日本オペレーションズ・リサーチ学会
2022年 春季研究発表会

楕円分類器と変数選択を評価できる bAUC最大化モデルによる格付分類問題への適用

05000167　慶應義塾大学大学院理工学研究科　*田中 克弘　TANAKA Katsuhiro
02702020　慶應義塾大学理工学部　山本 零　YAMAMOTO Rei

1.　概要

分類モデルの性能を測る指標に Area Under Curve(AUC) がある．AUC が高いほどモデルの性能が高いことを意味するが，AUC は離散型の非凸型関数であるため，直接最大化することは難しい．そこで，[1] は凸型モデルに定式化できる Generalized bufferd AUC(bAUC) を提案し，更に bAUC 最適化は AUC 最適化を代用できることも示している．

bAUC 最大化モデルの利点は二つある．一つは，SVM に近いモデルの構造であることから多重共線性の影響を受けにくい点である．もう一つは，ハイパーパラメータが一つの分布の超分位点と示せることから，統計的な解釈が可能な点である．よって，ハイパーパラメータに理論的な意味を持たない SVM などと異なり，水準設定が容易になる．

このように近年 bAUC 最大化モデルの特徴が示されているが、実証分析の報告は限定的であり，単純なサンプルデータを用い，ハイパーパラメータが 0 のケースに限り計算時間を評価した程度である．

そこで本研究では汎化性能を高めるため離散変数を用いた変数選択および楕円分類器を用いた bAUC 最大化モデルを提案した上でその実証を行う．提案モデルは大規模な混合整数半正定値計画問題（MISDP）として定式化されるため、既存アルゴリズムでは現実的な解を得られない．そこで [2] が提案したヒューリスティック解法により，短い時間でその解を得られることを示す．

2.　問題設定

2.1.　bAUC の最大化

定式の準備として各記号を定義する．所与とするデータセット $(\{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_m, y_m)\})$ は，$\boldsymbol{x}_i \in \mathbf{R}^{\mathrm{P}}$ をサンプル i に対する説明変数，$y_i \in \{\pm 1\}$ をそのラベルとする．更に $i \in \mathcal{I} := \{1, \ldots, m\}$ かつ $p \in \mathcal{P} := \{1, \ldots, P\}$ とする．各ラベルが属する群の集合を $\mathcal{I}_+ = \{i \mid y_i = 1, i \in I\}$, $\mathcal{I}_- = \{i \mid y_i = -1, i \in I\}$ とし，$m_+ = |\mathcal{I}_+|, m_- = |\mathcal{I}_-|$ とする．なおここでは $\mathcal{I} = \mathcal{I}_+ \cup \mathcal{I}_-, m = m_+ + m_-$ となる．

係数 $\boldsymbol{w} \in \mathbf{R}^{\mathrm{P}}$ を導入して，次の線形分類器 (Linear classifier) を定義する．

$$R(\boldsymbol{w}|\boldsymbol{x}_i) = \boldsymbol{w}^{\mathbf{T}}\boldsymbol{x}_i, (i \in \mathcal{I}). \tag{1}$$

このとき Area under curve (AUC) を $R(\boldsymbol{w}|\boldsymbol{x}_{i_+}), (i_+ \in \mathcal{I}_+)$ が $R(\boldsymbol{w}|\boldsymbol{x}_{i_-}), (i_- \in \mathcal{I}_-)$ が上回る確率としてランキングエラー関数 $\xi(\boldsymbol{w}|\boldsymbol{x}_{i_+}, \boldsymbol{x}_{i_-}) = -\{R(\boldsymbol{w}|\boldsymbol{x}_{i_+}) - R(\boldsymbol{w}|\boldsymbol{x}_{i_-})\}, (i_+ \in \mathcal{I}_+, i_- \in \mathcal{I}_-)$ を用いて以下のように定義する．

$$\mathrm{AUC} = 1 - \frac{1}{m_+ m_-} \sum_{i_+ \in \mathcal{I}_+} \sum_{i_- \in \mathcal{I}_-} \mathbb{1}_{\xi(\boldsymbol{w}|\boldsymbol{x}_{i_+}, \boldsymbol{x}_{i_-})}, \tag{2}$$

