G-20

ブランド和牛の肉質予測における地域間共通特徴量の効率的選択手法

Finding Common Features Among Multiple Groups in Sparse Feature Selections: A Case Study with Wagyu Data

東口 奈那美† 本廣 多胤† 池上 春香‡
Nanami Higashiguchi Masatsugu Motohiro Haruka Ikegami
松橋 珠子‡ 松本和也‡ 吉廣卓哉§
Tamako Matsuhashi Kazuya Matsumoto Takuya Yoshihiro

1. はじめに

和牛は日本の高級ブランド牛肉として知られており，食肉には脂肪が交雑しているため，柔らかく口溶けが良いことが特徴である．日本には著名な和牛の生産地が多数あり，それらの地域では高品質な牛肉を生産するために，それぞれが特色のある方法で種雄牛と雌牛を肥育している．これまでは，この地域間での協力と競争により肉牛の肥育方法は改善されてきた．しかし，肉質改善はほとんどが血統に関する統計分析や肥育経験に基づいた方法によって行われてきたため，同様の方法で改善を続けることは，改善速度の面で大きな制約となる．

最近では，ゲノミクスやプロテオミクスにおける包括的な分析方法が開発されている．例えば，現在では，膨大な遺伝子やたんぱく質の発現量を表す発現量プロファイルを安価かつ少ない手間で入手でき，これを用いた分析が行われるようになってきた．具体的には，各牛のサンプルから血中成分値やタンパク質発現量を取得し，これらのデータから肉牛肥育の初期段階で各肉牛の肉質を予測することが考えられる．この方法は，肉質を向上させるために，肉牛を肥育する革新的な方法論の開発を促進する可能性があると考えられる．

この手法における大きな問題は，遺伝子やタンパク質プロファイルの変数（遺伝子やタンパク質の数）が非常に多いため，牛肉の品質を予測するのに最適な変数集合を選択することが難しいことである．第二の問題は，異なる地域の肉牛は地域独特の異なった傾向を持っているため，これらの地域差を考慮した分析方法が必要になることである．

第 1 の問題に関しては，最近ではスパース分析が利用されており，少ない計算時間で最適に近い特徴変数を選択できる．特に重回帰分析を行う場合は，LASSO（Least Absolute Shrinkage and Selection Operator）[1]がよく利用される．LASSO は，MSE（Mean Squared Error）を線形重回帰式の形で最小化する．LASSO は重回帰式に L1 ノルムを加えることにより，ほとんどの係数をゼロに収束させることができ，元データの次元数が非常に多い場合でも，実行可能時間内に最適に近い少数の変数集合を選択することができる．しかし，和牛データを分析するにあたっては，LASSO はブランド地域の傾向を考慮した特徴選択ができないという問題点がある．

最適な変数選択を行う際に，複数の目的関数を考慮できる Multi-task-LASSO [2]が提案されている．Multi-task-LASSO は，ペナルティ項として L1 / L2 ノルムを適用することで，複数の目的関数を説明する変数集合を選択する．つまり，各ブランド和牛地域の肉質を目的関数とすることで，すべての対象地域の傾向を説明する地域間共通の変数を選択することができ，Wagyu ブランドの各地域の傾向を説明できる．しかし，LASSO と Multi-task-LASSO のいずれにおいても，選択された変数が常に最適であるとは限らないため，和牛の目標形質を高精度に説明する最適な変数集合を選択することは難しいことが知られている[3]．複数の地域の傾向を同時に説明する十分に最適な変数集合を選択するためには，複数の和牛ブランド地域を考慮しながら最適な変数集合を選択する手法が必要である．

本論文では，この問題に対する解決策を提案する．すなわち，実現可能な計算時間内に複数の和牛ブランド地域間で共通に作用する最適な説明変数の集合を選ぶ変数選択法を提案する．各タンパク質と形質の相関係数，およびそれらの値のばらつきの小ささを測る fairness-index を利用して，膨大にあった変数から少数の変数を説明変数候補として取り出す．次に，重回帰分析で候補変数のすべての組合せを調べることによって，十分に最適な変数集合を取得する．

