7964
58	MNIST	785	70000	10	0.9695	<b>0.9731</b>	<b>0.9731</b>	0.9493
59	MicroMass	1300	360	10	0.9333	0.9222	<b>0.9722</b>	0.9556
	分類精度	平均			0.8054	0.7945	<b>0.8275</b>	0.8069
		p 値			0.17619	0.17619	0.17619	-

## 8. む す び

本論文では、本田らの手法を ANB 学習に拡張させることで、従来より大規模な BNC を学習する手法を提案した。また、提案手法が ANB について漸近一致性を有することを示した。ベンチマークデータセットを用いた実験の結果、RAI-GBN は、データ数が少ないとき CI テストの精度が低く、目的変数のマルコフブランケットが小さくなるため、分類精度が低いことがわかった。一方、提案手法では ANB を学習することで分類精度が安定し、厳密解探索アプローチに劣らない分類精度が得られた。また、提案手法は厳密解探索アプローチでは学習することができない数千変数の大規模ネットワークを学習することが可能であり、他の BNC 手法に対して有意に高い分類精度が得られることを示した。また、一般的な分類器と提案手法の分類精度の比較も行った。結果として、提案手法の平均分類精度の方がランダムフォレスト、サポートベクターマシンの平均分類精度より高かったものの、深層ニューラルネットワークの平均分類精度より低かった。しかし、提案手法は確率モデルであり、モデルの説明可能性など深層ニューラルネットワークより優れた点も多い。更に、提案手法はモデル平均などの手法を用いれば更に精度向上が期待できる。

上記以外の課題として、以下が挙げられる。大規模構造学習ではデータがスパースになるためパラメータの分類精度や CI テストの信頼性が低下する傾向がある。この問題に対し、磯崎らは Minimum Free Energy (MFE) に基づくパラメータ推定法と CI テストを提案している [42]~[44]。この MFE に基づくパラメータ推定法と CI テストを組み込むことで大規模構造において学習精度と分類精度の向上が見込める。

### 文 献

- [1] 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.
- [2] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Machine Learning*, vol.29, no.2, pp.131-163, 1997.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] S. Sugahara, M. Uto, and M. Ueno, "Exact learning augmented naive Bayes classifier," *International Conference on Probabilistic Graphical Models*, vol.72, pp.439-450, 2018.
- [7] 菅原聖太, 植野真臣, "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," *信学論 (D)*, vol.103, no.4, p.301-313, April 2020.
- [8] J. Pearl, *Models, Reasoning, and Inference*, Cambridge University Press, 2000.
- [9] P. Spirtes, C. Glymour, and R. Scheines, *Causation, Prediction, and Search*, MIT press, 2000.
- [10] J. Cheng, R. Greiner, J. Kelly, D. Bell, and W. Liu, "Learning bayesian networks from data: An information-theory based approach," *Artificial Intelligence - AI*, vol.137, pp.43-90, May 2002.
- [11] I. Tsamardinos, L.E. Brown, and C.F. Aliferis, "The max-min hill-climbing Bayesian network structure learning algorithm," *Machine Learning*, vol.65, no.1, pp.31-78, 2006.
- [12] R. Yehezkel and B. Lerner, "Bayesian network structure learning by recursive autonomy identification," *Journal of Machine Learning Research*, vol.10, pp.1527-1570, 2009.
- [13] K. Natori, M. Uto, Y. Nishiyama, S. Kawano, and M. Ueno, "Constraint-based learning bayesian networks using bayes factor," *Proceedings of the Second International Workshop on Advanced Methodologies for Bayesian Networks - Volume 9505*, pp.15-31, AMBN 2015, Springer-Verlag New York, Inc., 2015.
- [14] K. Natori, M. Uto, and M. Ueno, "Consistent learning bayesian networks with thousands of variables," *Proc. Machine Learning Research*, vol.73, pp.57-68, 2017.
- [15] 名取和樹, 宇都雅輝, 植野真臣, "Bayes factor を用いた RAI アルゴリズムによる大規模ベイジアンネットワーク学習," *信学論 (D)*, vol.101, no.5, pp.754-768, May 2018.
- [16] 本田和雅, 名取和樹, 菅原聖太, 磯崎隆司, 植野真臣, "推移性を利用した大規模ベイジアンネットワーク構造学習," *信学論 (D)*, vol.102, no.12, pp.796-811, Dec. 2019.
- [17] D.M. Chickering, "Optimal structure identification with greedy search," *J. Machine Learning Research*, vol.3, pp.507-554, Jan. 2002.
- [18] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol.6, no.2, pp.461-464, 1978.
- [19] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol.14, no.5, pp.465-471, 1978.
- [20] M. Ueno, "Learning networks determined by the ratio of prior and data," *Proc. Twenty-Sixth Conf. Uncertainty in Artificial Intelligence*, p.598-605, AUAI Press, Arlington, Virginia, USA, 2010.
- [21] W. Buntine, "Theory refinement on bayesian networks," *Proc. Seventh Conf. Uncertainty in Artificial Intelligence*, pp.52-60, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1991.
- [22] R.G. Cowell, "Efficient maximum likelihood pedigree reconstruction," *Theoretical Population Biology*, vol.76, pp.285-291, 2009.
- [23] M. Koivisto and K. Sood, "Exact bayesian structure discovery in bayesian networks," *J. Machine Learning Research*, vol.5, pp.549-

