A-028

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‡ 大谷郁美‡ 渡邉理子‡

1. はじめに

相続では，様々な要因から一定の割合で紛争が発生する．さらに，少子高齢化に伴い遺産分割事件数も増加している．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・傾向を分析し紛争を予防・回避することは重要である．

採点システムとは，回帰モデルの説明変数を質問項目，偏回帰係数を得点とし，利用者がいくつかの小さな数字を足したり引いたりするだけで予測ができる手法である．これは予測モデルを簡単に表現し，専門的知識がなくても，電卓や機械を使用せずに簡単に予測ができるという利点を持つ．採点システムでは，説明変数をなるべく少なく，かつ小さな整数の偏回帰係数をとる回帰モデルの使用が重要になる．

混合整数最適化を利用した採点システムでは，整数制約の下で L0・L1 正則化を加えて目的関数とし，病気であるか否かの分類予測を行う研究 [1] や，病気である確率を予測する研究 [2] がある．ロジスティック回帰モデルの変数選択問題では目的関数が非線形の凸関数となるが，Sato et al. [3] では接線に基づく区分線形近似を用いて混合整数線形最適化問題に帰着する方法が提案されている．

本発表では，上記の先行研究 [1, 2, 3] に基づいてロジスティック回帰モデルを用いたリスク採点システムを提案する．さらに本研究では，リスク採点システムにグループ変数選択を導入する．質的変数を複数のダミー変数に変換して利用する際に，元の質的変数を１項目として採点システムを構築することが可能となる．これより，質問項目を増やさずに採点システムが利用可能な情報を増やすことができる．また，本研究は相続に関わる紛争のリスク予測を行う新規的な研究になる．

本稿では，連続する正整数の集合を $[n] := \{1, 2, \ldots, n\}$ と表記する．

Risk scoring system for lengthy inheritance processes via mixed-integer optimization

Moe Shiina, Yuichi Takano, Tomoka Usami, Ikumi Otani, Riko Watanabe
† 筑波大学 ‡ 株式会社ルリアン
†University of Tsukuba, Tsukuba-shi, Ibaraki 305-8573, Japan
‡Lelien, Ltd., Kyoto-shi, Kyoto 604-8151, Japan

2. リスク採点システム

2.1. ロジスティック回帰モデル

ロジスティック回帰モデルは 2 値の目的変数に対する確率値を予測する統計的回帰モデルである．

n 個のデータ $\{(y_i, \boldsymbol{x}_i) \mid i \in [n]\}$ に対し，それぞれ p 次元説明変数 $\boldsymbol{x}_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,p})^\top$，2 値を取る目的変数であるクラスラベル $y_i \in \{-1, 1\}$ が与えられている．また，切片 β_0，p 次元の偏回帰係数ベクトル $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^\top$ を推定すべきパラメータとする．このときクラスラベル y_i が生起する確率は以下のように書ける．

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

ここで，ロジスティック損失関数を $f(v) = \log(1 + \exp(-v))$ としたとき，対数尤度は以下のように書ける．

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

以上のロジスティック損失関数の総和を小さくすることで対数尤度関数が大きくなり，実測値と予測値の誤差が少ないモデルを導くことができる．

2.2. 区分線形近似

ロジスティック損失関数（非線形凸関数）を接線に基づいて区分線形近似することで，大規模な混合整数非線形最適化問題を近似的に解くことが可能になる．

ロジスティック損失関数 $f(v) = \log(1 + \exp(-v))$ における離散点集合を $V = \{v_k \mid k \in [m]\}$ とする．接点 v_k における $f(v)$ の接線 $g_k(v)$ は，接点の近傍において $f(v)$ を近似している．$f(v)$ は凸関数であるため，すべての接線に対して以下の不等式が成り立つ．

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

これより，ロジスティック損失関数は各点 v で最大値を取る接線で構成される区分線形関数で近似するこ

Copyright © 2022 by
The Institute of Electronics, Information and Communication Engineers and
Information Processing Society of Japan All rights reserved.

