

# ベイジアンネットの段階的構造学習法に対する確率的枝刈りを用いた高速化について

北越 大輔<sup>1,a)</sup> 東 悟大<sup>1,b)</sup> 鈴木 雅人<sup>1,c)</sup>

**概要:** ベイジアンネット (Bayesian Network: BN) は確率変数間の依存関係を表現した確率モデルとして知られる。本稿では、BN の適切な結合構造を高速に獲得可能な、確率的枝刈りを用いた段階的構造学習法に注目し、その構造学習効率のさらなる改善を図る。段階的構造学習法ではノードをクラスタに分割し、クラスタ間学習を段階的に行うことで探索空間を抑制する。確率的枝刈りは、クラスタ間類似度を基準に学習対象を限定することで構造学習のさらなる高速化を図るが、不適切なクラスタの枝刈りによって、構造の妥当性を大きく損なう可能性がある。本稿では、複数種類の多変量データを対象とした計算機実験を通して、確率的枝刈りを適切に実施可能な設定について検討する。

## Improving Learning Speed in Stepwise Structure Learning Method for Bayesian Networks by using Probabilistic Pruning

DAISUKE KITAKOSHI<sup>1,a)</sup> GODAI AZUMA<sup>1,b)</sup> MASATO SUZUKI<sup>1,c)</sup>

### 1. はじめに

多変量データにおける個々の要素間の確率的依存関係や、データ全体としての特徴を発見したり、発見した関連性を用いて予測や分析に活用する、データマイニングに関する研究が注目を集めている。ビッグデータというキーワードが一般的に使われるようになる以前から、データマイニングツールの一つであるベイジアンネット (Bayesian Network: BN) [6] をはじめとした確率モデルに関する研究は盛んに行われており、その重要性は年々高まっている。BN は要素間の依存関係が非循環性有向グラフとして表現された知識表現系モデルであり、その結合構造を用いた様々な確率推論法も提案されているため、多様な分野への適用事例が報告されている。著者らはこれまで、多変量データをもとに適切な結合構造を比較的短時間で獲得可能な段階的構造学習法 (Stepwise Structure Learning: SSL)

[2] を提案した上で、学習対象となる結合構造空間の効果的な削減によって学習時間を短縮するため、当該手法に確率的枝刈り処理を導入しその特徴について評価してきた [1]。

本論文では、提案する構造学習法 (SSL-Pro) における枝刈りの“起こりやすさ”を特徴づける確率に注目し、データに対する妥当性を維持しながら高速に構造を獲得可能な設定について、計算機実験により検討する。具体的には、特徴的な結合構造および確率的特徴を有する複数種類の多変量データを対象に、複数の設定を適用した提案手法による構造学習を実施する。実験を通して、適切な枝刈り確率計算法や、枝刈り確率を特徴づけるパラメータの変化に伴い得られる結合構造の特徴等について考察する。

### 2. 準備

本節では、提案手法の適用対象となるベイジアンネット (Bayesian Network: BN)、および BN の結合構造を多変量データから決定するため提案されてきた、典型的な構造学習法について簡潔に紹介する。

<sup>1</sup> 東京工業高等専門学校  
National Institute of Technology, Tokyo College, Hachioji,  
Tokyo 193-0997, Japan

a) kitakosi@tokyo-ct.ac.jp

b) s11003@tokyo-ct.ac.jp

c) suz@tokyo-ct.ac.jp

## 2.1 ベイジアンネット (BN)

BN は、同時確率分布  $P(X_1, \dots, X_n)$  をもとに、各確率変数 (ノード) 間の依存関係を非循環性有向グラフとして表現した知識表現系モデルとして知られている。2 ノード間に確率的依存関係が存在する場合、当該ノード間に有向リンクが描かれるとともに、ノード間の関係性は条件付確率で表される。BN 全体に対応する同時確率分布は、 $P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | pa(X_i))$  として表される。ここで  $pa(X_i)$  は、ノード  $X_i$  への有向リンクが存在するノード (親ノード) の集合である。

BN は、変数間の依存関係が視覚的に表現されることから、データマイニングツールとしてしばしば応用されている。また、その結合構造および各ノードに割り当てられた条件付確率表 (CPT) を用いた様々な確率推論法が提案されており、不確実性を伴う事象の予測・分析ツールとしても利用されている。

## 2.2 BN の構造学習に用いられる手法

利用者が知りたい対象に関する BN の結合構造や CPT が、事前に与えられていることは多くの場合稀であり、対象に関する多変量データからネットワークの結合構造と CPT を決定 (構造学習) することが一般的である。

BN における適切な結合構造の学習は組み合わせ最適化問題の一つとして知られ、近似解法として知られる遺伝的アルゴリズムや、確率的反復改善探索法として知られる焼きなまし法を応用した学習法等に加え [5], TPDA [4] のように、BN の構造学習に特化したアルゴリズムも提案されている。本研究における提案手法の基盤となる段階的構造学習法 (Stepwise Structure Learning: SSL) [2] は、TPDA と同様、BN の結合構造学習に特化した手法の一つであり、データを構成する全確率変数 (ノード) を小集合に分割し、それらに対応する小 BN の構造学習に引き続き、小 BN 同士の構造学習を段階的に行うことで、全体として妥当な構造を高速に獲得することを目指す。当該手法の特徴評価のため、いくつかの実験を実施した結果、望ましいクラスタ分割やクラスタ内学習の方法に関する設定指針が明らかになった一方、適切なクラスタ間学習の手法についてはさらなる検討が必要であることや、ノード数の増加に伴い、クラスタ間学習にかかる計算量の増大が手法の全体的な計算量に大きな影響を与えることが確認された。

