

ベイジアンネットワーク学習アルゴリズム PBIL-RS の性能に関する一検討

On the Performance of An Algorithm to Learn Bayesian Networks Based on Probability Vectors

山中 優馬[†] 藤木 生聖[†] 福田 翔[†] 吉廣 卓哉[‡]
 Yuma Yamanaka Takatoshi Fujiki Sho Fukuda Takuya Yoshihiro

1. はじめに

ベイジアンネットワークは事象間の因果関係を視覚化する確率モデルとして広く用いられている。その利用は様々な分野に渡り、バイオインフォマティクス、医学、ドキュメント分類、情報検索、意思決定支援など、様々な分野でベイジアンネットワークが活用される。近年、計算機とデータベースの発展により、大規模なデータ分析が可能になった。ベイジアンネットワークは様々な研究分野で有用な、大規模なデータ分析技術の一つとして有望視されている [14]。

しかしながら、ベイジアンネットワークの学習は NP 困難であることが明らかになっている [1]。ベイジアンネットワークは含まれる事象が増えるに従い、事象と因果関係の組み合わせが指数的に増加する。膨大な組み合わせの中から最適なベイジアンネットワークモデルを見つけるために、全ての組み合わせを試行するような単純な方法では、実用的な時間で解を得ることは不可能である。このため、近似解を高速に得るための方法がいくつか提案されている。K2[2] と呼ばれるアルゴリズムは Cooper らによって提案された。このアルゴリズムは事象間の因果関係に順序があると仮定し、順序を制約とすることで、解の候補を削減し、高速にベイジアンネットワークを求める。ここで順序制約とは、例えば、全ての事象が発生した時刻がわかっている場合には、時間的に前に発生した事象のみが、後に発生した事象の原因に成り得る、つまり、ある事象 X より過去の事象だけが事象 X の原因となり得るといった制約である。しかし現実の実用場面では、順序制約が仮定できない場合が多く、利用できる場面が限られてしまう問題がある。

順序制約を用いないベイジアンネットワーク学習アルゴリズムも存在する [3][4][5]。その多くは遺伝的アルゴリズムを用いた方法であり、高速にある程度優れたベイジアンネットワークモデルが得ることができる。しかしながら、近年、より大規模なデータを扱う要求が高まり、より効率の良いアルゴリズムが求められる。

これに対し、EDA (Estimation of Distribution Algorithm) という遺伝的アルゴリズムの一種に基づいたアル

ゴリズムが規模の大きいベイジアンネットワーク学習において、良い性能を発揮するという報告がある [6][7][8]。EDA とは個体生成の基となる確率分布を進化させる遺伝的アルゴリズムの一手法である。各世代で個体を生成し、優れた個体から確率分布を進化させることで、優れた個体を生成する確率が高くなる。このような動作を何世代も繰り返し、確率的な探索効率を向上させる手法である。ベイジアンネットワーク学習の問題においては EDA の一種である PBIL が最も適しているという報告がある [7]。

Blanco らは初めて、PBIL をベイジアンネットワーク学習に適用した [9]。彼らのアルゴリズムには、学習の過程で局所解に陥った際に、局所解を脱するとができず解の探索が終了する欠点があった。この欠点に対して、突然変異によって局所解を回避する試みが多く提案された。半田らは、生成したベイジアンネットワークモデルに対して、1 ビットの変異を確率的に加える、つまり、モデルに含まれる因果関係を削除、あるいは追加する動作を一定確率で行う Bitwise Mutation を提案した [6]。Kim らは、ベイジアンネットワークに特化して、因果関係を反転させる Transpose Mutation を提案した [7]。福田らは、学習する確率分布そのものに突然変異を起こす Probability Mutation を提案した [8]。これらは、局所解に陥ることを防ぎ、PBIL に基づいたベイジアンネットワーク学習アルゴリズムの改良に寄与する。しかし、突然変異は一定確率で発生するため、優れた解が存在する領域を十分に探索する前に、突然変異により探索領域が変わってしまう欠点があった。

これに対して、我々の研究グループでは、突然変異に代わる新たな仕組みを導入することで、従来よりも優れた解を短時間で探し出すアルゴリズム PBIL-RS (PBIL-Repeated Search) を提案した [15]。PBIL-RS は、始めに PBIL によって探索領域が優れた解が存在しそうな領域に狭まる。次に、探索領域が十分に狭まり、その局所領域を十分な探索を行ったと判断すると、探索領域を拡張して再探索を行う。この戦略により、PBIL-RS は既存手法よりも効率的に探索し、より優れたベイジアンネットワークを学習することが可能である。しかしながら、PBIL-RS が既存手法よりも優れた性能を持つことは示す結果はあるものの [15]、その詳細な評価と優れた性能を発揮する根拠となる挙動は明らかにされていない。

[†] 和歌山大学大学院 システム工学研究科, Graduate School of System Engineering, Wakayama University

[‡] 和歌山大学システム工学部, Faculty of System Engineering, Wakayama University

本論文では、PBIL-RS に対して詳細な評価を行い、PBIL-RS がどのような挙動で解の探索を行い、優れたベイジアンネットワークを探し出すのか、その特性と優れた性能を明らかにする。

