

ロジスティック回帰モデル族における層別変数がある場合の 近似ベイズ最適な予測法に関する研究

情報数理応用研究

5216C002-6 荒井琢充
指導教員 後藤正幸

A Study of Approximate Bayesian Prediction Algorithm using Stratification Variables for Logistic Regression Model Class

ARAI Takumi

1 研究背景・目的

説明変数からそのデータが属するカテゴリを予測する分類問題は広い適用範囲を持つ課題として古くから統計学の分野で扱われてきた。多数存在する分類問題の中で、病気の発症の有無、薬剤投与の効果の有無、商品のヒットの可能性などの出来事の発生有無など2値で生起する事象を予測する問題が存在する。そのような問題に対しては、ロジスティック回帰モデルが最もよく適用される方法である [1]。ロジスティック回帰モデルとは、目的変数と説明変数の間にある種の線形関係が想定できる場合に当てはまりが良いとされる。しかしながら、ある層別変数の取りうる値によって、他の説明変数と目的変数の線形構造が異なる問題も考えられる。例えば、病気の発症の有無を予測する問題では、性別により心拍数や血糖値など他の説明変数と目的変数の関係が異なる可能性がある。ある層別変数の値により線形関係が異なると判断される場合には、その変数を層別したもとのロジスティック回帰モデルを構築することが適当であると考えられる。そこで本研究では、ロジスティック分布族の中で、層別変数の候補が与えられている場合に対し、2値の目的変数を予測する問題を対象とする。

また、機械学習の分野において、説明変数の値によって層別し予測を行う代表的な手法として、決定木モデルがある。決定木モデルとは、説明変数の値をもとに木構造を用いて目的変数を予測する手法であり、CART, CHAID, ID3 など様々な決定木生成アルゴリズムが提案されている。更に、決定木モデルの各ノードに対して、ロジスティック回帰を当てはめたモデルも提案されている [2]。この手法は、CART アルゴリズムを用いてデータの層別を行ったもとの、それぞれのノードに所属するデータに対して、ロジスティック回帰式のパラメータを推定する手法である。しかし、層別を行った上で複数のロジスティック回帰式を推定するため、モデル全体の自由度が高い一方、1つのパラメータ推定に対して用いることのできるデータの数が相対的に減るため、その推定精度が低下してしまう可能性がある。そのため、このモデルでは学習データが与えられたもとの過度に自由度が高くない適切なモデルを1つ選択する必要があった。しかしながら、未観測のデータを予測する際には、必ずしも1つのモデルを選択する必要はない。すなわち、複数のモデルを構築したもとの、それらの予測結果を確率的に混合することが有効であると考えられる。

ここで、決定木モデルの分野においては、複数のモデルを混合する様々なアンサンブル手法が既に提案されている。中でも、性能がよい手法としてバギングやランダムフォレストが知られている。これらのモデルは、全データセットの中から複数のサブデータセットを作成し、そのサブデータセットが最も予測できるよう決定木を構築し、それらの予測を混合することで予測精度の向上を目指している。すなわち、学習データを分割することで弱学習器を多数構築し、それらのアンサンブルによりモデル全体の推定精度の向上を目指している。一方で、Bayesian Model Averaging の枠組みでは、弱学習器を用意するのではなく、考えられる全てのモデルで学習データを学習さ

せ、得られたモデルの予測とそのモデルの事後確率の積和によって予測を行う [3]。しかしながら、全てのモデルの混合を考えた場合、説明変数の増加によりモデル数が膨大になってしまう。そのため、実用上は何らかの方法でモデルを絞って決定木を構築し、近似的にベイズ最適な予測を構築することが行われている。これに対して、決定木モデルに限定した厳密なベイズ最適な予測法として、ベイズ基準のもとで平均予測誤り率を最小にするベイズ予測アルゴリズムが提案されている [4]。このアルゴリズムは予め層別変数が与えられたもとの、その並びのもとの考えられる決定木モデルを全て混合可能である。

本研究では、層別変数を木構造に組み込むか否かの判断が大きく予測結果に影響すると考え、層別変数に関してベイズ予測アルゴリズムを援用することで厳密ベイズ最適な手法の提案を目指している。しかし、従来のベイズ予測アルゴリズムは、葉ノードに多項分布が付与されたモデルであるため、ベイズ予測はラプラス推定量で与えられる。そのため、全ての決定木モデルを考える際も、計算量の削減を行うことができる。しかしながら、ロジスティック回帰モデルは厳密にベイズ推論を導入することが出来ないため、何らかの方法で近似を行う必要がある。さらに、ベイズ予測アルゴリズムは、データごとに逐次的に確率推定を行うモデルであるため、ロジスティック回帰の予測に際しても毎データに対して近似計算を行う必要があり、膨大な計算量となってしまう。