ここで関数 $\mathbb{1}_{\bullet}$ は Indicator function である．

AUC は $R(\boldsymbol{w}|\boldsymbol{x}_{i_-}) - R(\boldsymbol{w}|\boldsymbol{x}_{i_+})$ の範囲で，$m_+ \times m_-$ の数の計算を行う．表記を簡単にするためランキングエラー関数の全組み合わせを以下のベクトルで定義する．

$$H(\boldsymbol{w}) = \xi(\boldsymbol{w}|\boldsymbol{x}_{i_+}, \boldsymbol{x}_{i_-}), (i_+ \in \mathcal{I}_+, i_- \in \mathcal{I}_-). \tag{3}$$

これより (2) 式の AUC は以下のように書ける．

$$\mathrm{AUC} = 1 - P(H(\boldsymbol{w}) \geqq 0). \tag{4}$$

次に AUC を拡張した bAUC を説明する．ある分位点 $Q_\gamma(H(\boldsymbol{w}))$ を確率水準 $\gamma \in \{0, 1\}$ を導入して次のとおり定義する:

$$Q_\gamma(H(\boldsymbol{w})) = \inf\{a \in \mathbf{R} | P(H(\boldsymbol{w}) > a) \leqq 1 - \gamma\}. \tag{5}$$

更にある超分位点 $q_\gamma(H(\boldsymbol{w}))$ を次のとおり定義する:

$$\begin{aligned} q_\gamma(H(\boldsymbol{w})) &= E[H(\boldsymbol{w})|H(\boldsymbol{w}) > Q_\gamma(H(\boldsymbol{w}))], \\ &= \min_{\alpha} \left\{ \alpha + \frac{E[H(\boldsymbol{w}) - \alpha]^+}{1 - \gamma} \right\}, \end{aligned} \tag{6}$$

ここで $[\cdot]^+ = \max\{\cdot, 0\}$ を指す．

(6) 式の最適解 α^* より，$\alpha^* = Q_\gamma(H(\boldsymbol{w}))$．更に超分位点に相当する $q_\gamma(H(\boldsymbol{w}))$ の逆関数の確率 $p_{\tilde{q}}(H(\boldsymbol{w}))$ を次に定義する．

$$\begin{aligned} p_{\tilde{q}}(H(\boldsymbol{w})) &= \max\{1 - \gamma \mid q_\gamma(H(\boldsymbol{w})) \geqq \tilde{q}\} \\ &= \min\{1 - \gamma \mid q_\gamma(H(\boldsymbol{w})) \leqq \tilde{q}\}, \end{aligned} \tag{7}$$

ここで $\tilde{q} \leqq 0$ はハイパーパラメータを表す．

[1] は，この $p_{\tilde{q}}(H(\boldsymbol{w}))$ を用いることで bAUC を以下の通りに定義している:

$$\mathrm{bAUC}_{\tilde{q}} = 1 - p_{\tilde{q}}\ (H(\boldsymbol{w})). \tag{8}$$

ここで，(5),(6),(7) および $\tilde{q} \leqq 0$ より，次を示せる．

$$P(H(\boldsymbol{w}) \geqq 0) \ \leqq \ P(H(\boldsymbol{w}) \geqq \tilde{q}) \ \leqq \ p_{\tilde{q}}(H(\boldsymbol{w})). \tag{9}$$

更に [1] は (9) から次式を示している．

$$\begin{aligned} 1 - P(H(\boldsymbol{w}) \geqq 0) \ &\geqq \ 1 - p_{\tilde{q}}\ (H(\boldsymbol{w})) \\ \Leftrightarrow \ \mathrm{AUC} \ &\geqq \ \mathrm{bAUC}_{\tilde{q}} \end{aligned} \tag{10}$$

(10) 式は，AUC 最大化の代用としてｂ AUC 最大化を

利用できることを意味している.

