

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

青見 樹^{†a)} 菅原 聖太^{†b)} 植野 真臣^{†c)}

Model Averaging Bayesian Network Classifier by Ensemble Learning

Itsuki AOMI^{†a)}, Shouta SUGAHARA^{†b)}, and Maomi UENO^{†c)}

あらまし ベイジアンネットワーク分類器 (Bayesian Network Classifier: BNC) について, 説明変数を所与とした目的変数の条件付き確率をモデル化した BNC (識別モデル) が, 周辺ゆう度を用いて学習した BNC (生成モデル) に対し, 分類精度が高くなると報告されている. これは, 周辺ゆう度は漸近一致性をもつため, データ数が少ない場合に生成モデルの各構造の事後確率が低くなり, 真の確率分布をうまく近似できなくなってしまうからであると考えられる. このような場合, 全てのモデル候補を用いて推論するモデル平均が有効であることが知られている. しかし, ベイジアンネットワークの候補構造は変数数に対して指数的に増加し, 計算が困難である. この計算困難性を解決するために, 構造を確率の高い順に K 個選択する k -best 法が知られているが, 選択される構造が類似してしまう傾向があり, 分類精度が劣化してしまう可能性がある. 本研究では, 選択される構造が類似しないように, データセットからアンサンブル法により構造学習した複数の構造についてモデル平均を行う BNC を提案する. 評価実験により, モデル平均を行う構造間の類似度が低下する場合に分類精度が向上することを示し, 更に k -best 法とアンサンブル法を組み合わせた手法が, 小規模データで有効かつデータ規模にかかわらず既存手法と比べ分類精度が高くなることを示す.

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, アンサンブル学習

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, ノード間の依存関係を非循環有向グラフ (Directed Acyclic Graph: DAG) で表現した確率的グラフィカルモデルの一つである. このモデルは DAG を仮定することで, 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.

ベイジアンネットワークはデータから DAG 構造を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. 構造学習において, 構造の予測性の高さを示す学習スコアがある. 全ての候補構造から, 学習スコアを最適にする構造を探索するスコアベースアプローチが, ベイジアンネットワークの構

造学習で従来から行われてきた. 一般に学習スコアとして, 構造を所与とした学習データの周辺ゆう度を用いる. 周辺ゆう度はパラメータの事前分布にディリクレ分布を仮定すると, 閉形式で表すことができる.

ベイジアンネットワークにおいて, あ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]. 周辺ゆう度を用いて学習したベイジアンネットワークは, 全変数の同時確率を生成モデルとしてモデル化したものであり, BNC として用いる場合 General Bayesian Network (GBN) と呼ぶ. BNC の学習において Friedman らは, 説明変数を所与とした目的変数の条件付きゆう度を最大化する識別モデルが, 生成モデルである GBN と比べ分類精度が高くなると報告している [1].

しかし, 周辺ゆう度は漸近一致性を有しており, 推定される構造は漸近的に真の構造に収束するため, 識別モデルと比べて分類精度が低くならないはずである. Friedman らの報告に対し, Sugahara らが行った GBN と識別モデルの分類精度の比較実験では, デー

[†] 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: aomi@ai.lab.uec.ac.jp

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

c) E-mail: ueno@ai.is.uec.ac.jp

DOI:10.14923/transinfj.2019JDP7059

タ数が多いデータセットでは周辺ゆう度を用いて学習した GBN が識別モデルと比べ分類精度が高くなるが、データ数が少ないデータセットでは分類精度が低くなるという結果が報告されている [2]. Sugahara らはデータ数が少ない場合、GBN の目的変数をもつ親変数が多くなると目的変数のパラメータ推定が不安定となるため、分類精度が低くなってしまふと分析している [2]. これを解決するために、Sugahara らは目的変数が全ての説明変数の子にもつ制約をもった Augmented naive Bayes classifier (ANB) の厳密学習を提案している [2]. しかし、そもそも周辺ゆう度による厳密学習は漸近的に真の構造に収束するが、データ数が少ない場合、推定される構造の事後確率が低くなり、分類精度が劣ってしまうことが挙げられる。

一方、各構造の事後確率が低い場合は、モデル平均を行うことにより推論精度が向上できることが知られている [3]. ただし、モデル平均を BNC にそのまま適用すると、学習データから考えられる全ての候補構造について計算する必要があるため、現実的に計算が困難となる。この計算困難性を解決する先行研究として、学習スコアが最適順に K 個の構造を選択し、得られた K 個の構造のモデル平均を行う k -best 法が知られている [4], [5]. しかし、この手法では類似度の高い構造を選択してしまうため、得られた構造でモデル平均を行っても推論精度の向上が限定されてしまう。そこで、本研究では類似した構造の学習をなるべく減らすために、アンサンブル法 (非復元抽出) により構造学習した複数の構造についてモデル平均を行う BNC を提案する。

リポジトリ・データベースを適用した結果、提案手法は、

(1) 小規模データでは、GBN, ANB, 他モデル平均手法と比べ最も分類精度が高い

(2) BNC のモデル平均では、用いられる複数の構造間の類似度が低い方が分類精度が向上することを示した。また、これまで最も高い分類精度を示していた生成モデルとしての ANB 構造の厳密学習手法と比較しても、GBN のモデル平均が分類精度が高くなった。ANB 構造を仮定したモデル平均では、GBN のモデル平均と比べて構造の仮定によって構造間が類似してしまい、結果として分類精度の向上が限定されていることがわかった。

ただし、実験から提案したアンサンブル法が k -best 法に劣る場合も観測された。これを踏まえ、本研究で

は更にアンサンブル法により T 個のデータを発生させ、それぞれのデータに k -best 法を適用して $T \times K$ 個の構造を発生させ、それらのモデル平均を行うアンサンブル k -best 手法を提案する。実験により、この手法がデータ数が少ないデータで更に分類精度を改善し、データ規模にかかわらず識別モデル、GBN と比べ有意に分類精度が高くなることを示した。

2. ベイジアンネットワークと分類器

2.1 ベイジアンネットワーク

ベイジアンネットワークは、離散確率変数をノードとして、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される確率的グラフィカルモデルである。今、 $N+1$ 個の離散確率変数集合 $\{X_0, X_1, \dots, X_N\}$ について、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、変数 X_i が値 k をとるとき $X_i = k$ と書く。このとき、各変数 X_i の親変数集合を Π_i としたときの、ベイジアンネットワークの確率構造 G における同時確率分布 $p(X_0, \dots, X_N | G)$ は次のように書ける。

$$p(X_0, \dots, X_N | G) = \prod_{i=0}^N p(X_i | \Pi_i, G) \quad (1)$$

ベイジアンネットワークはデータから DAG 構造を推定する必要があり、この問題はベイジアンネットワークの構造学習と呼ばれる。今、学習データとしてデータ数 n のデータ $D = \langle \mathbf{x}^1, \dots, \mathbf{x}^n \rangle$ が与えられるとする。親変数集合 Π_i が j 番目のパターンをとったとき ($\Pi_i = j$ と書く)、 $X_i = k$ となる条件付き確率 $p(X_i = k | \Pi_i = j, G)$ を示すパラメータを θ_{ijk} とする。ベイジアンネットワークの構造学習では、このパラメータ θ_{ijk} の推定値として、期待事後確率推定値 (Expected a Posteriori: EAP) が最もよく用いられる。事前分布としてディリクレ分布を仮定すれば、EAP は次のとおりとなる [6].