そこで、本研究ではロジスティック回帰モデルに対してベイズ予測アルゴリズムを適用するため、学習データごとの確率推定に対して、新たな2つの近似計算方法を導入することで解決を図る。具体的には、全ての学習データによる予測分布を代用する方法と一部の計算に学習データごとに確率推定を行うことが容易な0-1損失関数を代用する方法である。以上より、本研究ではロジスティック回帰の一手法であるベイズロジスティック回帰とアンサンブルの一手法であるベイズ予測アルゴリズムの考え方を援用した、近似的にベイズ最適となる予測アルゴリズムを構築する。さらに木構造の特徴を活かし、計算量を削減した効率的な計算アルゴリズムも提案する。ベンチマークデータセット [5], [6] を用いた実験を行うことで、提案手法の有効性を検証する。

2 準備

2.1 対象問題

本研究では、新規データの説明変数からその目的変数を予測する問題を扱う。これは与えられたデータから説明変数に対し、目的変数への適切な写像関数を求める問題といえる。いま、連続変数と離散変数が混在した K 個の説明変数を $\mathbf{x} = (x_1, x_2, \dots, x_k, \dots, x_K)$ 、目的変数を $y \in \{0, 1\}$ と定義する。さらに i 番目のデータの説明変数 x_i 、目的変数 y_i の組み合わせを $\mathbf{z}_i = (x_i, y_i)$ とする。また、 n 個のデータからなるデータ集合の説明変数を $\mathbf{x}^n = x_1 x_2 \dots x_n$ 、目的変数を $\mathbf{y}^n = y_1 y_2 \dots y_n$ とする。その組み合わせを $\mathbf{z}^n = z_1 z_2 \dots z_n$ とする。本研究では、学習データとして \mathbf{z}^n が与えられたもとの、 $n+1$ 番目の説明変数 x_{n+1} に対応する目的変数 y_{n+1} を予測する問題を扱う。

2.2 決定木モデル

決定木モデルとは、枝とノードからなる木構造を用いて識別規則を表現するモデルである。説明変数のうち、2値の離散値をとる説明変数を $x_d \in \{0, 1\} (d = 1, 2, \dots, D)$ とし、各枝には x_d が割り付けられ、葉ノードには目的変数 y が割り付けられた決定木モデルを考える。以下では、決定木に割り当てる説明変数は2値の説明変数であるとし、決定木は2分木を考える。

2.3 ベイズ予測アルゴリズム

前節で述べた決定木モデルは、学習データが与えられたもとで考えられ得る決定木モデルの中から Gini 係数、MDL や BIC など何らかの選択基準を用い、1つの決定木モデルを選択している。しかしながら、学習データが与えられたもとで未観測のデータを予測する問題を考えた場合、必ずしも1つのモデルを選択する必要はない。そこで、須子ら [4] は考えられる全ての決定木モデルを混合した予測モデルを提案している。ただし、考える決定木モデルは、説明変数 x_1, x_2, \dots, x_d が順番に与えられるとし、 x_1 は深さ1の枝に割り当てられるように、その順番で割り付けられるものとする。また、 \mathbf{x} と d が与えられたもとで、一意に定まる状態を s^d とする。ここで、 m を決定木モデル、 M を最大深さが同じ決定木モデル m の集合とする。さらに、 $P(m)$ をモデル m の生起確率、 $\theta(m)$ をモデル m のパラメータとしたとき、 $n+1$ 番目の目的変数 y_{n+1} のベイズ最適な生起確率は式 (1) で与えられる。

$$P(y_{n+1} | \mathbf{x}_{n+1}, \mathbf{z}^n) = \sum_{m \in M} \int_{\theta(m)} P(y_{n+1} | \mathbf{x}_{n+1}, \mathbf{z}^n, m, \theta(m)) \times P(\theta(m) | \mathbf{x}_{n+1}, \mathbf{z}^n, m) P(m | \mathbf{x}_{n+1}, \mathbf{z}^n) d\theta(m) \quad (1)$$

