

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

2019年2月26日

情報数理工学コース

学籍番号 1511003

青見樹

指導教員 植野真臣

目次

1	まえがき	3
2	ベイジアンネットワークと分類器	5
2.1	ベイジアンネットワーク	5
2.2	ベイジアンネットワーク分類器	6
2.3	Augmented naive Bayes Classifier	8
3	BNC のモデル平均	10
3.1	k -best 法を用いた手法	10
3.2	アンサンブル法を用いた手法	11
4	評価実験	12
4.1	実験概要	12
4.2	結果と考察	14
4.3	BNC のモデル平均における構造の類似性と分類精度に関する分析	15
5	アンサンブル k -best 手法	18
5.1	アンサンブル k -best 手法の概要	18
5.2	アンサンブル k -best 手法の評価実験	18
6	むすび	20

表目次

1	実験で用いたデータセット	13
2	各 BNC の分類精度	15
3	モデル平均 BNC の構造間平均 SHD	17
4	既存手法との分類精度の比較	19

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は、構造を所与とした学習データの周辺尤度を用いる。周辺尤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 (Friedman et al. 1997)。周辺尤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 et al. (1997) は、説明変数を所与とした目的変数の条件付き尤度を最大化する識別モデルが、生成モデルである GBN に対し分類精度が高くなると報告している。

しかし、周辺尤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良いはずである。Friedman et al. (1997) の報告に対し、Sugahara et al. (2018) が行った GBN と識別モデルの分類精度の比較実験では、サンプル数が多いデータセットでは周辺尤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、サンプル数が少ないデータセットでは分類精度が劣化するという結果が報告されている。Sugahara et al. (2018) はサンプル数が少ない場合、GBN の目的変数が持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、分類精度が劣化してしまうと分析している。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持つような制約を持った Augmented naive Bayes classifier (ANB) の厳密学習を提案している (Sugahara et al. 2018)。また、周辺尤度による構造推定は漸近的に真の構造に収束するが、サンプル数が少ないデータセットに対しては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と推測できる。

一方、一つ一つの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 (Chickering and Heckerman 2000). 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しなければいけないため、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、学習スコアが最適な順に 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 (Tian et al. 2012, Chen and Tian 2014). しかしながら、この手法では類似性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、本論文では類似した構造を減らすために、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、アンサンブル法を用いた手法は、

1. 大規模データでは、生成モデルである GBN とともに他手法よりも分類精度が高い.
2. 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.
3. 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、これまで最も良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た.

さらに、実験から提案したアンサンブル法が k -best 法に劣る場合が観測された. これを踏まえて、本研究ではさら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 $T \times K$ 個の構造を発生し、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、この手法がサンプル数が少ないデータでさらに分類精度を改善し、データ規模に関わらず識別モデル、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。

$$p(X_0, \dots, X_N | G) = \prod_{i=0}^N p(X_i | \Pi_i, G) \quad (2.1)$$

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、学習データとしてサンプル数 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 は次の通りとなる (Buntine 1991)。

$$\hat{\theta}_{ijk} = \frac{\alpha_{ijk} + n_{ijk}}{\alpha_{ij} + n_{ij}} \quad (2.2)$$

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。一般にその学習スコアとして周辺尤度が用いられてきた。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 $\Theta = \{\theta_{ijk}\}$, ($i = 0, \dots, N, j = 1, \dots, q_i, k = 1, \dots, r_i$) 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.

$$p(\Theta | G) = \prod_{i=0}^N \prod_{j=1}^{q_i} \frac{\Gamma(\sum_{k=1}^{r_i} \alpha_{ijk})}{\prod_{k=1}^{r_i} \Gamma(\alpha_{ijk})} \prod_{k=1}^{r_i} \theta_{ijk}^{\alpha_{ijk}-1} \quad (2.3)$$

このとき、構造 G の周辺尤度は次式で表される.

$$p(D | G) = \prod_{i=0}^N \prod_{j=1}^{q_i} \frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij} + n_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(\alpha_{ijk} + n_{ijk})}{\Gamma(\alpha_{ijk})} \quad (2.4)$$