次に bAUC 最大化を考える. これは $p_{\tilde{q}}(H(\boldsymbol{w}))$ の最小化なので, 以下のように定式化できる:

$$\min_{\boldsymbol{w}} p_{\tilde{q}}(H(\boldsymbol{w})) = \left| \begin{array}{ll} \min\limits_{\gamma,\alpha,\boldsymbol{w}} & 1-\gamma \\ \text{s.t.} & \alpha + \dfrac{E[H(\boldsymbol{w})-\alpha]^+}{1-\gamma} \leqq \tilde{q}. \end{array} \right.$$

次にベクトル $\boldsymbol{w}$ のノルムを固定し書き直す.

$$\left| \begin{array}{ll} \min\limits_{\alpha,\boldsymbol{w}} & \dfrac{E[H(\boldsymbol{w})-\alpha]^+}{\tilde{q}-\alpha} \\ \text{s.t.} & \|\boldsymbol{w}\|_1^1 = 1, \end{array} \right. \tag{11}$$

ここで $\|\boldsymbol{w}\|_1^1 = \sum_{p\in\mathcal{P}} |w_p|$ を意味する.

最後に $\boldsymbol{w}$ を基準化するため $\boldsymbol{b} \in \mathbf{R}^{\mathrm{P}}$ を次に定義する:

$$\boldsymbol{b}^{\mathbf{T}} = \frac{\boldsymbol{w}^{\mathbf{T}}}{\tilde{q}-\alpha}. \tag{12}$$

これにより問題 (11) は次のとおりに書き直せる:

$$\min_{\boldsymbol{b}} \left\{ E[H(\boldsymbol{b}) - (\|\boldsymbol{b}\|_1^1)\tilde{q} + 1]^+ \right\}. \tag{L}$$

2.2. 楕円分類器と変数選択の追加

まず楕円分類器を導入するため, 二次係数行列 $\boldsymbol{B} \in \mathbf{R}^{\mathrm{P}\times\mathrm{P}}$ を置き, 分類器を次のとおり定義する:

$$r(\boldsymbol{b},\boldsymbol{B}|\boldsymbol{x}_i) = \boldsymbol{b}^{\mathbf{T}}\boldsymbol{x}_i + \boldsymbol{x}_i^{\mathbf{T}}\boldsymbol{B}\boldsymbol{x}_i,$$
$$h(\boldsymbol{b},\boldsymbol{B}) = -\{r(\boldsymbol{b},\boldsymbol{B}|\boldsymbol{x}_{i_+}) - r(\boldsymbol{b},\boldsymbol{B}|\boldsymbol{x}_{i_-})\}.$$

これにより二次分類器用の問題 (Q) に拡張できる:

$$\left| \begin{array}{ll} \min\limits_{\boldsymbol{B},\boldsymbol{b}} & E[h(\boldsymbol{b},\boldsymbol{B}) - (\|\boldsymbol{b}\|_1^1 + \|\boldsymbol{B}\|_1^1)\tilde{q} + 1]^+ \\ \text{s.t.} & \boldsymbol{B}^{\mathbf{T}} = \boldsymbol{B}. \end{array} \right. \tag{Q}$$

更に半正定値制約 $\boldsymbol{B} \succeq \boldsymbol{O}$ を (Q) に次の通り追加する:

$$\left| \begin{array}{ll} \min\limits_{\boldsymbol{B},\boldsymbol{b}} & E[h(\boldsymbol{b},\boldsymbol{B}) - (\|\boldsymbol{b}\|_1^1 + \|\boldsymbol{B}\|_1^1)\tilde{q} + 1]^+ \\ \text{s.t.} & \boldsymbol{B} \succeq \boldsymbol{O}, \boldsymbol{B}^{\mathbf{T}} = \boldsymbol{B}. \end{array} \right. \tag{E}$$

次に離散変数 $\boldsymbol{z} \in \{0,1\}^{\mathrm{P}}$, ユーザーパラメータ S を導入し変数選択制約を書く.

$$\sum_{p_1\in\mathcal{P}} z_{p_1} \leqq S, (p_1, \in \mathcal{P}). \tag{13}$$

$$z_{p_1} = 0 \Rightarrow b_{p_1} = 0 \quad (p_1, \in \mathcal{P}), \tag{14}$$

$$z_{p_1} + z_{p_2} = \{0,1\} \Rightarrow B_{p_1,p_2} = 0 \ (p_1, p_2 \in \mathcal{P}). \tag{15}$$