## 3. 確率的枝刈りを用いた段階的構造学習法

ノード数増に伴い計算量が大幅に増加し得るという SSL における問題点の改善を目的として、著者らのグループは、構造学習の対象空間を確率的に限定可能な確率的枝刈り処理を導入した手法 (Stepwise Structure Learning with Probabilistic pruning: SSL-Pro) を提案した。

確率的枝刈りとは、あるクラスタ  $C_a$  と、 $C_a$  の併合候補

クラスタ  $C_b$  との“確率的依存関係の意味での類似度”が低い場合、構造学習を実施しても各クラスタ間にリンクが付加される可能性が低いと判断し、 $C_b$  を  $C_a$  の併合候補から“確率的に除外する”処理に相当する。上記のような 2 クラスタ間での学習を実施しないことで、データに対する結合構造の妥当性を大きく損なうことなく、クラスタ間学習にかかる計算時間を抑制可能となることが期待できる。

確率的枝刈りの効率的な実施のため、BN 内のノードクラスタリングは階層的クラスタリング [3] の概念にもとづいて行われる。階層的クラスタリングでは当初、 $N$  個の要素を  $N$  個のクラスタと見なしたうえで、クラスタ間類似度が最大となる 2 クラスタを 1 つのクラスタに併合し、クラスタ数が所望の値となるまで当該処理を繰り返す。SSL-Pro におけるクラスタ間類似度は、階層的クラスタリングの一手法として知られる群平均法で用いる、クラスタ  $C_x$  と  $C_y$  の間の類似度関数  $s(C_x, C_y)$  によって計算される。

$$s(C_x, C_y) = \frac{1}{|C_x||C_y|} \sum_{X \in C_x} \sum_{Y \in C_y} I(X; Y) \quad (1)$$

(1) 式における  $I(X; Y)$  は、一般的な群平均法では 2 クラスタ内の個々の要素間の類似度を表し、提案手法では 2 ノード (確率変数) 間の確率的依存関係の強さを特徴づける相互情報量に相当する。従って SSL-Pro では、相互情報量をノード間の類似度を測る尺度とした階層的クラスタリング (群平均法) を実施しながら、クラスタ間学習を実施してもノード間にリンクが付く見込みが少ないクラスタを確率的に除外することで、BN のデータに対する妥当性の低下を抑制しながら学習時間の短縮を図る。

### 3.1 SSL-Pro の適用手順

提案手法を適用するための準備として、 $t$  回目の繰り返し処理におけるクラスタ  $C_i^t$  の添え字集合  $A^t$  ( $i \in A^t$ )、および、併合候補となる 2 クラスタの添え字の組み合わせ集合  $S = \{(x, y) | x, y \in A^t\}$  を用意する。 $N$  ノード ( $X_1, X_2, \dots, X_N$ ) の BN に対して提案手法を適用する手順を以下に示す。

- (1)  $C_i^0 = \{X_i\}$ ,  $A^0 = \{1, 2, \dots, N\}$ ,  $S = \{(i, j) | \forall i, j \in A^0, i \neq j\}$  とする
- (2) 類似度最大となる 2 クラスタ  $C_j^t, C_k^t$  ( $(j, k) \in S$ ) を選択し、それらの間の結合構造を学習
- (3)  $C_j^{t+1} = C_j^t \cup C_k^t$ ,  $C_m^{t+1} = C_m^t$  ( $m \in A^t - \{j, k\}$ )
- (4)  $C_j^{t+1}, C_m^{t+1}$  に対して以下の 3 条件にもとづき確率的枝刈り処理を実施 ( $\forall m \in A^t - \{j, k\}$ )
  - (a)  $(j, m) \in S$  かつ  $(k, m) \in S$  ならば、確率  $p_1$  で  $S \leftarrow S \setminus \{(j, m)\}$
  - (b) 併合前の 2 クラスタのうち、一方のみが  $m$  の併合候補となっている場合は、以下のいずれかに従う (ここで  $j^*$  は、クラスタ  $j$  と  $k$  が併合されてでき

た新たなクラスタを表す)

(i)  $(j, m) \in S$  のみが成り立つならば, 確率  $p_2$  で  
 $S \leftarrow S \setminus \{(j^*, m)\}$

(ii)  $(k, m) \in S$  のみが成り立つならば, 確率  $1-p_2$   
で  $S \leftarrow S \cup \{(j^*, m)\}$

(c) 上記以外の場合,  $S \leftarrow S \setminus \{(j, m)\}$ ,

(5)  $A^{t+1} = A^t \setminus \{k\}$ ,  $S \leftarrow S \setminus \{(k, n) | \forall n \in A^t\}$

(6)  $|A^t| = 1$  もしくは  $S = \phi$  ならば終了. そうでなければ  
 $t \leftarrow t+1$  として (2) へ戻る

手順 (4) に示した確率的枝刈りの 3 条件 (a) ~ (c) に伴う処理はそれぞれ, (a) 併合前の 2 クラスタ双方がクラスタ  $m$  の併合候補であった場合, (b) 2 クラスタ中の一方のみが  $m$  の併合候補であった場合, (c) 2 つのクラスタ双方が  $m$  の併合候補でなかった場合, 確率  $p_1$ ,  $p_2$ , および 1.0 で, 併合後クラスタを  $m$  の併合候補から外すことに対応する. また, 確率  $p_1$  と  $p_2$  は, クラスタ間類似度関数 ((1) 式) を用いて, 以下のように計算される.