Heckerman ら Heckerman et al. (1995) 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尤度等価原理を、ベイジアンネットワーク学習に導入した. この尤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.

$$\alpha_{ijk} = \alpha p(X_i = k, \Pi_i = j | G^h) \quad (2.5)$$

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. さらに、Buntine (1991) 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 et al. (1995) や Ueno (2011, 2010) の報告によれば、ユーザが事前知識を持たない場合、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. 周辺尤度は漸近一致性を有しているため、周辺尤度を用いて構造を推定すれば漸近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確率的因果構造を推定したい場合には周辺尤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 (Chickering 1996). 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 (Bartlett and Cussens 2013). 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 (Natori et al. 2015, 2017, 名取和樹 et al. 2018).

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 (Friedman et al. 1997). 今, 目的変数を X_0 , 説明変数を X_1, \dots, X_n とする BNCG 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.

$$\begin{aligned}
\hat{c} &= \arg \max_{c \in \{1, \dots, r_0\}} p(c \mid \mathbf{x}, G) \\
&= \arg \max_{c \in \{1, \dots, r_0\}} \frac{p(c, \mathbf{x} \mid G)}{p(\mathbf{x} \mid G)} \\
&= \arg \max_{c \in \{1, \dots, r_0\}} p(c, \mathbf{x} \mid G) \\
&= \arg \max_{c \in \{1, \dots, r_0\}} \prod_{i=0}^n \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} (\theta_{ijk})^{1_{ijk}} \\
&= \arg \max_{c \in \{1, \dots, r_0\}} \left[\prod_{j=1}^{q_0} \prod_{k=1}^{r_0} (\theta_{0jk})^{1_{ijk}} \right. \\
&\quad \left. \times \prod_{i: X_i \in \mathbf{C}} \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} (\theta_{ijk})^{1_{ijk}} \right] \tag{2.6}
\end{aligned}$$

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, \mathbf{C} は目的変数 X_0 の子変数集合である.

一般に, 周辺尤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, 前述の通り GBN の構造学習は NP 困難問題である (Chickering 1996) 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. これに対し BNC としては, ベイジアンネットワークの下位モデルであり, 構造学習の必要が無い Naive Bayes (Maron and Kuhns 1960, Minsky 1961) が知られている. Naive Bayes は全説明変数が目的変数のみを親として持つ構造となっている. さらに, 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) (Friedman et al. 1997) が提案されている. TAN は, 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. また, Friedman et al. (1997) は BNC の分類精度向上のためには周辺尤度ではなく, 説明変数を所与とした目的変数の条件付き尤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尤度 (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 は分解可能ではないため, CLL を近似する手法がいくつか提案されている.

Grossman and Domingos (2004) は、構造の探索方法として hill-climbing アルゴリズムを用いることで、近似的に CLL を最大化する手法を提案している。また、Carvalho et al. (2013) 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。識別モデルについては、Friedman et al. (1997) 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。

しかし、GBN の構造学習の学習スコアとして周辺尤度を用いた場合、この結果は奇異である。周辺尤度は漸近一致性を有しているため、周辺尤度を用いて推定される構造は漸近的に真の構造に収束する。結果として、サンプル数が十分に大きいときには GBN が識別モデルよりも分類精度が良いと推測される。Friedman et al. (1997) の報告に対し、Sugahara et al. (2018) が行った GBN と識別モデルの分類精度の比較実験では、サンプル数が多いデータセットでは周辺尤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、サンプル数が少ないデータセットでは分類精度が劣化するという結果が報告されている。Sugahara らは学習データのサンプル数が少ない場合、学習データがスパースであり、結果として GBN の目的変数が持つ親変数が多くなっていくことで式 (2.6) 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条件を満たす Augmented naive Bayes Classifier (ANB) の厳密学習を提案している (Sugahara et al. 2018)。

2.3 Augmented naive Bayes Classifier

ANB は各説明変数が目的変数を親に持つ制約を持つ BNC である。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 (Sugahara et al. 2018)。具体的には、Silander and Myllymäki (2006) の動的計画法を用いた GBN の厳密学習アルゴリズムの候補構造として ANB 構造のみを用いる。今、全ての ANB 構造の候補構造集合を \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.