この論文は以下のように構成される．第 2 節では，ブランド和牛について説明する．第 3 節では，本研究の前提知識として，LASSO と Multi-task-LASSO を説明する．第 4 節で提案手法を示し，第 5 節で評価結果を示す．最後に，第 6 節でまとめる．

2. ブランド和牛

黒毛和牛は，優れた遺伝的資質を持つ肉牛であり，神戸牛や松阪牛など，日本では複数の地域でブランド和牛として肥育されている．ブランド和牛の肥育方法はそれぞれの地域で異なるため，地域毎に肉牛の特色が異なる．生産された肉牛がブランド和牛として認定されるかどうかは，各地域で独自の基準に基づいて判断されるが，主に牛の肥育地，肥育方法，枝肉成績等が用いられる．この中で最も重要な項目が枝肉成績であり，これは屠殺時に出荷される牛肉の品質を評価した成績である．枝肉成績は複数の測定項目から構成されるが，中でも主要 6 形質と呼ばれる 6 項目は重要視されている[4]．その 6 項目は，CW（枝肉重量），BMS（牛脂肪交雑基準），YE（歩留り等級），RT（バラの厚さ），SFT（皮下脂肪厚），および REA（ロース芯面積）である．一般的には，これら主要 6 形質を基準として，

†和歌山大学大学院システム工学研究科，Graduate School of Systems Engineering, Wakayama University
‡近畿大学大学院生物理工学研究科，Graduate School of Biology-Oriented Science and Technology, Kinki University
§和歌山大学システム工学部，Faculty of Systems Engineering, Wakayama University

市場における牛肉の経済的価値が決定されるため，和牛を肥育する農家は，高品質な牛肉を生産するために主要 6 形質を向上する努力を重ねてきた．

和牛農家は質の高い牛肉を安定的に生産するために様々な方法を用いている．現在，最も重要視されている方法は，主要 6 形質の成績が良くなるように血統を制御する育種改良である．牛の血統は主要 6 形質と密接な関係があることが知られており，遺伝的に優秀な牛同士を交配し，それらの中からより優秀な牛を選別して子孫を残すことで，効率的に和牛の遺伝能力を改良できる．各ブランド和牛の生産地域では通常，優れた遺伝的能力を有する種雄牛と呼ばれる牛を肥育している[5] [6]．種雄牛から精子を取り，それらを凍結し，農家に売ることにより，優れた種雄牛の遺伝子が各農家に配布され，優れた血統から何千もの子牛を生み出すことができる．和牛においては，各肉牛の父親を「一代祖」と呼び，肉牛の父親の能力は主要 6 形質を予測するための最も重要な指標であるとみなされている．

先祖にあたる肉牛の枝肉成績を用いて，ある肉牛の主要 6 形質を予測するための統計的指標として，育種価が広く知られている．育種価は各主要 6 形質のそれぞれについて計算され，集団内の平均的形質値と各牛の相対的形質値の優劣を表す．従って，育種価は各形質に対するその牛の能力を表す．育種価には，推定育種価と期待育種価の 2 種類がある．前者は，屠殺時の主要 6 形質の成績を保持する子孫を持つ種雄牛に対して計算され，その牛の主要 6 形質に関する推定能力値を表す．後者は，育種価を推定するのに十分な数の子孫を持たない各肉牛に対して計算される，屠殺後の肉質を推定するために用いられる．育種価の計算に利用される血統モデルは複数種類ある．現在，最も頻繁に利用されているモデルはアニマルモデルであり，これは同じ母親を持つ牛の兄弟を含むすべての遺伝的な関係を考慮している．BLUP 法は血統に基づいて育種価を計算する統計的手法であり，血統により受け継がれた遺伝的能力として育種価を計算する[7].