説明変数の数が増えた場合、考える木の深さ D が深くなり、考慮すべきモデルの数が指数的に増大してしまい、計算が非常に困難になる。そこで、須子ら [4] は、ベイズ符号化アルゴリズム [7] を援用することで、定められたモデルクラスのもとで、全ての決定木モデルを考慮した混合モデルを提案した。説明変数の適切な並び x_1, x_2, \dots, x_d が与えられた場合、最大深さ D の完全木からなる全ての部分木をモデルクラスとしてベイズ最適な予測が可能である。ここで、 s を完全木における葉ノードの状態、 S を状態 s の集合、 $\theta(s)$ を状態 s のパラメータと定義する。また、複数の決定木モデルに対して同一の状態 s が含まれる場合、パラメータ $\theta(s)$ の事前分布 $P(\theta(s))$ は同一の状態 s であれば全て等しいものとする、モデル m の事前確率 $P(m)$ を状態 s の事前確率 $P(s)$ で置き換えることができる。上記の変換を行うことで、全ての決定木モデルを考える代わりに、1つの混合モデルを用いた効率的な予測アルゴリズムを構成することが可能となる。このアルゴリズムで計算される予測分布は、式 (2) に示すパス上のノードで混合を取る混合モデルで与えられる。

$$P(y_{n+1} | \mathbf{x}_{n+1}, \mathbf{z}^n) = \sum_{s \in S} \int_{\theta(s)} P(y_{n+1} | \mathbf{x}_{n+1}, \mathbf{z}^n, s, \theta(s)) \times P(\theta(s) | \mathbf{x}_{n+1}, \mathbf{z}^n, s) P(s | \mathbf{x}_{n+1}, \mathbf{z}^n) d\theta(s) \quad (2)$$

2.4 ロジスティック回帰

ロジスティック回帰モデルとは、2値の離散事象の発生確率を予測する手法で、予測結果は0から1の値の間を取る。ある現象が生起する場合の目的変数を $y = 1$ としたとき、 $n+1$ 番目の説明変数 \mathbf{x}_{n+1} に対応する目的変数 y_{n+1} が1になる確率 $P(y_{n+1} = 1)$ を用い、式 (3) のモデルが仮定される。ただし、 n 個のデータ \mathbf{z}^n を学習した時の回帰係数を $\beta_n = (\beta_n^1, \beta_n^2, \dots, \beta_n^k, \dots, \beta_n^K)$ とすると、 β は、一般的に最尤推定法により推定される。

$$P(y_{n+1} = 1) = \frac{1}{1 + \exp(\beta_n^0 + \beta_n \mathbf{x}_{n+1}^T)} \quad (3)$$

ロジスティック回帰モデルでは、パラメータ推定の際、最尤推定法を用いるため、過学習を起こす恐れがある。そのため、ベイズ的枠組みを用いたベイズロジスティック回帰モデルが提案されている。しかしながら、ロジスティック回帰に対して、厳密なベイズ推論を計算することは難しく、事後分布をガウス分布で近似したラプラス近似により近似事後分布を算出している。 β^{MAP} を \mathbf{z}^n が与えられたもとでの β_n の事後確率最大推定量、 S^N を β_n の事後分散共分散とすると、ラプラス近似を用いた予測分布の算出式は式 (4)–(7) で与えられる。

$$P(y_{n+1} = 1 | \mathbf{x}_{n+1}) = \frac{1}{1 + \exp(\kappa(\sigma_{n+1}^2) \mu_{n+1})} \quad (4)$$

$$\kappa(\sigma_{n+1}^2) = (1 + \pi \sigma_{n+1}^2 / 8)^{-1/2} \quad (5)$$

$$\sigma_{n+1}^2 = \mathbf{x}_{n+1}^T \mathbf{S}_n^N \mathbf{x}_{n+1} \quad (6)$$

