事例研究［レター］

混合整数最適化による相続工程の長期化リスク採点システム

椎名　萌，高野　祐一，宇佐美　朋香，山西　康孝，藤巻　米隆

1. はじめに

死亡した人（被相続人）に対し，その財産を一定の身分関係にある人（相続人）に承継させる制度を相続という．相続の権利や遺産分割の基準については民法（第5編 相続）で規定され，被相続人の遺言によって共同相続人の相続分を定めることができる．遺言は遺留分の規定（民法 第1028条）に違反しない限り，民法による法定相続分よりも優先されるが，遺言の内容は相続人全員の同意があれば変更が認められている [1].

相続では，遺言内容が不平等，相続人同士の関係が希薄，財産が不動産のみで分割できない，などのさまざまな要因によって，一定の割合で紛争が発生する．司法統計年報（家事事件編）によれば，少子高齢化を背景に遺産分割事件数は年々増加している．今後訪れる多死社会を悲しみだけで終わらせないために，相続紛争の原因・傾向を分析し，予防・回避することは重要である．

本論文では，紛争が生じるリスクを事前に予測するための手法として，採点システム [2] に着目する．この手法では，質問への回答結果から簡単な演算で予測を行うことが可能であり，専門的知識がなくても計算機を使用せずにリスクを算出できる方法として，医療や刑事司法の分野で利用されている [2]．線形回帰モデルを用いて採点システムを作成する場合には，説明変数（質問項目）を少なく，かつ偏回帰係数（得点）を小さな整数とすることが要求される．

しいな　もえ
筑波大学大学院システム情報工学研究群
たかの　ゆういち
筑波大学システム情報系
〒305–8573 茨城県つくば市天王台 1–1–1
うさみ　ともか，やまにし　やすたか，ふじまき　よねたか
株式会社ルリアン
〒604–8151 京都府京都市中京区蛸薬師通烏丸西入ル橋弁慶町 227 第12長谷ビル9階
受付 23.8.30　採択 24.1.17

一方で，既存の採点システムの多くは経験に基づいて手作業で作成されている [3]．また，一般的な推定手法を用いて実数の偏回帰係数を求めることは可能だが，実数の偏回帰係数を用いると採点システムの解釈性が低下する [4]，偏回帰係数を整数に丸めると採点システムの最適性が損なわれる [3]，などの問題点がある．

近年は，混合整数最適化を用いた変数選択の研究が増えつつあり [5, 6]，採点システムの作成にも利用されている．Ustun and Rudin [3] は2値分類問題を対象として，偏回帰係数の整数制約の下で3種類の目的関数（誤分類数，説明変数の数，偏回帰係数の絶対値）の加重和を最小化する最適化モデルを提案している．さらに Ustun and Rudin [7] は上記の手法をロジスティック回帰に応用し，リスクの確率を算出できる最適化モデルを提案している．しかし，ロジスティック回帰では損失関数が非線形となるため，最適化ソルバーを用いて直接求解することは難しい．この難点を解消するために，区分線形近似 [8–10] や切除平面法 [7, 11, 12] などの方法が提案されている．

本論文では，ロジスティック回帰に基づく混合整数最適化モデル [7] と区分線形近似 [8] を組み合わせて，相続紛争の予防・回避のためのリスク採点システムを作成する．先行研究 [3, 7, 13] では目的関数の加重和によって説明変数の数が最小化されているが，質的変数を複数のダミー変数に変換した場合のように，複数の説明変数が一つの質問項目に相当する場合には質問項目の数を制御できていない．そこで，本論文ではグループ変数選択 [14–16] を導入し，質問項目の数を制限して採点システムを作成する．

本論文では，相続手続きに関して顧客と専門家をつなぐプラットフォーム事業を展開している株式会社ルリアン[1]から提供されたデータセットを利用する．ただし，相続紛争の発生を記録したデータの入手は困難で

1 https://le-lien.co.jp/

あるため，本論文では紛争発生の代理指標として遺産分割協議の長期化を目的変数として予測する．

数値実験の結果，ロジスティック回帰モデルの偏回帰係数に整数変換処理を適用した場合と比較して，提案手法は優れた予測精度を示した．また，ランダムフォレストなどの非線形機械学習手法と比較して，提案手法の予測精度の低下率は小さいことを検証した．さらに，提案手法で作成した採点システムについて，質問項目に対する選択肢や得点の妥当性を考察した．

