The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

アイテムセットを用いたスパースベイズ学習

Sparse Bayesian Learning for Itemset Data

矢船僚一朗 *1　西郷浩人 *2
Ryoichiro Yafune　Hiroto Saigo

*1*2 九州大学 大学院システム情報科学府
Graduate School of Information Science and Electrical Engineering, Kyushu University

Sparse bayesian learning can learn sparse solution for linear classification / regression problem. Although it has a number of advantages over non-bayesian approach, extension of it to non-linear model is non-trivial. In this paper, we employ itemset mining, and consider building sparse bayesian model on the binary occurrence matrix of items. We propose an iterative algorithm that can efficiently extract non-linear features while avoiding the entire enumeration. In computational experiments based on simulated dataset, our approach could correctly identify non-linearity in the dataset. In experiments using HIV dataset, we demonstrate the effectiveness of bayesian approach by rejecting samples with large estimated variance.

1. はじめに

スパースモデリングとは多数の変数から特に重要なものをシステマティックに選んでモデル構築を行う手法で，過去数 10 年の間に計算生物学，信号処理，コンピュータビジョンなど多くの分野で成功を収めてきた．特に LASSO [Tib96] として知られる L1 正則化問題は幅広く研究されている．LASSO の利点は，凸形式で定式化され高速計算ができることにある．これに対して私たちは事後分布を与える回帰モデルを構築するスパースベイズ学習 [TF03] を用いる．LASSO は点推定であるが，ベイジアンな手法を用いることでターゲット値の推定値に加えて，推定の不確かさなどの追加情報が得られる．今回対象とする問題はアイテムセットに対する回帰学習である．アイテムセットとはアイテムの組み合わせのことである．各アイテムの有無を $\{0,1\}$ のバイナリベクトルで表すことによりアイテムセットの有無を $\{0,1\}$ で表現する．これは化合物，ガンの遺伝子型，ウイルスの突然変異などさまざまな種類の実データの表現が可能である．単独のアイテムを用いた場合は線形モデルであるが，アイテムセットを用いることで相互作用を利用した非線形モデルが構築される．また生物のデータに対しては選ばれた変数から，どの組み合わせが反応に影響を与えているかを解釈できる．これは創薬において重要な情報である．アイテムセットの次元数は単独なアイテムの次元数と比較して膨大であるがスパースモデリングを行うことで次元数を削減できる．この論文ではアイテムセットに対してスパースベイズ学習を適用する方法とアイテムセットの探索を効率的に行う定理を紹介する．そして，人工データと実データを用いた計算機実験を行い手法の利点について説明する．

2. 準備

2.1 アイテムセットによる回帰学習

データベース上に含まれるアイテム $item_1, \cdots, item_n$ の集合を E とする．アイテムとは「商品」などのものである．E の部分集合 $z \subseteq E$ をアイテムセット (Itemset) という．トレーニングデータの各サンプルはアイテムセット $z_i \subseteq E$ とターゲット値 $t \in \mathbb{R}$ で構成される．このとき，各サンプルのアイテムセット z をトランザクションという．与えられたアイテムセットから，指定した出現数やサイズを満たすアイテムセットを出力する方法をアイテムセットマイニングという．アイテムセットマイニングのアルゴリズムは LCM(Linear time Closed itemset Mining) [UAUA03] を用いた．この論文ではテストデータのトランザクション z_{test} を入力して，ターゲット値 t_{test} の予測値 $f(z_{test})$ を出力する回帰モデルを考える．単純なモデルとしてはアイテム単体の有無に重みをつけた回帰モデルがある．

$$f(z_i) = w_1\phi(item_1 \subseteq z_i) + \cdots + w_n\phi(item_n \subseteq z_i) \qquad (1)$$