$$\begin{aligned}
G^* &= \arg \max_{G \in \mathcal{G}_{ANB}} Score(G) \\
&= \arg \max_{G \in \mathcal{G}_{ANB}} \left\{ Score(G_0(\Pi_0)) + \sum_{i: X_i \in \mathbf{X}} Score(G_i(\Pi_i)) \right\} \\
&= \arg \max_{G \in \mathcal{G}_{ANB}} \sum_{i: X_i \in \mathbf{X}} Score(G_i(\Pi_i)) \\
&= \arg \max_{G \in \mathcal{G}_{ANB}} Score_{ANB}(G) \tag{2.7}
\end{aligned}$$

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, $Score(G_i(\Pi_i))$ はローカルスコアと呼ばれる.

Sugahara et al. (2018) は ANB の厳密学習によって得られた ANB を用いて, サンプル数が少ないデータに対しての分類精度の改善を実験によって示している. しかし, サンプル数が少ないデータにおいて分類精度が劣化する原因は, 周辺尤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. これに対し, 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モデル平均が有効であることが知られている (Chickering and Heckerman 2000). 本研究では, サンプル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化を行う。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 D とする。モデル平均を式 (2.6) 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。

$$\begin{aligned}\hat{c} &= \arg \max_{c \in \{1, \dots, r_0\}} \sum_{G \in \mathcal{G}} \frac{p(G | D)}{\sum_{G' \in \mathcal{G}} p(G' | D)} p(c | \mathbf{x}, D) \\ &= \arg \max_{c \in \{1, \dots, r_0\}} \sum_{G \in \mathcal{G}} p(D | G) p(c | \mathbf{x}, D)\end{aligned}\quad (3.1)$$

しかし、式 (3.1) 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 (Tian et al. 2012, Chen and Tian 2014)。

3.1 k -best 法を用いた手法

Tian et al. (2012) 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。 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 $Score(G_i(\Pi_i))$ 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 (3.1) で目的変数の推定を行う。また、Chen and Tian (2014) の学習スコアが等価になってしまう DAG を除いて、 K 個の構造を選択するようにした先の k -best 法を改良した手法 (以降、 k -bestEC と呼ぶ) も用いる。 k -bestEC は通常の k -best と比べて、式 (3.1) 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より乖離する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。いくつかの分類器の結合により行われる予測は、多くの場合で最も優れた単一の分類器による予測よりも正しいということが経験的に示されている (Hansen and Salamon 1990)。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 (2006) の動的計画法を用いた構造学習を行う。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 (3.1) より次式で目的変数値 \hat{c} を推定する。

$$\hat{c} = \arg \max_{c \in \{1, \dots, r_0\}} \sum_{t=1}^T p(D_t | G_t) p(c | \mathbf{x}, D_t) \quad (3.2)$$

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、サンプル数の少ない学習データに対応するねらいがある。また、得られる複数の構造それぞれが類似することを防ぐことで k -best 法よりもさらに式 (3.2) の $p(D_t | G)$ が多様となることが期待できる。以降、この手法を $\text{Ens}(\text{GBN})$ 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 et al. (2018) の厳密学習手法で構造学習した 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 et al. (2014) と同様に, 各説明変数毎の中央値を区切りとして 2 値をとるカテゴリデータに変換し, 欠損値の含むサンプルは取り除いた. また, GBN の構造学習では周辺尤度として BDeu スコアを用い, 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 (Ueno 2011, 2010).

