F-010

深さ優先分枝限定法によるベイジアンネットワーク分類器学習
Learning Bayesian network classifiers by depth-first branch and bound algorithm

加藤 弘也 † 菅原 聖太 † 植野 真臣 †
Koya Kato Shouta Sugahara Maomi Ueno

1. まえがき

分類はデータ分析やパターン認識の基礎であり, 機械学習分野における中心的問題である. 分類精度の高い最も有力な分類器の一つとしてベイジアンネットワーク分類器(Bayesian Network Classifier: BNC)が知られている[1, 2].

現在, 最も分類精度の高い BNC として, 菅原ら[3]は, 目的変数が親変数を持たない制約の下で, 真の同時確率を表現可能な構造のうち, 目的変数パラメータ数(Number of Class variable Parameters: NCP)のみ最小にする BNC 学習法を提案している.

一方, 従来のベイジアンネットワーク(Bayesian Network: BN)構造学習は, 全ての構造の中で, 周辺尤度(Marginal Likelihood: ML)を最大にする構造を推定し, 分類に影響しない変数を含めた同時確率分布を推定する. 菅原ら[3]の手法は分類に影響する目的変数に関わるパラメータ数のみを最小化し分類確率の推定のみを最適化しようとするもので, 新たな学習アルゴリズムを必要とした. そこで, 彼らは, 以下の二つのステップからなる新しいアルゴリズムを提案している. 第一ステップでは, 目的変数から始まる全ての変数順序について, ML を最大化する構造をそれぞれ求める. 第二ステップでは, 第一ステップで求めた構造のうち NCP を最小にする構造を探索する. 第二ステップの探索はグラフの最短パス探索問題として定式化され, 彼らのアルゴリズムは幅優先探索により最短パスを探索するが, 膨大な計算時間が必要となる.

本論では, 以上の問題を緩和するために, 新しいアルゴリズムを提案する. 菅原ら[3]のアルゴリズムの第二ステップは, 計算量が第一ステップに比べ大きく, 大規模な構造学習のために改良する必要がある. そこで, 本論では, 第二ステップにおける探索を枝刈りにより効率化することを考える. 菅原ら[3]の用いている幅優先探索は逐次的に最適な構造を更新できないため, 枝刈りを適用しても, その効果が限定的である. 本論では, 幅優先探索ではなく, 逐次的に最適な構造を更新する深さ優先探索に枝刈りを適用する新しい探索アルゴリズムを提案する. このアルゴリズムは深さ優先分枝限定法における枝刈りを行うことで, 従来手法より計算時間の削減が可能になる.

複数のベンチマークによる比較実験で, 従来手法では 20 変数程度の構造学習が限界であったが, 提案手法では 58 変数の構造学習を実現することを示す.

2. 目的変数パラメータ数最小化による BNC 学習

ベイジアンネットワーク分類器(以降, BNC)は, ベイジアンネットワーク(以降, BN)の一つであり, 確率変数をノードとし, ノード間の依存関係をエッジで表す非循環有向グラフGと, 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集合Θで表現される[1,2]. BNC では, Gにおいて一つのノードを目的変数, その他のノードを説明変数として扱い, 目的変数は親変数を持たないとする.

今, Gは離散確率変数集合 $\mathbf{V} = \{X_0, \dots, X_n\}$の各変数をノードとして持つとする. また, X_0を目的変数, $X_1, \dots, X_n$ を説明変数とする. さらに, Gにおける変数X_iの親変数集合を$Pa(X_i,\ G)$と表すとする. Gの各変数を要素とするベクトル$\boldsymbol{\sigma}$に対し, $\boldsymbol{\sigma}$のi番目の要素をX_{σ^i}で表すと, $\forall\, i, Pa\left(X_{\sigma^i}, G\right) \subseteq \bigcup_{j=1}^{i-1}\{X_{\sigma^j}\}$が成り立つとき, $\boldsymbol{\sigma}$をGの変数順序という.

菅原ら[3]は, 現在, 最も分類精度の高い BNC として, 真の同時確率を表現可能な構造のうち, 目的変数パラメータ数(以降, NCP)のみ最小にする BNC 学習法を提案している. また, 彼らは$\mathbf{V}$からなる全ての変数順序集合を$\sigma(\mathbf{V})$としたとき, 次の定理を証明した.

[定理 2.1] $\forall\, \sigma \in \sigma(\mathbf{V})$について, σを所与として周辺尤度を最大化する構造は, σに従い, 真の分類確率に漸近収束する構造の中で NCP が最小の構造に漸近的に一致する.

この定理に基づき, 彼らの手法は以下の二つのステップから構成される. 第一ステップでは, 目的変数から始まる全ての変数順序について, 周辺尤度を最大化する構造をそれぞれ求める. 第二ステップでは, 第一ステップで得られた構造の中で NCP 最小の構造を探索する. 第一ステップでは$O(2^n)$, 第二ステップでは$O(n2^n)$の計算量が必要となる.