2. 提案手法

本節では，リスク採点システムの基礎となるロジスティック回帰モデルについて説明し，区分線形近似とグループ変数選択を用いた混合整数線形最適化モデルを提示する．本論文では，連続する正整数の集合を $[n] := \{1, 2, \ldots, n\}$ と表記する．

2.1 ロジスティック回帰モデル

n 個の事例で構成される標本 $\{(y_i, \boldsymbol{x}_i) \mid i \in [n]\}$ が与えられ，$y_i \in \{+1, -1\}$ は事例 i の目的変数（2 値ラベル）のデータ，$\boldsymbol{x}_i := (x_{ij})_{j \in [p]} \in \mathbb{R}^p$ は事例 i の p 個の説明変数のデータとする．ロジスティック回帰モデルでは，目的変数 $y_i \in \{+1, -1\}$ の生起確率を以下のように記述する．

$$\Pr(y_i \mid \boldsymbol{x}_i) := \frac{1}{1 + \exp(-y_i(\boldsymbol{\beta}^\top \boldsymbol{x}_i + \beta_0))}$$

ここで，$\beta_0 \in \mathbb{R}$ は切片，$\boldsymbol{\beta} := (\beta_j)_{j \in [p]} \in \mathbb{R}^p$ は偏回帰係数である．

ロジスティック回帰モデルの対数尤度（標本生起確率の対数）は以下のように書ける．

$$\begin{aligned} &\log \left(\prod_{i=1}^{n} \Pr(y_i \mid \boldsymbol{x}_i) \right) \\ &= -\sum_{i=1}^{n} \log(1 + \exp(-y_i(\boldsymbol{\beta}^\top \boldsymbol{x}_i + \beta_0))) \\ &= -\sum_{i=1}^{n} f(y_i(\boldsymbol{\beta}^\top \boldsymbol{x}_i + \beta_0)) \end{aligned} \tag{1}$$

ただし，

$$f(v) := \log(1 + \exp(-v)) \tag{2}$$

はロジスティック損失関数とよばれ，その 2 階微分は常に正であるため非線形で滑らかな凸関数である．

2.2 区分線形近似

ロジスティック損失関数 (2) の区分線形近似に用いる接点集合を $\{(v_k, f(v_k)) \mid k \in [m]\}$ とすると，接点における $f(v)$ の接線

$$g_k(v) := f'(v_k)(v - v_k) + f(v_k)$$

は，接点の近傍で $f(v)$ を近似している．また，$f(v)$ は凸関数であるため，以下の不等式が成り立つ．

$$f(v) \geq g_k(v) \quad (k \in [m])$$

以上より，各点 v で最大値を取る接線によって構成される区分線形関数を用いて，$f(v)$ は以下のように近似できる．

$$\begin{aligned} f(v) &\approx \max\{g_k(v) \mid k \in [m]\} \\ &= \min\{t \mid t \geq g_k(v) \quad (k \in [m])\} \end{aligned} \tag{3}$$

文献 [8, 10] では，区分線形近似において有効な接点を選択するためのアルゴリズムが提案されている．

2.3 グループ変数選択

質問項目 $s \in [q]$ に関連するグループ変数（説明変数の集合）を $J_s \subseteq [p]$ とする．たとえば，被相続人の不動産財産額について，「300 万円未満」「300 万円〜1,000 万円未満」「1,000 万円以上」の 3 種類の説明変数（ダミー変数）を定義する場合には，これらの説明変数をまとめてグループ変数「不動産財産額」とする．不動産財産額を質問すれば，これらの説明変数にまとめて回答を得ることができる．

グループ変数の選択を表す 0–1 決定変数 $\boldsymbol{z} := (z_s)_{s \in [q]} \in \{0, 1\}^q$ を以下のように定義する．

$$z_s := \begin{cases} 1 & \text{（グループ変数 } J_s \text{を選択）} \\ 0 & \text{（それ以外）} \end{cases} \quad (s \in [q])$$