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, 学習スコアとして相互情報量を用いた (Friedman et al. 1997).
 - 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 (Carvalho et al. 2013).
 - BNC2P: 各変数の最大親変数数を 2 とした構造を仮定し, 学習スコアを CLL として hill-climbing アルゴリズムで貪欲学習した (Grossman and Domingos 2004).
- GBN: 学習スコアとして BDeu を用いて, 動的計画法 (Silander and Myllymäki 2006) で構造学習した.
- ANB: 学習スコアとして BDeu を用いて, Sugahara et al. (2018) の厳密学習手法で構造学習した.
- モデル平均手法
 - kBest(GBN): k -best 法 (Tian et al. 2012) により, $K = 10$ 個の構造を選択してモデル平均を行った.
 - kBestEC(GBN): k -bestEC (Chen and Tian 2014) により, $K = 10$ 個の構造

表 1: 実験で用いたデータセット

	学習データ名	サンプル数
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

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
 - kBestEC(ANB): k -bestEC(Chen and Tian 2014) 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 et al. (2018) の ANB の厳密学習手法で ANB を構造学習してモデル平均を行った。

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 分割し内 1 つをテストデータ、残り 9 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。さらに、サンプル数下位 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。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。

まず、モデル平均を行った手法とモデル平均を行わない識別モデル、GBN、ANB との結果の比較を行う。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も GBN が最も高い結果となった。前述のように、GBN は漸近一致性を有しており、データ数が大きくなると真の構造に近づいて分類精度が向上していることがわかる。この結果は、Sugahara et al. (2018) の結果とも一致する。一方小データ群では、本提案の GBN のアンサンブル法によるモデル平均である Ens(GBN) が最も高い分類精度を示している。さらに Ens(GBN) は k -best 法と比べても大きく分類精度が向上している。特にデータ 2 や 4 では大きく分類精度が向上しており、アンサンブル法によるモデル平均手法はサンプル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。また、Ens(ANB) と比較しても、Ens(GBN) は分類精度が高いことがわか