本論文の構成は以下の通りである。第2節では、本研究で対象とする、遺伝的アルゴリズムの一種であるPBILに基づくベイジアンネットワーク学習アルゴリズムについて述べる。第3節では我々の研究グループが提案するPBIL-RSを紹介する、第4節でPBIL-RSの詳細な性能評価を行い、第5節で本論文をまとめる。

2. 関連研究

2.1 ベイジアンネットワーク

ベイジアンネットワークモデルは事象間の因果関係を可視化したグラフ表現である。事象をノード、因果関係をリンクによって表す。簡単なベイジアンネットワークの例を図1に示し、これを用いて説明する。ノード X_1, X_2, X_3 はそれぞれが事象を表しており、それぞれ対応する事象が生起すれば値が1、事象が生起しなければ値が0をとる。リンク $X_1 \rightarrow X_3, X_2 \rightarrow X_3$ のそれぞれは因果関係を表し、これによって、 X_3 は X_1 と X_2 の値に依存する条件付確率によって値が定まることを表す。 $X_1 \rightarrow X_3$ のようなリンクが存在するとき、 X_1 を親ノード、 X_3 を子ノードと呼ぶ。親ノードを持たないノード X_1 と X_2 はそれぞれ自身が持つ生起確率 $P(X_1), P(X_2)$ によって値が定まる。一方で、 X_3 は X_1, X_2 という2つの親ノードを持ち、その生起確率は条件付き確率 $P(X_3|X_1, X_2)$ であり、この確率に従い X_3 の値が定まる。例えば、 X_1 と X_2 の両方の値が1をとるとき、 X_3 の生起確率は0.890である。また、図1のモデルにおいて X_3 のみが観測された場合に、ベイズ推定を用いることで X_3 の親ノードである X_1 と X_2 の事後確率を求められる。このように、ベイジアンネットワークモデルを学習することにより、事象の原因となる事象を推定することが可能である。ベイジアンネットワークでは、事象を観測することにより得られるデータからモデルを学習する。観測の数を S として、観測データを $O = \{o_j\}, 1 \leq j \leq S$ で表す。事象の数を N としたとき、 j 番目の観測は $o_j = (x_{j1}, x_{j2}, \dots, x_{jN})$ で表せる。ここで x_{ji} はすべての $i(1 \leq i \leq N)$ は事象 X_i に対する j 番目の観測を表す。我々はこのように与えられた観測データから優れたモデル θ を学習することを試みる。ここで言う優れたモデルとは、より観測データ O に合致したモデル、つまり、その確率モデルに従って事象を生起させた場合に、 O に近いデータが生成されるようなモデルのことである。モデルがどれほどデータに合致しているかの評価基準として、ベイジアンネットワーク学習問題を取り扱う上でよく用いられる、情報量基準

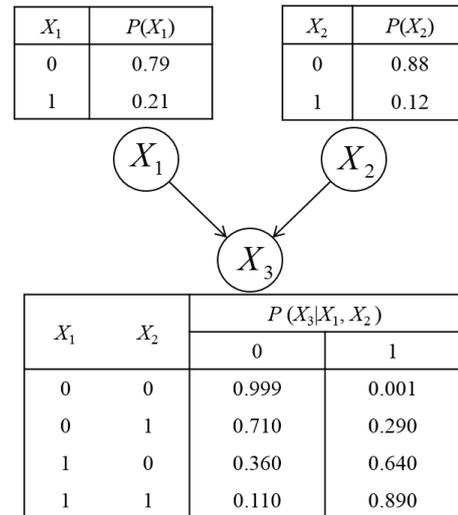


図1: ベイジアンネットワークモデルの例

の一種であるAIC (Akaike's Information Criterion) [10]を用いる。我々が考える問題は、観測データセットに対して最も評価関数AICが低い値をとるベイジアンネットワークモデルを求めることである。

2.2 ベイジアンネットワーク

最近、ベイジアンネットワーク学習問題を解くために、EDA (Estimation Distribution Algorithm) と呼ばれるアルゴリズムに基づいた手法がさかんに提案され、優れたベイジアンネットワークを効率よく学習できることが報告されている [6][7][8]。EDAの一種であるPBIL [9] は遺伝的アルゴリズムの一種であり、1994年にBalujaらによって提案された。後に、BlancoらがPBILをベイジアンネットワークの学習に適用し、PBILがベイジアンネットワーク学習問題において、効率よくベイジアンネットワークモデルを学習することが可能であることを示した [13]。PBILでは、ひとつの個体をベクトル $s = \{v_1, v_2, \dots, v_L\}$ で定義する。 $v_i (1 \leq i \leq L)$ は0または1の値をとり、全体が L 個の要素で構成される個体の i 番目の要素を表す。また、個体を生成する確率ベクトル $P = \{p_1, p_2, \dots, p_L\}$ を定義する。ここで、 $p_i (1 \leq i \leq L)$ は $v_i = 1$ である確率である。この時PBILの動作は以下のように記述される。