$$\mu_{n+1} = \beta^{MAP^T} \mathbf{x}_{n+1} \quad (7)$$

3 提案手法

3.1 概要

本研究では、 K 個の説明変数の中で離散変数と連続変数が混在し、そのうちある離散説明変数が層別変数の候補となる場合に対して、目的変数を予測する問題を対象とする。ここで、ノード s において、 \mathbf{x}_{n+1} のうち上位のノードにて層別変数に用いられている説明変数を $\mathbf{x}_{n+1,s}^{DI}$ 、それ以外のロジスティック回帰に用いる説明変数を $\mathbf{x}_{n+1,s}^{CO}$ とする。つまり、末端のノードにおいては、 $\mathbf{x}_{n+1,s}^{DI}$ の要素が木の深さと同じ d 個であり、 $\mathbf{x}_{n+1,s}^{CO}$ の要素の数は $(K-d)$ 個となる。本手法では木構造に組み込む d 個の説明変数は、層別に用いるべきか否かの判断が大きく予測結果に影響すると考え、用いるか否かの不確実性をモデルに組み込む。そこで、既に分かっている説明変数の重要順をもとにした考えられる全ての決定木モデルに関してはベイズ基準に基づきモデルを足し合わせる。一方、ロジスティック回帰モデルに組み込むその他の説明変数に関しては、計算量の問題で全ての可能性を考えることが困難なため、ラプラス推定を用いて近似計算する。このとき、損失関数をロジスティック損失として、式 (8) で定義する。

$$Loss = \log \left\{ \frac{y_{n+1} + (1 - y_{n+1}) \exp(-P(y_{n+1} = 1))}{1 + \exp(-P(y_{n+1} = 1))} \right\} \quad (8)$$

式 (8) から考えられる平均予測誤り率を最小にするような最適予測を考えると、 $n+1$ 番目の目的変数 y_{n+1} の生起確率はベイズ予測アルゴリズムとベイズロジスティック回帰を用い、式 (9) で計算できる。

$$P(y_{n+1} | \mathbf{x}_{n+1}, \mathbf{z}^n) = \sum_{m \in M} \int_{\beta_m} P(y_{n+1} | \mathbf{x}_{n+1}, \mathbf{z}^n, m, \beta_m) \times P(\beta_m | \mathbf{z}^n, m) P(m | \mathbf{z}^n) d\beta_m = \sum_{m \in M} \int_{S_m^N} P(y_{n+1} | \mathbf{x}_{n+1}, \mathbf{z}^n, m, \beta_m^{MAP}, S_m^N) \times P(\beta_m^{MAP}, S_m^N | \mathbf{x}_{n+1}, \mathbf{z}^n, m) P(m | \mathbf{x}_{n+1}, \mathbf{z}^n) dS_m^N = \sum_{m \in M} f(y_{n+1} | \mathbf{x}_{n+1}, \mathbf{z}^n, \beta_m^{MAP}, S_m^N, m) \times P(m | \mathbf{x}_{n+1}, \mathbf{z}^n) \quad (9)$$

3.2 効率的なベイズ最適な予測法

説明変数の数が増えた場合、考える木の最大深さ D が深くなり、考慮すべきモデルの数が指数的に増大してしまうため、計算が非常に困難になる。すなわち、説明変数が増大した場合、式 (9) の計算は現実的ではない。そこで、モデルクラスを制限したもとの考えられる全ての決定木モデルを混合した混合モデルによる予測法を提案する。以下では、須子らの手法 [4] にならい、説明変数の並びが与えられたもとの全ての決定木モデルを考慮する混合モデルを提案する。ここで、 $n+1$ 番目の学習データを学習した時の深さ d のノードの状態を s_{n+1}^d とすると、 s_{n+1}^d の事前分布を式 (10) により定義するものとする。

$$P(s_{n+1}^d | \mathbf{z}^n) = q(s_{n+1}^d | \mathbf{z}^n) \prod_{r=0}^{d-1} (1 - q(s_{n+1}^r | \mathbf{z}^n)) \quad (10)$$

これにより、各モデルに持たせていた事前確率を各ノードに持たせることができる。すなわち、目的変数の生起確率 $P(y_{n+1} | \mathbf{x}_{n+1}, \mathbf{z}^n)$ は、混合モデルを用いることで、式 (10) を用いずに式 (11), (12) により計算できる。

$$\begin{aligned} P(y_{n+1} | \mathbf{x}_{n+1}, \mathbf{z}^n) &= \sum_{s \in \mathcal{S}_{n+1}} \int_{\boldsymbol{\theta}(s)} f(y_{n+1} | \mathbf{x}_{n+1}, \mathbf{z}^n, \boldsymbol{\theta}(s), s) \\ &\quad \times P(\boldsymbol{\theta}(s) | \mathbf{x}_{n+1}, \mathbf{z}^n, s) P(s | \mathbf{x}_{n+1}, \mathbf{z}^n) d\boldsymbol{\theta}(s) \\ &= \tilde{P}(y_{n+1} | \mathbf{x}_{n+1}, \mathbf{z}^n, s_{n+1}^0) \end{aligned} \quad (11)$$