$$\hat{\theta}_{ijk} = \frac{\alpha_{ijk} + n_{ijk}}{\alpha_{ij} + n_{ij}} \quad (2)$$

ここで、 n_{ijk} は変数 X_i の親ノード変数集合 Π_i が j 番目のパターンをとり、 $X_i = k$ となったときの頻度を表し、 $n_{ij} = \sum_{k=1}^{r_i} n_{ijk}$ である。データ数 n は $n = \sum_{j=1}^{q_i} n_{ij}$, ($i = 0, \dots, N$) となる。また、 α_{ijk} はディリクレ分布におけるハイパーパラメータで、

$\alpha_{ij} = \sum_{k=1}^{r_i} \alpha_{ijk}$ である。ベイジアンネットワークの構造 G をデータから生成モデルとして学習するために、候補構造から最適な学習スコアをも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。一般にその学習スコアとして周辺ゆう度が用いられてきた。周辺ゆう度を最大化することは、式 (3) に示すように事前分布 $p(G)$ が一様分布であるときの事後確率を最大化することである。

$$\begin{aligned} p(G | D) &= \frac{p(D | G)p(G)}{p(D)} \\ &\propto p(D | G)p(G) \\ &\propto p(D | G) \end{aligned} \quad (3)$$

ここで、ベイジアンネットワークの条件付き確率パラメータ集合を $\Theta = \{\theta_{ijk}\}, (i = 0, \dots, N, j = 1, \dots, q_i, k = 1, \dots, r_i)$ として、パラメータの事前分布として次のディリクレ分布 $p(\Theta | G)$ を仮定する。

$$p(\Theta | G) = \prod_{i=0}^N \prod_{j=1}^{q_i} \frac{\Gamma(\sum_{k=1}^{r_i} \alpha_{ijk})}{\prod_{k=1}^{r_i} \Gamma(\alpha_{ijk})} \prod_{k=1}^{r_i} \theta_{ijk}^{\alpha_{ijk}-1} \quad (4)$$

このとき、構造 G の周辺ゆう度は次式で表される。

$$p(D | G) = \prod_{i=0}^N \prod_{j=1}^{q_i} \frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij} + n_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(\alpha_{ijk} + n_{ijk})}{\Gamma(\alpha_{ijk})} \quad (5)$$

Heckerman ら [7] は、二つの構造がマルコフ等価であれば、それらのパラメータ同時確率分布は同一でなければならないというゆう度等価原理を、ベイジアンネットワーク学習に導入した。このゆう度等価に矛盾しないディリクレ分布の条件として次のハイパーパラメータが提案されている。

$$\alpha_{ijk} = \alpha p(X_i = k, \Pi_i = j | G^h) \quad (6)$$

ここで、 α は Equivalent Sample Size (ESS) と呼ばれる事前知識の重みを示す疑似データである。また、 G^h はユーザが事前に仮定するネットワーク構造であり、その構造 G^h を所与として ESS を α_{ijk} に分配する。このスコアは Bayesian Dirichlet equivalent (BDe) と呼ばれている。更に、Buntine [6] は ESS をパラメータ数で除した $\alpha_{ijk} = \alpha / (r_i q_i)$ としたスコアを提案している。このスコアは BDe の特殊系とみなすことができ、Bayesian Dirichlet equivalent uniform (BDeu)

と呼ばれる。Heckerman ら [7] や Ueno [8], [9] の報告によれば、ユーザが事前知識をもたない場合、無情報事前分布を用いた BDeu を用いるのが望ましいとしている。周辺ゆう度は漸近一致性を有しているため、周辺ゆう度を用いて構造を推定すれば漸的に真の構造に収束する。すなわち、構造学習において真の生成モデルを推定したい場合には周辺ゆう度最大化による方法は非常に有用である。

スコアベースアプローチによる構造学習は、探索する構造の数が変数数に対し指数的に増加する NP 困難問題である [10]。より多くの変数を扱うための方法としては整数計画法を用いた手法が提案されており、60 変数程度の学習が可能である [11]。また、近年では Bayes factor を用いた学習手法が提案されており、条件付き独立性を用いたエッジ刈りによって、千変数以上のネットワークの構造学習が実現されている [12]～[14]。

2.2 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のある一つのノードを目的変数、その他の変数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、目的変数を X_0 、説明変数を X_1, \dots, X_N とする構造 G の BNC を考える。このとき、分類を行うデータとして説明変数に対応するデータサンプル列 $\mathbf{x} = \langle x_1, \dots, x_N \rangle$ が与えられるとすると、目的変数の推定値 \hat{c} は次で求めることができる。

$$\begin{aligned} \hat{c} &= \arg \max_{c \in \{1, \dots, r_0\}} p(c | \mathbf{x}, G, \Theta) \\ &= \arg \max_{c \in \{1, \dots, r_0\}} \frac{p(c, \mathbf{x} | G, \Theta)}{p(\mathbf{x} | G, \Theta)} \\ &= \arg \max_{c \in \{1, \dots, r_0\}} p(c, \mathbf{x} | G, \Theta) \\ &= \arg \max_{c \in \{1, \dots, r_0\}} \prod_{i=0}^n \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} (\theta_{ijk})^{1_{ijk}} \\ &= \arg \max_{c \in \{1, \dots, r_0\}} \left[\prod_{j=1}^{q_0} \prod_{k=1}^{r_0} (\theta_{0jk})^{1_{ijk}} \right. \\ &\quad \left. \times \prod_{i: X_i \in C} \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} (\theta_{ijk})^{1_{ijk}} \right] \end{aligned} \quad (7)$$

ここで、 1_{ijk} はサンプル \mathbf{x} において $X_i = k$ かつ $\Pi_i = j$ であるとき 1 をとり、その他では 0 をとる変数である。また、 C は目的変数 X_0 の子変数集合である。

一般に、周辺ゆう度を用いて学習される BNC は、

全変数の同時確率を生成モデルとしてモデル化したものであり、General Bayesian Network (GBN) と呼ばれる。しかしながら、前述のとおり GBN の構造学習は NP 困難問題である [10] ため、規模の大きいネットワークの学習には膨大な時間を要する。一方、古くから説明変数が目的変数のみを親にもつという最も単純な構造を仮定した BNC として Naive Bayes [15], [16] が知られている。更に、Naive Bayes の構造では説明変数間の相関が考慮できないため、これを考慮した Tree-Augmented Naive Bayes (TAN) [1] が提案されている。TAN は、全説明変数が目的変数を親にもち、説明変数間で木構造をとることを仮定した構造であり、多項式時間で構造推定ができる。

