サポートベクトルに着目した ECOC-SVM による多値分類

1X12C112-1　山極 綾子
指導教員　後藤 正幸

1　研究背景・目的

近年，データの大規模化により効率的な知識獲得のための自動分類技術の重要性が増しており，その多くはカテゴリ数が 3 以上の場合を対象とした多値判別問題である．多値判別問題に対しては様々なアプローチがあるが，本研究では複数の二値判別器を組合わせる手法の 1 つである Error Correcting Output Coding 多値判別法 [1](以下 ECOC 法) に注目する．ECOC 法は，強力な二値判別器を利用して少ない計算コストで複雑な識別境界を表現可能とする手法であり，特に二値判別器にソフトマージン Support Vector Machine[2](以下 SVM) を用いるものを ECOC-SVM と呼ぶ．この手法では，判別器の構成を表現した符号表の生成法と，学習後の二値判別器の出力結果から新規データの所属カテゴリを推定するための統合方法が重要である．後者の統合方法としては，符号語との距離を用いるものや損失を定義するものなどがあり，二値判別器の性能を最も活かす手法が必要とされている．

本研究では ECOC 法のうち，学習データの誤判別による損失が最も少なくなるような出力の扱い方を学習する Optimized Weight Decoding[3](以下 OW) に着目する．OW はカテゴリ推定方法として損失を定義する手法のうちの 1 つであり，学習データの判別精度に対して最適であると考えられる損失の定義を学習する．ここで学習データの中には，識別境界付近に位置しており学習によって判別精度の向上が見込めるデータと，外れ値のような判別精度を低下させるデータが存在するが，OW は全学習データを同等に扱っており，後者の外れ値のようなデータを考慮することが，精度の低下を招いている可能性がある．

そこで本研究では，OW の重み付け方法を改良し，判別性能向上に有効な学習データについて重点的に学習を行う方法を提案する．二値判別器の学習により得られるパラメータを利用し，重視すべき学習データに重み付けすることで，判別精度の向上に有効なデータに対して重点的に学習が行われ，判別精度が向上すると考えられる．また，提案手法をベンチマークデータに適用しその有効性を示す．

2　準備

2.1　ECOC-SVM

ECOC 法とは，符号理論で用いられる誤り訂正技術を自動分類に応用し，カテゴリが未知の入力データに対し，複数の二値判別器を組み合わせて所属カテゴリを推定する手法である．カテゴリ数を P，二値判別器数を Q とする．このとき符号表 $\boldsymbol{M}$ は以下のように表される．

$$\boldsymbol{M} = [\mathbf{m}_1, ..., \mathbf{m}_P]^{\mathrm{T}} \tag{1}$$

ここで縦ベクトル $\mathbf{m}_p \in \{-1, 1\}^Q$ はカテゴリ p の符号語であり，符号表の各列は，要素がそれぞれ -1 と 1 のカテゴリ集合を判別する二値判別器を表している．

また，ECOC-SVM で判別器として用いられるソフトマージン SVM は，識別境界と最近傍データの距離であるマージンを最大化する手法であり，識別境界との距離が最大マージンと一致するデータをサポートベクトルと呼ぶ．この手法では，識別境界との距離が最大マージンより小さいデータにおいて，スラック変数と呼ばれるペナルティ項が正の値をとる．

2.2　損失を用いたカテゴリ推定

ECOC 法におけるカテゴリの推定段階において，各カテゴリの理想の出力である符号語と，データ出力結果のずれを測る方法の 1 つに損失を用いる手法がある．この手法では，テストデータは重み行列 $\boldsymbol{W}$ を用いて算出される損失が最小となるカテゴリに識別される．ここで，重み行列とは各判別器の各カテゴリの判別に対する重要度を示す $P \times Q$ 行列であり，p 番目の行を ${\mathbf{w}_p}^{\mathrm{T}}$ で表す．新規入力データの Q 個の二値判別器の出力を縦ベクトル $\mathbf{x}_t \in \{0, 1\}^Q$ で表し，その予測カテゴリ $\hat{y}_t$ を以下の式 (2) で求める．ただし，$\circ$ はベクトルの要素ごとの積を表す．

$$\hat{y}_t = \arg\min_p \left(-\mathbf{w}_p^{\mathrm{T}} \left(\mathbf{m}_p \circ \mathbf{x}_t \right) \right) \tag{2}$$

3　従来手法

ECOC 法に損失の考え方を導入した OW は，誤判別が最も少なくなるような重み行列 $\boldsymbol{W}$ を学習するという手法である．いま，学習データ $\boldsymbol{\rho}_i$ の出力符号語を $\mathbf{x}_i \in \{-1, 1\}^Q$ とすると，カテゴリラベルが y_i である $\boldsymbol{\rho}_i$ をカテゴリ p に誤判別する際の損失 ξ_i^p は，以下の式 (3) で定義される．

$$\xi_i^p = \max \left(0, -\mathbf{w}_{y_i}^{\mathrm{T}} \left(\mathbf{m}_{y_i} \circ \mathbf{x}_i \right) + \mathbf{w}_p^{\mathrm{T}} \left(\mathbf{m}_p \circ \mathbf{x}_i \right) \right) \tag{3}$$