$$p_1 = \alpha \frac{s(C_j^{t+1}, C_m^{t+1})}{\bar{I}} \quad (2)$$

$$p_2 = \alpha \frac{s(C_z^t, C_m^t)}{s(C_j^t, C_k^t)} \quad (3)$$

ここで,  $\bar{I}$  は全ノード間で計算した相互情報量の平均値,  $\alpha$  は確率的枝刈りの発生しやすさを決定するパラメータであり ( $\alpha \in [0, 1)$ ),  $\alpha$  の増加に伴い対象クラスタが併合候補から外れる確率が増加する. また, (3) 式における  $z$  は,  $(j, m) \in S$  の場合には  $z = j$ ,  $(k, m) \in S$  の場合には  $z = k$  とする. (2) 式における  $p_1$  は,  $\bar{I}$  と  $s(C_j^{t+1}, C_m^{t+1})$  との比を指数とすることで, 当該 BN における平均的な確率的依存関係の強さを基準に  $C_j^{t+1}$  と  $C_m^{t+1}$  との類似度を考慮して求められている. 一方, (3) 式における  $p_2$  は,  $t$  の時点で  $C_z^t$  の併合候補である  $C_m^t$  との類似度  $s(C_z^t, C_m^t)$  が, この時点での最大類似度  $s(C_j^t, C_k^t)$  に対してどの程度近いかを考慮して計算される.

また, 手順 (2) における 2 クラスタ間の結合構造学習手法としては, 一方のクラスタを親ノード (集合) 候補, 他方を子ノード (集合) 候補とした greedy 探索を採用しており, 構造の評価関数としては MDL 規準を用いている.

文献 [1] において我々は, SSL-Pro における階層的クラスタリングの高速化を図るとともに, 2 種類の多変量データを対象に構造学習を行い, SSL や焼きなまし法を適用した場合との比較を通して, 提案手法の特徴や獲得された結合構造の妥当性を評価した. 実験の結果, SSL-Pro は高速な構造学習を実現可能である一方, リンクを多く有する構造を学習する傾向があることに加え, 主に枝刈り確率  $p_2$  の影響で必要以上の枝刈りが発生し, 学習速度の向上以上に構造の妥当性が低下してしまう場合があることが確認された.

表 1 枝刈り確率  $p_2$  の新たな計算法

Table 1 New options for parameter  $p_2$ .

Method	Expression
Average	$\alpha \frac{\{s(C_j^t, C_m^t) + s(C_k^t, C_m^t)\}/2}{s(C_j^t, C_k^t)}$
RMS	$\alpha \frac{\sqrt{\{s(C_j^t, C_m^t)\}^2 + \{s(C_k^t, C_m^t)\}^2}/2}}{s(C_j^t, C_k^t)}$
Equivalent to $p_1$	$\alpha \frac{s(C_j^{t+1}, C_m^t)}{\bar{I}}$

### 3.2 適切な枝刈り確率計算法の検討

前節 (3) 式で示される確率  $p_2$  は, クラスタ  $C_j^{t+1}$  を構成する 2 クラスタ  $C_j^t, C_k^t$  のうち, 一方のみがクラスタ  $C_m^t$  との併合候補となっている場合に枝刈りを実施するかどうかを判定するが, 枝刈り判定の基準として当該時点での最大類似度  $s(C_j^t, C_k^t)$  を用いているため, 多くの場合で枝刈りが実施されることとなる. 結果として, 本来であれば枝刈りすべきではない (クラスタ間学習を実施することが望ましいと考えられる) 2 クラスタ間の学習が未実施となり, 学習時間の短縮というメリット以上に妥当なリンクが学習されないデメリットが顕著になる傾向があった.

そこで本研究では,  $p_2$  のより効果的な設定について検討するため, 表 1 に示す 3 種類の  $p_2$  計算法を SSL-Pro における確率的枝刈りに導入し, 各手法により計算された  $p_2$  を適用した提案手法による構造学習の特性, および性能を評価する. 従来採用されていた  $p_2$  では, クラスタ  $C_j^{t+1}$  を構成する  $C_j^t, C_k^t$  のうち一方のみが考慮される形式となっていたため, 表 1 における Average では, 双方のクラスタと  $C_m^t$  との類似度の平均値をパラメータ計算に用いることとした. RMS でも同様に,  $C_j^t$  および  $C_k^t$  と,  $C_m^t$  との類似度に注目するが, 枝刈りの必要以上の発生を抑制するために二乗平均平方根を取ることで, クラスタ間類似度が小さい場合の  $p_2$  全体に与える影響を抑制する. これら二つの計算では,  $\alpha$  の指数部における分母の値は従来の  $p_2$  と同様, 当該時点での最大類似度  $s(C_j^t, C_k^t)$  としているため,  $p_2$  の最小値 =  $\alpha$  である. 一方で, 相互情報量の平均値  $\bar{I}$  を採用したパラメータ  $p_1$  については [1] で実施した実験において比較的良好に機能していたことから,  $p_2$  の計算法を  $p_1$  と同等とする (Equivalent to  $p_1$ ) 改良案についても併せて検討することとする (ここで, 3.1 節の手順 (3) に記載した通り,  $C_m^{t+1} = C_m^t$ ).

## 4. 計算機実験

SSL-Pro における確率的枝刈りの効率的実施を可能とする設定について検討するため, 3.2 節で紹介した 3 種類の  $p_2$  計算法を適用して BN の構造学習を実施し, 学習時間や得られる結合構造の妥当性に関して評価する.