あるグループ変数を選択しない場合には，そのグループ変数に含まれる説明変数の偏回帰係数をすべてゼロにする．したがって，グループ変数の選択は以下のように記述できる．

$$\begin{aligned} &z_s = 0 \;\Rightarrow\; \beta_j = 0 \quad (j \in J_s,\ s \in [q]) \\ &\sum_{s=1}^{q} z_s \leq \theta \end{aligned}$$

ただし，$\theta \in [q]$ は選択するグループ変数（質問項目）の数を指定するパラメータとする．

2.4 混合整数線形最適化問題

区分線形近似に用いる決定変数を $\boldsymbol{t} := (t_i)_{i \in [n]} \in \mathbb{R}^n$ とし，偏回帰係数の絶対値を求めるための非負の決定変数を $\boldsymbol{\beta}^+ := (\beta_j^+)_{j \in [p]} \in \mathbb{R}_+^p$，$\boldsymbol{\beta}^- := (\beta_j^-)_{j \in [p]} \in \mathbb{R}_+^p$ とする．以下の混合整数線形最適化問題を解いて，採点システムを作成する．

表 1　説明変数一覧

グループ変数	質問項目	選択肢（ダミー変数）
都道府県	被相続人の住所（都道府県）	「小人口」「中人口」「大人口」
性別	被相続人の性別	「男性／女性」
年代	被相続人の年齢	「70 歳未満」「70 代」「80 代」「90 代」「100 歳以上」
相続人数	相続人に該当する人数	「1 人」「2 人」「3 人」「4 人以上」
子ども	被相続人に子どもがいるか	「いる／いない」
配偶者	被相続人に配偶者がいるか	「いる／いない」
単身配偶者	被相続人の配偶者が単身生活か	「単身／同居」
不動産財産額	相続される不動産財産の現金換算額	「300 万円未満」「300 万円〜1,000 万円未満」「1,000 万円以上」
金融財産額	相続される金融財産の現金換算額	「300 万円未満」「300 万円〜1,000 万円未満」「1,000 万円以上」
その他財産	相続されるその他財産があるか	「ある／ない」
マイナス分財産	相続されるマイナス分財産があるか	「ある／ない」
相続財産合計額	相続される財産の現金換算額合計	「1,000 万円未満」「1,000 万円〜2,500 万円未満」「2,500 万円以上」
相続人年代	相続人の平均年齢	「50 歳未満」「50 代」「60 代」「70 代」「70 歳以上」
同居	被相続人に同居家族がいたか	「いた／いない」
養子	相続人に養子が含まれるか	「含まれる／含まれない」
相続順位構成	各順位の相続人が存在するか	「第 1 順位」「第 2 順位」「兄妹」「甥姪」
財産構成	相続される財産の構成	「不動産のみ」「金融のみ」「両方」
最近相続人	被相続人と最も近い相続人の住所	「同市区町村」「同都道府県」「同地域」「別地域」
最遠相続人	被相続人と最も遠い相続人の住所	「同市区町村」「同都道府県」「同地域」「別地域」
関係地域数	被相続人と相続人の住所の地域数	「1 地域」「2 地域」「3 地域以上」
海外相続人	相続人に海外在住者が含まれるか	「含まれる／含まれない」

$$\min \quad \frac{1}{n}\sum_{i=1}^{n} t_i + \varepsilon \sum_{j=1}^{p}(\beta_j^+ + \beta_j^-) \tag{4}$$

$$\text{s.t.} \quad t_i \geq g_k(y_i(\boldsymbol{\beta}^\top \boldsymbol{x}_i + \beta_0)) \quad (i \in [n],\ k \in [m]) \tag{5}$$

$$-Mz_s \leq \beta_j \leq Mz_s \quad (j \in J_s,\ s \in [q]) \tag{6}$$

$$\sum_{s=1}^{q} z_s \leq \theta \tag{7}$$

$$\beta_j = \beta_j^+ - \beta_j^- \quad (j \in [p]) \tag{8}$$

$$\beta_0 \in \mathbb{R}, \quad \boldsymbol{\beta} \in \mathbb{Z}^p, \quad \boldsymbol{\beta}^+, \boldsymbol{\beta}^- \in \mathbb{R}_+^p, \quad \boldsymbol{t} \in \mathbb{R}^n, \quad \boldsymbol{z} \in \{0,1\}^q \tag{9}$$