3. 深さ優先分枝限定法による BNC 学習法の提案

本論では, 菅原ら[3]のアルゴリズムの第二ステップにおける探索を枝刈りにより効率化する新たなアルゴリズムを提案する. 彼らの手法は第二ステップにおける探索にグラフを用いていないため, 枝刈りを適用することができない. そこで, 本論では, 第二ステップにおける探索をグラフを用いた最短パス探索問題として定式化する.

3.1 NPC リバースオーダグラフ

本節では, 枝刈りを導入するために菅原ら[3]の手法の第二ステップを NPC リバースオーダグラフを用いた最短パス探索問題として定式化する. NPC リバースオーダグラフは, $X_0 \in \mathbf{U}$となる任意の変数集合$\mathbf{U} \subseteq \mathbf{V}$に対応するノード集合と, ノード集合の各ノード$\mathbf{U}$からノード$\mathbf{U} \setminus \{X_i\}$に引かれたエッジ集合からなる有向グラフである. 4 変数に対する NPC リバースオーダグラフ例を図 1 に示す.

ここで, ノード$\mathbf{U}$からノード$\mathbf{U} \setminus \{X_i\}$へのエッジは変数集合$\mathbf{U}$からなる構造において, X_iが子変数を持たず, X_iの親変数集合が$\mathbf{U} \setminus \{X_i\}$を親候補集合としたときの最適な親変数集合であることを意味する. そして, ノード$\mathbf{U}$からノード$\mathbf{U} \setminus \{X_i\}$へのエッジがもつ NCP をコストとして定義する. NPC リバー

† 電気通信大学大学院情報理工学研究科 Graduate school of Informatics and Engineering, The University of Electro-Communications

Copyright © 2023 by
The Institute of Electronics, Information and Communication Engineers and
Information Processing Society of Japan All rights reserved.

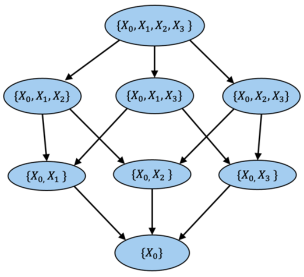

図 1. NPC リバースオーダグラフ

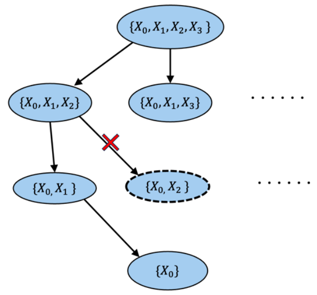

図 2. 深さ優先分枝限定法の動作例

表 1. 大規模データセットにおける各手法の分類精度

No.	Variables	Naive Bayes	幅優先探索	深さ優先分枝限定法
1	31	0.9139	TO	0.9350
2	37	0.6640	TO	0.9061*
3	42	0.7877	TO	0.7735
4	44	0.8333	TO	0.8444
5	58	0.8794	TO	0.9109*
	average	0.8156	-	0.8740

スオーダグラフの始点$\mathbf{V}$から終点$\{X_0\}$へのノードの連なりであるパスは目的変数を先頭にもつ各変数順序に対応する. また, パスのコストをそのパスを始点から終点まで辿ったときのコストの総和とすると, 最短パスは任意のパスのコスト以下であるパスと定義される. 最短パスを探索することで, 真の同時確率分布を表現可能な構造の中で NCP が最小の構造が得られる.

3.2 深さ優先分枝限定法を用いた最短パス探索

菅原ら[3]の第二ステップの手法は枝刈りによる効率化を行わないため探索グラフを用いていないが, NPC リバースオーダグラフの最短パスを幅優先探索する手法に一致する. 幅優先探索は NPC リバースオーダグラフの始点$\mathbf{V}$から探索を開始し, 幅優先的に全てのノードの展開を行う. 幅優先探索は変数数の増加に伴い膨大な計算時間が必要になるため, 20 変数程度の構造学習が限界である.

本論では, NPC リバースオーダグラフを用いた探索を枝刈りにより効率化することを考える. 幅優先探索は, 最適構造を逐次的に更新できないため, 枝刈りを適用しても, その効果が限定的である. 本論では, 最適構造の更新が可能な深さ優先探索に枝刈りを適用する.

図 2 は深さ優先分枝限定法の動作例を表している. 深さ優先分枝限定法は NPC リバースオーダグラフの始点 $\mathbf{V}$ から探索を開始し, 深さ優先的にノードの展開を行う.