実験では 6 種類の特徴的な多変量データを用意し, 各データに対してパラメータ  $p_2$  の計算法として, 表 1 で示した (1) Average (AVE), (2) RMS (RMS), (3) Equivalent

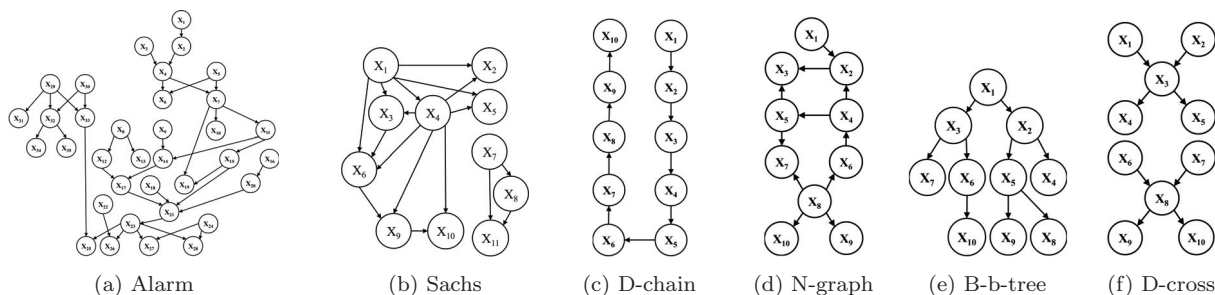


図 1 Structures of BNs used to generate data for structure learning.

表 2 構造学習データ生成元 BN の特徴

Table 2 Characteristics of BNs for generating data.

BN	ノード数	リンク数	自由度	生成元 BN	ノード数	リンク数	自由度
Alarm	35	39	$1.93 \times 10^{15}$	N-graph	10	11	59049
Sachs	11	17	177147	B-b-tree	10	9	59049
D-chain	10	9	59049	D-cross	10	8	59049

to  $p_1$  (EP1) を適用した場合に加え、従来の  $p_2$  (EXT) を適用した場合について、 $\alpha$  の値を変化させながら構造学習を実施し、学習結果について比較する。また、6 種類の BN (図 1) における CPT をもとに確率的に生成された多変量データを構造学習用データとして用いる (データ数 = 2,000,000)。図中の sachs ネットワークについては、研究目的で利用される Bayesian Network Repository [7] で公開された結合構造と CPT をそのまま使用しており、alarm については [7] における構造をもとに、無向ループ数が減少するよう一部ノードを削除し、CPT を修正した構造を採用している。他のネットワークは著者らが結合構造を設計したうえで、CPT については [7] で紹介される BN の各ノードに割り当てられた CPT の値を参考に設定した。各 BN の特徴を表 2 に示す。また、表 3 に記載された  $R_D$  は“BN から生成されたデータに対応する確率分布の同時エントロピー/同一状態数の分布における同時エントロピーの最大値”を表している。 $0 \leq R \leq 1$  であり、ある分布に対応する同時エントロピーが最大値の時、当該分布は一様分布のため、 $R = 1$  に近づくほど分布の偏りは少なく、0 に近づくほど偏りが大きい (分布が特徴的である) ことを意味する。各データに対する  $R_D$  を比較すると、alarm ネットワークから生成されたデータの  $R_D$  が最小であり、特徴的である (単純でない) データであることが示唆される一方、D-chain や D-cross に対応するデータの  $R_D$  は比較的大きく、alarm と比較して分布の偏りが少ないことが分かる。

実験では各多変量データに対して、新たに提案した 3 種類の  $p_2$  計算法を導入した SSL-Pro(AVE), SSL-Pro(RMS), SSL-Pro(EP1), および、従来の  $p_2$  を採用した SSL-Pro(EXT) をそれぞれ 20 回ずつ適用して構造学習を実施し、得られた結合構造をデータ生成元 BN の構造と比較する。また、獲得された BN の結合構造のそれぞれに対

表 3 構造学習用データの  $R_D$

Table 3 Values of  $R_D$  in data for structure learning.

生成元 BN	Alarm	Sachs	D-chain
$R_D$	0.598	0.772	0.784
生成元 BN	N-graph	B-b-tree	D-cross
$R_D$	0.764	0.629	0.795

して確率推論によって  $P(Q|E)$  を計算し、真の値との誤差を比較する。ここで、 $Q$  は推論の対象となるノード (Query ノード) 1 つを表し、 $E$  は  $Q$  を計算する前提として与えられる任意の数、種類の同時確率変数 (Evidence ノード) に対応する (同時確率変数の要素数は、1 つ以上、BN を構成する全ノード数 -1 以下とする)。Query ノードの種類と値、および Evidence ノードの数、種類と値はそれぞれ、確率推論の実施ごとにランダムに設定される。BN の結合構造を用いた確率推論アルゴリズムとして、本研究ではサンプリング手法の一つである Likelihood Weighting [8] を用いる。当該手法において 1 回あたり 1,000,000 ずつサンプルを生成して推論値を計算し、 $t$  回目までの生成サンプルによる結果と  $t-1$  回目の結果との誤差が  $10^{-4}$  未満となった値を推論結果として採用する。一方、確率推論の比較対象となる真の値は、構造学習用に生成した多変量データそのものから計算する。獲得された個々の結合構造に対して 50 回ずつ確率推論を実施し、推論誤差の値を比較することで BN 構造のデータに対する妥当性を評価する。

## 5. 結果と考察

本節ではまず、SSL-Pro による構造学習の特徴と、確率的枝刈りが学習後の結合構造に与える影響について考察する。続いて、データの特徴に応じて適切な枝刈り確率  $p_2$  を決定するためのパラメータ  $\alpha$  の設定法、および、 $p_2$  の適切な計算法について議論する。

表 4 各パラメータ計算法を用いた場合の構造学習結果  
Table 4 Results of structure learning using each parameter setting.