- 573, 2004.
- [24] A. Singth and A. Moore, "Finding optimal bayesian networks by dynamic programming," Technical Report, Carnegie Mellon University, 2005.
- [25] T. Silander and P. Myllymäki, "A simple approach for finding the globally optimal bayesian network structure," Proc. Uncertainty in Artificial Intelligence, pp.445–452, 2006.
- [26] B.M. Malone, C. Yuan, E.A. Hansen, and S. Bridges, "Improving the scalability of optimal bayesian network learning with external-memory frontier breadth-first branch and bound search," Proc. Uncertainty in Artificial Intelligence, pp.479–488, 2011.
- [27] C. Yuan, H. Lim, and T.-C. Lu, "Most relevant explanation in bayesian networks," J. Artificial Intelligence Research, vol.42, no.1, pp.309–352, 2011.
- [28] M. Barlett and J. Cussens, "Advances in bayesian network learning using integer programming," Proc. Twenty-Ninth Conference on Uncertainty in Artificial Intelligence, pp.182–191, 2013.
- [29] F.V. Jensen and T.D. Nielsen, Bayesian Networks and Decision Graphs, 2nd edition, Springer Publishing Company, Incorporated, 2007.
- [30] T.M. Mitchell, Machine Learning, 1 edition, McGraw-Hill, 1997.
- [31] M. Minsky, "Steps toward artificial intelligence," Proc. IRE, vol.49, pp.8–30, 1961.
- [32] M.G. Madden, "On the classification performance of TAN and general Bayesian networks," Knowledge-Based Systems, pp.489–495, 2009.
- [33] H. Steck and T.S. Jaakkola, On the dirichlet prior and Bayesian regularization, pp.697–704, MIT Press, 2002.
- [34] M. Ueno, "Robust learning Bayesian networks for prior belief," Proceedings of Uncertainty in Artificial Intelligence, pp.689–707, 2011.
- [35] J.S. Ide and F.G. Cozman, "Random generation of bayesian networks," Brazilian Symposium on Artificial Intelligence, pp.366–376, Springer, 2002.
- [36] J.S. Ide, F.G. Cozman, and F.T. Ramos, "Generating random bayesian networks with constraints on induced width," Proc. 16th European Conference on Artificial Intelligence, IOS Press, pp.353–357, 2004.
- [37] M. Lichman, "UCI Machine learning repository," 2013. <http://archive.ics.uci.edu/ml>
- [38] G. Hommel, "A stagewise rejective multiple test procedure based on a modified bonferroni test," Biometrika, pp.383–386, 1988.
- [39] C.X. Ling and H. Zhang, "The representational power of discrete bayesian networks," J. Mach. Learn. Res., pp.709–721, 2003.
- [40] A. Choi, R. Wang, and A. Darwiche, "On the relative expressiveness of bayesian and neural networks," International Journal of Approximate Reasoning, vol.113, pp.303–323, 2019.
- [41] 青見 樹, 菅原聖太, 植野真臣, "アンサンブル学習によるモデル平均ベイジアンネットワーク分類器," 信学論 (D), vol.103, no.3, pp.183–193, March 2020.
- [42] T. Isozaki, N. Kato, and M. Ueno, "Minimum free energies with "data temperature" for parameter learning of bayesian networks," 2008 20th IEEE Int. Conf. Tools with Artificial Intelligence, vol.1, pp.371–378, 2008.
- [43] T. Isozaki, N. Kato, and M. Ueno, ""Data temperature" in mini-

mum free energies for parameter learning of bayesian networks," Int. J. Artificial Intelligence Tools, vol.18, pp.653–671, Oct. 2009.

- [44] 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, 2009.

(2020年1月23日受付, 7月29日再受付,  
9月25日早期公開)



菊谷 成慎

2019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。



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。



名取 和樹 (正員)

2014年電気通信大学情報理工学部卒。2016年同大学院情報システム学研究科社会知能情報学専攻博士前期課程修了。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後期課程入学、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** A Bayesian network classifier (BNC) is known as a probabilistic classifier for discrete variables. Previous research indicates the exact learning of BNC with the Augmented Naïve Bayes (ANB) structure constraint outperforms the approximation methods. However, the exact learning cannot learn the huge BNC. On the other hand, there is a constraint-based algorithm that can learn 3500 nodes Bayesian networks by the RAI algorithm with the transitivity. This paper proposes a method that extends this algorithm for the ANB learning, and proves that the proposed method has an asymptotic consistency for ANB structures. The experimental results show that the proposed method outperforms the other methods for the huge BNC learning.

**Key words** Bayesian networks, probabilistic graphical models, machine learnings, classifiers, constraint-based approach