表 2: 各 BNC の分類精度

学習データ名	識別モデル				k-best 法		アンサンブル法		ANB+モデル平均		アンサンブル k-best 法		
	NB	TAN	aCLL- TAN	BNC2P	GBN	ANB	kBest (GBN)	kBestEC (GBN)	Ens (GBN)	kBestEC (ANB)	Ens (ANB)	Ens-kBestEC (GBN)	Ens-kBestEC (ANB)
1 lenses	0.6250	0.7083	0.7083	0.6042	0.8125	0.8750	0.8750	0.8333	0.8333	0.6250	0.8750	0.8333	0.6667
2 mux6	0.5469	0.6094	0.5938	0.5313	0.4531	0.3125	0.3594	0.3594	0.6094	0.5781	0.4063	0.6250	0.6250
3 post	0.6552	0.6322	0.5977	0.6839	0.7126	0.7126	0.7126	0.7126	0.7126	0.6552	0.7126	0.7126	0.6667
4 zoo	0.9901	0.9406	0.9505	0.8218	0.9426	0.9604	0.9426	0.9505	0.9604	0.9703	0.9604	0.9505	0.9703
5 Hayes-Roth	0.8106	0.6439	0.6742	0.6212	0.6136	0.8333	0.6136	0.7652	0.6136	0.7955	0.8333	0.7727	0.7879
6 iris	0.7133	0.8267	0.8200	0.7667	0.8267	0.8067	0.8267	0.8267	0.8267	0.8267	0.8267	0.8267	0.8267
7 wine	0.9270	0.9213	0.9157	0.8652	0.9438	0.9270	0.9494	0.9438	0.9551	0.9270	0.9213	0.9438	0.9270
8 glass	0.5421	0.5467	0.6215	0.5000	0.5607	0.5280	0.5607	0.5841	0.5701	0.5888	0.5234	0.5748	0.6075
9 Congressional Voting Records	0.9095	0.9526	0.9224	0.9353	0.9612	0.9526	0.9612	0.9612	0.9698	0.9569	0.9569	0.9698	0.9526
10 heart	0.8296	0.8333	0.8148	0.7519	0.8296	0.8444	0.8296	0.8333	0.8407	0.8222	0.8407	0.8370	0.8222
11 Breast Cancer	0.7365	0.7220	0.6968	0.7004	0.7076	0.6751	0.7076	0.7220	0.7004	0.7148	0.6787	0.7220	0.7112
12 cleve	0.8311	0.8243	0.8446	0.6959	0.8074	0.8142	0.8074	0.8074	0.8108	0.8209	0.8142	0.8176	0.8277
13 liver	0.6464	0.6609	0.6522	0.6058	0.5768	0.6058	0.5681	0.6087	0.6174	0.6783	0.6261	0.6232	0.6696
14 MONK's Problems	0.7500	1.0000	1.0000	0.7824	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	0.7090	0.8691	0.8672	0.9375	0.9277	0.8906	0.8926	0.8789	0.9023	0.8945
小データ群 (1~15) の平均値	0.7543	0.7794	0.7802	0.7050	0.7745	0.7810	0.7768	0.7891	0.7941	0.7901	0.7903	0.8074	0.7970
16 Balance Scale	0.9168	0.8688	0.8624	0.7440	0.9168	0.9168	0.8896	0.9040	0.9168	0.8704	0.9168	0.9040	0.8848
17 crx	0.8392	0.8515	0.8453	0.8293	0.8392	0.8622	0.8392	0.8392	0.8499	0.8392	0.8652	0.8499	0.8453
18 Breast Cancer Wisconsin	0.9766	0.9678	0.9473	0.9407	0.9766	0.9766	0.9766	0.9736	0.9766	0.9751	0.9751	0.9736	0.9751
19 Australian	0.8348	0.8290	0.8478	0.8188	0.8565	0.8580	0.8565	0.8565	0.8464	0.8362	0.8594	0.8464	0.8377
20 pima	0.7057	0.7188	0.7031	0.7018	0.7253	0.7188	0.7266	0.7292	0.7227	0.7201	0.7161	0.7266	0.7135
21 Tic-Tac-Toe	0.6889	0.7599	0.7192	0.6952	0.8549	0.8445	0.8539	0.8539	0.8466	0.8518	0.8445	0.8518	0.8455
22 banknote authentication	0.8433	0.8819	0.8761	0.8448	0.8812	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	0.8409	0.8431	0.8431	0.8431	0.8431	0.8431	0.8236	0.8431	0.8431	0.8265
24 Contraceptive Method Choice	0.4644	0.4725	0.4650	0.4511	0.4549	0.4270	0.4528	0.4487	0.4521	0.4623	0.4270	0.4487	0.4766
25 Car Evaluation	0.8623	0.9392	0.9427	0.7338	0.9416	0.9416	0.9306	0.9277	0.9172	0.9433	0.9172	0.9161	0.9398
26 led7	0.7288	0.7309	0.7347	0.4500	0.7288	0.7288	0.7288	0.7294	0.7284	0.7281	0.7284	0.7309	0.7306
27 shuttle-small	0.9383	0.9567	0.9538	0.8459	0.9693	0.9716	0.9693	0.9693	0.9693	0.9714	0.9702	0.9693	0.9705
28 Nursery	0.9028	0.9249	0.9188	0.6777	0.9345	0.9174	0.9270	0.9269	0.9305	0.9152	0.9143	0.9265	0.9164
29 EEG	0.5774	0.6298	0.6138	0.5850	0.6844	0.6895	0.6862	0.6887	0.6881	0.6931	0.6955	0.6899	0.6968
30 HTRU2	0.8966	0.9141	0.9141	0.9082	0.9141	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	0.7260	0.7859	0.7879	0.7855	0.7861	0.7859	0.7877	0.7880	0.7860	0.7879
大データ群 (16~31) の平均値	0.7938	0.8137	0.8081	0.7371	0.8317	0.8299	0.8288	0.8295	0.8293	0.8258	0.8285	0.8286	0.8277
平均値	0.7747	0.7971	0.7946	0.7215	0.8040	0.8062	0.8036	0.8099	0.8122	0.8085	0.8100	0.8184	0.8128

る。k-best 法, Ens(ANB) のの双方で共通しているのは, Ens(GBN) のモデル平均で用いられる構造に比べて構造間の類似度が高くなっている可能性があることである。k-best 法は一つのスコアを大きくするため類似構造を選択しやすく, ANB 構造は目的変数にすべての説明変数へのエッジが引かれる分, GBN に比較してアンサンブル法で得られる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性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性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性を調査するために,