一方，高品質な肉牛を安定的に生産するための肥育方法も研究されてきた．しかし，その大半が農家の経験に依存しており，科学的な知見や実際のデータに基づいていない．例えば，和牛の農家は，高価値な肉牛を育てるための方法を長年の経験から学び，ノウハウとして蓄積してきた．これらのノウハウは，牛の肥育方法などの直接的な知識から，牛舎の構造などの間接的な知識までを指し，経験に基づいているゆえに，ほかの農家や後継者への伝承が困難であり，高品質な和牛を効率的に生産するには至っていない．従って，科学的根拠に基づいた効率的な肥育方法の確立が求められている．

主要 6 形質を向上させる肥育方法に関する研究として，いくつかの文献がある．例えば，ビタミン A の濃度を調節することにより BMS の値を改善できる可能性を示したものある[8]．しかし現状では，肥育農家がそれを活用し，優れた高品質な肉牛を生産する水準には至っていない．

3. LASSO と Multi-task-LASSO

LASSO[1]は，線形回帰モデルに基づき，多数の相関の高い特徴から少数の特徴を選択するための手法としてよく知られている．与えられたサンプルの集合をS，特徴の集合をFとし，$x_{sf}(s=1,\ 2,\ \ldots,\ |S|,\ f=1,\ 2,\ \ldots,\ |F|)$はサンプル$s$の特徴$f$を測定した特徴量とする．ここで，$|S|$および$|F|$は，$S$および$F$の要素数である．以下，簡潔にするために$|S|$および$|F|$を$S$および$F$と書くことがある．$x_f = (x_{1f},\ x_{2f},\ \ldots,\ x_{Sf})^T$を各特徴$f \in F$について測定されたベクトルとし，$X = [x_1,\ x_2,\ \ldots,\ x_F]$は特徴量全体を表す行列とする．各サンプル$s$について測定された形質値を$y_s$，形質ベクトルを$\mathrm{y} = (y_1,\ y_2,\ \ldots y_S)$とする．すると，LASSO は次のように定式化される．

$$\hat{B} = arg\min_B \left(\left\| y - \sum_{f \in F} \beta_f X \right\| + \lambda \sum_{f \in F} |\beta_f| \right) \quad (1)$$

ここで，λは非負の正則化パラメータで$B = (\beta_1,\ \beta_2,\ \ldots \beta_F)^T$は，$X$の係数ベクトルである．λはいわゆるL1ノルムペナルティの係数で，L1ノルムの大きさ，すなわちペナルティの重さを表す．ペナルティ項により，最適解を計算する間にβ_fの大部分はゼロに収束する．結果として，少数の非ゼロ係数だけが残り，LASSOが多数の特徴変数からの特徴選択を実現したことを意味する．

Multi-task-LASSO [2]はLASSOを拡張したものであり，複数の目的関数を扱う．タスクの集合すなわち，目的関数の集合を Tとする．さらに$x_{sf}^{(t)}$をタスク$t \in T$の測定特徴量とする特徴行列とし，また$y_s^{(t)}$をタスク$t \in T$の測定形質値とする特徴ベクトルを定義する．同様に，$x_f^{(t)}$，$X^{(t)}$，$y^{(t)}$と書く．ただし各タスクのサンプル数$S^{(t)}$は異なる可能性がある．このとき，Multi-task-LASSOは次のように表現される．

$$\widehat{W} = \arg\min_W \left(\sum_{t \in T} \left\| y^{(t)} - \sum_{f \in F} \beta_f^{(t)} X^{(t)} \right\| + \lambda \sum_{f \in F} \|w_f\| \right) \quad (2)$$

ここで，Wはすべての係数ベクトルを組み合わせた係数行列で$W = (B^{(1)},\ B^{(2)},\ \ldots B^{(T)})$と表し，$w_f$は各タスク内のL1ノルムで，$w_f = \sum_{t \in T} \left| \beta_f^{(t)} \right|$として定義される．正則化項は，図1に示すように，L1ノルムとL2ノルムの組合せである．Multi-task-LASSOでは係数$\beta_f^{(t)}$は各$f \in F$と$t \in T$に対して定義される．Multi-task-LASSOはL1/L2正則化の組み合わせを利用しており，最初に同じ特徴の係数のL1ノルムが計算されることにより，w_fを形成する．次にそれらw_fのL2ノルムを正規化項で利用することにより，最初にすべてのタスクに対して共通して有効な機能を選択し，次に各タスク内で選択した機能の係数を最適化することができる．