$$\begin{aligned} \tilde{P}(y_{n+1} | \mathbf{x}_{n+1}, \mathbf{z}^n, s_{n+1}) &= \begin{cases} f(y_{n+1} | \mathbf{x}_{n+1, s_{n+1}}^{CO}, \mathbf{z}^n, s_{n+1}) & (s = s_{n+1}^K) \\ q(s_{n+1} | \mathbf{x}_{n+1, s_{n+1}}^{DI}, \mathbf{z}^n) f(y_{n+1} | \mathbf{x}_{n+1, s_{n+1}}^{CO}, \mathbf{z}^n, s_{n+1}) \\ \quad + (1 - q(s_{n+1} | \mathbf{x}_{n+1, s_{n+1}}^{DI}, \mathbf{z}^n)) \tilde{P}(y_{n+1} | \mathbf{x}_{n+1}, \mathbf{z}^n, s'_{n+1}) & (\text{otherwise}) \end{cases} \\ & \quad (12) \end{aligned}$$

ここで、 $f(y_{n+1} | \mathbf{x}_{n+1}, s)$ は式 (4) により与えられ、 s' はノード s の子ノードとする。また、説明変数を 3 個としたときの本研究で構築しているモデルのイメージを図 1 に示す。説明変数 x_1, x_2, x_3 のうち、深さ 1 の枝に x_1 を深さ 2 の枝に x_2 を割り当てた時の提案手法のイメージを示す。対象ノードに至るまでの層別に用いた説明変数は重複して利用してしまうため、そのノードのロジスティック回帰には取り込まない。

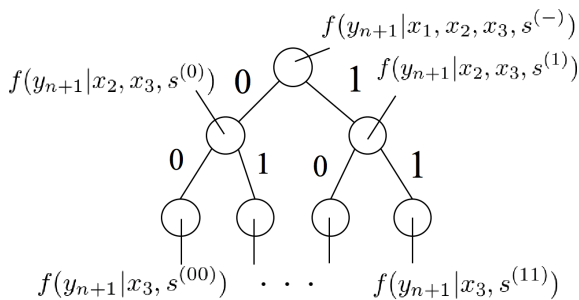


図 1: 提案手法イメージ

3.3 ノードの重みパラメータ算出方法

須子らの手法 [4] を援用した場合、 i 番目の学習データを学習した際のノード s_i の重みパラメータ $q(s_i | \mathbf{x}_{i, s_i}^{DI})$ は、式 (13) で与えられ、学習データごとに更新する必要がある。

$$q(s_i | \mathbf{x}_{i, s_i}^{DI}) = \frac{q(s_{i-1} | \mathbf{x}_{i, s_i}^{DI}) f(y_i | \mathbf{x}_{i, s_i}^{CO}, \mathbf{z}^{i-1}, s_i)}{\tilde{P}(y_i | \mathbf{x}_i, \mathbf{z}^{i-1}, s_i)} \quad (13)$$

この場合、学習データを 1 つ読み込むごとロジスティック回帰の予測値を求めるため、その予測分布を推定する必要があり、膨大な計算量となってしまう。そこで本研究では、計算量の増加を避けるため、近似的な値を用いることによりこれを回避する。この近似的な計算法について以下で 2 つの提案を述べる。

まず第一の提案として、重みパラメータ算出の際のみ、ロジスティック回帰式の予測値 $f(y_i | \mathbf{x}_i, \mathbf{z}^{i-1}, s_i)$ を、式 (14) に示す目的変数の生起回数に置き換え、式 (12) の計算を行う方法である。

$$f(y_i | \mathbf{x}_i, \mathbf{z}^{i-1}, s_i) = \frac{n(y_i | \mathbf{x}_i, s_i) + \beta(y_i)}{n(0 | \mathbf{x}_i, s_i) + n(1 | \mathbf{x}_i, s_i) + \beta(0) + \beta(1)} \quad (14)$$

ただし、 $n(y_{i+1} | \mathbf{x}_i)$ は、 i 番目の学習データまでにおける目的変数 y_i の生起回数、 $\beta(\cdot)$ はベータ分布の既知のパラメータとする。これにより、目的変数の生起回数のみで計算を行えるため、式 (13) と比べ、大幅に計算量の削減が可能となる。