ここで，$\phi(k \subseteq z_i) = 2I(k \subseteq z_i) - 1 \in \{-1, 1\}$ である．つまり ϕ はアイテム k がトランザクション z_i に存在するときに 1 を，存在しないときに -1 を返す指示器である．これを発展させ，次のような適当に選んだアイテムセット $p_1, \cdots, p_m$ の有無に重みをつけた回帰モデルを導入する．

$$f(z_i) = w_1\phi(p_1 \subseteq z_i) + \cdots + w_n\phi(p_m \subseteq z_i) \qquad (2)$$

アイテムセットによる回帰学習の利点は非線形効果を取り入れることによってより複雑なモデルを構築できることにある．一方で，アイテムセットの数は膨大でありアイテムセット空間の全てに重みをつけることは過学習を起こす可能性がある．そのため，適切なアイテムセットにのみ重みをつけるスパースベイズ学習を使用する．

2.2 スパースベイズ学習

Tipping [TF03] によるスパースベイズ学習の概要を説明する．アイテムセット $\mathbf{x_n}$ とターゲット値 t_n の組み合わせからなるデータセット $\{\mathbf{x_n}, t_n\}$ が与えられたとき，ターゲットベクトル $\mathbf{t} = (t_1, \cdots, t_N)^T$ を近似ベクトルを $\mathbf{y} = (y(\mathbf{x}_1), \cdots, y(\mathbf{x}_N))^T$ と誤差ベクトル $\epsilon = (\epsilon_1, \cdots, \epsilon_N)^T$ の和として表現する:

$$\mathbf{t} \quad = \quad \mathbf{y} + \epsilon = \mathbf{\Phi w} + \epsilon \qquad (3)$$

ここで，$\mathbf{w}$ はパラメータベクトルである．$\mathbf{\Phi} = [\Phi_1, \cdots, \Phi_M]$ は $N \times M$ の表現行列であり．列ベクトルは各データに対してアイテムセットの有無を示す $\{0,1\}$ の基底ベクトルを構成する．

連絡先: 矢船僚一朗： yafune.ryoichiro.549@s.kyushu-u.ac.jp

スパースベイズモデルは，誤差が平均 0，分散 σ^2 の独立したガウス分布として確率的に形作られるという仮定を行う：$p(\boldsymbol{\epsilon}) = \prod_{n=1}^{N} N(\epsilon_n|0,\sigma^2)$．パラメータ σ^2 は前もって知っていれば定めることが出来るが一般的にはデータから推定される．したがって，誤差のモデルはターゲットベクトル $\mathbf{t}$ に対しての多変量正規分布を意味し，これは尤度関数となる:

$$p(\mathbf{t}|\mathbf{w},\sigma^2) = (2\pi)^{-N/2}\sigma^{-N}\exp\{-\frac{\|\mathbf{t}-\mathbf{y}\|^2}{2\sigma^2}\}. \tag{4}$$

この尤度関数はパラメータの事前確率によって補完される:

$$p(\mathbf{w}|\alpha) = (2\pi)^{-M/2}\prod_{m=1}^{M}\alpha_m^{1/2}\exp(-\frac{\alpha_m w_m^2}{2}). \tag{5}$$

ここでハイパーパラメータ，$\boldsymbol{\alpha} = (\alpha_1,\cdots,\alpha_M)^T$ を導入した．これはそれぞれが個々に事前確率の強さを制御する．$\boldsymbol{\alpha}$ が与えられたとき事後確率分布はベイズの定理のもとで尤度と事前確率を結合することで与えられる:

$$p(\mathbf{w}|\mathbf{t},\alpha,\sigma^2) = \frac{p(\mathbf{t}|\mathbf{w},\sigma^2)p(\mathbf{w}|\boldsymbol{\alpha})}{p(\mathbf{t}|\boldsymbol{\alpha},\sigma^2)}. \tag{6}$$

これはガウス分布 $N(\boldsymbol{\mu},\boldsymbol{\sigma})$ に従い，平均と分散は次のように表される:

$$\boldsymbol{\Sigma} = (\mathbf{A}+\sigma^{-2}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{-1} \quad \boldsymbol{\mu} = \sigma^{-2}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{t}, \tag{7}$$