我々の問題にMulti-task-LASSOを適用することができると考えられる．ブランド和牛の各地域tにおいて，別個のサンプル，すなわち各地域で牛が成長した結果，各地域tについて特徴集合$X^{(t)}$を測定したと考えられる．形質に関しては，BMSなどの同じ形質を適用するが，各地域には独自の牛が存在するため，データは$y^{(t)}$として表される．上記の$X^{(t)}$と$y^{(t)}$を用いてMulti-task-LASSOを解くことにより，Multi-task-LASSOの枠内で複数の地域間における共通に有効な特徴集合を得ることができる．しかし問題は，Multi-task-LASSOは最適性が欠けているため，最適でない特徴集合が頻繁に選択されることである．本論文では，$y^{(t)}$と$x_f^{(t)}$の相関係数とfairness-indexを利用して最適性を改善することを試みる．

$$W = \begin{bmatrix} \beta_1^{(1)} & \cdots & \beta_F^{(1)} \\ \vdots & & \vdots \\ \beta_1^{(T)} & \cdots & \beta_F^{(T)} \end{bmatrix}$$

$$\Downarrow$$

$$(\|w_1\|_2 \ldots \|w_F\|_2)$$

図1:Multi-task-LASSOのL1/L2正則化の図

4. 提案手法

提案手法では，複数の測定された集合$X^{(t)}$，$t \in T$の間で一般的に有効な特徴の集合をFから選択することを目指す．LASSOは重回帰分析に基づいているので，集合ごとの同等に高い重相関係数を導くような特徴集合を得る．これに対して我々の戦略は，重相関係数ではなく，各特徴と形質との間における相関係数を利用する．これは一組の特徴における重相関係数が高い場合，各特徴も高い単相関係数を有するという傾向があることを前提にしている．したがって，数百または数千の特徴から少数の特徴を選択するために，我々は最初に各特徴の単相関係数を計算し，この値を用いて事前の選抜を行う．事前選抜によって数十の特徴に絞り込んだ後，我々は全ての組み合わせをテストし，全組合せから最も高い重相関係数を持つ特徴の最良の組み合わせを見つける．

和牛ブランドの各地域を$t \in T$，特徴測定データセットを$X^{(t)}$，測定形質データセットを$y^{(t)}$と表す．提案した方法の計算ステップは以下の通りである．

1. 和牛ブランドの各地域において，特徴それぞれに対して，形質との相関係数を計算する．
2. ステップ1で計算した地域ごとの相関係数を用いて地域間で相関係数に差があるかどうかを調べるためにfairness-indexを計算する．
3. 各特徴と形質の相関係数を，すべての地域のサンプルを用いて計算する．
4. 計算されたfairness-indexがある閾値Jを下回る場合はFからその特徴を取り除き，部分集合$F' \subset F$を取得する．
5. $|F'|$が事前に定義した数Nより大きい場合は，次のように選択する．F'個の特徴の中からステップ3で計算した相関係数の高い順に，N個の特徴を選択し，部分集合$F'' \subseteq F'$を得る．
6. 最終的に重回帰分析を行う際の重回帰項の数をM個と決める．F''内の特徴をM個に絞り、M個の特徴の全ての組合せに対して重相関係数を計算し，最も高い値を有する組合せを調べる．

まず，本稿で使用する変数の定義を表1に示す．ステップ1では，各特徴と地域ごとにおける目的形質の相関係数を計算する．具体的には，各$f \in F$および$t \in T$について$x_f^{(t)}$と$y^{(t)}$の相関係数を計算し，これを$C_f^{(t)}$と表す．

ステップ2は，計算された複数の地域における相関係数の公平性を調べる．ここでは，Jain's fairness-index[9]を利用して公平性を測定する．この値はすべての値が同じ場合は1を取り，最悪の場合は，n^{-1}を取る．ここでnは変数の数である．地域間の相関係数のfairness-indexをとるので，fairness-indexは次のように定義できる．

$$J_f = J\left(C_f^{(1)},\ C_f^{(2)},\ \ldots,\ C_f^{(T)}\right) = \frac{\left(\sum_{t \in T} C_f^{(t)}\right)^2}{n \sum_{t \in T} \left(C_f^{(t)}\right)^2} \qquad (3)$$