式 (3) は，誤ったカテゴリ p への所属し易さを評価する損失関数である．$\boldsymbol{\rho}_i$ が正しく判別されるとき全ての p について $\xi_i^p = 0$ となる．式 (3) を用い，すべての誤判別による損失の和を最小化する重み行列 $\boldsymbol{W}$ を，以下の最適化問題を解くことにより求める．ただし，学習データ数を n とする．

$$\min_{\boldsymbol{W}} \sum_{i=1}^{n} \sum_{p=1}^{P} \xi_i^p \tag{4}$$

$$\text{s.t.} \quad -\mathbf{w}_{y_i}^{\mathrm{T}} \left(\mathbf{m}_{y_i} \circ \mathbf{x}_i \right) + \mathbf{w}_p^{\mathrm{T}} \left(\mathbf{m}_p \circ \mathbf{x}_i \right) \leq \xi_i^p \tag{5}$$

$$\xi_i^p \geq 0 \tag{6}$$

式 (4) は，式 (3) で定義した ξ_i^p の全てのデータ，全てのカテゴリについての和の最小化である．また，式 (5) は ξ_i^p が誤判別による損失以下の値を取らないようにし，式 (6) は誤判別による損失が非負であることを定義している．

しかし，この最適化問題においてすべての i と p の組み合わせについて考慮すると，空間，時間計算量が膨大になってしまう．そこで，インデックス g_i^p を導入して用いる制約の数を制限し，計算量を削減する．インデックス g_i^p は，$g_i^p = 1$ のとき「i 番目の学習データをカテゴリ p に誤判別しない」という制約の追加を意味し，0 のとき制約を考慮しないことを意味している．また，反復回数を t とする．

STEP1) 符号表 $\boldsymbol{M}$ を生成し，二値判別器を学習する．

STEP2) それぞれの二値判別器において，各カテゴリごとに学習データの正解率を求め，各行の和を 1 となるように正規化した値を重み行列 $\boldsymbol{W}^{(0)}$ の初期値とする．

STEP3) $t = 0$ とし，全ての i と p について $g_i^p = 0$ とする．

STEP4) $\boldsymbol{W}^{(t)}$ における損失 ξ_i^p の最大値を与える i と p の組み合わせ i^* と p^* を求め，$g_{i^*}^{p^*} = 1$ として制約を追加する．すでに $g_{i^*}^{p^*} = 1$ であった場合，アルゴリズムを終了する．

STEP5) 以下の式 (7) により重み行列を更新する．

$$\boldsymbol{W}^{(t+1)} = \arg\min_{\boldsymbol{W}} \sum_{i=1}^{n} \sum_{p=1}^{P} g_{i,p} \circ \xi_i^p \tag{7}$$

STEP6) $t = t + 1$ とし，STEP4 に戻る．

4　提案手法

4.1　概要

複数カテゴリの領域の境界部分は誤判別が起きやすく，この境界の学習を適切に行うことは，学習において重要である．しかし，従来手法では重み行列の学習のとき，識別境界付近の判別性能向上に有効な学習データを外れ値を含む他の学習データと同等に扱っており，そのことが判別精度の向上を妨げていると考えられる．

そこで，本研究では誤判別による損失について，SVM の学習で得られるパラメータを活用し，識別境界付近に存在するデータの重みを大きくし，誤判別による損失の最小化を行う方法を提案する．これにより，識別境界付近の学習データについて重点的に学習を行うことが可能になり，判別精度の向上が期待できる．

4.2　学習データの重みの計算

識別境界に近いデータを重視した学習を行うためには，そのようなデータに大きな重みを付与すればよい．そこで，識別境界との近さを評価する変数としてソフトマージン SVM を学習する際に得られるスラック変数 α を用いることを考える．SVM では，最大マージンが 1 になるように正規化しており，α が 1 より大きいとき，そのデータが誤ったカテゴリの領域にあることを意味している．具体的には，$\boldsymbol{\rho}_i$ の重み ω_i を，q 番目の判別器におけるスラック変数 α_i^q の値を用い，以下の式 (8) で定義する．ここで，$\sum_{q=1}^{Q}|m_{y_i,q}|$ は $\mathbf{m}_{y_i}$ の要素のうち，値が 1 か -1 の個数を示している．

$$\omega_i = \frac{\sum_{q=1}^{Q}\hat{\alpha}_i^q}{\sum_{q=1}^{Q}|m_{y_i,q}|} \tag{8}$$

$$\hat{\alpha}_i^q = \begin{cases} 2-\alpha_i^q & (1<\alpha_i^q<2) \\ 0 & (otherwise) \end{cases} \tag{9}$$