- (1) 確率ベクトル P を $p_i = 0.5 (i = 1, 2, \dots, L; L$ は個体を構成する要素の個数) として初期化する。
- (2) 確率ベクトル P に基づき $v_i (1 \leq i \leq L)$ の値を決定し、個体 s を生成する。つまり v_i の値は確率 p_i に基いて定まる。これを C 回繰り返し、 C 個の個体集合 S を生成する。
- (3) 各個体 $s \in S$ に対して評価関数を用いて評価値を算出する。

- (4) 評価値の高い上位 C' 個 ($C' \leq C$) の個体を選択し、個体集合 S から選択された C' 個の個体集合 S' について、以下の計算式を用いて確率ベクトルを更新する：

$$p_i^{new} = ratio(i) \times \alpha + p_i \times (1 - \alpha), \quad (1)$$

ここで、 p_i^{new} は更新後の確率ベクトル、 p_i は更新前の確率ベクトルを表す。 $ratio(i)$ は個体集合 S' に含まれるリンクの割合であり、 α は学習率と呼ばれるパラメータである。

- (5) (2) に戻る

PBIL は評価値の上位 C' 個の個体を用いて、学習率 α の割合で確率ベクトルを学習することで、より優れた個体が生成されやすい確率ベクトルに進化していく。PBIL は他の遺伝的アルゴリズムと異なり個体同士の交叉ステップは含まない。代わりに生成した個体を親として確率ベクトルを進化させていく。

2.3 ベイジアンネットワーク

この節では、PBIL によるベイジアンネットワーク学習アルゴリズムを記述する。我々が扱うベイジアンネットワーク学習問題を解くことは、前節で説明した PBIL の説明と少し異なり、それに合わせた PBIL の修正が必要である。このアルゴリズムでは、PBIL における各個体が、それぞれのベイジアンネットワークモデルに相当する。すなわち、事象の数を N とすると、ひとつの個体を、ベクトル $s = \{v_{11}, v_{12}, \dots, v_{1N}, v_{21}, v_{22}, \dots, v_{N1}, v_{N2}, \dots, v_{NN}\}$ で定義する。ここで v_{ij} は X_i から X_j へのリンクに相当する。例えば、もし、 $v_{ij} = 1$ であれば、 X_i から X_j へのリンクが存在することを表し、 $v_{ij} = 0$ であれば、 X_i から X_j へのリンクが存在しないことを表す。同様に、個体となるベイジアンネットワークモデルを生成する元となる確率ベクトル P を $P = \{p_{11}, p_{12}, \dots, p_{1N}, p_{21}, p_{22}, \dots, p_{N1}, p_{N2}, \dots, p_{NN}\}$ と定義する。 p_{ij} は X_i から X_j へのリンクが存在する確率である、つまり、個体 s の要素 v_{ij} の値は、確率ベクトル P の要素 p_{ij} によって決定される。ここで注意したいのは、ベイジアンネットワークは自己循環リンクを持つことは許されない制約があるので、 $i = j$ を満たす p_{ij} は、すべて値が 0 をとることである。確率ベクトル P は確率表とみなすことができ、それを図 2 に示す。PBIL によるベイジアンネットワーク学習アルゴリズムの動作は、基本的には PBIL の手順と同じであり、以下のように記述される

- (1) 確率ベクトル P を初期化する。この時 $i = j$ ならば $p_{ij} = 0$ 、それ以外は $p_{ij} = 0.5$ として初期化する。

- (2) 確率ベクトル P を用いて C 個のベイジアンネットワークモデルを生成し、これを個体集合 S とする。
- (3) 個体集合 S に含まれるすべての個体 $s \in S$ のそれぞれに対して、評価値を算出する。
- (4) 評価値の高い上位 C' 個 ($C' \leq C$) の個体を選択する。
- (5) 個体集合 S から選択された C' 個の個体集合 S' を用いて式 (1) を用いて確率ベクトルを更新する。((4), (5) の動作を図 3 に示す)
- (6) (2) に戻る。

PBIL と同様に、PBIL によるベイジアンネットワーク学習アルゴリズムは、確率ベクトルが、より優れたベイジアンネットワークモデルが生成されやすい値に進化していく。しかし、ベイジアンネットワークモデルは特有の制約がある。それは非循環構造であるという制約である。その制約を考慮した、詳しい (2) のステップを以下に記述する。

- (2a) ノード X_i と X_j のペアを $pair(i, j)$ と定義する。 $1 \leq i, j \leq N$ かつ、 $i \neq j$ 。である全てのペアに対してランダムに $pair(i, j)$ の順序を生成する。
- (2b) (2a) で生成された順序で、 $pair(i, j)$ を順に選び、確率ベクトル P に基づいて v_{ij} の値を決定する。毎回の v_{ij} の値を決定する際に、もし $v_{ij} = 1$ ならば、既に決定されたリンク集合に、この X_i から X_j へのリンクを追加することで、循環構造が生成されるかどうかを調べる。この時、循環構造を生成してしまうリンクであると判断された場合は $v_{ij} = 0$ とする。
- (2c) すべての $pair(i, j)$ に対してリンクの有無を決定するまで (2a) へ戻る。