第二の近似算出法は、各ノードのベイズロジスティック回帰式の予測分布を算出することに計算コストがかかるため、逐次学習の際にも全てのデータ \mathbf{z}^n を用いて算出した予測分布を用いる。すなわち、式 (13) を式 (15) に代えて重みパラメータの算出を行う。

$$q(s_i | \mathbf{x}_{i, s_i}^{DI}) = \frac{q(s_{i-1} | \mathbf{x}_{i, s_i}^{DI}) f(y_i | \mathbf{z}^n, s_i)}{\tilde{P}(y_i | \mathbf{x}_{i, s_i}^{DI}, \mathbf{z}^n, s_i)} \quad (15)$$

この提案においても、最も計算コストのかかるベイズロジスティック回帰式の予測分布の算出は一度のみに抑えることで、計算量の削減を図る。

3.4 提案学習アルゴリズム

提案手法のアルゴリズムを重みパラメータの算出方法別に以下に示す。

提案手法 1: 式 (14) により生起回数を代用

STEP1 $i = 1$ とする。

STEP2 i 番目の学習データに対し、決定木モデルの識別規則により、層別を行う。

STEP3 i 番目のデータに対して、式 (14) を計算したもとの、式 (12), (13) を算出。

STEP4 $i = n$ なら終了。 $i \neq n$ なら STEP2 へ

提案手法 2: 全データによる予測分布を代用

STEP1 n 件の学習データを独立に決定木モデルの識別規則により、層別を行う。各ノードへ層別されたデータを用いて、ロジスティック回帰式の回帰係数を算出 ($i = 1$ とする)。

STEP2 i 番目のデータに対して、式 (15) を計算したもとの、式 (12), (13) を算出。

STEP3 $i = n$ なら終了。 $i \neq n$ なら STEP1 へ

4 実証実験

提案手法の有効性を示すため、Wooldridge data set 中の mroz データ [5] と UCI データセット中の Bank Marketing データ [6] を用いて 100 回の繰り返し実証実験を行った。用いた実験セットの詳細は表 1 の通りである。

表 1. 実験データセット詳細

	mroz	Bank
説明変数	22	21
学習データ	565	3087
テストデータ	188	1030

また、評価指標としては、式 (16) に示される予測誤り率を用いた。

$$\text{予測誤り率} = 1 - \frac{\text{正しく予測されたテストデータ数}}{\text{テストデータ数}} \quad (16)$$

さらに、比較手法 1 として、全ての変数を用いたロジスティック回帰式を求める手法、比較手法 2 として決定木モデルの葉ノードにロジスティック回帰式を付与する手法を用いた。また木構造に用いる説明変数は変数の特徴から予め選択するものとし、本実験では、各データに対して 2 個の説明変数を木構造に当てはめる。すなわち、深さ 2 の決定木モデルを考える。実験結果を表 2 に示す。

表 2. 実験結果 (予測誤り率)

	提案手法 1	提案手法 2	比較手法 1	比較手法 2
mroz	0.0038	0.0038	0.0041	0.0038
Bank	0.087	0.081	0.089	0.089

また、用いる学習データ数による精度の違いを検証するため、用いる学習データ数を全体の 20% から 100% までの 20% 刻みで変化させ、目的変数が未知なテストデータに対して予測を行う。その実験結果を図 2, 3 に示す。