Friedman らはリポジトリ・データを用いた実験を行い、GBN と比べより単純な Naive Bayes や TAN のほうが分類精度が高い場合があることを指摘した [1]。その理由として、BNC の分類精度向上のためには周辺ゆう度ではなく、説明変数を所与とした目的変数の条件付きゆう度を最大化するスコアを用いるべきであると主張した。このスコアとして、説明変数を所与とした目的変数の条件付き対数ゆう度 (Conditional Log Likelihood: CLL) を提案している。このようなスコアを最大化するモデルを識別モデルと呼ぶ。しかし、CLL は計算量が膨大であるため、CLL を近似する手法が幾つか提案されている。例として Carvalho らは、CLL スコアそのものを分解可能となるように近似した aCLL (approximate Conditional Log Likelihood) スコアを提案し、TAN の学習に用いている [17]。こういった識別モデルについては、Friedman らによる数値実験において、識別モデルが生成モデルである GBN より高い分類精度があると報告されている [1]。

しかし、そもそも周辺ゆう度は漸近一致性を有しているため、推定される構造は漸近的に真の構造に収束する。そのため、データ数が十分に大きいときに GBN の分類精度が低いことは奇異である。このような考察から Sugahara らは GBN と識別モデルの分類精度の比較実験を多くのリポジトリ・データを用いて行っている [2]。その結果、データ数が多いデータセットでは周辺ゆう度を用いて学習した GBN が識別モデルと比べ分類精度が高くなった一方で、データ数が少ないデータセットでは分類精度が低いことがわかった。Sugahara らは更に学習データのデータ数が少ない場合、学習データがスパースであり、結果として GBN の目的変数をもつ親変数が多くなってしま

ことで式 (7) の θ_{0jk} の推定が不安定となるため、分類精度が劣化してしまうと分析している。この分析から、Sugahara らは目的変数の親変数が存在しないような構造の条件を満たす Augmented naive Bayes Classifier (ANB) の生成モデルとしての厳密学習を提案している [2]。

2.3 Augmented naive Bayes Classifier

本節では Sugahara らの提案手法 [2] を紹介する。ANB は各説明変数が目的変数を親にもつ制約をもつ BNC である。ANB は従来、識別モデルとして学習されてきたが、Sugahara らが提案した厳密学習手法は、識別モデルとしての学習ではなく、生成モデルである GBN の学習に ANB 構造を制約として与えた手法である [2]。具体的には、Silander and Myllymäki の動的計画法を用いた GBN の厳密学習アルゴリズム [18] の候補構造として ANB 構造のみを用いる。今、全ての ANB 構造の候補構造集合 \mathcal{G}_{ANB} から学習スコアを最大にする構造 G^* を探索する。このとき、構造 G の学習スコアを $Score(G)$ 、説明変数集合を \mathbf{X} 、とすると、次式のようにして G^* を求める。

$$\begin{aligned} G^* &= \arg \max_{G \in \mathcal{G}_{ANB}} Score(G) \\ &= \arg \max_{G \in \mathcal{G}_{ANB}} \left\{ Score_0(G_0(\Pi_0)) \right. \\ &\quad \left. + \sum_{i: X_i \in \mathbf{X}} Score_i(G_i(\Pi_i)) \right\} \\ &= \arg \max_{G \in \mathcal{G}_{ANB}} \sum_{i: X_i \in \mathbf{X}} Score_i(G_i(\Pi_i)) \\ &= \arg \max_{G \in \mathcal{G}_{ANB}} Score_{ANB}(G) \end{aligned} \quad (8)$$

ここで、 $G_i(\Pi_i)$ は X_i 、 Π_i をノード集合、 Π_i の各変数から X_i に向かうエッジをエッジ集合とした DAG であり、 $Score_i(G_i(\Pi_i))$ はローカルスコアと呼ばれる。

Sugahara らは ANB の厳密学習によって得られた ANB を用いて、データ数が少ないデータに対する分類精度の改善を実験によって示している [2]。しかし、データ数が少ないデータにおいて分類精度が劣化する原因は、周辺ゆう度最大化によって推定される生成モデルの構造の事後確率が低く、真の構造とは大きく異なることが本質である。このような場合、一つ一つの構造の事後確率が低い場合にでも推論精度を向上できる複数の構造のモデル平均が有効であることが知られている [3]。本研究では、データ数が少ないデータ

の分類精度の劣化に対し、BNC のモデル平均を用いることで分類精度改善を行う。

3. BNC のモデル平均

ベイジアンネットワークにおけるモデル平均では、考えられる全ての候補構造について平均して対象の推論を行う。本研究ではこれを BNC における目的変数の推論に用いる。今、その候補構造の集合を \mathcal{G} 、与えられる学習データを D とする。モデル平均を式 (7) に導入すると、それぞれの候補構造 $G \in \mathcal{G}$ に対し、 $p(G)$ が一様であることを仮定して、次のように \hat{c} を推定する。

$$\begin{aligned} \hat{c} &= \arg \max_{c \in \{1, \dots, r_0\}} \sum_{G \in \mathcal{G}} \frac{p(G | D)}{\sum_{G' \in \mathcal{G}} p(G' | D)} p(c | \mathbf{x}, G, D) \\ &= \arg \max_{c \in \{1, \dots, r_0\}} \sum_{G \in \mathcal{G}} p(D | G) p(c | \mathbf{x}, G, D) \quad (9) \end{aligned}$$

しかし、式 (9) における \mathcal{G} の構造数は、学習データの変数数に対して超指数的に増加するため、計算時間が現実的ではない。この計算困難性を解消する先行研究として、学習スコアが最適な順に K 個の構造を選択し、得られた K 個の構造についてモデル平均を行う k -best 法が知られている [4], [5]。

3.1 k -best 法を用いた手法

Tian らによって提案されている k -best 法は、モデル平均において計算する構造の数を限定するために、学習スコアを用いて最適な順に K 個の構造を選択する手法である [4]。 k -best 法では、次の手順で K 個の構造を選択する。

(1) 考えられる全ての $G_i(\Pi_i)$ のローカルスコア $Score_i(G_i(\Pi_i))$ を計算する。

(2) 各変数 X_i について親ノード候補 $\mathcal{C}(\mathcal{C} \subseteq X \setminus \{X_i\})$ を所与としたときの、上位 K 個の最適親ノード集合を選択する。