この手順を踏むことで PBIL の枠組みで、優れたベイジアンネットワークモデルを探索する問題を扱うことが可能になる。上記の、PBIL によるベイジアンネットワーク学習アルゴリズムは、突然変異操作を含まず、学習によって局所解に陥る欠点がある。これを回避するために、PBIL に適用する突然変異操作が複数提案されている。半田らは、生成した個体に対して、1 ビットの変異を確率的に加える Bitwise Mutation を提案した [6]。Kim らは、ベイジアンネットワークに特化して、因果関係を反転させる Transpose Mutation を提案した [7]。福田らは、学習する確率分布そのものに突然変異を起こす Probability Mutation を提案した [8]。これらは、局所解に陥ることを防ぎ、PBIL によるベイジアンネットワークの学習を、時間をかければ優れた解が見つかる、より信頼性の高い方法へと改良した。

P		親ノード					
		X_1	X_2	...	X_i	...	X_N
子ノード	X_1	0.0	0.5	...	p_{i1}	...	0.5
	X_2	0.5	0.0	...	p_{i2}	...	0.5
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	X_j	p_{1j}	p_{2j}	...	p_{ij}	...	p_{Nj}
	X_N	0.5	0.5	...	p_{iN}	...	0.0

図 2: 確率ベクトル

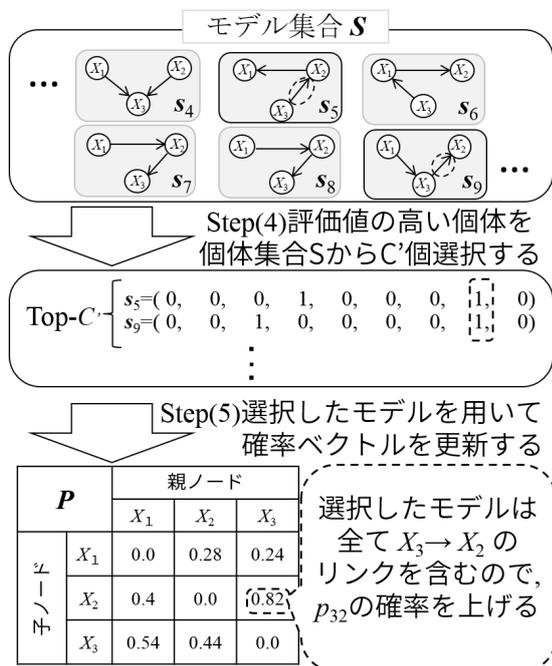


図 3: 確率ベクトルの更新

3. PBIL-RS

PBIL-RS (PBIL-Repeated Search) は PBIL に基づいたベイジアンネットワーク学習アルゴリズムであり、福田らによって提案された [15]。このアルゴリズムは突然変異に変わる新たな仕組みを導入することで、さらに効率的な探索を実現する。突然変異は一定の確率で発生するため、優れた解が存在する領域を十分に探索する前に突然変異により探索領域が移る傾向がある。その結果、優れた解が多数存在する領域を深く探索できず、探索効率が低下することになる。これに対して PBIL-RS は、探索領域が十分に狭まり、その領域を十分に探索したと判断した場合にのみ探索領域を広げ、異なる探索領域に移る。この方法により、優れた解が存在しそうな領域を十分に探索すると同時に、局所解に陥ることなく効率的に探索する。

PBIL-RS の動作の概要を示す。ベイジアンネットワーク探索問題の解空間には複数の局所解が存在する。類似した構造を持つモデルは近い評価値を持つ傾向があるため、優れた解は偏った領域に分布すると考えられる。そこで PBIL-RS は、初期状態では探索空間全体を対象とする。世代が進むと、優れた解が存在すると推定される領域に探索領域が収束していく。十分に探索領域が収束し、その領域を十分に探索したと判断されると、探索領域を拡大し、同様に異なる局所領域を探索する。ただし、探索領域を拡大しても、新たな探索領域が十分に大きくなければ、再び同じ局所領域に収束する恐れがある。これを避けるために、探索領域の拡大を段階的に行う。つまり、始めは、近くの局所領域を探索するために拡大幅を小さくしておき、各収束時に優れた解が発見できない場合には、より離れた探索領域を探索するために拡大幅を段階的に増大させる。

PBIL-RS は探索領域の大きさを確率ベクトル P によって制御する。確率ベクトルの各要素 p_{ij} は各リンクの生成確率を表している。これらの値はベイジアンネットワークの探索が進むに連れて 0 か 1 に近づいていく。これは生成されるモデルが特定の形状に偏ることを意味する。言い換えれば、確率ベクトルの各要素の値が 0.5 に近いほど探索領域が大きく、0 か 1 に近いほど探索領域が小さいと見なせる。すなわち、確率ベクトルは生成されるモデルの形状と偏りを制御している。これに基づいて、確率ベクトルに対して、収束度 S を次のように定義する。

$$S = \frac{\sum_{i,j(i \neq j)} \{0.5 - |p_{ij} - 0.5|\}}{N(N-1)} \quad (2)$$

収束度 S は、確率ベクトルの $i \neq j$ の全要素に対して、0 または 1 からの差の和の平均である。つまり、この値が小さいほど探索領域が狭く、収束していることを表す。PBIL-RS の確率ベクトルの更新では、通常、世代が進むにつれて収束度の値が小さくなる。また、PBIL-RS が十分に探索領域は収束したと判断するために、探索回数 k を導入する。つまり、収束度 S が過去 k 世代にわたって小さくならない場合には、十分に収束したと見なす。ただし、収束度 S はその直前の世代よりも大きくなる場合があるため、正確には、 k 世代前の収束度が、それ以降の収束度の最小値を更新しない場合に、収束したと見なす。