ここで $\boldsymbol{A}$ は $\mathrm{diag}(\alpha_1,\cdots,\alpha_M)$ と定義される．タイプ II 最尤法を用いることで点推定 $\boldsymbol{\alpha}_{mp}$ が得られる．すなわち，スパースベイズ学習は周辺尤度 $\boldsymbol{\alpha}$ に関する局所的な最大化問題として定式化する．これは対数尤度 $L(\boldsymbol{\alpha})$ の最大化問題に等しい:

$$\begin{aligned} L(\boldsymbol{\alpha}) &= \log p(\mathbf{t}|\boldsymbol{\alpha},\sigma^2) = \log\int_{\infty}^{\infty} p(\mathbf{t}|\mathbf{w},\sigma^2)p(\mathbf{w}|\boldsymbol{\alpha})d\mathbf{w} \\ &= -\frac{1}{2}[N\log 2\pi + \log|C| + \boldsymbol{t}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{t}], \qquad (8) \\ \mathbf{C} &= \sigma^2\mathbf{I} + \boldsymbol{\Phi}\mathbf{A}^{-1}\boldsymbol{\Phi}^{\mathrm{T}}. \qquad (9) \end{aligned}$$

$\boldsymbol{\alpha} = \boldsymbol{\alpha}_{MP}$ で (6) 式を評価することによってパラメータに対する点推定 $\boldsymbol{\mu}_{MP}$ が得られる．これは最終的な近似値 $\mathbf{y} = \boldsymbol{\Phi}\boldsymbol{\mu}_{MP}$ を与える．

2.3 周辺尤度最大化

周辺尤度最大化のアルゴリズムを紹介する．単一のハイパパラメータ α_i，$i \in \{1\cdots M\}$ で構成される項ができるように，$L(\boldsymbol{\alpha})$ の $\mathbf{C}$ を書き換える:

$$\begin{aligned} \mathbf{C} &= \sigma^2\mathbf{I} + \sum_{m\neq i}\alpha_m^{-1}\phi_m\phi_m^{\mathrm{T}} + \alpha_i^{-1}\phi_i\phi_i^{\mathrm{T}} \\ &= \mathbf{C}_{-i} + \alpha_i^{-1}\phi_i\phi_i^{\mathrm{T}} \qquad (10) \end{aligned}$$

ここで $\mathbf{C}_{-i}$ は $\mathbf{C}$ から基底ベクトル i の寄与が除かれた行列である．$L(\boldsymbol{\alpha})$ の記述のために行列式と逆行列が必要になるため下に示す:

$$\begin{aligned} |\mathbf{C}| &= |\mathbf{C}_{-i}||1+\alpha_i^{-1}\phi_i^{\mathrm{T}}\phi_i| \qquad (11) \\ \mathbf{C}^{-1} &= \mathbf{C}_{-i}^{-1} - \frac{\mathbf{C}_{-i}^{-1}\phi_i\phi_i^{\mathrm{T}}\mathbf{C}_{-i}^{-1}}{\alpha_i+\phi_i^{\mathrm{T}}\mathbf{C}_{-i}^{-1}\phi_i} \qquad (12) \end{aligned}$$

これより $L(\alpha)$ を次のように書くことができる:

$$\begin{aligned} L(\alpha) &= -\frac{1}{2}[N\log(2\pi) + \log|\mathbf{C}_{-i}| + \mathbf{t}^{\mathrm{T}}\mathbf{C}_{-i}^{-1}\mathbf{t} \\ &\quad - \log\alpha_i + \log(\alpha_i + \phi_i^{\mathrm{T}}\mathbf{C}_{-i}^{-1}\phi_i) - \frac{(\phi_i^{\mathrm{t}}\mathbf{C}_{-i}^{-1}\mathbf{t})^2}{\alpha_i+\phi_i^{\mathrm{T}}\mathbf{C}_{-i}^{-1}\phi_i}] \qquad (13) \\ &= L(\boldsymbol{\alpha}_{-i}) + \frac{1}{2}[\log\alpha_i - \log(\alpha_i+s_i) + \frac{q_i^2}{\alpha_i+s_i}] \\ &= L(\boldsymbol{\alpha}_{-i}) + l(\alpha_i) \end{aligned}$$