ただし，$\varepsilon \in \mathbb{R}_+$ は偏回帰係数の絶対値の重み，$M \in \mathbb{Z}_+$ は偏回帰係数の範囲を指定するパラメータとする．

式 (4), (5) により区分線形近似 (3) に基づく対数尤度関数 (1) を最大化し，式 (4), (8) により偏回帰係数の絶対値和を最小化する．式 (6), (7) でグループ変数の数と偏回帰係数の範囲を制限する．式 (9) は決定変数の一覧であり，特に偏回帰係数 $\boldsymbol{\beta}$ は整数変数とする．

3. 数値実験

本節では，数値実験の結果から提案手法の有効性を検証する．数値実験では，株式会社ルリアンから提供された相続業務の実際のデータセットを利用する．

3.1 説明変数

提供されたデータセットは匿名加工情報であり，個人を特定できない範囲で被相続人の住所，財産額，年齢，性別，配偶者・子どもの有無や，相続人の住所，年齢，性別，被相続人との間柄などの情報が抽出できる．ここから相続紛争に関連すると考えられる情報を選定して，説明変数を作成した（表 1）．

質的変数はダミー変数に変換した．量的変数は数値の分布を確認し，業務上の利便性も考慮して区間を設定し，複数のダミー変数に変換した．都道府県は小人口（150 万人未満）・中人口（150 万人〜600 万人未満）・大人口（600 万人以上）の 3 種類に分類した．相続人の住所の地域は北海道・東北・関東・中部・近畿・中国・四国・九州の 8 種類とした．以上より，グループ変数の数は $q = 21$，説明変数（ダミー変数）の数は $p = 53$ となる．

3.2 目的変数

株式会社ルリアンの業務は相続に関する相談や法務支援であり，法的権限をもたない立場で相続紛争に関与することはできない．また，紛争は他人に隠される傾向があるため，相続紛争の観測データの入手は困難である．

そこで本論文では，紛争発生の代理指標として遺産分割協議の長期化に着目する．相続では財産の分割内

表 2　3 分割交差確認による AUC

手法		$M=1$			$M=2$		
		$\theta=4$	$\theta=6$	$\theta=8$	$\theta=4$	$\theta=6$	$\theta=8$
提案手法		**0.6345**	**0.6387**	0.6312	**0.6393**	**0.6437**	**0.6403**
L1 正則化法	単位重み	0.5883	0.6006	0.6277	—	—	—
	拡大縮小	0.5845	0.5965	0.5935	0.5849	0.5967	0.6085
EN 正則化法	単位重み	0.6022	0.6176	**0.6459**	—	—	—
	拡大縮小	0.5929	0.6039	0.6105	0.6054	0.6114	0.6322
変数増加法	単位重み	0.5960	0.6346	0.6252	—	—	—
	拡大縮小	0.5255	0.5255	0.5255	0.5255	0.5255	0.5255

容を決定するために，不動産や金融などの財産を調査して財産目録を作成し，相続人同士の遺産分割協議後に協議書を作成する．この遺産分割協議では，企業は介入せずに相続人全員が分割内容に同意したうえで捺印する必要がある．したがって相続の際に紛争が発生すれば，必然的に遺産分割協議が長期化すると考えられる．また，株式会社ルリアンの業務効率の観点からも，遺産分割協議の長期化は望ましくない．

遺産分割協議期間に対して，紛争の有無を識別するような明確な閾値を設定することは難しい．そこで本論文では，遺産分割協議期間について上位 20%の長期化案件の目的変数は $y_i = +1$，下位 60%の案件は $y_i = -1$ と設定し，残りの 20%の案件は分類が困難であると考えてデータを削除した．また，紛争が発生した案件について株式会社ルリアンでアンケート調査を実施し，それらの案件を正例として追加した．以上より，事例数は $n = 1076$，正例 $(y_i = +1)$ の割合は 23.4%となる．