生成元 BN	$\alpha$ \ Method	None		EXT		AVE		RMS		EP1	
		Time [s]	MAE	Time [s]	MAE	Time [s]	MAE	Time [s]	MAE	Time [s]	MAE
Alarm	0.10			762.28	0.056	718.05	0.055	711.245	0.055	951.006	0.054
	0.20	4370.79	0.031	635.30	0.057	615.65	0.058	647.195	0.057	780.308	0.055
	0.30			566.06	0.058	566.48	0.060	627.292	0.057	753.281	0.056
Sachs	0.10			3.885	0.049	3.769	0.054	3.952	0.036	3.975	0.034
	0.20	7.302	0.018	3.533	0.055	3.550	0.057	3.223	0.069	4.080	0.037
	0.30			3.059	0.074	3.049	0.076	3.279	0.068	3.701	0.047
D-chain	0.10			0.565	0.038	0.452	0.041	0.634	0.034	0.652	0.032
	0.20	2.163	0.016	0.569	0.037	0.507	0.038	0.545	0.037	0.552	0.036
	0.30			0.520	0.040	0.457	0.042	0.502	0.039	0.552	0.039
N-graph	0.10			0.408	0.043	0.352	0.055	0.394	0.040	0.363	0.041
	0.20	1.138	0.021	0.282	0.069	0.287	0.061	0.286	0.061	0.352	0.050
	0.30			0.250	0.071	0.243	0.070	0.281	0.065	0.324	0.053
B-b-tree	0.10			0.249	0.072	0.191	0.081	0.257	0.073	0.223	0.076
	0.20	0.691	0.034	0.211	0.078	0.169	0.086	0.205	0.080	0.242	0.074
	0.30			0.180	0.083	0.144	0.091	0.161	0.089	0.229	0.078
D-cross	0.10			0.601	0.041	0.706	0.030	0.711	0.032	0.679	0.031
	0.20	2.080	0.012	0.559	0.042	0.626	0.039	0.619	0.039	0.622	0.036
	0.30			0.506	0.050	0.554	0.045	0.610	0.040	0.638	0.036

表 5 各パラメータ計算法における学習効率 (RDA) の比較  
Table 5 Comparison in Ratio of Degradation in Accuracy (RDA) for each computing method for parameter  $p_2$ .

	EXT	AVE	RMS	EP1	平均
Alarm	0.765	0.862	0.918	0.826	0.843
Sachs	1.187	1.139	1.576	1.302	1.301
D-chain	0.973	1.036	0.905	1.007	0.980
N-graph	1.031	0.868	1.178	1.171	1.062
B-b-tree	0.832	0.845	0.762	1.049	0.872
D-cross	1.021	1.185	1.070	1.076	1.088
平均	0.968	0.989	1.068	1.072	1.024

なお、実験結果として以降で紹介する数値は、特に断らない限り各設定において構造学習を実施した 20 回の結果に関する平均値を示している。確率推論の結果については、20 回の学習で得られた構造のそれぞれに対して 50 回の推論を実施した結果の“平均値の平均値”を掲載している。

### 5.1 SSL-Pro の特性について

はじめに、6 種類の BN から生成した各データに対して、確率的枝刈りを実施せず階層的クラスタリング (群平均法) のみを適用した場合 (None), および 4 種類の  $p_2$  計算法を適用した場合のそれぞれに対する結果の概要について、表 4 にまとめた。表中の MAE は、獲得された BN を用いた確率推論結果の真の値に対する平均絶対誤差 (Mean Absolute Error) を表している。表より、None では枝刈りが一切行われないため (すなわち、 $\alpha = 0.00$  のため)、全データにおいて枝刈りを適用した場合よりも多くの計算時

間を要しているが、その分、結合対象となる全ノードに対する構造学習を実施するため、MAE は最小の値となった。一方、SSL-Pro において 4 種類の  $p_2$  計算法を適用した場合に注目すると、 $\alpha$  の増加に伴い枝刈りが発生しやすくなることから学習にかかる時間が減少し、その反面、MAE が増加するという傾向がほぼ全ての場合において確認できた。続いて、 $\alpha$  を 0.10 ~ 0.30 まで変化させた際の、学習時間の短縮に伴う MAE の増加率 (Ratio of Degradation in Accuracy : RDA) について比較する (表 5)。RDA=1.0 の時、学習の高速化率と誤差の増加率は同等であり、値が小さいほど“誤差の増加を抑制しつつ高速な学習を実現可能”であることを意味する。データ生成元の各 BN について RDA の値を確認すると、Alarm では全般的に RDA < 1.0 となっており、どのパラメータ計算法を適用しても効率の良い学習ができていると考えられる。B-b-tree においても、EP1 を除き RDA の値は 1.0 未満となっている。その一方、Sachs や D-cross では全般的に RDA > 1.0 の値を示し、今回の実験で適用したどのパラメータ計算法を適用しても、 $\alpha$  の増加に伴い計算量は減少するものの、それ以上に推論誤差の増加率が大きくなる傾向があることが分かる。

$\alpha$  の増加に伴う学習時間、および MAE の変化の傾向については、今回取り上げた 6 種類のデータにおいてある程度共通の傾向が確認されたが、RDA の値については、 $p_2$  の計算法における相違のみならず、構造学習対象となるデータの確率的特徴が少なからず影響を与えているものと考えられる。ここで、多変量データにおける確率的特徴の指標である  $R_D$  (表 3) に注目して、今回使用した 6

表 6  $\alpha = 0.10$  における構造学習の特徴比較

Table 6 Comparison of features in structure learning for  $\alpha = 0.10$ .