ここで s_i，q_i の値を次のように定義した:

$$s_i \triangleq \phi_i^{\mathrm{T}}\mathbf{C}_{-i}^{-1}\phi_i \quad q_i \triangleq \phi_i^{\mathrm{t}}\mathbf{C}_{-i}^{-1}\mathbf{t} \tag{14}$$

q_i は $q_i = \sigma^{-2}\phi_i^{\mathrm{t}}(\mathbf{t}-\mathbf{y}_{-i})$ と書ける．目的関数 $L(\boldsymbol{\alpha})$ は計算のために ϕ_i の項を含まない $L(\boldsymbol{\alpha}_{-i})$ と ϕ_i を含む項である $l(\alpha_i)$ に分割された．$l(\alpha)$ を計算することで，$L(\boldsymbol{\alpha})$ は α_i に関して次の値のときに最大値を取ることが分かった:

$$\alpha_i = \frac{s_i^2}{q_i^2 - s_i} \quad , \quad q_i^2 > s_i \text{のとき} \tag{15}$$

$$\alpha_i = \infty \quad , \quad q_i^2 < s_i \text{のとき} \tag{16}$$

全ての基底ベクトル (アイテムセット) について，q_i と s_i を計算することは $S_i = \phi_i^{\mathrm{T}}\mathbf{C}^{-1}\phi_i$，$Q_i = \phi_i^{\mathrm{t}}\mathbf{C}^{-1}\mathbf{t}$ とおいたときに以下の計算式を用いることで比較的簡単にできる:

$$s_m = \frac{\alpha_m S_m}{\alpha_m - S_m}, \quad q_m = \frac{\alpha_m Q_m}{\alpha_m - S_m} \tag{17}$$

上の式 (15)，(16) はアルゴリズム上で次の手続きを行うことを意味している:

- ϕ_i がモデルに含まれていて $(\alpha_i < \infty)$ かつ $q_i^2 \leq s_i$ であるとき，ϕ_i を消去する $(\alpha_i = \infty)$.
- ϕ_i がモデルに含まれていなくて $(\alpha_i < \infty)$ かつ $q_i^2 > s_i$ であるとき，ϕ_i を追加する $(\alpha_i = s_i^2/(q_i^2 - s_i))$.

3. 手法

3.1 アルゴリズム

アイテムセットを使った周辺尤度最大化のアルゴリズムを Algorithm 1 に示す．アルゴリズムは全ての $\log\alpha$ が 10^{-6} 以下になり，かつすべての $\theta \leq 0$ で追加が出来ない状態で停止させている．これはモデルに追加できるアイテムセットが存在せず，かつ α の再計算でも周辺尤度があまり向上しない状態である．しかし，テストデータに適用した際にこの状態が最もよい性能を出すとは限らない．そのため後に記述する実験ではバリデーション用のデータを用いて，そのデータでの性能が最も良くなった状態のモデルをテストデータに採用した．テストデータの新しいアイテムセット $\mathbf{x}$ に対する予測 t を出力するために，次のように定義される予測分布を評価する [Bis06].

$$p(t|\mathbf{t},\alpha,\sigma^2) = \int p(t|\mathbf{w},\sigma^2)p(\mathbf{w}|\mathbf{t},\alpha,\sigma^2)d\mathbf{w} \tag{18}$$

ただし，$\mathbf{t}$ はトレーニングデータのターゲット値からなるベクトルである．また表記を簡単にするため，予測するターゲット値に対応する入力を右辺の条件部分から省いた．式 (18) の右辺の 2 つの確率分布はそれぞれターゲット値の条件付き分布，重