Structural Hamming Distance(SHD)(Tsamardinos et al. 2006)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性の指標として用いる。

表3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3の太字は、データ毎のモデル平均手法で最も分類精度が高いもの、最もSHDの値が高いものをそれぞれ表している。

結果より、小規模データであるデータ4, 7, 9のようにアンサンブル法の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、アンサンブル法は表2よりGBN, ANB, k -best法以上の分類精度があり、構造間の大きな乖離は分類精度向上に好影響を与えることが考えられる。さらにEns(ANB)に着目して、表3のデータ4, 9, 29を見ると、kBestEC(ANB)と比べ明らかにSHDが高い。これにより、Ens(ANB)はkBestEC(ANB)と比べて分類精度が高くなっている。しかし、Ens(GBN)には分類精度は劣っているのは、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。また、 k -best法、アンサンブル法ともに候補構造としてANB構造を仮定すると、仮定しない場合に比べ全体的にSHDの値が減少していることがわかる。さらに、 k -best法に着目する。 k -bestECは、各データ内でSHDが最も大きいケースは多いが、アンサンブル法と比べると特筆して大きい値にはならず、データ1, 13のように分類精度がGBNと比べ向上するものの、大きく向上はしない。アンサンブル法が k -best法と比べ分類精度が優れているのは、データによっては構造間に大きな乖離を起こすことが出来るためであると考えられる。しかし、データ5, 11のようにアンサンブル法では構造の乖離が発生しない場合であっても、 k -best法はSHDが高くなり、分類精度の向上に貢献す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さらに精度を向上させるかもしれない。そこで、本研究ではさらに、 k -best法とアンサンブル法を組み合わせた手法を提案する。

表 3: モデル平均 BNC の構造間平均 SHD

	<i>k</i> -best 法		アンサンブル法	ANB+モデル平均	
	kBest (GBN)	kBestEC (GBN)	Ens (GBN)	kBestEC (ANB)	Ens (ANB)
1 lenses	1.39	2.61	1.07	2.30	0.33
2 mux6	2.92	3.24	0.61	1.96	0.89
3 post	0.98	2.33	0.42	1.92	0.42
4 zoo	1.71	7.03	32.56	7.41	12.98
5 Hayes-Roth	1.44	2.35	0.00	2.39	0.07
6 iris	2.25	5.09	1.87	2.86	1.54
7 wine	1.53	7.57	11.85	2.56	5.50
8 glass	0.55	3.71	4.76	4.33	5.31
9 Congressional Voting Records	0.00	6.37	23.27	6.60	24.25
10 heart	0.11	4.48	7.19	2.09	6.42
11 Breast Cancer	0.00	2.36	1.02	2.20	1.02
12 cleve	0.00	3.74	5.95	2.16	5.07
13 liver	2.56	4.98	4.27	2.08	4.17
14 MONK's Problems	1.18	1.80	0.00	2.99	2.43
15 threeOf9	2.39	3.16	4.33	3.44	3.36
小データ群 (1~15) の平均値	1.27	4.05	6.61	3.15	4.92
16 Balance Scale	0.43	2.00	0.00	2.21	0.00
17 crx	5.17	9.36	8.68	3.70	6.29
18 Breast Cancer Wisconsin	0.00	5.73	0.32	1.80	0.00
19 Australian	4.59	6.25	10.79	3.97	9.39
20 pima	0.71	5.05	3.30	3.15	1.97
21 Tic-Tac-Toe	0.89	7.86	8.06	7.18	5.85
22 banknote authentication	2.60	5.54	0.00	3.74	0.00
23 Solar Flare	0.48	5.20	3.24	4.38	2.58
24 Contraceptive Method Choice	0.83	5.57	4.64	2.79	3.60
25 Car Evaluation	0.86	4.04	0.32	1.98	0.38
26 led7	0.02	3.58	0.00	1.80	0.00
27 shuttle-small	4.56	6.23	1.80	5.83	5.05
28 Nursery	1.36	2.15	0.70	2.30	0.28
29 EEG	2.51	9.04	7.99	12.07	15.40
30 HTRU2	0.00	6.03	0.32	5.01	0.32
31 MAGIC Gamma Telescope	2.92	9.02	3.82	7.14	1.36
大データ群 (16~31) の平均値	1.75	5.79	3.37	4.32	3.28
平均値	1.51	4.95	4.94	3.75	4.07