3.3　実験設定

提案手法の混合整数線形最適化問題 (4)–(9) は最適化ソルバー Gurobi Optimizer[2] 9.5.0 を用いて求解した．区分線形近似では 17 個の接点（文献 [8] の接点集合 V_3）を使用し，偏回帰係数の絶対値の重みは $\varepsilon = 10^{-4}$ と設定した．

提案手法の比較対象として，ロジスティック回帰モデルに対する L1 正則化法，EN（エラスティックネット）正則化法，変数増加法を考える．これらの手法では説明変数を徐々に増加させて，グループ変数の数が $\theta+1$ 個となったら，その直前のモデルを選択する．ただし，これらの手法によって得られる偏回帰係数は実数値であるため，文献 [7] を参考に以下の整数変換処理を適用して，採点システムとしての性能を評価する．

単位重み　偏回帰係数 β_j が正値の場合は $+1$，負値の場合は -1 に変換する．

拡大縮小　$\gamma := (M+0.49)/\max_{j\in[p]} |\beta_j|$ として，すべての偏回帰係数を $\gamma\beta_j$ に変換した後，$-M$ から M の間の最も近い整数に丸める．

正則化法は Python 言語の scikit-learn ライブラリの LogisticRegression 関数，変数増加法は R 言語の step 関数を用いて実装した．

予測精度の参考として，非線形機械学習手法であるランダムフォレスト，勾配ブースティング決定木，サポートベクトルマシンを考える．これらの手法は Python 言語の RandomForestClassifier 関数，LGBMClassifier 関数，SVC 関数を用いて実装した．また，超パラメータ（決定木の深さ，決定木の数，正則化項の重み，RBF カーネルのパラメータ）は訓練データを用いた交差確認によるグリッドサーチで調整した．

3.4　提案手法の性能検証

偏回帰係数の範囲 $M \in \{1,2\}$，グループ変数の数 $\theta \in \{4,6,8\}$ の設定で 3 分割交差確認を実施した．各手法の予測精度を表す AUC（ROC 曲線下の面積）の結果を表 2 に示す．$M=1$，$\theta=8$ の場合を除いて提案手法の予測精度が最も優れており，提案手法は $M=2$，$\theta=6$ の場合に最も良い精度を示した．$M=1$，$\theta=8$ の場合には EN 正則化法（単位重み）の予測精度が優れていたが，提案手法（$M=2$，$\theta=6$）と比較すると，提案手法は少ないグループ変数で EN 正則化法（単位重み）とほぼ同等の予測精度を達成している．

3 分割交差確認における 3 回の訓練時間の平均を表 3 に示す．提案手法は厳密解法であるため，ほかの手法（計算時間は 0.1 秒未満）と比較すると計算時間は長いが，$M=2$，$\theta=8$ の場合を除けば計算時間は 1 時間以内であった．

3 分割交差確認における非線形機械学習手法の性能を表 4 に示す．これらの手法の性能は提案手法よりも

[2] https://www.gurobi.com/

表 3　3 分割交差確認の計算時間（秒）

手法		$\theta = 4$	$\theta = 6$	$\theta = 8$
提案手法	$(M = 1)$	273.4	2686.6	2034.4
	$(M = 2)$	320.2	1900.5	39836.2
L1 正則化法		< 0.1	< 0.1	< 0.1
EN 正則化法		< 0.1	< 0.1	< 0.1
変数増加法		< 0.1	< 0.1	< 0.1

表 4　3 分割交差確認における非線形機械学習手法の性能

手法	AUC	計算時間（秒）
ランダムフォレスト	0.6670	0.4
勾配ブースティング決定木	0.6625	0.1
サポートベクトルマシン	0.5969	0.2

表 5　提案手法 $(M = 2,\ \theta = 6)$ による採点システム

質問項目	選択肢	得点
1. 相続人数	1 人	-2 点
	2 人以上	$+2$ 点
2. 不動産財産額	300 万円未満	$+1$ 点
3. その他財産	ある	$+1$ 点
4. 相続財産合計額	2,500 万円以上	$+1$ 点
5. 相続順位構成	第 2 順位	-1 点
	甥姪	$+1$ 点
6. 財産構成	不動産のみ	$+1$ 点
	合計：	