探索領域が収束したと判断すると、PBIL-RS は、探索領域を拡大する。探索領域の拡大幅を H とする。探索領域を拡大するために、収束度 S が H まで拡大されるように確率ベクトルに対して操作を行う。確率ベクトルの操作としては、 $i, j (i \neq j)$ をランダムに選択し、 $p_{ij} = 0.5$ とする。この操作を $S \leq H$ である限り繰り返す。さらに、先述したように、再び同じ局所領域に収束すること

を避けるために、探索領域の拡大を段階的に行う。これは拡大幅 H を動的に変動させることで実現する。具体的には、拡大幅 H はあらかじめ設定された初期値から始まり、拡大操作をする度に一定値 H_{inc} だけ増大する。一方で、それまでで最も優れた解が発見された場合（つまり、最適解が更新された場合）に、 H を初期値に戻す。これにより、優れた解が発見されにくい局所領域を早期に離れ、より広い探索領域内で探索を行う事ができる。上記の動作を形式的に記述すると、以下ようになる。これら (i)-(iv) の処理を、2.3 節で述べたステップ (5) と (6) の間に挿入する。

- (i) もし、ベイジアンネットワークが最適解を更新したら拡大幅 H の値を初期値 H_{min} とする。
- (ii) 確率ベクトル P の要素 p_{ij} を無作為に選択し、 $p_{ij} = 0.5$ とする。
- (iii) もし $S < H$ であるならば手順 (ii) に戻る。
- (iv) 拡大幅 H を拡大増加量 H_{inc} だけ加算する。

4. 評価

4.1 評価目的と方針

福田らは、PBIL-RS に対して、基礎的な評価を行った。その評価から、PBIL-RS が既存手法よりも優れたベイジアンネットワークモデルを学習できる手法であることを明らかにした [15]。

本研究では、福田らの評価に加えて、各種パラメータの影響評価をする、そして、ノード数によらず PBIL-RS が有効であることを示す。様々なパラメータが PBIL-RS の性能に与える影響を評価し、適切なパラメータの値を明らかにする。これを明らかにすることより、PBIL-RS を用いてベイジアンネットワークモデルを探索する際、パラメータ設定を行う大きな手助けになるものと考えられる。また、福田らの評価では、最大で 37 ノードのベイジアンネットワークを用いて評価をしている。PBIL-RS がより大きい規模のベイジアンネットワークモデルにおいて有効であることを示し、大規模なベイジアンネットワークモデルを学習においても、実用的な手法であることを示す必要がある。

4.2 既存手法との性能比較

実験の手順としては、はじめに、観測データの生成に使用するベイジアンネットワークモデルを用意する。今回の評価実験に使用したモデルは Alarm Network [11] と呼ばれ、ベイジアンネットワークモデル学習問題を取り扱う際によく用いられるモデルと Pathfinder [12] と呼ばれる外科病理学者がリンパ節疾患の診断をする際に支援するために学習されたベイジアンネットワークモデルであ

る。ベイジアンネットワークは各ノードが、そのノードのとりうる値を決定する確率をもっている。親ノードは自身の生起確率によって値を決定し、子ノードは親ノードの値に依存する条件付き確率によって値が決定する。はじめに親ノードから値を決定する。次に親ノードの値によって条件付き確率が決まった子ノードの値が決定する。これを繰り返すことで全てのノードの値が決定する。このノードの値の組み合わせを 1 サンプルとする。これを 1000 回繰り返すことにより、サンプル数 1000 のデータセットが出来上がる。これをベイジアンネットワークモデルを学習するための観測データとする。このように用意した観測データから PBIL-RS の性能評価を行う。評価指標は AIC [10] を用いており、この値が低いほど、優れたベイジアンネットワークモデルである。

既存手法と比較して PBIL-RS の性能評価を行う。PBIL-RS と比較するために突然変異を含まない PBIL 手法と複数の突然変異を含む手法、そして K2 [3] に遺伝的アルゴリズムを適用し、順序制約を進化させながら学習するアルゴリズムの計 5 つとの比較を行う。PBIL-RS のパラメータ設定は使用モデルを Alarm Network と Pathfinder の 2 つを用いて評価した。パラメータは各手法がほとんど AIC の更新の起こらなくなる世代数と学習率に設定した。各手法のパラメータ設定は表 1 の通りで、世代数を 500 としたパラメータを使用した、ただし比較手法は k , S , H の値は使用しない。突然変異の発生確率は、Alarm Network において、BM は [0.001:0.2] の範囲で最も性能が良かった 0.005、TM は [0.05:0.5] の範囲で最も良かった 0.1 を使用し、PM は [0.001:0.009] の範囲で最も性能が良かった 0.002 を使用した。実験の試行回数は 10 回としその平均スコアにより評価した。

