

ECOC Multi-Valued Classification using MCE Training
for Cost-Sensitive Learning

YASUDA Naoki

1 研究背景・目的

近年の情報技術の発達に伴い、大規模データに対する自動分類技術など有用な手法が広く用いられるようになってきている。中でも、複数カテゴリのデータを対象とした多値分類問題に対する高い頑健性を実現する手法として、精度の高い複数の二値分類器を組み合わせて多値分類を行う ECOC 多値分類手法 [1] が提案されている。一般に分類問題においては、データの誤分類による損失を全て同等に扱い、分類対象データにおける誤分類率の最小化を目的として分類器を学習する場合が多い。しかし、「ある患者が健康なのか、もしくはいずれかの疾病に該当するか」を予測する疾病診断等、実問題では誤りの方向も考慮しなければならない場合も多い。このような場合を想定し、コスト考慮型学習と呼ばれる方法が研究されている [2]-[4]。コスト考慮型学習は、多値分類問題において、各カテゴリに属するデータがどのカテゴリに誤分類されるかによって損失が異なる場合を想定し、その平均損失を小さくするように分類器を学習するための方法である。

ECOC 多値分類手法にコスト考慮型学習を導入することを考える。ECOC 多値分類手法では、複数のカテゴリを 2 つのカテゴリ集合に分割し、その 2 つのカテゴリ集合に属するデータを分割する二値分類器の学習を行う。そのため、ECOC 多値分類手法を用いてコスト考慮型学習を考える際には、カテゴリの分割の仕方が異なる各二値分類器に対して、各誤分類による損失を考慮して学習しなければならない。さらに、ECOC 多値分類手法では、複数の二値分類器を組み合わせて最終的な多値分類を行うため、事前に設定した損失にもとづいて各二値分類器を学習しても、全体に対して設定した損失を最小化する結果が得られるとは限らないといった問題も考えられる。

ここで、機械学習の様々な分類手法に対して精度の高い分類器を学習する手法として、分類誤り数損失を平滑な関数で近似しその損失の最小化を行う Minimum Classification Error (以下、MCE) 学習法 [5] が提案されている。ECOC 多値分類手法の各二値分類器の学習において、各誤分類に対して異なる損失を設定しながら MCE 学習法を適用することで、使用する分類手法を問わないコスト考慮型学習が実現できると考えられる。また、最終的な多値分類のために各二値分類器を組み合わせる際に、各分類器に対し適切な重みを与えることで、誤分類による損失をより減少できる可能性も考えられる。

そこで、本研究では、ECOC 多値分類手法に MCE 学習法を援用することで、コスト考慮型学習を実現する手法の提案を行う。具体的には、ECOC 多値分類手法の各二値分類器の学習と複数の二値分類器を組み合わせる際の 2 段階で、MCE 学習法を用いて適切なパラメータを求めることで、より一般的なコスト考慮型学習の手法を提案する。さらに、実験により提案手法の有効性を示す。

2 準備

2.1 分類問題

いま、学習データを (\mathbf{x}_i, y_i) ($i = 1, \dots, N$) とする。ただし、 $\mathbf{x}_i \in \mathbb{R}^d$ は特徴量をその要素にもつ d 次元ベクトルとし、 $y_i \in \mathcal{C}$ を所属カテゴリとする。ただし、 $\mathcal{C} = \{c_1, \dots, c_K\}$ はカテゴリ集合である。分類問題とはカテゴリの情報を所持する N 個の学習データを用いて分類器の学習を行い、カテゴリが未知の新規データ \mathbf{x} に対し、カテゴリ集合 \mathcal{C} から所属カテゴリを推定する問題のことである。また、 $K \geq 3$ の場合を多値分類問題という。

2.1.1 コスト考慮型学習

実問題においては、疾病診断（どの疾患かを予測する問題）などのように、所属カテゴリと誤分類先のカテゴリによって、損失の大きさが異なる状況が存在する。そこで、誤分類による損失の大きさがそれぞれ異なる場合を想定し、分類器を学習するコスト考慮型学習が研究されている。コスト考慮型学習では、データを誤分類した際の損失を各誤分類に対し事前に設定し、その損失が最小となるような分類を実現する分類器の学習を考える。

2.2 Support Vector Machine

Support Vector Machine (以下、SVM) [6] は、統計的学習理論の枠組みで提案された二値分類手法である。誤分類の最小化とともに、マージン最大化の概念を目的関数に加えることで、汎化能力の高い識別関数を求める。