(3) スコアが上位 K 個のベイジアンネットワークを推定する。

この手法を用いて学習データ D に対し K 個の構造を選択し、得られた複数の構造に対し式 (9) で目的変数の推定を行う。また、等価スコアの DAG を除いて、 K 個の構造を選択するよう改良した k -best 法 [5] (以降、 k -bestEC と呼ぶ) も用いる。 k -bestEC は通常の k -best と比べて、式 (9) の $p(D | G)$ が各構造 G について等価な構造が除かれる分、より多様となることが知られている。

しかし、 k -bestEC でさえも類似した構造を選択してしまいがちで、得られた構造でモデル平均を行っても推論精度の向上が限定的となってしまう可能性がある。そこで、本研究では得られる構造が更に類似しないようにアンサンブル法を用いた手法を提案する。

3.2 アンサンブル法を用いた手法

アンサンブル法は同じ問題を解くための複数の分類器について訓練を行う方法である。幾つかの分類器の結合により行われる予測は、最も優れた単一の分類器による予測よりも正しいことが多いと経験的に示されている [19]。今、もとの学習データから複数の学習データを生成し、それぞれについて構造学習を行うことを考える。本研究では、もとの学習データから非復元抽出を行い新しくデータを生成することを複数回繰り返し、複数の学習データを得る方法を用いる。この方法で得られた T 個の学習データそれぞれに対し、Silander and Myllymäki の動的計画法を用いたアルゴリズム [18] を用いて GBN の構造学習を行う。学習データを D_1, \dots, D_T とし、それぞれを構造学習して得られる構造を G_1, \dots, G_T とすれば、式 (9) より次式で \hat{c} を推定する。

$$\hat{c} = \arg \max_{c \in \{1, \dots, r_0\}} \sum_{t=1}^T p(D_t | G_t) p(c | \mathbf{x}, G_t, D_t) \quad (10)$$

この手法はデータを複数生成することで、一つのデータに対する過学習を防ぎ、データ数の少ない学習データに対応するねらいがある。以降、この手法を Ens(GBN) と呼ぶ。

4. 評価実験

4.1 実験概要

UCI リポジトリ・データベース^(注1)から表 1 に示す 31 個の学習データを用いて、識別モデルである BNC と生成モデルである GBN, Sugahara らの厳密学習手法 [2] で構造学習した ANB, 提案手法を含むモデル平均手法の比較実験を行った。なお、構造学習の際に候補構造を ANB 構造と仮定すればモデル平均手法が適用可能であるため、ANB とモデル平均手法の組み合わせ手法も比較対象として加え、実験を行った。ここで、データベースに含まれる数値的データは De Campos らと同様に、各説明変数ごとの中央値を区切りとして 2 値をとるカテゴリーデータに変換し、欠損値を含む

(注1) : <https://archive.ics.uci.edu/ml/index.php>

表 1 実験で用いたデータセット
Table 1 The datasets for the experiment.

#	学習データ名	データ数
1	lenses	24
2	mux6	64
3	post	87
4	zoo	101
5	Hayes-Roth	132
6	iris	150
7	wine	178
8	glass	214
9	Congressional Voting Records	232
10	heart	270
11	Breast Cancer	277
12	cleve	296
13	liver	345
14	MONK's Problems	432
15	threeOf9	512
16	Balance Scale	625
17	crx	653
18	Breast Cancer Wisconsin	683
19	Australian	690
20	pima	768
21	Tic-Tac-Toe	958
22	banknote authentication	1372
23	Solar Flare	1389
24	Contraceptive Method Choice	1473
25	Car Evaluation	1728
26	led7	3200
27	shuttle-small	5800
28	Nursery	12960
29	EEG	14980
30	HTRU2	17898
31	MAGIC Gamma Telescope	19020

サンプルは取り除いた [20]. 式 (9) に示すようにモデル平均は周辺ゆう度によって平均される. 本研究では, 周辺ゆう度として最もよく知られている BDeu を用いる. なお, BDeu におけるハイパーパラメータである ESS α は 1.0 とした. これは, $\alpha = 1.0$ のときデータの影響を最大化するために最適であるという報告に基づく [8], [9].

以下に比較する手法を示す.

- 識別モデル
 - NB: Naive Bayes
 - TAN: 候補構造に TAN を仮定し, 学習スコアとして条件付き相互情報量を用いた [1].
 - aCLL-TAN: 候補構造に TAN を仮定し, 学習スコアとして aCLL を用いた [17].
 - GBN: 学習スコアとして BDeu を用いて, 動的計画法を用いたアルゴリズム [18] で構造学習した.
 - ANB: 学習スコアとして BDeu を用いて, Sugahara らの厳密学習手法 [2] で構造学習した.

- AdaBoost(TAN): AdaBoost [21] を用いた BNC 学習の先行研究として, 条件付き相互情報量を用いて TAN の学習に適用しているものが既に提案されている [22]. 本研究では, この手法 [22] を用いて TAN を 10 回反復学習した.

- AdaBoost(GBN): 提案手法と条件を揃えるために BDeu で動的計画法 [18] を用いて AdaBoost を適用して GBN を 10 回反復学習した.

- モデル平均手法
 - kBest10(GBN): k -best 法 [4] により, $K = 10$ 個の構造を選択してモデル平均を行った.

- kBest100(GBN): k -best 法 [4] により, $K = 100$ 個の構造を選択してモデル平均を行った.

- kBestEC10(GBN): k -bestEC [5] により, $K = 10$ 個の構造を選択してモデル平均を行った.

- kBestEC100(GBN): k -bestEC [5] により, $K = 100$ 個の構造を選択してモデル平均を行った.

- Ens(GBN): 学習データの 90% のサイズになるように $T = 10$ 個のデータをランダムに生成し, それぞれに対し GBN を構造学習してモデル平均を行った.

- ANB とモデル平均の組み合わせ手法
 - kBestEC10(ANB): k -bestEC [5] を用いたもので, 候補構造に ANB 構造の制約を加えた. $K = 10$ 個の構造を選択してモデル平均を行った.

- Ens(ANB): 学習データの 90% のサイズになるように $T = 10$ 個のデータをランダムに生成し, それぞれに対し Sugahara らの ANB の厳密学習手法 [2] で ANB を構造学習してモデル平均を行った.

なお, k -best 法における K は, Chen and Tian の実験 [5] に従い $K = 10, 100$ で実験を行った. また, 本実験では解析の妥当性を示すために, 各 BNC の構造学習と分類に対して 10 分割交差検証を行った. 10 分割交差検証では, データを 10 分割した後, 一つをテストデータ, 残り九つを学習データとして学習と分類を計 10 回繰り返す. テストデータの各説明変数データに対し, 推定された目的変数の値が正しく分類されている割合を算出し, 10 分割交差検証におけるその平均を分類精度とした.

4.2 結果と考察

表 2 は各 BNC のそれぞれのデータセットにおける分類精度とその平均値を示したものである. 更に, データ数下位 15 個すなわちデータ 1~15 を小データ群, 他 16 個すなわちデータ 16~31 を大データ群と分け, それぞれについての分類精度の平均値も示した.