図 4 は 37 ノードによる結果、図 5 は 135 ノードによる結果である。縦軸を AIC の値、横軸を世代数としており、その世代数までに発見した最も良いベイジアンネットワークモデルの AIC の値を示している。AIC の遷移を見ると PBIL に基づいた手法の中で PBIL-RS が最も早く収束している。この理由は、PBIL は一度収束するとそこで探索を打ち切ってしまうこと（37 ノードでは 160 世代、135 ノードでは 302 世代で探索を終了している）。そして、BM や TM、PM は深く探索する前に探索領域が変化する傾向があるからである。ここで、Pathfinder において、K2-GA は膨大な時間がかかり計算完了しなかったため結果が出ていない。K2-GA は 45 世代の計算に 650 時間が必要であった、これに対し PBIL-RS は 45 世代を 3 時間程度で実行する。

また、表 2 に最終的に得られた AIC の値をまとめた。表 2 は行が各手法を指し、列が使用したモデルを指し、各手法が使用モデル毎に得られた最良のベイジアンネッ

表 1: PBIL-RS のパラメータ設定

パラメータ	値
サンプル数	1000
1 世代中のモデル生成数 (C)	1000
選択するモデル数 (C')	10
学習率 (α)	0.1
探索回数 (k)	10
初期拡大幅 (H_{min})	0.4
拡大幅増加量 (H_{inc})	0.1
評価指標	AIC

表 2: 500 世代における AIC の値

手法	サンプルデータの基モデル	
	Alarm (37 ノード)	Pathfinder (135 ノード)
PBIL-RS	8536.484	30138.66
PBIL	8627.213	30243.75
PBIL + BM	8563.084	35240.99
PBIL + TM	8654.337	34784.23
PBIL + PM	8582.991	33003.01
K2-GA	13347.706	-

トワークモデルの AIC の値をまとめた表である。表 2 によると 37 ノードと 135 ノード規模の両方のベイジアンネットワークにおいて既存手法よりも PBIL-RS が最も良い性能を示している。

4.3 学習率 α の影響

最も学習に影響を与える学習率 α の影響を調べる。性能への影響を明らかにするため、異なる値によって実行し、ベイジアンネットワークモデルの学習を行い、求めた最良のベイジアンネットワークモデルの AIC の値を比較する。

学習率を変化させ、PBIL-RS が 500 世代までに探し出したベイジアンネットワークモデルの AIC のスコアの遷移を図 6 に示した。図 6 は探索開始から探索終了までのベイジアンネットワークモデルの AIC の値の遷移を示している。図 6 は縦軸がベイジアンネットワークモデルの AIC の値を、横軸が経過した世代数を表しており、その世代数までに発見した最も良いベイジアンネットワークモデルの AIC の値を示している。図 6 によると学習率が低いほど収束が緩やかになる傾向があることがわかる。また、各学習率において、初めて収束した世代数を表 3 にまとめた。表 3 から、学習率が低くなるに従い、収束にかかる世代が指数的に増加することがわかる。

PBIL-RS が探索した最も良い AIC の値が、学習率により、どのような影響を受けたのかを図 7 に示す。図 7 の

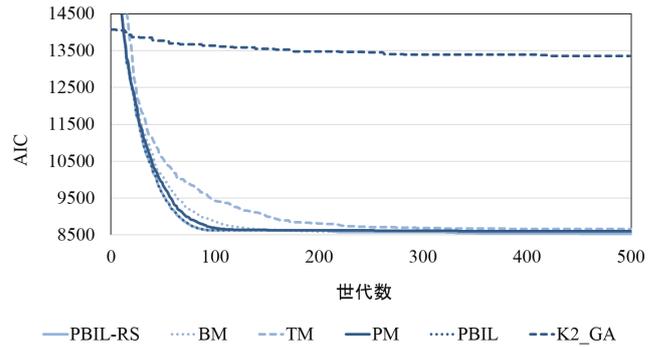


図 4: ALARM における手法毎の AIC の遷移

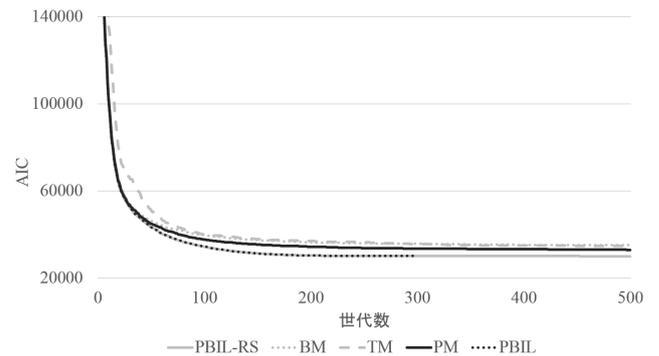


図 5: Pathfinder における手法毎の AIC の遷移

縦軸の AIC の値は試行回数 30 回の平均値である。図 7 から、学習率が低い時には、最終的な AIC の値が低い値をとることがわかる。つまり、学習率が低いほど収束する速度は緩やかであるが、最終的に得られるベイジアンネットワークモデルは優れている。すなわち、学習率は世代数とのバランスによって最適な値が決まることがわかる。