Algorithm 1 Sparse Bayesian Learning for Itemset Data

$\sigma^2 = var[t] \times 0.1$
for all itemsets ϕ_i **do**
  $\alpha_i = \infty$
  $\delta L_i = \|\phi_i^{\mathrm{T}}\mathbf{t}\|^2/\|\phi_i\|^2$
end for
$j = \mathrm{argmax}|\delta \mathrm{L}_i|$
$\alpha_j = \frac{\|\phi_j\|^2}{\|\phi_j^{\mathrm{T}}\mathbf{t}\|^2/\|\phi_j\|^2 - \sigma^2}$
Initialize $\boldsymbol{\Sigma}, \boldsymbol{\mu}$
while **do**
  Compute δL_i for all itemset from (13)
  $j = \mathrm{argmax}|\delta \mathrm{L}_i|$
  $\theta_j \triangleq q_j^2 - s_j$
  if $\theta_i > 0$ and $\alpha_i < \infty$ **then**
    Re-estimate α_j
  else if $\theta_j > 0$ and $\alpha_j = \infty$ **then**
    Add ϕ_j to the model with updated α_j
  else if $\theta_j \leq 0$ and $\alpha_j < \infty$ **then**
    Delete ϕ_j from the model and set $\alpha_j = \infty$
  end if
  Update $\sigma^2 = \|\mathbf{t} - \mathbf{y}\|^2/(N - M + \Sigma_m \alpha_m \Sigma_{mm})$
  Recompute/Update $\boldsymbol{\Sigma}, \boldsymbol{\mu}$
  if $\log \alpha_i < 10^{-6}$ and $\theta_i < 0$ for all basis **then**
    break
  end if
end while

みの事後分布である．ターゲット値の条件付き分布 $p(t|\mathbf{w}, \sigma^2)$ は式 (4) で与えられ，重みの事後分布 $p(\mathbf{w}|\mathbf{t}, \alpha, \sigma^2)$ は式 (6) の多次元ガウス分布で与えられる．これを使って式 (18) は次のようなガウス分布で表される．

$$p(t|\mathbf{x}, \mathbf{t}, \alpha, \sigma^2) = N(t|\boldsymbol{\mu}^{\mathrm{T}}\phi(\mathbf{x}), \beta(\mathbf{x})) \tag{19}$$

ただし，予測分布の分散 $\beta(\mathbf{x})$ は

$$\beta(\mathbf{x}) = \frac{1}{\sigma^2} + \phi(\mathbf{x})^{\mathrm{T}}\boldsymbol{\Sigma}\phi(\mathbf{x}) \tag{20}$$

で与えられる．(20) の第 1 項はデータに含まれるノイズを表し，一方第 2 項は $\mathbf{w}$ に関する不確かさを反映している．ノイズとパラメータ $\boldsymbol{\mu}$ の分布は独立なガウス分布であるからこれらの分散は加法的である．

3.2 枝刈りの定理

Algorithm 1 を適用するとき，手続き Add の候補となる全ての $\alpha_t = \infty$ のアイテムセット t について $\theta = q_t^2 - s_t$ を計算している．この節はアイテムセットの探索空間の木構造を利用した定理を紹介する．

定理 1 (枝刈りの定理) 探索木のノード t に到達したとき，もし次の条件を満たすなら以降のノードには *Add* の候補となるアイテムセットは存在しない．

$$\begin{aligned} &4\sum_j \sum_k \mathbf{M}_{jk} I(t \subseteq T_j) I(t \subseteq T_k) \\ &- 4\sum_{M_j < 0} \mathbf{M}_j I(t \subseteq T_j) + \sum_j \sum_k \mathbf{M}_{jk} \leq 0 \end{aligned} \tag{21}$$

Max pattern	number of nodes
1	21
2	211
3	1351
4	6191
5	21255
6	50605
∞	107607

表 1: 最大アイテムセット数を変化させたときのノードの総数の変化．サンプル数 200，特徴数 20 で固定した．

ここで，$\mathbf{M} = (\mathbf{C}^{-1}tt'\mathbf{C}^{-1} - \mathbf{C}^{-1})$ であり，$\mathbf{M}_j$ は $\mathbf{M}$ の j 列目を表しており，T_j は j 番目のトランザクションを表している．