このfairness-indexは各$f \in F$について計算される．

ステップ3では，対応する形質値を有する全地域のサンプルにおいて各特徴と各形質の相関係数を単純に計算する．ベクトル$x_f^{(all)} = \left(x_f^{(1)},\ x_f^{(2)},\ \ldots,\ x_f^{(T)}\right)$，$y^{(all)} = \left(y^{(1)},\ y^{(2)},\ \ldots,\ y^{(T)}\right)$と定義すると，$x_f^{(all)}$と$y^{(all)}$の相関係数を計算し，$C_f^{(all)}$を求めることができる．$C_f^{(all)}$は$f \in F$ごとに計算する．

ステップ4では，fairness-index J_fに閾値Jを適用し，部分集合$F' \subset F$を得る．閾値Jは事前に設定されており，閾値JよりJ_fが大きい値であるfの集合F'を作成する．fairness-indexを利用することによって，T内の全ての地域について公平な相関値を有する特徴を選択することを意図している．

ステップ5では，$C_f^{(all)}$に関してF'の上位N個の特徴を選択する．選択した新しい集合をF''とする．これは，ステップ6の計算負荷を，Nで表される実行可能な水準に減らすことを意図している．最終的な特徴集合の候補となるためには，重相関係数と公平性が高くなければならない．そこで本稿では，まず公平性に閾値を適用し，次に重相関係数を利用して特徴数を制限することを提案する．

ステップ6では，最後にF''から特徴の最適な組み合わせを選択する．F''内のM個の特徴のあらゆる組み合わせに対して重相関係数を計算し，最良のものを出力する．これは，最良の重相関係数を持つM個の特徴の組合せを求めることを意味する．

表1:変数の定義

記号	説明
F	入力データの特徴fの集合
T	和牛ブランド地域tの集合
$x_f^{(t)}$	地域tで特徴fを測定した特徴量
$y^{(t)}$	地域tで測定した形質値
λ	LASSOにおけるペナルティ項の係数
$\beta_f^{(t)}$	地域tで特徴fを測定した$x_f^{(t)}$の係数
w_f	地域tに関しての$\beta_f^{(t)}$のL1ノルム
$C_f^{(t)}$	$x_f^{(t)}$と$y_f^{(t)}$の相関係数
$C_f^{(all)}$	全サンプルで特徴と形質の相関係数を計算した値
J_f	特徴fにおけるJain's fairness-indexを算出した値
J	Jain's fairness-indexの閾値
N	ステップ5で選択した特徴の数
M	最終的に選択した特徴の数

表2:評価結果

主要6形質	手法	重相関係数				Fairness-index
		全地域	地域A	地域B	地域C	
CW(枝肉重量)	提案手法	0．4489	0．5242	0．4956	0．5370	0．9989
	MT-LASSO	0．3692	0．4104	0．6140	0．5706	0．9736
REA(ロース芯面積)	提案手法	0．4596	0．2607	0．6739	0．5785	0．8907
	MT-LASSO	0．4233	0．4567	0．7330	0．3847	0．9243
RT(バラの厚さ)	提案手法	0．5558	0．5835	0．8145	0．5138	0．9609
	MT-LASSO	0．4913	0．5191	0．7978	0．4389	0．9354
SFT(皮下脂肪厚)	提案手法	0．4691	0．4081	0．3861	0．6147	0．9541
	MT-LASSO	0．4150	0．3102	0．3884	0．5843	0．9322
YE(歩留り等級)	提案手法	0．4487	0．3793	0．5202	0．4802	0．9837
	MT-LASSO	0．3639	0．4309	0．6276	0．3082	0．9230
BMS(牛脂肪交雑基準)	提案手法	0．4951	0．5350	0．7020	0．3990	0．9509
	MT-LASSO	0．4847	0．6104	0．7699	0．2619	0．8694