4.4 探索領域の拡大幅による影響

PBIL-RS が初めて探索領域の拡大を行う時、その拡大幅を初期拡大幅 H_{min} によって決定する。まずは、初期拡大幅 H_{min} の影響をみる。

図 8 は最終的に得られる AIC が初期拡大幅 H_{min} によってどのように変化するかを表している。縦軸は AIC を表し、横軸は設定した初期拡大幅 H_{min} を表している。また、パラメータは表 1 から、初期拡大幅 H_{min} を変更した値を用いた。図 8 によると、最大値と最低値の差はあまりなく、初期拡大幅 H_{min} は学習に大きな影響を与えないことがわかる。

次に探索の様子をみる。図 9 は左側の主縦軸が散布図の AIC の更新幅を取り、右側の第二縦軸は折れ線で示す収束度の値を表す。横軸は経過した世代数を表す。また、主縦軸の AIC の更新幅は 1 つ前までの世代で算出した最も良い値から、その世代までに算出した最も良い値を減算した値である。図 9 によると PBIL-RS は確率ベク

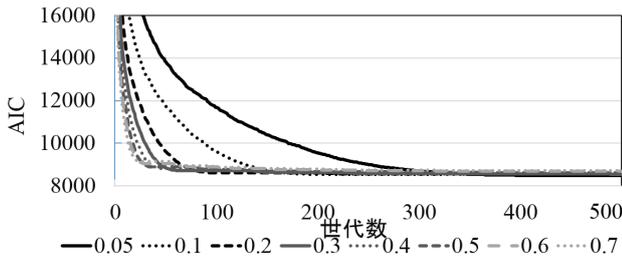


図 6: 学習率別のスコア遷移

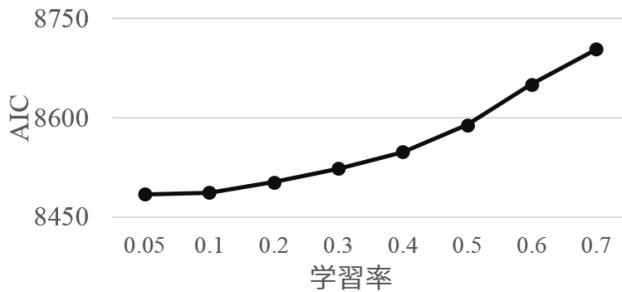


図 7: 学習率 (α) 別のスコア

トルが収束するたびに、適切に探索領域の拡大を行い、優れたベイジアンネットワークを探し出していることがわかる。また、一度探索領域の拡大をしたあとで、AIC の値が更新される世代の収束度がいずれも近い値をとることがわかる。

そこで、AIC が更新された時の収束度を図 10 に示す。図 10 は縦軸が収束度を示し、横軸は AIC の更新幅を表す。探索領域の拡大を行った世代から 1000 世代まで毎世代の AIC の更新幅を求めてグラフに描画した。パラメータは表 1 を用いて、試行回数は 10 回で行った。図 10 によると、収束度が [0.001:0.03] でのみ優れたベイジアンネットワークモデルを探し出し、AIC の値の更新が起こっており、深い探索が良いベイジアンネットワークモデルを見つけるために効果的であることを示している。

4.5 探索領域の段階的拡大の効果

PBIL-RS は探索領域を拡大する大きさを段階的に大きくしている。この動作の効果を調べるために PBIL-RS を 4000 世代で実行した、パラメータは $H_{min} = 0.1$ 、 $H_{inc} = 0.1$ とした。図 11 によると 1200 世代から 2400 世代にかけて探索領域を段階的に拡大していくことで優れたベイジアンネットワークモデルを探し出しており、探索領域の段階的拡大が効果的に働いている様子が見える。

5. おわりに

本研究では、ベイジアンネットワーク学習アルゴリズム PBIL-RS に対して詳細な評価を行った。PBIL-RS は既存手法に対して、ベイジアンネットワークモデルの規模によらず、最も優れた性能を持つベイジアンネットワー

表 3: 各学習率が初めて収束する世代数

学習率	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7
世代数	544	308	153	103	78	63	52	43

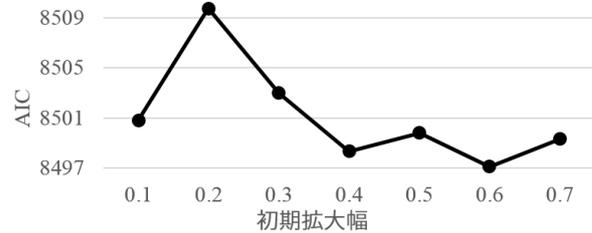


図 8: 初期拡大率 (H_{min}) 別のスコア

ク学習アルゴリズムであることを示した。学習率の影響評価では、学習率の値は世代数とのバランスによって適切な値が決まることがわかった。また、優れたベイジアンネットワークモデルが探し出される収束度には、規則性があることが明らかになった。さらに、PBIL-RS の探索領域の段階的な拡大動作は、この性質を捉えて有効に働いていることを示した。