表 2 各 BNC の分類精度
Table 2 Accuracies of each BNCs.

# 学習データ名	既存手法					モデル平均 (構造数:10)					モデル平均 (構造数:100)					
	識別モデル					k-best 法					アンサンブル法					
	NB	TAN	TAN	GBN	ANB	AdaBoost (TAN)	AdaBoost (GBN)	kBest10 (GBN)	kBestEC10 (GBN)	Ens (GBN)	Ens (GBN)	kBest10 (GBN)	Ens (GBN)	kBest100 (GBN)	kBestEC100 (GBN)	Ens (GBN)
1 lenses	0.6250	0.7083	0.7083	0.8125	0.8750	0.6667	0.8125	0.8750	0.8333	0.8333	0.6250	0.8750	0.8333	0.8333	0.8333	0.8333
2 mux6	0.5469	0.6094	0.5938	0.4531	0.3125	0.5938	0.4531	0.3594	0.3594	0.6094	0.5781	0.4063	0.3750	0.4219	0.6250	
3 post	0.6552	0.6322	0.5977	0.7126	0.7126	0.6552	0.7126	0.7126	0.7126	0.7126	0.6552	0.7126	0.7126	0.7126	0.7126	
4 zoo	0.9901	0.9406	0.9505	0.9426	0.9604	0.9901	0.9406	0.9426	0.9505	0.9604	0.9703	0.9604	0.9505	0.9505	0.9505	
5 Hayes-Roth	0.8106	0.6439	0.6742	0.6136	0.8333	0.6970	0.6136	0.6136	0.7652	0.6136	0.7955	0.8333	0.7955	0.7803	0.7727	
6 iris	0.7133	0.8267	0.8200	0.8267	0.8067	0.8267	0.8200	0.8267	0.8267	0.8267	0.8267	0.8267	0.8267	0.8200	0.8267	
7 wine	0.9270	0.9213	0.9157	0.9438	0.9270	0.9326	0.9213	0.9494	0.9438	0.9551	0.9270	0.9213	0.9438	0.9438	0.9438	
8 glass	0.5421	0.5467	0.6215	0.5607	0.5280	0.5981	0.5701	0.5607	0.5841	0.5701	0.5888	0.5234	0.5794	0.5748	0.5748	
9 Congressional Voting Records	0.9095	0.9526	0.9224	0.9612	0.9526	0.9310	0.9655	0.9612	0.9612	0.9698	0.9569	0.9569	0.9612	0.9655	0.9698	
10 heart	0.8296	0.8333	0.8148	0.8296	0.8444	0.8333	0.8074	0.8296	0.8333	0.8407	0.8222	0.8407	0.8296	0.8333	0.8370	
11 Breast Cancer	0.7365	0.7220	0.6968	0.7076	0.6751	0.7148	0.7509	0.7076	0.7220	0.7004	0.7148	0.6782	0.7184	0.7329	0.7220	
12 cleve	0.8311	0.8243	0.8446	0.8074	0.8142	0.8176	0.7939	0.8074	0.8074	0.8108	0.8209	0.8142	0.8108	0.8176	0.8176	
13 liver	0.6464	0.6609	0.6522	0.5768	0.6058	0.6638	0.5971	0.5681	0.6087	0.6174	0.6783	0.6261	0.6174	0.6261	0.6232	
14 MONK's Problems	0.7500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
15 threeOf9	0.8008	0.8691	0.8906	0.8691	0.8672	0.8789	0.9063	0.9375	0.9277	0.8906	0.8926	0.8789	0.8809	0.9434	0.9023	
小データ群 (1~15) の平均値	0.7543	0.7794	0.7802	0.7745	0.7810	0.7886	0.7777	0.7768	0.7891	0.7941	0.7901	0.7903	0.7890	0.7971	0.8074	
16 Balance Scale	0.9168	0.8688	0.8624	0.9168	0.9168	0.8784	0.9168	0.8896	0.9040	0.9168	0.8704	0.9168	0.9104	0.9152	0.9040	
17 crx	0.8392	0.8515	0.8453	0.8392	0.8622	0.8331	0.8591	0.8392	0.8392	0.8499	0.8392	0.8652	0.8392	0.8484	0.8499	
18 Breast Cancer Wisconsin	0.9766	0.9678	0.9473	0.9766	0.9766	0.9795	0.9575	0.9766	0.9736	0.9766	0.9751	0.9751	0.9766	0.9751	0.9736	
19 Australian	0.8348	0.8290	0.8478	0.8565	0.8580	0.8333	0.8638	0.8565	0.8565	0.8464	0.8362	0.8594	0.8565	0.8478	0.8464	
20 pima	0.7057	0.7188	0.7031	0.7253	0.7188	0.7083	0.7240	0.7266	0.7292	0.7227	0.7201	0.7161	0.7266	0.7331	0.7266	
21 Tic-Tac-Toe	0.6889	0.7599	0.7192	0.8549	0.8445	0.7505	0.9123	0.8539	0.8539	0.8466	0.8518	0.8445	0.8549	0.8486	0.8518	
22 banknote authentication	0.8433	0.8819	0.8761	0.8812	0.8812	0.8754	0.8776	0.8812	0.8812	0.8812	0.8812	0.8812	0.8812	0.8812	0.8812	
23 Solar Flare	0.7804	0.7970	0.8200	0.8431	0.8431	0.8143	0.8431	0.8431	0.8431	0.8431	0.8236	0.8431	0.8431	0.8431	0.8431	
24 Contraceptive Method Choice	0.4644	0.4725	0.4650	0.4549	0.4270	0.4779	0.4399	0.4528	0.4487	0.4521	0.4623	0.4270	0.4481	0.4616	0.4487	
25 Car Evaluation	0.8623	0.9392	0.9427	0.9416	0.9416	0.9444	0.9340	0.9306	0.9277	0.9172	0.9433	0.9172	0.9259	0.9294	0.9161	
26 led7	0.7288	0.7309	0.7347	0.7288	0.7288	0.7300	0.7288	0.7288	0.7294	0.7284	0.7281	0.7284	0.7281	0.7303	0.7309	
27 shuttle-small	0.9383	0.9567	0.9538	0.9693	0.9716	0.9681	0.9662	0.9693	0.9693	0.9693	0.9714	0.9702	0.9693	0.9693	0.9693	
28 Nursery	0.9028	0.9249	0.9188	0.9345	0.9174	0.9209	0.9509	0.9270	0.9269	0.9305	0.9152	0.9143	0.9258	0.9239	0.9265	
29 EEG	0.5774	0.6298	0.6138	0.6844	0.6895	0.6031	0.6906	0.6862	0.6887	0.6881	0.6931	0.6955	0.6893	0.6885	0.6899	
30 HTRUC2	0.8966	0.9141	0.9141	0.9141	0.9141	0.9102	0.9073	0.9141	0.9141	0.9141	0.9141	0.9141	0.9141	0.9141	0.9141	
31 MAGIC Gamma Telescope	0.7447	0.7769	0.7656	0.7859	0.7879	0.7734	0.7849	0.7855	0.7861	0.7859	0.7877	0.7880	0.7871	0.7871	0.7860	
大データ群 (16~31) の平均値	0.7938	0.8137	0.8081	0.8317	0.8299	0.8126	0.8348	0.8288	0.8295	0.8293	0.8258	0.8285	0.8298	0.8310	0.8286	
平均値	0.7747	0.7971	0.7946	0.8040	0.8062	0.8000	0.8072	0.8036	0.8099	0.8122	0.8085	0.8100	0.8100	0.8146	0.8184	