4. 人工データの実験

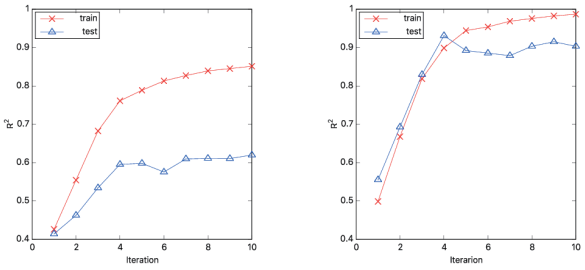

図 1: 単独のアイテム (左)，アイテムセット (右) について，トレーニングデータとテストデータの学習がどのように変化するかを示す学習曲線を書いた．横軸はアルゴリズムの反復回数で縦軸は決定係数 R^2 を示している．

アイテムセットを用いたスパースベイズの効果を確認するために，シミュレーションデータを作成し実験を行なった．データは [TTS18] を参考に次のように作成した．まずベルヌーイ分布 ($q = 0.6$) に従う確率変数を使って $\mathbf{X} \in \{0, 1\}^{n \times p}$ の表現行列を生成した．次に表現行列からランダムに 5 個の特徴を選択し，さらに 2^5 の組み合わせの中からサイズ 2 以上の組み合わせを含むものを 5 個選択する．つまり単独のアイテム 5 個とサイズ 2 以上のアイテムセット 5 個の計 10 個を選択した．選択した特徴にそれぞれ独立な一様分布 $U_{[0,1]}$ から重み $\mathbf{w}$ を与え，応答変数 $\mathbf{y} = \mathbf{X}\mathbf{w}$ を計算した．まず，アルゴリズムの繰り返しによる精度の変化を調査した．サンプル数 100，特徴数 50 のデータをトレーニング用とテスト用にそれぞれ用意して，学習曲線を書いた．精度として以下の式で定義される決定係数 $R^2 = 1 - \sum_i (y_i - f_i)^2 / \sum_i (y_i - \overline{y})^2$ を用いた．図 1 は学習曲線を表している．左が最大アイテムセット数を 1 にした学習で，右が 5 にした場合ある．トレーニングデータの精度は反復を繰り返すにつれてどちらも上昇しており正しく学習が出来ていることが確認された．またアイテムセットを取り入れることによって精度が向上しているのが分かる．次にデータを変化させたときの計算時間の比較を行なった．図 2 は最大アイテムセット数とトレーニングデータのサンプル数を変化させたときの計算時間の変化を表している．特徴数は $p = 20$ とした．3 つの線はアイテム単独で回帰を行なった最大アイテム数

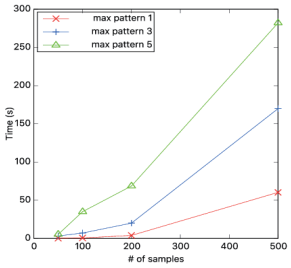

図 2: サンプル数を増やしたときの計算時間の変化．最大アイテムセットサイズ 1，3，5 をそれぞれ指定したときを比較した．特徴の数は 20 で固定した．

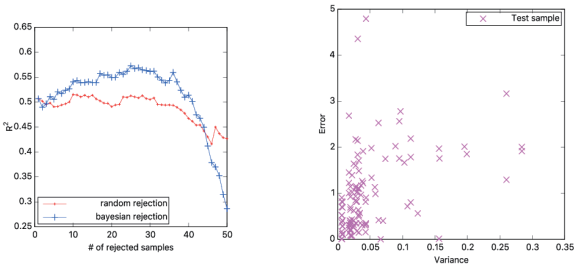

図 3: 棄却による精度の変化 (左) 分散と誤差の相関関係 (右)