謝辞

本研究は、日本中央競馬会畜産振興事業の支援を得た。ここに記して謝意を示す。

参考文献

- [1] D.M. Chickering, D. Heckerman, C. Meek, "Large-Sample Learning of Bayesian Networks is NP-Hard," *Journal of Machine Learning Research*, Vol.5, pp.1287-1330 (2004).
- [2] G.F. Cooper, and E. Herskovits, "A Bayesian Method for the Induction of Probabilistic Networks from Data," *Machine Learning*, Vol.9, pp.309-347 (1992).
- [3] W.H. Hsu, H. Guo, B.B. Perry, and J.A. Stilson, "A Permutation Genetic Algorithm for Variable Ordering in Learning Bayesian Networks from Data," In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO'02)*, pp.383-390 (2002).
- [4] O. Barrière, E. Lutton, P.H. Willemin, "Bayesian Network Structure Learning Using Cooperative Coevolution," In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO'09)*, pp.755-762 (2009).
- [5] A.P. Tonda, E. Lutton, R. Reuillon, G. Squillero, and P.H. Willemin, "Bayesian Network Structure Learn-

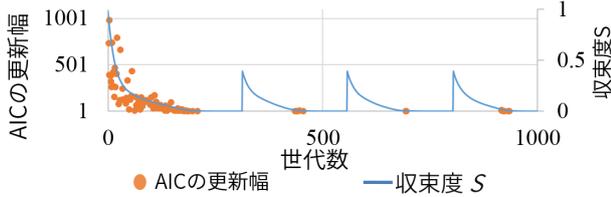


図 9: 探索領域の幅と AIC の更新幅

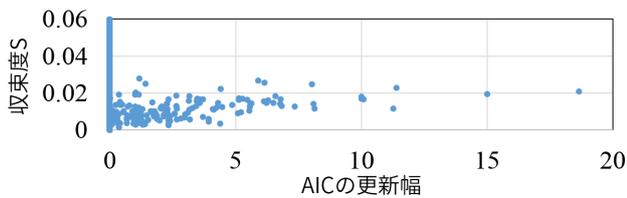


図 10: 収束度と AIC の更新幅

ing from Limited Datasets through Graph Evolution,” In Proceedings of the 15th European conference on Genetic Programming (EuroGP '12), pp.254-265 (2012).

- [6] H. Handa, “Estimation of Distribution Algorithms with Mutation,” Lecture Notes in Computer Science, Vol.3448, pp.112-121 (2005).
- [7] D.W. Kim, S. Ko, and B.Y. Kang, “Structure Learning of Bayesian Networks by Estimation of Distribution Algorithms with Transpose Mutation,” Journal of Applied Research and Technology, Vol.11, pp.586-596 (2013).
- [8] S. Fukuda, Y. Yamanaka, and T. Yoshihiro, “A Probability-based Evolutionary Algorithm with Mutations to Learn Bayesian Networks,” International Journal of Artificial Intelligence and Interactive Multimedia, Vol.3, No.1, pp.7-13 (2014).
- [9] R. Blanco, I. Inza, P. Larrañaga, “Learning Bayesian Networks in the Space of Structures by Estimation of Distribution Algorithms,” International Journal of Intelligent Systems, Vol.18, pp.205-220 (2003).
- [10] H. Akaike, “Information theory and an extension of the maximum likelihood principle,” Proceedings of the 2nd International Symposium on Information Theory, pp.267-281 (1973).
- [11] I.A. Beinlich, H.J. Suermondt, R.M. Chavez, G.F. Cooper, “The ALARM Monitoring System: A Case Study with Two Probabilistic Inference Techniques for Belief Networks,” In Proc. of Second Euro-

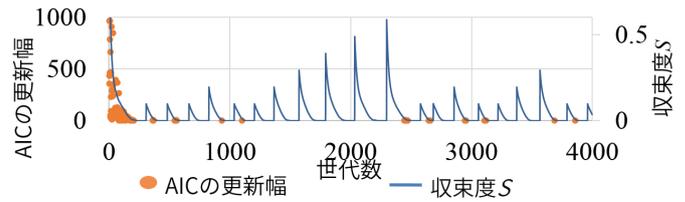


図 11: 4000 世代での探索領域の幅と AIC の更新幅

pean Conference on Artificial Intelligence in Medicine, Vol. 38, pp.247-256 (1989).

- [12] D.E. Heckerman, E.J. Horvitz, B.N. Nathwani, “Towards Normative Expert Systems: Part I - the Pathfinder Project,” Methods of Information in Medicine, pp. 90-105 (1992).
- [13] S. Baluja, “Population-Based Incremental Learning: A method for Integrating Genetic Search Based Function Optimization and Competitive Learning,” Technical Report CMU-CS-94-163, Carnegie Mellon University (1994).
- [14] 繁柝算男, 植野真臣, 本村陽一: ペイジアンネットワーク概説, 培風館 (2006) .
- [15] 福田 翔, 吉廣 卓哉, “ 遺伝的アルゴリズムによる確率学習に基づいたペイジアンネットワークモデル構築手法 ”和歌山大学大学院コミュニケーション科学クラス修士論文 (2014) .