なお、表 2 中の太字は各学習データで最も分類精度が高かったものを表している。

大データ群での分類精度の平均値では識別モデル, ANB, 提案手法を含んだモデル平均手法と比べて, AdaBoost(GBN), GBN の順で高い結果を得た。ただし, AdaBoost(GBN) と GBN の結果には有意差はない。AdaBoost は漸近一致性をもたないが準最適性を持ち, データ数が大きい場合に最適に近い結果が得られることが知られている [23]。また, 漸近一致性を有している GBN はデータ数が大きくなると真の構造に近づいたため, データ数が大きい場合で分類精度が向上していることがわかる。この結果は, Sugahara らの結果 [2] と一致する。一方小データ群では, モデル平均手法である KBestEC100(GBN), Ens(GBN) の順で, 分類精度が高い結果となった。特にデータ 2, 6, 7, 15 で, これらのモデル平均手法が既存手法と比べ分類精度が向上しており, モデル平均手法はデータ数が少ないデータにおいて, 有効な BNC であると考えられる。しかし, これらのモデル平均手法に対して, 大データ群では分類精度が最も良かった AdaBoost(GBN) が小データ群では分類精度が劣っている。この理由を分析するために, 表 3 に各手法で学習された GBN の目

表 3 目的変数の親変数の数
Table 3 The number of parents' nodes of the objective variable.

	GBN	AdaBoost (GBN)	kBest10 (GBN)	kBestEC10 (GBN)	Ens (GBN)
小データ群の平均値	1.97	3.45	1.87	1.85	1.97
大データ群の平均値	1.30	3.62	1.32	1.32	1.14
平均値	1.65	3.53	1.60	1.60	1.57

的変数の親変数の数を示した。GBN, 他のモデル平均手法と比べて AdaBoost(GBN) で得られた構造は目的変数の親変数の数が多くなっている。このような場合, データ数が少ないときにパラメータ推定の精度が悪くなってしまい, 分類精度に悪影響を及ぼすことが Sugahara らによって述べられている [2]。

また, モデル平均手法において, 構造数が 10 の場合で比べた場合には Ens(GBN) の分類精度が kBestEC10(GBN) を上回っており, 特にデータ 2, 4, 7 などの小データにおいて差が見られる。更に, Ens(GBN) と Ens(ANB) を比較した場合 Ens(GBN) が平均分類精度が高く, kBestEC10(GBN) と kBestEC10(ANB) を比較した場合 kBestEC10(ANB) が平均分類精度が高いという結果が得られた。すなわち, 構造学習において ANB

の構造を候補構造に仮定しないほうが、分類精度が高い。これらの結果から共通して考えられることは、分類精度が下回った手法は比較対象の手法に対して構造間の類似度が高く、モデル平均の利点を十分に発揮できていない可能性がある。k-best 法は類似構造を選択しやすく、ANB 構造は目的変数に全ての説明変数へのエッジが引かれる分、GBN に比較して構造間の類似度が高くなってしまふ。このことから、BNC のモデル平均において、k-best 法やアンサンブル法で得られる構造間の類似度が分類精度に影響している可能性が示唆される。

4.3 BNC のモデル平均における構造の類似度と分類精度に関する分析

本節では、前節の仮説に従い、モデル平均で用い

られる構造の類似度と分類精度の関連性についての分析を行う。ベイジアンネットワークの構造の類似度を調査するために、Structural Hamming Distance (SHD) [24] を用いる。SHD は構造間の距離を示すもので、ここでは提案手法によって生成した複数の構造に対しそれぞれの SHD を算出し、それらの平均値を構造の類似度の指標として用いる。

表 4 は表 2 のデータに対するモデル平均 BNC の構造間平均 SHD を記したものである。なお、表 4 の太字は、データごとのモデル平均手法で最も SHD の値が高いものを表している。

表 4 より、小データであるデータ 4, 7, 9 のように Ens(GBN) の SHD の値が k-best 法と比べて明らかに大きくなっているものがある。このとき、Ens(GBN) は

表 4 モデル平均 BNC の構造間平均 SHD
Table 4 Average SHD of model averaging BNCs.

学習データ名	構造数:10					構造数:100	
	k-best 法		アンサンブル法	モデル平均 + ANB		k-best 法	
	kBest10 (GBN)	kBestEC10 (GBN)	Ens (GBN)	kBestEC10 (ANB)	Ens (ANB)	kBest100 (GBN)	kBestEC100 (GBN)
1 lenses	1.39	2.61	1.07	2.30	0.33	3.09	4.11
2 mux6	2.92	3.24	0.61	1.96	0.89	4.33	4.56
3 post	0.98	2.33	0.42	1.92	0.42	2.31	3.32
4 zoo	1.71	7.03	32.56	7.41	12.98	5.43	10.51
5 Hayes-Roth	1.44	2.35	0.00	2.39	0.07	2.87	3.78
6 iris	2.25	5.09	1.87	2.86	1.54	4.23	6.12
7 wine	1.53	7.57	11.85	2.56	5.50	4.04	9.40
8 glass	0.55	3.71	4.76	4.33	5.31	3.33	5.36
9 Congressional Voting Records	0.00	6.37	23.27	6.60	24.25	3.03	8.91
10 heart	0.11	4.48	7.19	2.09	6.42	3.61	7.64
11 Breast Cancer	0.00	2.36	1.02	2.20	1.02	0.76	4.67
12 cleve	0.00	3.74	5.95	2.16	5.07	2.50	7.34
13 liver	2.56	4.98	4.27	2.08	4.17	4.45	6.63
14 MONK's Problems	1.18	1.80	0.00	2.99	2.43	2.78	3.61
15 threeOf9	2.39	3.16	4.33	3.44	3.36	2.59	3.79
小データ群 (1~15) の平均値	1.27	4.05	6.61	3.15	4.92	3.29	5.98
16 Balance Scale	0.43	2.00	0.00	2.21	0.00	2.57	3.70
17 crx	5.17	9.36	8.68	3.70	6.29	7.18	10.63
18 Breast Cancer Wisconsin	0.00	5.73	0.32	1.80	0.00	1.26	6.98
19 Australian	4.59	6.25	10.79	3.97	9.39	5.60	9.82
20 pima	0.71	5.05	3.30	3.15	1.97	4.63	7.10
21 Tic-Tac-Toe	0.89	7.86	8.06	7.18	5.85	7.19	10.54
22 banknote authentication	2.60	5.54	0.00	3.74	0.00	3.77	6.88
23 Solar Flare	0.48	5.20	3.24	4.38	2.58	4.10	7.19
24 Contraceptive Method Choice	0.83	5.57	4.64	2.79	3.60	3.63	6.67
25 Car Evaluation	0.86	4.04	0.32	1.98	0.38	4.98	5.75
26 led7	0.02	3.58	0.00	1.80	0.00	1.21	5.49
27 shuttle-small	4.56	6.23	1.80	5.83	5.05	5.56	6.87
28 Nursery	1.36	2.15	0.70	2.30	0.28	2.90	3.67
29 EEG	2.51	9.04	7.99	12.07	15.40	6.92	10.26
30 HTRU2	0.00	6.03	0.32	5.01	0.32	0.80	9.56
31 MAGIC Gamma Telescope	2.92	9.02	3.82	7.14	1.36	4.28	12.47
大データ群 (16~31) の平均値	1.75	5.79	3.37	4.32	3.28	4.16	7.72
平均値	1.47	4.82	4.98	3.64	4.16	3.72	6.70