例えば, 図 2 では, まず$\{X_0, X_1, X_2, X_3\}$を展開し, $\{X_0, X_1, X_2\}, \{X_0, X_1\}, \{X_0\}$の順で探索する. 次に, $\{X_0, X_1, X_2\}$から$\{X_0, X_2\}$への展開を行う. ここで, 始点$\{X_0, X_1, X_2, X_3\}$から$\{X_0, X_2\}$へのコストと, $\{X_0, X_2\}$から終点$\{X_0\}$へのコストの下限値の和がこれまでの最適構造のコストを上回っている場合, $\{X_0, X_1, X_2\}$から$\{X_0, X_2\}$への展開は行われず, 枝刈りされる. 次節では, コストの下限値の推定法を提案する.

3.3 Naïve Bayes を用いた下限

コストの下限値の計算には, ヒューリスティック関数を用いる. 本論では, 以下の定理 3.1 を証明し, 新しいヒューリスティック関数を提案する.

[定理 3.1] 任意の変数集合$\mathbf{V}$に対して, 真の同時確率を表現可能な構造の中で NCP を最小にする BNC を$G^*(\mathbf{V})$とし, $G^*(\mathbf{V})$における目的変数の子変数集合$\mathbf{V}_c$を説明変数とする Naïve Bayes を$G^{NB}(\mathbf{V}_c)$とすると以下が成り立つ.

$$G^{NB}(\mathbf{V}_c)\text{の NCP} \leq G^*(\mathbf{V})\text{の NCP}$$

定理 3.1 より, $G^*(\mathbf{V})$の目的変数の子変数集合$\mathbf{V}_c$を得ることで, $G^*(\mathbf{V})$の NCP の下限値を推定できる. したがって, 本論では, NPC リバースオーダグラフにおけるノード$\mathbf{U}$から終点までの下限値を計算するために, 次式で表される新しいヒューリスティック関数を提案する.

$$h^*(\mathbf{U}) = \sum_{\mathrm{X_i} \in (\mathbf{U} \cap \mathbf{V_c})} \mathrm{X_i}\text{の親が}\mathrm{X_0}\text{のみであったときの NCP}$$

4. 比較実験

本章では, 菅原ら[3]のアルゴリズムでは学習できない 20 変数以上の BNC 学習を評価する.

この実験では, 説明変数が目的変数のみを親に持つと仮定する Naïve Bayes と, 菅原ら[3]の深さ優先探索, 提案手法である, 深さ優先分枝限定法を比較する.また, 31 ~ 58 変数の 5 種類のベンチマークデータセットを用いて実験を行った. さらに, 構造学習は, 6 時間の制限時間を設け, 超過する場合は学習を打ち切った.

各手法の分類精度を表 1 に示す. 表 1 における”TO”は制限時間内に学習できなかったことを表す. また, 表 1 の深さ優先分枝限定法の結果において “*”はメモリオーバによって打ち切りが発生し, それまでに得た最も NCP の小さい構造を用いたときの分類精度に付与されている.

表 1 より, 深さ優先分枝限定法は全てのデータセットで学習構造を得ることができた. 一方, 幅優先探索は全てのデータセットで学習が打ち切られ, 構造を得ることができなかった. また, 深さ優先分枝限定法の分類精度は 3 番を除く全てのデータセットで Naïve Bayes の分類精度を上回っていることが確認できる.

5. むすび

本論では, 深さ優先分枝限定法による NCP を最小にして真の分類確率に漸近収束するベイジアンネットワーク分類器の構造学習法を提案した. 具体的には, (1) Naïve Bayes の NCP がその下限値であることを証明し, (2) 深さ優先分枝限定法のための NPC リバースオーダグラフを提案して, (1)で示した下限値を用いた分枝限定法を導入した.

提案手法は以下の利点を持つ. (1) 従来の幅優先探索を用いた手法よりも計算時間を大幅に削減する. (2) 実行途中にメモリ等のリソースが不足してもそれまでの最適な構造を得ることが可能である.

ベンチマークネットワークによる実験により, 提案手法は従来手法では学習できない 58 変数の構造学習を実現できることを示した.

参考文献

[1] S. Sugahara, and M. Ueno, “Exact learning augmented naïve Bayes classifier”, Entropy, vol. 23, no. 12, p.1703, 2021.
[2] S. Sugahara, W. Kishida, K. Kato, and M. Ueno, “Recursive autonomy identification-based learning of augmented naïve Bayes classifiers,” International Conference on Probabilistic Graphical Models, PMLR, pp.265-276, 2022.
[3] 菅原聖太, 植野真臣, “分類影響パラメータ数を最小化するベイジアンネットワーク分類器学習”, 電子情報通信学会論文誌 D, Vol.105, No.11, pp.679-690, 2022.

Copyright © 2023 by
The Institute of Electronics, Information and Communication Engineers and
Information Processing Society of Japan All rights reserved.