1 と最大アイテムセット数 3,5 のサンプル数を変えたときの計算時間の変化である．サンプル数が増え，最大アイテムセット数が増えるほど計算時間が伸びることが確認された．表 1 は特徴数 20 のデータで最大アイテムセット数を変化させた時の探索空間のノードの数を示している．このときトレーニングデータのサンプル数は 200 とした．データの性質やサンプル数によっても変わるが，最大アイテムセット数を増やすことで指数関数的にノードが増加する．

5. 実データの実験

この節では，提案手法を HIV-1 drug dataset [RGK+03] に適用した結果を紹介する．このデータは市販の抗 HIV 薬に対する生体外感受性検査の結果をまとめたものであり，各薬剤に対するウイルスの耐性が野生型のそれと比較して倍率変化 (fold-change) として記録されている．この記録は各ウイルスの薬剤耐性を表す応答値 y と野生型と異なる遺伝子型のアイテムセットからなる．例えば，あるウイルスが参照した配列の 1 番目と 6 番目に突然変異して元のアミノ酸からアルギニンとシステインに変化しているときアイテムセットは $\{1A, 6C\}$ となる．スパースベイズによるテストデータの予測は平均と分散からなる．分散の効果を調べるために次の実験を行なった．サンプル数 630 個のジドブジン (AZT) のデータセットをトレーニング 60%，バリデーション 20%，テスト 20% に分割した．バリデーションデータの精度が最も良い反復回数のモデルをテストデータに適用した．適用して得られた分散が大きいものから棄却した場合 (ベイジアン棄却) とランダムに棄却した場合 (ランダムの棄却) での精度の変化を比較した．図 3 の左はベイジアン棄却とランダムの棄却による精度の変化を 5 回の平均で表している．ランダムの棄却と比較した場合にベイジアン棄却は前半はランダムの棄却よりも精度が上がり，後半は精度が下がっている．これよりベイジアンモデルの分散が予測の正確さと相関関係があると考えることができる．図 3 の右はクロスバリデーションのある 1 回の結果について分散と誤差の相関関係を示した図である．分散は (20) の第 2 項のサンプルによって変化する部分，誤差は予測値と正解の差の絶対値を表している．図より一定の相関関係があり，相関係数は 0.51332 であった．

6. おわりに

アイテムセットに対して，スパースベイズ学習を適用するアルゴリズムを説明した．スパースベイズ学習は，与えたデータの変数の数よりもパラメータの数を少なくし，結果がガウス分布で得られる特徴を持っている．アイテムセットを変数に持たせることで，単独のアイテムのみを変数にするよりも複雑な非線形モデルを作ることが可能になる．また分布を推定することで推定値だけでなく，推定の不確かさが計算でき，LASSO などの点推定よりも多くの情報を得ることができた．ベイジアン棄却の実験を行なって，分散の高いデータを棄却させると精度が上昇することが分かった．

参考文献

[Bis06] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006.

[RGK+03] Soo-Yon Rhee, Matthew J Gonzales, Rami Kantor, Bradley J Betts, Jaideep Ravela, and Robert W Shafer. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic acids research*, 31(1):298–303, 2003.

[TF03] Michael E. Tipping and Anita C. Faul. Fast marginal likelihood maximisation for sparse bayesian models. In *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics, AISTATS 2003, Key West, Florida, USA, January 3-6, 2003*, 2003.

[Tib96] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[TTS18] Mirai Takayanagi, Yasuo Tabei, and Hiroto Saigo. Entire regularization path for sparse nonnegative interaction model. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 1254–1259. IEEE, 2018.

[UAUA03] Takeaki Uno, Tatsuya Asai, Yuzo Uchida, and Hiroki Arimura. LCM: an efficient algorithm for enumerating frequent closed item sets. In *FIMI '03, Frequent Itemset Mining Implementations, Proceedings of the ICDM 2003 Workshop on Frequent Itemset Mining Implementations, 19 December 2003, Melbourne, Florida, USA*, 2003.