最後に, (E) に変数選択制約 (13),(14),(15) を追加し, 次の MISDP を定式化する:

$$\left| \begin{array}{ll} \min\limits_{\boldsymbol{B},\boldsymbol{b},\boldsymbol{z}} & E[h(\boldsymbol{b},\boldsymbol{B}) - (\|\boldsymbol{b}\|_1^1 + \|\boldsymbol{B}\|_1^1)\tilde{q} + 1]^+ \\ \text{s.t.} & \boldsymbol{B} \succeq \boldsymbol{O}, \boldsymbol{B}^{\mathbf{T}} = \boldsymbol{B}, (13), (14), (15). \end{array} \right. \tag{$\mathbb{E}$}$$

2.3. MISDP の解法

MISDP を解くアルゴリズムとして ($\mathbb{E}$) を分割して解くアルゴリズムを提案する.

ヒューリスティックアルゴリズム (HA)

Step1 (E) を解く. その後, 解 $\hat{\boldsymbol{B}}$ を保存する.

Step2 変数 $\lambda \geqq 0$ を導入した次の問題を解き, その後, 解 $\boldsymbol{z}^*$ を用いて, $\mathcal{J} = \{p | z_p^* = 1, p \in \mathcal{P}\}$, $x_{i,s\in\{1,\cdots,S\}} := x_{i,p\in\mathcal{J}}, i \in \mathcal{I}$ として入力データ $\boldsymbol{x}_i \in \mathbf{R}^{\mathrm{S}}$ を削減する.

$$\left| \begin{array}{ll} \min\limits_{\boldsymbol{B},\boldsymbol{b}} & E[h(\boldsymbol{b},\boldsymbol{B}) - (\|\boldsymbol{b}\|_1^1 + \|\boldsymbol{B}\|_1^1)\tilde{q} + 1]^+ \\ \text{s.t.} & z_{p_1} + z_{p_2} = \{0,1\} \Rightarrow \lambda\hat{B}_{p_1,p_2} = 0, \\ & \delta \geqq \lambda \geqq 0, (13), (14). \end{array} \right.$$

Step3 再び (E) を解く.

Step2 でランク P の $\hat{\boldsymbol{B}}$ が半正定値行列なら, ランク S の解 $\boldsymbol{B}^*$ は, 半正定値性の維持が期待できる. 又この問題は離散変数を含んだ線形計画問題に定式化できる.

3. 数値結果

使用データは東証一部上場企業で R&I 社が格付を付与している 2012 年の 405 社 ($m_+ = 269$, $m_- = 124$) である. 格付データは R&I の定義から $\mathcal{I}_+$ は AAA+〜AAA, $\mathcal{I}_-$ は BBB+〜CCC+に相当すると設定し, また財務データは Yeo-Johnson 変換でスケーリングを実施した. また $\epsilon = 1.0\mathrm{e}^{-3}, \delta = 2$ とした.

ここでは提案した HA と, 先行研究で提案されている既存アルゴリズム (CPA) を比較した.

$P = 30$ および S={5,10,15} の際の計算時間の比較結果は次のとおりで, HA は相対的に速く解を得られた. なお “—” は 720 分以内に解けないことを意味する.

表 1: CPA と HA による計算時間 (単位: 分)

	CPA			HA		
$\tilde{q}\backslash S$	5	10	15	5	10	15
-0.4	12	29	23	2	2	9
-0.35	24	53	46	3	3	3
-0.3	36	96	72	3	3	3
-0.25	84	—	218	3	3	3
-0.2	360	—	—	2	4	4
-0.15	—	—	—	3	5	5
-0.1	—	—	—	4	24	46
-0.05	—	—	—	48	187	115

またの各アルゴリズムによる目的関数の相対誤差は, 最大で 0.5%と誤差は小さいことを確認している.

当日は他のケース及び汎化性能結果も報告する.

参考文献

[1] Norton, M. and Uryasev, S. “Maximization of AUC and Buffered AUC in binary classification”. *Mathematical Programming*, 174, 575-612, (2019).

[2] Tanaka, K. and Yamamoto, R. “Identification of Best Discrimination Surface by Mixed-Integer Semi-Definite Programming for Support Vector Machine”. *International Journal of Financial Engineering*, (to appear), (2021).