生成元 BN	Criteria	Time [s]	MAE	削除 リンク数	付加 リンク数	反転 リンク数	相互情報量の 変化量
Alarm	EXT	762.279	0.056	11.85	29.40	14.45	3.669
	AVE	718.047	0.055	11.35	27.70	13.75	3.300
	RMS	711.245	0.055	11.40	27.75	13.35	3.487
	EP1	951.006	0.054	10.75	33.60	14.20	3.796
Sachs	EXT	3.885	0.049	1.85	9.65	7.90	0.954
	AVE	3.769	0.054	2.10	9.35	7.55	0.870
	RMS	3.952	0.036	1.05	9.15	7.50	0.983
	EP1	3.975	0.034	1.00	9.75	8.05	1.012
D-cross	EXT	0.601	0.041	2.15	3.55	2.55	0.335
	AVE	0.706	0.030	1.45	4.10	3.05	0.415
	RMS	0.711	0.032	1.55	4.60	3.45	0.319
	EP1	0.679	0.031	1.55	4.05	3.30	0.356

種類のデータの RDA を比較すると,  $R_D$  の小さい Alarm や B-b-tree における RDA の値は  $p_2$  計算法の違いによる影響を受けにくく, ほぼ全般的に小さな値を示している (Alarm における RDA の平均値は 0.843 であり, B-b-tree では 0.872). 逆に,  $R_D$  の値が 0.70 を上回る他の 4 種類のデータに SSL-Pro を適用した際の RDA の平均値は約 1 かそれ以上の値を示した.

確率的枝刈りの効果も RDA と同様に, SSL-Pro の適用対象となるデータの特徴に依存する面が大きいと予想される. また, 今回使用したデータのうち, Alarm と Sachs から生成されたデータへ提案手法を適用した場合における計算時間は数秒 ~ 数百秒と, 構造学習に比較的多くの時間を要しているのに対し, これらを除く BN から生成したデータへ提案手法を適用した場合における計算時間は 1 秒未満であり, 構造学習に要する計算時間の差が大きい. RDA は計算時間の短縮割合に対する誤差の増加率に関する指標であるため, 学習に多くの時間を要する Alarm や Sachs に対応するデータと比較して, 他のデータへ手法を適用した場合, 推論精度の増減が RDA の値に与える影響が大きい. 従って今後は, 学習時間やノード数, 自由度といった特徴量を一定範囲内に抑え, 個々の結果を容易に比較可能な状態としたうえで, 確率的特徴が顕著な多変量データを用意し提案手法の適用対象とすることで, 手法の特性をより詳細に検証していく必要がある.

## 5.2 適切なパラメータ設定および枝刈り確率の計算法

確率的枝刈りの起こりやすさ  $p_2$  を特徴づける  $\alpha$  の変化, および  $\alpha$  を用いた  $p_2$  計算法の違いにより得られる結合構造の相違について考察し, 適切な  $\alpha$  の設定法, および  $p_2$  計算法について検討する.

[ $\alpha$  の妥当な設定法について]

今回実験で採用したデータ生成元 BN のうち, ノード

数, リンク数, および  $R_D$  における相違が大きい Alarm, Sachs, D-cross から生成されたデータについて,  $\alpha = 0.10$  における構造学習結果を比較する (表 6). 表中の削除, 付加, 反転リンク数はそれぞれ, データ生成元 BN の結合構造として存在していたにも関わらず消失したリンク数, 当該 BN には存在しなかったが構造学習により付加されたリンク数, 当該 BN における向きとは反対方向に付与されたリンク数を意味している. また, リンクの削除, 付加, 反転に伴い, データ生成元 BN と同一構造を有するネットワークの全ノード間における相互情報量の総和と比較して, 学習後の BN における相互情報量がどの程度変化したかについても併せて示す.

相互情報量の変化量は, データ生成対象となった各 BN の結合構造に対して, “最終的に付加された” リンク数 (付加リンク数 - 削除リンク数) との相関が見受けられる. 表 6 から分かる通り, データ生成元 BN の構造として, 今回の実験中最多である 35 個のノードと, 39 本のリンク数を有する Alarm (表 2) の場合, 生成データを用いた構造学習の結果, 平均で約 18 本リンク数が増加しており, 相互情報量の変化量は紹介する 3 種類のデータの中で最大である. 一方, Sachs と D-cross から生成されたデータを用いた結合構造を比較した場合, 削除リンク数は同程度であり, 反転リンク数と付加リンク数が双方とも D-cross の 2 倍以上の値を示す Sachs における相互情報量の変化量は, D-cross の 2 ~ 3 倍程度の大きさを示していることが分かる. 表 6 の結果に加え, 実験結果全体を通して比較すると, 反転リンク数が同程度であった場合, 付加リンク数 - 削除リンク数の値が大きなネットワークほど相互情報量の変化量が大きな値を示すことも確認された. また, 表における個々のデータ生成元 BN に注目すると, 相互情報量の増加量が大きな BN ほど, MAE の値が小さいことも分かる.  $\alpha = 0.20, 0.30$  の場合における結果は紙面の都合上掲載し

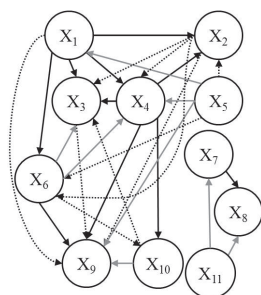


図 2 SSL-Pro で獲得された結合構造の例 (データ生成元 BN: Sachs)