5. 評価

5．1　評価データ

提案手法をMulti-task-LASSOの結果と比較して評価する．データは，地域A，B，およびCを指す3つのブランドの和牛の地域からなり，各地域は，データ中に51，10，および35の肉牛，すなわちサンプルを有する．各肉牛はいずれかの地域で育ち，屠殺前に主要6形質が測定され，その後牛肉として売られる．主要6形質とはCW（枝肉重量），REA（ロース芯面積），RT（バラの厚さ），SFT（皮下脂肪厚），YE（歩留り等級），およびBMS（牛脂肪交雑基準）である．結果として，データは3つの地域からなり，各サンプルに対して6つの形質値を持つ．

特徴データセットは，血清プロテオーム発現プロファイルである．各肉牛について，血清を3〜4ヶ月の間隔で採取し，それを独自の前処理法を用いてSWATH-MS[10]で分析する．SWATH-MSとは高精度で網羅的にタンパク質を定量する方法であり，サンプル中にどのタンパク質がどれだけ含まれているかがわかる．この方法を利用し，各サンプルについて137個のタンパク質の発現量を得た．その結果，6時期で137個のタンパク質発現値が得られたので，3つの地域からの各サンプルについて137×6=822の特徴があることになる．そこから欠損値を持つ特徴を取り除いた後，我々は提案された方法を適用する580の特徴を得た．

5．2　評価方法

提案手法とMulti-task-LASSO（MT-LASSO）を上記のデータセットに適用した．LASSOは重回帰分析における最適化に基づいているため，我々の評価基準は選択された特徴を用いて計算された重相関係数が最適であると考える．提案手法と比較するのに最適な手法はMT-LASSOである．公平性の観点から最適な特徴集合を取得しようとしているので，各地域の相関係数を計算し，Jain’s fairness-indexを利用してそれらの変動を調べる[9]．fairness-indexが高い値，すなわち1に近い値をとる場合，選択された特徴集合はすべての地域で形質との相関があるとみなされ，特徴集合の説明能力は地域に特有ではない，すなわち和牛の肉質に一般的に効果的な要素であることを意味する．これは本稿の我々の研究の目的を満たしている．

提案手法では，$M = 3$，$N = 50$，$J = 0.8$とする．すなわち，和牛の各対象形質を説明するために，最良の3つの特徴を選択する．F''の要素数が50を超える場合はF''の要素数を50に制限する．$M = 3$，$N = 50$の場合，50の特徴候補から3つの特徴を選び，その全組合せに対して重回帰分析を実行するときの計算時間は約1時間程度であり，実行可能範囲内である．なお，本評価はCPUが1.0GHz，メモリー容量が4.00GBの計算機で実施した．また，MT-LASSOでは，各形質に対して特徴が3つ残るようにパラメータを設定した．

提案手法は特別なアルゴリズムを用いず，一般的な統計処理で行った．MT-LASSOはPythonのライブラリscikit-learn [11]に含まれているが，このライブラリの入力形式が今回の和牛データに当てはまらないため，適用できなかった．MT-LASSOにおいてすべてのグループに共通な特徴を選択するのはペナルティにL1ノルムを用いて正則化する部分であるため，この部分はLASSOの機能と同じであると考えられる．ゆえに，今回はPythonのscikit-learn [11]で実装されているLASSOを利用して特徴選択を行った．

5．3　評価結果

結果を表2に示す．表2では，提案した方法とMT-LASSOの性能を6つの主要6形質について比較している．MT-LASSOの結果はMT-LASSOによって直接得られるものではなく，特徴を選択した後，重回帰分析を行い，得られた重相関係数を示した．

最初に，表2の全地域の重相関係数を見ると，6つの形質すべてにおいて，MT-LASSOよりも提案手法の方が高いことがわかる．これは提案手法が最適化性能の点でMT-LASSOより優れていることを意味する．さらにこれは，ステップ5で特徴を$N = 50$に絞ったときに，この中に優れた特徴が含まれていることを意味する．