いま、データ \mathbf{x} に対して SVM で学習する識別関数 $f(\mathbf{x})$ を、係数ベクトル \mathbf{w} 、バイアス項 b を用いて式 (1) で表す。

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \quad (1)$$

SVM では、以下の最適化問題を解くことにより、最適分離超平面を構成する識別関数を求める。

$$\text{minimize} \quad \frac{1}{2} \|\mathbf{w}\|^2 + \eta \sum_{i=1}^N \pi_i \quad (2)$$

$$\text{subject to} \quad t_i (\mathbf{w}^T \mathbf{x}_i + b) - 1 + \pi_i \geq 0 \quad (3)$$

$$\pi_i \geq 0 \quad (4)$$

ただし、 π_i は学習データ \mathbf{x}_i に対するスラック変数、 η はスラック変数に対するペナルティパラメータ、 $t_i \in \{-1, +1\}$ は学習データ \mathbf{x}_i のカテゴリ集合ラベルを表す。一般的に SVM では、上述の最適化問題を主問題とし、それに対応する以下の双対問題を解くことで識別関数を求める。

$$\text{maximize} \quad \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j t_i t_j \mathbf{x}_i^T \mathbf{x}_j \quad (5)$$

$$\text{subject to} \quad \sum_{i=1}^N \alpha_i t_i = 0 \quad (6)$$

$$0 \leq \alpha_i \leq \eta \quad (7)$$

ただし, α_i は各学習データ \mathbf{x}_i に対するラグランジュ未定乗数を表す. また, 分類に関しては, データ \mathbf{x} を式 (1) の識別関数に代入した値の正負により判定する.

2.3 ECOC 多値分類手法

誤り訂正符号 (ECOC) は, 情報系列に対し冗長な情報を付加することで, データ送信の際に多少のノイズが混入したとしても元の情報への訂正を可能とし, 通信の信頼性を高める技術である. Dietterich らは ECOC を利用し, 多値分類問題を複数の二値分類問題に分解することで解く枠組みを与えた [1].

ECOC 多値分類手法では, $K \times L$ 行列 \mathbf{W} を作成し, \mathbf{W} にもとづいて二値分類器の学習を行う. ここで, L は作成する二値分類器の個数を意味する. \mathbf{W} の各列は各二値分類器 f_l ($l = 1, \dots, L$) における分類の仕方を意味し, カテゴリ集合ラベル $\{-1, +1\}$ に従い分類器を学習する. また, 行列 \mathbf{W} の k 行目を L 次元ベクトル \mathbf{W}_{c_k} とし, カテゴリ c_k に対応する符号語とする. 二元の符号語を用いて全ての分割パターンを実現する Exhaustive 符号を用いた場合, 作成する二値分類器の個数は $L = 2^{K-1} - 1$ で与えられる.

$$\begin{array}{c} f_1 \quad f_2 \quad f_3 \quad f_4 \quad f_5 \quad f_6 \quad f_7 \\ \mathbf{W}_{c_1} \begin{pmatrix} +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ \mathbf{W}_{c_2} \begin{pmatrix} -1 & -1 & -1 & -1 & +1 & +1 & +1 \\ \mathbf{W}_{c_3} \begin{pmatrix} -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ \mathbf{W}_{c_4} \begin{pmatrix} -1 & +1 & -1 & +1 & -1 & +1 & -1 \end{pmatrix} \end{pmatrix} \end{pmatrix} \end{pmatrix} \end{array}$$

図 1. $K = 4$ の場合の Exhaustive 符号による分類器構成

ECOC 多値分類手法では, 符号語の設定に従って全カテゴリを 2 つのカテゴリ集合に分割し, 分割した 2 つのカテゴリ集合に属する学習データを分類する二値分類器を学習する. その後, それらの出力を組み合わせることで多値分類を実現する.

ECOC 多値分類手法における分類には, 式 (8) の尺度を用いるものとする.

$$F_{c_k}(\mathbf{x}) = \sum_{l=1}^L D_l(\mathbf{x}) g_{c_k, l} \quad (8)$$

$D_l(\mathbf{x})$ はデータ \mathbf{x} を分類器 f_l に入力した際の出力値, $g_{c_k, l} \in \{-1, +1\}$ は分類器 f_l におけるカテゴリ c_k に対するカテゴリ集合ラベルを表す. 各カテゴリに対し $F_{c_k}(\mathbf{x})$ を求め, その値が最大となるカテゴリへと分類を行うことで最終的な多値分類を行う.