Fig. 2 A typical example of BN structure acquired by SSL-Pro

ていないが、 $\alpha$  の増加に伴い確率的枝刈りが実施される頻度が増加し、付加リンク数より削除リンク数が大きくなる傾向が表れることによって、データ生成用の BN より総リンク数が減少するような状況も生じる。上記の傾向はデータ生成元 BN の違いや  $p_2$  計算法の違いに関わらず確認されており、結果として  $\alpha = 0.30$  における相互情報量の変化(増加)量は、 $\alpha = 0.10$  の場合と比較して平均で半分程度となり、それに伴い MAE の値も全般的に増加する。個々の計算法に注目してみた場合、 $\alpha$  の増加に伴う MAE の増大傾向は特に EXT (従来の計算法) および AVE において顕著である。一例として、N-graph から生成されたデータにおける結果(表 4)を確認すると、EXT と AVE の双方とも、 $\alpha = 0.30$  における MAE は 0.07 程度であり RMS や EP1 より大きな値を示している。 $\alpha = 0.10$  の場合からの増加量の面でも、EXT では 0.043 から 0.071 (0.028 増加)と、4 手法の中で最大の増加量となっている。

今回実施した実験における全般的な傾向として、 $\alpha$  を 0.20 程度まで増加した場合でも、構造学習後の BN の総リンク数がデータ生成元 BN の総リンク数より少なくなるような状況の発生頻度は低かった ( $\alpha \leq 0.20$  として実施した実験 54 種類のうち、付加リンク数 - 削除リンク数  $\leq -1.0$  となったのは 7 種類。 $\alpha = 0.30$  の場合、24 種類中 7 種類)。  $\alpha$  を 0.40 以上とする場合、計算速度のさらなる向上が見込めるものの、MAE が 0.10 を超える(推論結果に 10% の誤差が含まれる)状況も想定されるため、実用性が著しく低下することが予想される。実験で採用した  $\alpha \in [0.00, 0.30]$  において、SSL-Pro は [1] における実験結果と同様、“比較的ノード間に多くのリンクが付加された構造を学習する”傾向があることが確認された。今回の実験で獲得された結合構造の典型例として、Sachs ネットワークをデータ生成元、 $\alpha = 0.10$ 、 $p_2$  計算法として EP1 を適用した構造学習の結果を図 2 に示す。図の BN における黒、およびグレーの実線と点線はそれぞれ、データ生成元 BN に存在したリンク、向きが反転したリンク、および、データ生成元の構造には存在しておらず、構造学習の結果新たに付加されたリンクを表している。図に示した例では、削除リンク数 = 0

であったため、付加したリンク数 (= 10) の分だけ、データ生成元 BN よりもリンク数の多い構造が学習されたこととなる。BN に付加されるリンクの数が増えることで、構造学習の対象となったデータの特徴をより適切に表すことが可能となり、結果として当該 BN を用いた確率推論の精度向上に貢献することとなる。その一方で、BN が活用される主要な目的の一つであるデータマイニングの観点から考えると、“確率的依存関係の小さいリンクが多数付与される”よりも、(そのようなリンクは極力排除し)“顕著な依存関係のあるノード間のみ、正しくリンクが付与される”ことの方が重要である。そのような観点で改めて表 6 を見た場合、D-cross における結果に示される通り、削除リンク数、付加リンク数の双方ともが 0 に近い(データ生成元の BN に類似した構造を持つ)方が望ましいといえる。そのような観点も考慮した上で、確率推論の精度も担保することを意図すれば、今回の実験設定においては  $\alpha$  を 0.20 以下と設定することが妥当であると考えられる。

#### [適切な枝刈り確率計算法について]

確率的枝刈りに用いたパラメータ  $p_2$  に対する 4 種類の計算法の特徴について考察する(表 7)。表 2 や表 6 に示される通り、データ生成元として Alarm を対象とした結果については、学習時間や削除・付加・反転リンク数の値が他の BN と比較して大きくなる傾向がある事から、表 7 における各項目には中央値を記載している。表に示される通り、今回の実験設定下では AVE を用いた SSL-Pro が最も高速に BN の結合構造を学習しているのに対し、 $p_2$  の計算法を  $p_1$  と同様にした EP1 を適用することで、最もデータにフィットした(高精度な確率推論が可能)構造が獲得可能であることが分かる。AVE における個々の項目に注目すると、削除リンク数は 4 手法の中で最大である半面、付加リンク数も多く、付加リンク数 - 削除リンク数の値は最も 0 に近い。また、先述の通り計算時間は 4 手法中最小であるものの、MAE についても EP1 以外の 2 手法と同一の値を示している(平均値に基づく比較では EXT, RMS の双方より高精度)。EP1 を適用した場合においては、データ生成元 BN の結合構造に対する削除リンク数が 4 手法中最小である一方、付加リンク数が最大であることに加え、相互情報量の増加量も 4 手法中で最大値を示しており、この結果が高精度(MAE 最小)の推論の実現に起因していると考えられる。しかしながら、EP1 における RDA の値は 4 手法中で最大で最も学習効率が悪く、 $\alpha$  を増加させ学習を高速化した場合、計算時間の短縮率以上に誤差の増加率が増大する傾向があることも確認された。

これらの結果、特に表 7 を見る限り、高精度の推論結果が必要である場合には EP1、構造学習を高速に実施したい場合には AVE を  $p_2$  の計算法として適用するという方針が望ましいといえる。その反面、表 4 や表 6 でも示されている通り、構造学習の対象となるデータの特徴によって、

表 7 パラメータ  $p_2$  計算法の相違による BN 構造の特徴比較  
Table 7 Comparison of features in BN structure according to the difference in methods for computing  $p_2$ .