次に，表2のfairness-indexを見ると，主要6形質のうちの5つにおいて，MT-LASSOよりも提案手法の方が高いことがわかる．すなわち，MT-LASSOよりも提案手法のほうが，地域毎の重相関係数のばらつきが小さい．これは，提案手法で選択した特徴量が，3つの地域において同程度の説明能力があることを意味する．本研究の目的は地域に関係なく，和牛全体に共通する特徴を選択することであるため，MT-LASSOより提案手法の方が本研究の目的に適している．以上より，提案手法は選択された特徴の最適性だけでなく，異なる地域間での効果の共通性を表せる点でもMT-LASSOよりも優れていると結論付けられる．

6. おわりに

本論文では，膨大な特徴変数から，実現可能な計算時間内で，目的関数に対して複数のグループに共通に作用する最適な特徴集合を選択する手法を提案した．Multi-task-LASSOはこの問題に取り組むために提案された手法であるが，最適な特徴集合を常に選択せず，各グループに対する効果の公平性を考慮しないという問題がある．

この問題を解決するために，我々は前もって候補となる特徴を選択し，次に特徴の候補全てに対してどの組合せが最適でなおかつ公平性のある特徴集合かを調べて選択する方法を提案した．我々の提案手法は，最初に各特徴とグループごとにおける形質との相関係数を計算する．次に，計算した相関係数を利用して，グループ間の公平性と全グループを対象とした相関係数を計算する．これにより高い公平性と平均性の両方を持つ特徴を選択することができ，目的関数にたいして，すべてのグループに一般的に影響する候補特徴を得る．複数の地域からなるブランド和牛のプロテオームプロファイルデータセットを提案手法に適用すると，提案手法は最適性と複数地域間の公平性の両面でMulti-task-LASSOよりも優れていることがわかった．

和牛育種の観点からは，肉牛の育種方法などの固有に影響する要素，すなわち地域的影響を排除することにより，牛肉の品質に真に関連するタンパク質の集合を見つけることを試みる．提案手法を用いて，我々は従来のMulti-task-LASSO方式よりも効率的に主要6形質に共通に関連するタンパク質を選択することが可能であろう．

今後の研究としては，評価を追加し，提案手法の優秀さについてより多くの根拠を集める．また，最適な解を得るために，より洗練された手法が必要になる可能性があるため，Nの値が小さく，Mの値が大きい場合の提案手法の適用を検討する必要がある．

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] R. Tibshirani, Regression Shrinkage and Selection via the Lasso, Journal of the Royal Statistical Society, Series *B*, 58(1), pp.267–288 (1996).
[2] G. Obozinski, B. Taskar, and M. Jordan., Multi-task feature selection, In Technical Report, Department of Statistics, University of California, Berkeley, 2006.
[3] S. Hara and T. Maehara, “Enumerate Lasso Solutions for Feature Selection,” In Proc. AAAI2017, 2017.
[4] Japan Meat Grading Association, “The Manual of Standard in Dealing Pork and Beef,” 2001 (In Japanese).
[5] A Guide of Japanese Black Cattle Sires, http://liaj.lin.gr.jp/index.php/detail/data/m/803237096 (referred in May 2017) (In Japanese).
[6] Wagyu Registry Association, “Compliation of Sires of Japanese Black Cattle,” 2003 (In Japanese).
[7] N. D. Cameron, “Selection Indices and Prediction of Genetic Merit in Animal Breeding,” CAB International (1997).
[8] Oka, A., Dohgo, T., Juen, M., and Saito, T., “Effects of vitamin A on beef quality, weight gain, and serum concentration of thyroid hormones, insulin-like growth factor-I, and insulin in Japanese black steers,” Animal Science and Technology (1998).
[9] R. Jain, D.M. Chiu, W. Hawe, “A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems,” DEC Research Report TR-301, 1984.
[10] C. Ludwig, L. Gillet, G. Rosenberger, S. Amon, B.C. Collins, R. Aebersold, “Data- independent acquisition-based SWATH- MS for quantitative proteomics: a tutorial,” Molecular Systems Biology (2018) 14, e8126, DOI 10.15252/msb.20178126, 2018.
[11] Scikit-learn, https://scikit-learn.org/