優れているが，提案手法 ($M = 2,\ \theta = 6$) はグループ変数の数や偏回帰係数の値が厳しく制限されているにもかかわらず，ランダムフォレストに対して約 96.5% ($\approx 0.6437/0.6670$) の AUC を達成している．さらに次節で考察するように，提案手法は少数の質問から簡単な演算で予測を行なうことが可能であり，結果の解釈性の観点でも非常に優れている．

3.5　提案手法による採点システム

提案手法 ($M = 2,\ \theta = 6$) で作成した採点システムを表 5 に，対応するリスク換算表を表 6 に示す．利用者は 6 種類の質問項目（表 5）に対して該当する選択肢の合計得点を算出し，得点・リスク換算表（表 6）を参照することで，相続紛争（相続工程の長期化）のリスクを算出することができる．得点が正値の選択肢は紛争リスクを増加させ，負値の選択肢は紛争リスクを減少させると解釈できる．

表 5 の採点システムについて考察する．「1. 相続人数」に関しては，複数の相続人で財産を分割する場合

表 6　提案手法 $(M = 2,\ \theta = 6)$ の得点・リスク換算表

得点	≤ -1 点	0 点	1 点	2 点	3 点	4 点	≥ 5 点
リスク	≤ 1%	2%	4%	11%	25%	47%	≥ 71%

表 7　提案手法 $(M = 2,\ \theta = 6)$ の混同行列

	得点									
	-3	-2	-1	0	1	2	3	4	5	合計
正例	0	0	0	0	1	52	123	61	9	246
負例	2	19	19	1	7	323	377	77	5	830
合計	2	19	19	1	8	375	500	138	14	1076

にリスクが高くなる．「2. 不動産財産額」に関しては，300 万円未満の低額不動産が相続される場合にリスクが高くなり，その理由として低額不動産は換金して分割するのが難しいことが考えられる．「3. その他財産」と「4. 相続財産合計額」に関しては，金融・不動産以外の財産が存在し，財産額の合計が大きい場合にリスクが高くなる．「5. 相続順位構成」に関しては，相続人に第 2 順位（親）が含まれるとリスクが低くなる．つまり，子どもが亡くなって親が相続する場合には紛争が生じにくいと考えられる．また，甥や姪が相続人に含まれるとリスクが高くなり，それらの遠縁の相続人は被相続人やほかの相続人との関係性が希薄であることや，離れた場所に住んでいることが理由として考えられる．「6. 財産構成」に関しては，相続財産が不動産のみの場合にリスクが高くなる．不動産は分割ができず，また不可動であるため取り扱いが難しいことが理由として考えられる．

最後に，提案手法 ($M = 2,\ \theta = 6$) による混同行列の結果を表 7 に示す．採点システムによる得点の大部分は 2〜4 点の範囲に分布している．また，得点が 0 点以下の場合はすべて負例案件となっており，さらに得点が上がるにつれて負例案件に対する正例（長期化）案件の割合が単調に増加する．

4.　おわりに

本論文では，混合整数最適化モデルを利用して，相続紛争のリスクを予測する採点システムを作成した．ロジスティック損失関数に区分線形近似を適用し，グループ変数選択を導入して質問項目の数を制約する最適化モデルを，混合整数線形最適化問題として定式化した．採点システムを利用することで，専門的な知識がなくても計算機を使用せずに簡単に相続紛争のリス

クを予測することができる．

数値実験の結果，偏回帰係数の整数変換処理を伴うロジスティック回帰モデルと比較して，提案手法は予測精度が優れていることや，非線形機械学習手法に迫る予測精度を達成できることがわかった．また，提案手法による採点システムの妥当性を考察した．

株式会社ルリアンは相続手続きに関して顧客と専門家をつなぐプラットフォーム事業を展開しており，相続の案件ごとに異なる専門家を差配・紹介する役割がある．同社の相続支援では行政書士が大部分の役割を担うが，行政書士は紛争性のある案件を取り扱うことができず，案件の過程で紛争が発生した場合には，当該案件から撤退して弁護士に後を任せるなどの措置が必要となる．それゆえ相続人との初回の面談の際に簡便な方法で紛争リスクを予測することができれば，その水準に応じたガイドラインの準備と運用が容易になり，同社の顧客サービスと危機管理の両面において利点がある．一方で，相続人との関係性を構築できていない初回の限られた時間の面談の中で，紛争に関する細かな質問をすることは難しく，これまでは紛争リスクを分析して相続人に伝えることができなかった．したがって，初回の面談時に少数の質問から簡易的に紛争のリスクを算出できる採点システムは，相続人が遺産分割を円滑に進めるための支援や，二次相続時の紛争を防止するための助言などにも有用だと考えられる．