Criteria	Time [s]	MAE	削除 リンク数	付加 リンク数	反転 リンク数	相互情報量の 変化量	RDA
EXT	0.562	0.056	3.225	3.800	3.150	0.282	0.997
AVE	0.530	0.056	3.500	3.850	3.375	0.275	0.952
RMS	0.614	0.056	3.300	3.675	3.300	0.300	0.994
EP1	0.630	0.044	2.875	3.900	3.175	0.305	1.062

各計算法における学習時間や MAE の値にはばらつきが見られ、EXT や RMS が一般的に上記 2 手法より性能面で劣っていると結論付けることは、少なくとも今回の実験結果のみからは困難である。一方で、我々が実施した過去の実験 [1] において、BN の構造学習法としても一般的に使用される焼きなまし法を比較対象として SSL-Pro の学習性能を評価した際、提案手法は“同程度の学習時間を要する場合には SSL-Pro の方が高精度な推論が可能な構造を学習可能”かつ、“同程度の推論精度を実現する構造を、焼きなまし法より高速に学習可能”であることを確認していることから、提案手法は広く一般的なデータに対して適用可能な構造学習法として期待できる。確率的枝刈りの妥当性、およびその適切な適用法についてはデータの確率的特徴に依存し、枝刈りをどの程度積極的に適用すべきかを検討する必要がある。今回、実験対象としたデータ生成元 BN のうち、D-cross を除く全 BN では全てのノードが直接的、間接的に結合している構造を有している。例えば、データ生成元の BN 自体がいくつかの小 BN に分割でき、枝刈りの適切な実施が適切な結合構造の効率的な獲得に直結するような状況における、提案手法の特性評価はこれまで実施していない。そのため今後は、BN 中の複数のノード集合が互いに独立、ないしは依存関係の弱いクラスターを形成する構造から生成したデータに対する構造学習を通して、確率的枝刈りが適切なタイミングで適用されるか、および、枝刈りの適切な適用対象を選択できるか、といった観点に注目した実験を実施する必要があると考えられる。

## 6. おわりに

本論文では、BN の適切な結合構造を高速に獲得可能な手法である、確率的枝刈りを用いた段階的構造学習法 (SSL-Pro) における適切な設定について検討した。構造学習の適切さを可能な限り維持しながら学習時間の抑制を可能とする枝刈り確率の計算法を提案し、特徴の異なる複数の種類の多変量データを対象に計算機実験を実施した。

実験の結果、多変量データの特徴を適切に表現し、高精度の確率推論を実施できる構造学習の実現、および、妥当な推論精度を維持しながら高速な構造学習を実施可能とする、望ましい設定の方針について確認した。その一方で、

データの有する確率的特徴や構造学習にかかる計算時間等に応じて、獲得される構造の妥当性と、得られた構造を活用した確率推論精度の双方を考慮した適切な枝刈り確率の計算法について議論するためには、より多様な学習対象データを用意したうえで、多様な設定下での実験を通したさらなる評価、考察が必要であることも明らかになった。

今後はまず、提案手法における確率的枝刈りの特性について詳細に検証することを目的として、枝刈り確率計算法の違いによる構造学習過程の変化について考察すべく、顕著な確率的特徴を有する多様なデータを対象とした実験を実施していく。加えて、SSL-Pro の特徴が効率的な構造学習に貢献し得る適用範囲について検討するため、確率的枝刈りの適用が学習結果に直結するような特徴（例：小集合内の依存関係は強く、集合間の依存関係は弱い、ないしは互いに独立な構造）を有するデータを用いた特性・性能評価を実施するとともに、従来使用されている構造学習法との比較を通して、適切なパラメータ設定の指針、および適切な枝刈り確率計算法についても引き続き検討していく。

## 参考文献

- [1] 東悟大, 北越大輔, 鈴木雅人: クラスタリングと確率的枝刈りを用いたベイジアンネットワークの段階的構造学習法 - 確率的枝刈りの性能改善及び特性評価 -, 計測自動制御学会 SSI2015 講演論文集, pp. 666-671, (2015)
- [2] 福井広高, 北越大輔: 事前知識を活用した段階的なベイジアンネットワークの構造学習法, 信学技法, NC2008-91, pp.55-60, (2009)
- [3] 宮本定明: クラスタ分析入門, 森北出版株式会社, (1999)
- [4] J. Cheng, R. Greiner, J. Kelly, D. Bell, and W. Liu: Learning Bayesian networks from data: An information-theory based approach, Artificial Intelligence, Vol. 137, Iss. 1-2, pp. 43-90, 2002.
- [5] J. Lee, W. Chung, and E. Kim: Structure Learning of Bayesian Networks Using Dual Genetic Algorithm, IE-ICE Transactions on Information and Systems, Vol. E91-D, No. 1, pp. 32-43, (2008)
- [6] J. Pearl: Probabilistic Reasoning in Intelligent System: Networks of Plausible Inference, Morgan Kaufmann Pub. Inc., (1988)
- [7] M. Scutari: Bayesian Network Repository, <http://www.bnlearn.com/bnrepository/>, (2015)
- [8] R. D. Shachter, and M. A. Peot: Simulation approaches to general probabilistic inference on belief networks, Uncertainty in Artificial Intelligence, Vol. 5, pp. 221-231, (1990)