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 節で述べた k -best 法と 3.2 節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, それらすべ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.

$$\hat{c} = \arg \max_{c \in \{1, \dots, r_0\}} \sum_{t=1}^T \sum_{k=1}^K p(D_t | G_t^k) p(c | \mathbf{x}, D_t) \quad (5.1)$$

アンサンブル k -best 手法では, 通常の k -best 法 (Tian et al. 2012) と比べ分類精度が高かった k -bestEC (Chen and Tian 2014) を用いる.

5.2 アンサンブル k -best 手法の評価実験

4.2 節と同じ条件で, 前節で提案したアンサンブル k -best 手法, これに加えて候補構造を ANB 構造に仮定した手法を追加し, 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): k -bestEC (Chen and Tian 2014) 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 法による BNC.
- Ens-kBestEC(ANB): k -bestEC (Chen and Tian 2014) と ANB 構造を制約とした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 法による BNC.

なお, 二つのアンサンブル 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 -best 法で $K = 10$ 個の構造を選択した.

表 2 の「アンサンブル k -best 手法」が, 今回追加した 2 手法の分類精度である. 結果としては, k -bestEC とアンサンブル法を組み合わせたアンサンブル k -best 手法が, 比較

表 4: 既存手法との分類精度の比較

	aCLL-						Ens-kBestEC (GBN)
	NB	TAN	TAN	BNC2P	GBN	ANB	
分類精度	0.7747	0.7971	0.7946	0.7215	0.8040	0.8062	0.8184
p 値	<u>0.0016</u>	<u>0.0050</u>	<u>0.0087</u>	<u>0.0000</u>	<u>0.0434</u>	0.1703	-

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 4 である。なお、表の p 値は複合手法と他手法を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。

結果の分析を行う。 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 とアンサンブル法以上の分類精度となっている。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、さらに分類精度が向上することが確認された。一方、ANB 構造を制約とした場合のアンサンブル k -best 手法でも、同様の結果が得られている。しかしながら、ANB 構造を制約としないモデル平均と比べて分類精度が低い。前述の通り、ANB 構造の仮定による構造間の類似が分類精度向上の妨げとなっていることが考えられる。

以上より、提案したアンサンブル k -best 手法は、データ規模に関わらず既存手法と比べ分類精度の向上が確認された。

6 むすび

本研究ではまず、周辺尤度は漸近一致性を有しているのにも関わらず GBN が識別モデルに対し分類精度が如何にして劣るかについて、学習データのサンプル数が少ない場合に推定される構造の事後確率が低いために分類精度が劣化してしまうと分析した。そこで、本研究では BNC にモデル平均を導入した。モデル平均を用いれば、個々の構造の事後確率が低くても、推論精度向上が期待できる。さらに、本研究ではモデル平均が考えられる全候補構造についての計算による計算困難性を回避するために、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り以下を示した。

1. アンサンブル法はサンプル数が多い学習データに対して、他手法より分類精度が高い GBN と同等の分類精度を持つ分類器が構築できた。
2. アンサンブル法はサンプル数が少ない学習データに対して、GBN, ANB, 他モデル平均手法と比べ有効な分類器が構築できた。
3. BNC のモデル平均を行うとき、構造間の類似性が低い方が分類精度が向上した。

また、これまで最も良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 GBN のモデル平均と比べて分類精度の向上が限定された。

本研究ではさらに上記の実験から、 k -best 法とアンサンブル法についてそれぞれ分類精度が向上するデータが異なることを示し、この二つのモデル平均手法を組み合わせたアンサンブル k -best 法を提案した。この提案手法が既存の識別モデル、GBN に対しデータ規模に関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。