とができる．決定変数 t を導入し，以下のように近似する．

$$\begin{aligned} f(v) &\approx \max\{g_k(v) \,|\, k \in [m]\} \\ &= \min\{t \,|\, t \geq g_k(v)\ (k \in [m])\} \\ &= \min\{t \,|\, t \geq f'(v_k)(v - v_k) + f(v_k)\ (k \in [m])\} \end{aligned}$$

2.3. 定式化

接線に基づく近似手法を用いて，ロジスティック回帰モデルに基づく採点システムを定式化する．区分線形近似に用いる決定変数のベクトルを $\boldsymbol{t} = (t_1, t_2, \ldots, t_n)^\top$ と定義する．J_s を s 番目のグループに属する説明変数の集合とする．グループ変数の選択を表す 0-1 決定変数 $\boldsymbol{z} = (z_1, z_2, \ldots, z_q)^\top$ を以下のように定義する．

$$z_s = \begin{cases} 1 : s \text{ 番目のグループ変数が選択される場合} \\ 0 : s \text{ 番目のグループ変数が選択されない場合} \end{cases}$$

以下の混合整数最適化問題を解くことで係数ベクトル $\boldsymbol{\beta}$ と切片 β_0 の値を求める．

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

$$\text{s.t.} \quad t_i \geq f'(v_k)\left(y_i\left(\sum_{j=1}^{p}\beta_j x_{i,j} + \beta_0\right) - v_k\right) + f(v_k) \qquad (k \in [m],\, i \in [n]) \tag{2}$$

$$z_s = 0 \;\Rightarrow\; \beta_j = 0 \qquad (j \in J_s,\, s \in [q]) \tag{3}$$

$$\sum_{s=1}^{q} z_s \leq \theta \tag{4}$$

$$\beta_j = \beta_j^+ - \beta_j^- \qquad (j \in [p]) \tag{5}$$

$$-M \leq \beta_j \leq M \qquad (j \in [p]) \tag{6}$$

$$\boldsymbol{z} \in \{0,1\}^q,\ \beta_0 \in \mathbb{R},\ \boldsymbol{\beta} \in \mathbb{Z}^p,\ \boldsymbol{\beta}^+, \boldsymbol{\beta}^- \in \mathbb{R}_+^p,\ \boldsymbol{t} \in \mathbb{R}^n \tag{7}$$

式 (1) と式 (2) は損失関数の区分線形近似を表す．式 (3) はグループ変数 s が選択されなかったとき，J_s に属する変数の偏回帰係数を全て 0 にする．式 (1) と式 (5) では，偏回帰係数の絶対値和を最小化する L1 正則化を示し，偏回帰係数の値が小さいモデルを導く．ただし，ε は正則化項の重みを表すパラメータである．式 (4) より，採点システムに含まれるグループ変数の最大数を θ 個にする制約を表す．式 (6) は，変数 j の偏回帰係数の上下限を表す．3 種類のパラメータ ε，θ，M を適切に設定し，式 (7) の決定変数を最適化することで，採点システムに適した少ない説明変数，かつ小さな整数の偏回帰係数をもつロジスティック回帰モデルを推定する．

3. 評価

目的変数のクラスラベル y_i（相続紛争の発生）は相続工程の長期化から判断する．相続紛争において，企業は解決人としては認められておらず，家族内の秘密とする傾向が強いため紛争の発生を観測することは困難である．そこで，企業が介入しない遺産分割協議の遅延を相続紛争による影響とした．したがって，本研究では遺産分割協議の期間が閾値を越える事例を「相続紛争の発生」とみなす．

企業の顧客データをもとにデータセットを作成した．また，提案手法の精度検証のために，AUC を評価指標とする数値実験を行なった．実験結果の詳細は当日報告する．

参考文献

[1] B. Ustun, and C. Rudin, “Supersparse linear integer models for optimized medical scoring systems,” *Machine Learning,* vol. 102, no. 3 (2016), pp. 349-391.

[2] B. Ustun, and C. Rudin, “Learning optimized risk scores,” *Journal of Machine Learning Research,* vol. 20, no. 150 (2019), pp. 1-75.

[3] T. Sato, Y. Takano, Y. Miyashiro, and A. Yoshise, “Feature subset selection for logistic regression via mixed integer optimization,” *Computational Optimization and Applications,* vol. 64, no. 3 (2016), pp. 865-880.

Copyright © 2022 by
The Institute of Electronics, Information and Communication Engineers and
Information Processing Society of Japan All rights reserved.