表 2 より GBN, ANB, k -best 法以上の分類精度を示し, 構造間の類似度が低くなると分類精度向上に好影響を与えることが考えられる. また, kbestEC10(GBN) は, Ens(GBN) と比べると SHD が特筆して大きい値にならない. 同じ構造数においてアンサンブル法が k -best 法と比べ分類精度が高いのは, データ 4, 7, 9 のように場合によって構造間の類似度が低下するためであると考えられる. 次に, ANB とモデル平均手法の組み合わせ手法に着目する. Ens(ANB) を Ens(GBN) と比較すると全体的に SHD が低く, 平均値で比較しても Ens(ANB) の SHD が劣っている. Ens(ANB) が ANB 構造の仮定によって, モデル平均の構造間距離の大きさが限定的になってしまい, GBN のモデル平均と比べて分類精度の向上が低くなっていると考えられる. 実際に表 2 を確認すると, 小データ群と大データ群ともに SHD の平均値は Ens(GBN) が Ens(ANB) と比べて高く, 分類精度も高い. これは, k -bestEC においても同様のことが言える. 更に, データ 5, 11 のようにアンサンブル法では構造の類似度が高い場合であっても, kbestEC10(GBN) は類似度を低下させ, 分類精度が高い場合があることも確認された. 本結果より, k -best 法とアンサンブル法を組み合わせると更に精度を向上させるかもしれない. そこで, 本研究では更に, k -best 法とアンサンブル法を組み合わせ手法を提案する.

5. アンサンブル k -best 手法

5.1 アンサンブル k -best 手法の概要

本節では 3.1 で述べた k -best 法と 3.2 で提案したアンサンブル法を組み合わせたアンサンブル k -best 手法を提案する. アンサンブル k -best 手法では, 非復元抽出法によってもとの学習データから T 個の複数データを生成し, それらのデータそれぞれについて k -best 法で最適な順に K 個の構造を選択する. この手法により, $T \times K$ 個の構造が学習され, それら全ての構造についてモデル平均を行って目的変数の値を推定する. 学習データを D_1, \dots, D_T とし, それぞれのデータについて k -best 法で K 個の構造を学習するとき, データ D_t を用いて k -best 法で得られた k 番目の構造を G_t^k とすれば, 次式で目的変数を推定する.

$$\hat{c} = \arg \max_{c \in \{1, \dots, r_0\}} \sum_{t=1}^T \sum_{k=1}^K p(D_t | G_t^k) p(c | \mathbf{x}, G_t^k, D_t) \quad (11)$$

アンサンブル k -best 手法では, 通常の k -best 法 [4] と比べ分類精度が高かった k -bestEC [5] を用いる.

5.2 アンサンブル k -best 手法の評価実験

4.1 と同じ条件で, 前節で提案したアンサンブル k -best 手法を追加し, 分類精度の比較を行った. 追加した手法を以下に示す.

- Ens-kBestEC(GBN): GBN について k -bestEC [5] とアンサンブルモデル平均を組み合わせたアンサンブル k -best 手法による BNC.

なお, アンサンブル k -best 手法では, もとの学習データの 90% のサイズとなるように $T = 10$ 個のデータを生成し, それぞれについて k -bestEC で $K = 10$ 個の構造を選択した.

表 2 の「アンサンブル k -best 手法」が, 本節で追加した手法の分類精度である. 結果としては, アンサンブル k -best 手法が, 比較した BNC の中で分類精度の平均値が最も高くなった. これについて, 提案したアンサンブル k -best 手法と既存の識別モデル, GBN, ANB との分類精度の平均値で比較を行ったものが表 5 である. なお, 表の p 値は提案手法と他手法に対してウィルコクソンの符号順位検定を行い, 検定の多重性を考慮して hommel 法により補正したものを記した. この結果より, アンサンブル k -best 手法は, 有意水準 5% において既存の識別モデル, GBN との分類精度の有意な差が認められた.

k -bestEC が効果的であったデータ 5, 11 に着目すると, アンサンブル k -best 手法が, k -bestEC とアンサンブル法より高い分類精度を示している. これにより, 提案手法は k -best 法を用いた BNC のモデル平均で分類精度が向上する場合に, 更に分類精度が向上することが確認された. 以上より, 提案したアンサンブル k -best 手法は, データ規模にかかわらず既存手法と比べ分類精度の向上が確認された.

一方で, アンサンブル法, 提案手法の問題点として計算時間が増加する点が挙げられる. 表 6 の計算環境で, モデル平均手法の構造学習の計算時間を示したものが表 7 である. 提案手法は, 生成したデータセッ

表 5 既存手法との分類精度の比較
Table 5 Comparison of accuracies between the existing methods and the proposed method.

	aCLL-					Ens-kBestEC (GBN)
	NB	TAN	TAN	GBN	ANB	
分類精度	0.7747	0.7971	0.7946	0.8040	0.8062	0.8184
p 値	0.0016	0.0050	0.0087	0.0434	0.1703	-

表 6 計算環境
Table 6 Computational environment.

CPU	Intel(R) Xeon(R) E5-2630 v4 10 Cores, 2.20 GHz
System Memory	128 GB
ソフトウェア	Java 1.8

表 7 モデル平均手法の計算時間
Table 7 Computation time of model averaging
BNCs.