今後の課題として、さらに構造が多様となる構造の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。

謝辞

本論文を作成するにあたり、指導教員の植野真臣教授から、丁寧かつ熱心なご指導を賜りました。ここに感謝の意を表します。また、ゼミや日常の議論を通じて多くの示唆や知識を頂いた川野秀一准教授、西山悠助教、宇都雅輝助教、研究室の先輩・同期に感謝いたします。

参考文献

- Bartlett, M. and Cussens, J. (2013). Advances in Bayesian network learning using integer programming. *Proceedings of the Twenty-Ninth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-13)*, 182–191.
- Buntine, W. (1991). Theory refinement on Bayesian networks. *Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence*, 52–60.
- Carvalho, A. M., Adão, P., and Mateus, P. (2013). Efficient Approximation of the Conditional Relative Entropy with Applications to Discriminative Learning of Bayesian Network Classifiers. *Entropy*, **15**(7), 2716–2735.
- Chen, Y. and Tian, J. (2014). Finding the k -Best Equivalence Classes of Bayesian Network Structures for Model Averaging. *AAAI*, 2431–2438.
- Chickering, D. M. Learning Bayesian networks is NP-complete. In *Learning from data*, 121–130. Springer, (1996).
- Chickering, D. M. and Heckerman, D. (2000). A comparison of scientific and engineering criteria for Bayesian model selection. *Statistics and Computing*, **10**, 55–62.
- Campos, C.,d, Cuccu, M., Corani, G., and Zaffalon, M. (2014). Extended Tree Augmented Naive Classifier. *Lecture Notes in Computer Science*, **8754**, 176–189.
- Friedman, N., Geiger, D., and Goldszmidt, M. (1997). Bayesian Network Classifiers. *Machine Learning*, **29**(2), 131–163.
- Grossman, D. and Domingos, P. (2004). Learning Bayesian network classifiers by maximizing conditional likelihood. *Twenty-first international conference on Machine learning - ICML '04*, page 46.
- Hansen, L. K. and Salamon, P. (1990). Neural Network Ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**(10), 993–1001.
- Heckerman, D., Geiger, D., and Chickering, D. (1995). Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. *Machine Learning*, **20**, 197–243.
- Maron, M. E. and Kuhns, J. L. (1960). On Relevance, Probabilistic Indexing and Information Retrieval. *Journal of the ACM*, **7**(3), 216–244.
- Minsky, M. (1961). Steps toward Artificial Intelligence. *Proceedings of the IRE*, **49**(1), 8–30.
- Natori, K., Uto, M., Nishiyama, Y., Kawano, S., and Ueno, M. (2015). Constraint-

- Based Learning Bayesian Networks Using Bayes Factor. *Advanced Methodologies for Bayesian Networks*, **9505**, 15–31.
- Natori, K., Uto, M., and Ueno, M. (2017). Consistent Learning Bayesian Networks with Thousands of Variables. *Proceedings of Machine Learning Research*, **73**(1991), 57–68.
- Silander, T. and Myllymäki, P. (2006). A simple approach for finding the globally optimal Bayesian network structure. *Proceedings of Uncertainty in Artificial Intelligence*, 445–452.
- Sugahara, S., Uto, M., and Ueno, M. (2018). Exact learning augmented naive Bayes classifier. *Proceedings of Machine Learning Research*, **72**, 439–450.
- Tian, J., He, R., and Ram, L. (2012). Bayesian Model Averaging Using the k-best Bayesian Network Structures. *Proceedings of the Twenty-Sixth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-10)*, 589–597.
- Tsamardinos, I., Brown, L. E., and Aliferis, C. F. (2006). The max-min hill-climbing Bayesian network structure learning algorithm. *Machine Learning*, **65**(1), 31–78.
- Ueno, M. (2010). Learning networks determined by the ratio of prior and data. *Proceedings of the Twenty-Sixth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-10)*, 598–605.
- Ueno, M. (2011). Robust learning Bayesian networks for prior belief. *Proceedings of the Twenty-Seventh Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-11)*, 698–707.
- 名取和樹, 宇都雅輝, and 植野真臣. (2018). 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. 電子情報通信学会論文誌 *D*, **J101-D**(5), 754–768.