本研究の課題として，収集可能なデータの限界がある．本研究で使用したデータには，遺言書の有無や内容，異父兄弟などの血縁関係，介護や同居に関する情報など，相続紛争の要因となりえる変数が不足していた．これらの変数が加わることで予測精度を改善できる可能性がある．その他の課題として，データの信頼性の問題がある．本研究で使用したデータにはコロナ禍の案件が含まれており，本来であれば長期化しない案件でもコロナ禍の影響で長期化してしまった可能性がある．また，変数間の交互作用項を導入して予測精度を向上させることも，今後検討すべき重要な課題である．

謝辞　株式会社ルリアンとの共同研究の機会を与えていただいた大澤義明先生（筑波大学），貴重なデータを提供していただいた株式会社ルリアンの関係者の皆様に御礼申し上げます．また，大変有用なコメントをいただいた査読者と編集委員会に心より感謝申し上げます．

参考文献

[1] 川島志保，関ふ佐子，『家族と高齢社会の法』，NHK 出版，2017.

[2] C. Rudin and B. Ustun, “Optimized scoring systems: Toward trust in machine learning for healthcare and criminal justice,” *Interfaces*, **48**, pp. 449–466, 2018.

[3] B. Ustun and C. Rudin, “Supersparse linear integer models for optimized medical scoring systems,” *Machine Learning*, **102**, pp. 349–391, 2016.

[4] H. Hu et al., “Development and validation of risk models to predict the 7-year risk of type 2 diabetes,” *Journal of Diabetes Investigation*, **9**, pp. 1052–1059, 2018.

[5] 高野祐一，宮代隆平，“混合整数最適化による線形回帰モデルの最良変数選択，” 日本統計学会誌，**50**, pp. 343–362, 2021.

[6] C. Gambella, B. Ghaddar and J. Naoum-Sawaya, “Optimization problems for machine learning: A survey,” *European Journal of Operational Research*, **290**, pp. 807–828, 2021.

[7] B. Ustun and C. Rudin, “Learning optimized risk scores,” *Journal of Machine Learning Research*, **20**, pp. 1–75, 2019.

[8] T. Sato, Y. Takano, Y. Miyashiro and A. Yoshise, “Feature subset selection for logistic regression via mixed integer optimization,” *Computational Optimization and Applications*, **64**, pp. 865–880, 2016.

[9] M. Naganuma, Y. Takano and R. Miyashiro, “Feature subset selection for ordered logit model via tangent-plane-based approximation,” *IEICE Transactions on Information and Systems*, **102**, pp. 1046–1053, 2019.

[10] H. Saishu, K. Kudo and Y. Takano, “Sparse Poisson regression via mixed-integer optimization,” *Plos One*, **16**, e0249916, 2021.

[11] D. Bertsimas, J. Pauphilet and B. Van Parys, “Sparse classification: A scalable discrete optimization perspective,” *Machine Learning*, **110**, pp. 3177–3209, 2021.

[12] H. Saishu, K. Kudo and Y. Takano, “Cutting-plane algorithm for estimation of sparse Cox proportional hazards models,” *TOP, An Official Journal of the Spanish Society of Statistics and Operations Research*, 2024 (In Press).

[13] H. Zhang, Q. Morris, B. Ustun and M. Ghassemi, “Learning optimal predictive checklists,” *Advances in Neural Information Processing Systems*, **34**, pp. 1215–1229, 2021.

[14] M. Yuan and Y. Lin, “Model selection and estimation in regression with grouped variables,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **68**, pp. 49–67, 2006.

[15] D. Bertsimas and A. King, “OR forum—An algorithmic approach to linear regression,” *Operations Research*, **64**, pp. 2–16, 2016.

[16] 川野秀一，松井秀俊，廣瀬慧，『スパース推定による統計モデリング』，共立出版，2018.