構造数:10			
	kBest10 (GBN)	kBestEC10 (GBN)	Ens (GBN)
計算時間の平均値	3277 ms	4051 ms	25084 ms
構造数:100			
	kBest100 (GBN)	kBestEC100 (GBN)	Ens- kBestEC
計算時間の平均値	25345 ms	27657 ms	40609 ms

トそれぞれに対し構造学習を行うため計算時間を要する。これらに対し、並列コンピューティング等での計算時間の削減が今後の課題である。

6. むすび

本研究では、BNC にモデル平均を導入した。一般には考えられる全候補構造について構造の平均化を行うが、計算が現実的に困難である。本研究ではこの計算困難性を解消するためにアンサンブル法を用いた手法を提案した。リポジトリ・データベースを用いた実験によって、アンサンブル法は GBN で分類精度が劣化していたデータ数が少ない学習データについて分類精度の向上が確認できた。また、これまでに最も良い分類精度を示していた ANB 構造の厳密学習手法と比較しても、分類精度の向上が確認された。更に結果の分析を行ったところ、モデル平均手法では構造間の類似度が低い方が分類精度が向上することがわかった。ANB のモデル平均は構造間が類似してしまい、分類精度向上が限定的となった。提案したアンサンブル法は、従来手法である k -best 法と比べ構造間の類似度が低くなり、このときに分類精度が向上することが確認された。一方で、アンサンブル法で構造間が類似する場合であっても、 k -best 法は構造間の類似度が低くなり、分類精度が向上することもあった。この結果から、本研究では更にアンサンブル法と k -best 法を組み合わせたアンサンブル k -best 手法を提案した。本提案手法が既存の識別モデル、GBN に対しデータ規模にかかわらず有意に分類精度が高くなったことを実

験で示した。アンサンブル k -best 手法は k -best 法が有効であったデータに対して、更に分類精度が向上することも確認された。

今後の課題として、なるべく類似しない構造を選択する M-Modes アルゴリズム [25] の導入や、更に多様な構造を選択する探索アルゴリズムを考案し、BNC のモデル平均を行うことで本手法の有意性を検証する。また、アンサンブルのバイアスを考慮するために、EIC (Extended Information Criterion) [26] のようなスコアを周辺ゆう度を用いて導入することも考えたい。

文 献

- [1] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Mach. Learn.*, vol.29, no.2, pp.131–163, 1997.
- [2] S. Sugahara, M. Uto, and M. Ueno, "Exact learning augmented naive bayes classifier," *Proc. Ninth International Conference on Probabilistic Graphical Models*, vol.72, pp.439–450, 2018.
- [3] D.M. Chickering and D. Heckerman, "A comparison of scientific and engineering criteria for Bayesian model selection," *Statistics and Computing*, vol.10, pp.55–62, 2000.
- [4] J. Tian, R. He, and L. Ram, "Bayesian model averaging using the k -best Bayesian network structures," *Proc. Twenty-Sixth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-10)*, pp.589–597, 2010.
- [5] Y. Chen and J. Tian, "Finding the k -best equivalence classes of bayesian network structures for model averaging," *Twenty-Eighth AAAI Conference on Artificial Intelligence*, pp.2431–2438, 2014.
- [6] W. Buntine, "Theory refinement on Bayesian networks," *Proc. Seventh Conference on Uncertainty in Artificial Intelligence*, pp.52–60, 1991.
- [7] D. Heckerman, D. Geiger, and D. Chickering, "Learning Bayesian networks: The combination of knowledge and statistical data," *Mach. Learn.*, vol.20, pp.197–243, 1995.
- [8] M. Ueno, "Robust learning Bayesian networks for prior belief," *Proc. Twenty-Seventh Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-11)*, pp.698–707, 2011.
- [9] M. Ueno, "Learning networks determined by the ratio of prior and data," *Proc. Twenty-Sixth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-10)*, pp.598–605, 2010.
- [10] D.M. Chickering, "Leaning Bayesian networks is NP-complete," *Learning from Data*, pp.121–130, 1996.
- [11] M. Bartlett and J. Cussens, "Advances in Bayesian network learning using integer programming," *Proc. Twenty-Ninth Conference Annual Conference on Un-*

- certainty in Artificial Intelligence (UAI-13), pp.182–191, 2013.
- [12] K. Natori, M. Uto, Y. Nishiyama, S. Kawano, and M. Ueno, “Constraint-based learning bayesian networks using bayes factor,” *Advanced Methodologies for Bayesian Networks*, pp.15–31, 2015.
- [13] K. Natori, M. Uto, and M. Ueno, “Consistent learning bayesian networks with thousands of variables,” *Advanced Methodologies for Bayesian Networks (Proc. Machine Learning Research)*, vol.73, pp.57–68, 2017.
- [14] 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.J101-D, no.5, pp.754–768, May 2018.
- [15] M.E. Maron and J.L. Kuhns, “On relevance, probabilistic indexing and information retrieval,” *J. ACM*, vol.7, no.3, pp.216–244, 1960.
- [16] M. Minsky, “Steps toward artificial intelligence,” *Proc. IRE*, vol.49, no.1, pp.8–30, 1961.
- [17] 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.
- [18] T. Silander and P. Myllymäki, “A simple approach for finding the globally optimal Bayesian network structure,” *Proc. Twenty-Second Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-06)*, pp.445–452, 2006.
- [19] L.K. Hansen and P. Salamon, “Neural network ensembles,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.12, no.10, pp.993–1001, 1990.
- [20] C.P. De Campos, M. Cuccu, G. Corani, and M. Zaffalon, “Extended tree augmented naive classifier,” *European Workshop on Probabilistic Graphical Models*, pp.176–189, 2014.
- [21] Y. Freund and R.E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” *J. Computer and System Sciences*, vol.55, no.1, pp.119–139, 1997.
- [22] Y. Jing, V. Pavlović, and J.M. Rehg, “Boosted Bayesian network classifiers,” *Mach. Learn.*, vol.73, no.2, pp.155–184, 2008.
- [23] W. Jiang, “Process consistency for adaboost,” *The Annals of Statistics*, vol.32, no.1, pp.13–29, 2004.
- [24] I. Tsamardinos, L.E. Brown, and C.F. Aliferis, “The max-min hill-climbing Bayesian network structure learning algorithm,” *Mach. Learn.*, vol.65, no.1, pp.31–78, 2006.
- [25] C. Chen and C. Yuan, “Learning diverse bayesian network,” *Thirty-Third AAAI Conference on Artificial Intelligence*, pp.7793–7800, 2019.
- [26] M. Ishiguro, Y. Sakamoto, and G. Kitagawa, “Bootstrapping Log Likelihood and EIC, an Extension of

AIC,” *Annals of the institute of Statistical Mathematics*, vol.49, no.3, pp.411–434, 1997.

(2019年7月16日受付, 10月23日再受付,
11月26日早期公開)



青見 樹

2019年電気通信大学情報理工学部卒。同年、同大学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。