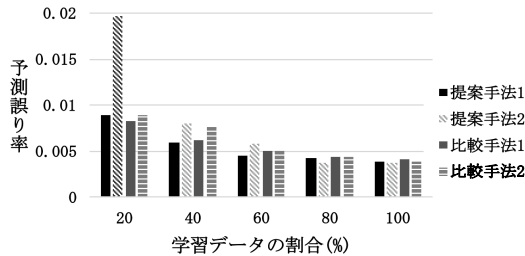


図 2: mroz の実験結果

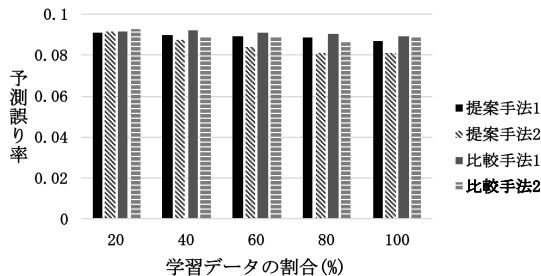


図 3: Bank の実験結果

まず、図 2 より、mroz データに関する結果と考察を述べる。比較手法 1 での精度が比較的安定して高く、説明変数により他の説明変数と目的変数の線形構造が大きく異なるデータであると考えられる。このようなデータに対しては、学習データの割合が少ない場合、提案手法 1, 2、比較手法 2 のような層別を行う手法は比較手法 1 と比べ、相対的にロジスティック回帰式のパラメータ推定に用いる学習データ数が少なくなるため、予測精度の面で劣る。特に、提案手法 2 は提案手法 1、比較手法 2 と比べ、推定精度の低いロジスティック回帰式を基にノードの重みパラメータの更新も行っているため、予測精度が低い。しかしながら、提案手法 1 は学習データの割合が増えるにつれ、データ不足の問題が解消されるため全ての手法とも同等の精度を示している。

続いて、図 3 に示した Bank データを用いた実験結果に関する考察を述べる。学習データ数に関わらず提案手法 1, 2 が比較手法 1, 2 に比べ、平均誤り率の観点で優れている。このデータは、木構造に用いた説明変数の値に

より、他の説明変数と目的変数の線形構造が大きく異なるデータであると言える。このような層別の必要性が高いデータにおいて、提案手法 1, 2 の有効性が際立つ。特に、提案手法 2 は提案手法 1、比較手法 1, 2 と比べ、学習データの割合が高い場合には予測精度の面で優れている。これは、推定精度が高いロジスティック回帰式のパラメータを用いて重みパラメータを行っているためと考えられる。

以上より、2 つの提案手法は、ロジスティック回帰式の適用性があるもとの、いくつかの説明変数による層別の必要性の判断が事前に難しい場合に有効性が高いと考えられる。このような問題設定において、比較手法 2 を用いた場合、説明変数において層別の有無を予め判断する必要があるが、提案手法 1, 2 は、可能性がある説明変数全てを取り込めば良い。一方で、ある説明変数により他の説明変数と目的変数の線形構造が大きく異なるデータに対しては、従来手法 1 でも十分であると考えられる。

5 まとめと今後の課題

本研究では、ロジスティック回帰モデル族のうちある説明変数の取りうる値によって、他の説明変数と目的変数の線形構造が異なる可能性があるような予測問題に対して、近似的なベイズ最適予測アルゴリズムを提案した。この手法は学習データが少なく、結果的に層別が必要であることが分かる問題に対し、予測精度の面で有効であることが分かった。また、説明変数が多く、どの説明変数で層別したらよいか分からない場合に、安定的に精度を担保できると考えられる。また、実証実験の結果より、予測精度の観点でも提案手法の有効性を示すことができた。

今後の課題としては、適切な木の深さの決定と多分木への応用が考えられる。現状では、データ数と説明変数の特性を鑑みて、木の深さは経験的に定める必要がある。更に、予測精度を安定させるためには適切な変数選択と木の深さの決定が必要であると考えられる。加えて、多値の離散変数も木構造へ取り込む応用として多分木への拡張が挙げられる。

参考文献

- [1] D. Cox, E. Snell, "Analysis of binary data," *CRC Press*, Vol. 32, 1989.
- [2] K. Chan, W. Loh, "LOTUS: An algorithm for building accurate and comprehensible logistic regression trees," *Computational and Graphical Statistics*, Vol. 13, No. 4, pp. 826–852, 2004.
- [3] B. Hernandez, A. Raftery, S. Pennington, A. Parnell, "Bayesian Additive Regression Trees using Bayesian Model Averaging," *Statistics and Computing*, pp. 1–21, 2017.
- [4] 須子統太, 野村亮, 松嶋敏泰, 平澤茂一, "決定木モデルにおける予測アルゴリズムについて," *信学技報, COMP*, Vol. 103, pp. 93–98, 2003.
- [5] T. Mroz, "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statical Assumptions," *Econometrica*, No. 55, pp. 765–799, 1987.
- [6] M. Lichman, "UCI Machine Learning Repository," <https://archive.ics.uci.edu/ml/datasets/bank+marketing>, 2013.
- [7] T. Matsushima, H. Inazumi, S. Hirasawa, "A Class of Distortionless Codes Designed by Bayes Decision Theory," *IEEE Trans. Inform. Theory*, Vol. IT-37, No. 5, pp. 1288–1293, 1991.