式 (9) はデータ $\boldsymbol{\rho}_i$ が誤判別され，かつ q 番目の判別器に対して識別境界との距離が 1 未満であるとき，識別境界との近さを評価する．距離が 1 以下の場合のみを対象とすることで，そのデータが識別境界付近に無く，外れ値である場合に，その距離が重みに反映されることを防ぐことが可能となる．

提案手法における，学習データ $\boldsymbol{\rho}_i$ をカテゴリ p に誤判別することによる損失 $\hat{\xi}_i^p$ を，以下の式 (10) で定義する．

$$\hat{\xi}_i^p = \omega_i \cdot \max\left(0, -\mathbf{w}_{y_i}^{\mathrm{T}}\left(\mathbf{m}_{y_i}\circ\mathbf{x}_i\right)+\mathbf{w}_p^{\mathrm{T}}\left(\mathbf{m}_p\circ\mathbf{x}_i\right)\right) \tag{10}$$

式 (10) により，識別境界に近いデータの損失の影響を大きくした学習が可能となる．$\boldsymbol{\rho}_i$ が外れ値であり大多数の識別境界から離れている場合でも，いずれか 1 つの識別境界に近ければ ω_i は 0 以外の値を持つが，ω_i は式 (8) で与えられるため，カテゴリ数がある程度以上の場合，二値判別器数 Q も大きくなりその影響は無視できると考えられる．

5　実験

提案手法の有効性を示すため，UCI 機械学習レポジトリのベンチマークデータセット 3 種類に対し，提案手法の判別精度の面での有効性および外れ値に対する頑健性の評価実験を行った．

5.1　実験条件

符号表の生成手法として，0 が一様に分布する符号表をランダムに生成する手法である Sparse Random，二値判別器の学習は線形ソフトマージン SVM を用いた．また，評価指標は式 (11) で計算される正解率を用いる．

$$\text{正解率} = \frac{\text{正しく分類されたテストデータ数}}{\text{総テストデータ数}} \tag{11}$$

なお，実験結果は分割数 5 の交差検定を 5 回行ったときの正解率の平均を示している．実験データセットの概要を表 1 に示す．

表 1．実験データの概要

実験データ名称	カテゴリ数	次元数	総データ数
Glass	7	9	214
Zoo	7	17	101
Yeast	9	8	1484

5.2　実験結果と考察

表 2 に，各データにおける正解率を示す．

表 2．実験結果 (分類精度)

実験データ	従来手法	提案手法
Glass	52.07	60.20
Zoo	96.97	97.18
Yeast	48.87	50.50

表 2 より，提案手法の有効性が示せた．これは，識別境界に近いデータに注目し学習を行ったことにより，判別に有効な識別境界を学習できたためであると考えられる．

次に図 1 に，Yeast のデータセットに対し，外れ値の割合を増やした場合の精度の推移を示し，片側 t 検定を行う．ただし，図 1 における*は 5%有意差，**は 1%有意差を示す．また，外れ値の割合が a% とは，もとの学習データ数 $\times a$% 個の外れ値を追加したことを示している．

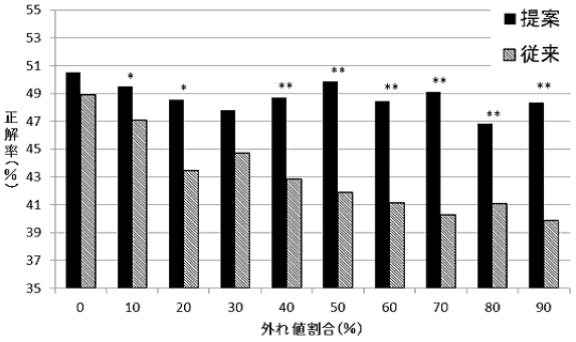

図 1．外れ値の割合と精度の関係 (Yeast)

図 1 より，外れ値の割合が増えるにしたがって従来手法では精度が大きく低下する一方，提案手法では精度の低下が緩やかになっている．この結果から提案手法は，外れ値が学習データに混入する場合においても，重点的に学習すべきデータを定めることで一定の精度を出すことが可能であり，頑健性の面でも有効であることが示された．

6　まとめと今後の課題

本研究では多カテゴリの分類問題を対象とし，従来手法では考慮出来なかった学習データの判別性能向上に対する影響を考慮し，各学習データに適切な重みを付けることによる判別手法を提案した．また実験結果より，全体的に判別精度が向上するだけではなく，外れ値に対する頑健性を持つことが示された．今後の課題として，二値判別器学習段階における重みの有効活用が挙げられる．

参考文献

[1] T. Dietterich and G. Bakiri, “Solving Multiclass Learning Problems via Error-Correcting Output Codes,” *J. Artificial Intelligence Research*, vol. 2, pp. 263-286, 1995.

[2] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, pp273-297, 1995.

[3] Xiao-Lei Zhang, “Heuristic Ternary Error-Correcting Output Codes Via Weight Optimization and Layered Clustering-Based Approach,” *IEEE Trans. on Cybernetics*, vol. 45, No. 2, pp289-301, 2015.