3 Minimum Classification Error 学習

MCE 学習法 [5] は, 分類誤り数損失を平滑な関数で近似しその最小化を目指す分類器を得ようとする学習手法である. 分類誤り数損失をシグモイド関数を用いて近似することで微分可能となり, パラメータ更新が容易であることが特徴といえる.

3.1 MCE 学習

MCE 学習では, 式 (9) の基準を用いて分類を行うことを考える.

$$\hat{c} = \arg \max_{c_k \in C} F_{c_k}(\mathbf{x}; \Lambda) \quad (9)$$

ここで $F_{c_k}(\mathbf{x}; \Lambda)$ はカテゴリ c_k に対する識別関数を表し, その値はデータ \mathbf{x} がカテゴリ c_k に属する度合いを表す. また, Λ は分類器のパラメータを表す.

いま, ある学習データ \mathbf{x}_i に対して, 式 (10) の誤分類尺度を考える. 式 (10) の値が正ならば誤分類, 負ならば

正分類を意味し, その絶対値はデータの分類における確信度と解釈することができる.

$$d(\mathbf{x}_i; \Lambda) = -F_{y_i}(\mathbf{x}_i; \Lambda) + \max_{c_k, c_k \neq y_i} F_{c_k}(\mathbf{x}_i; \Lambda) \quad (10)$$

パラメータの更新における計算を容易にするため, 式 (10) の関数に学習パラメータに関して微分可能なシグモイド関数を適用し, 正の定数 β を用いた式 (11) により, ある学習データの分類誤り数損失を平滑化分類誤り数損失として定式化する.

$$l(d(\mathbf{x}_i; \Lambda)) = \frac{1}{1 + \exp(-\beta d(\mathbf{x}_i; \Lambda))} \quad (11)$$

シグモイド関数のパラメータ β を大きく設定するほど, 図 2 のようにより 0-1 損失に近似する.

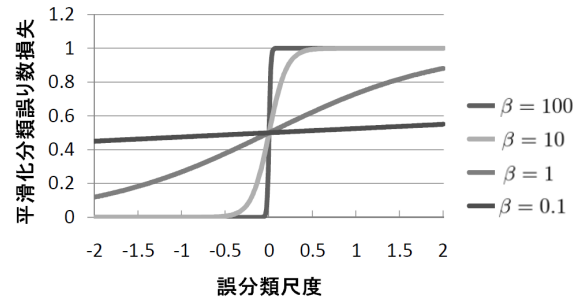


図 2. シグモイド関数

いま, N 個の学習データを利用したと仮定すると, MCE 学習法が最小化する目的関数は, 経験的平均損失として式 (12) で表される.

$$L(\Lambda) = \frac{1}{N} \sum_{i=1}^N l(d(\mathbf{x}_i; \Lambda)) \quad (12)$$

MCE では, 経験的平均損失 $L(\Lambda)$ の最小化を目指すことで, 有限個の学習データを持つ学習データセットにおける全体的な損失を減少させる. 具体的には, 式 (13) に基づいて Λ の逐次更新を行うことで分類器の学習を行う.

$$\Lambda^{(t+1)} = \Lambda^{(t)} - \frac{\varepsilon_t}{N} \sum_{i=1}^N \left. \frac{\partial l(d(\mathbf{x}_i; \Lambda))}{\partial \Lambda} \right|_{\Lambda = \Lambda^{(t)}} \quad (\varepsilon_t > 0) \quad (13)$$

4 提案手法

4.1 提案手法の概要

本研究では, 多値分類問題におけるより一般的なコスト考慮型学習のための手法として, ECOC 多値分類手法にコスト考慮型学習を導入することを考える. ECOC 多値分類手法は, カテゴリ集合の構成を変えた複数の二値分類器を組み合わせることで精度の高い多値分類を実現する手法であるため, コスト考慮型学習で各誤分類による損失の最小化を目指すためには拡張が必要となる. ECOC 多値分類手法は, 用いる二値分類器の種類を問わない手法である. そのため, ECOC 多値分類手法にコスト考慮型学習を導入する際には, 二値分類器の種類を問わず, 各二値分類器が多値分類に与える影響を考慮しながら, 誤分類による全体の損失の最小化を目指す必要がある. また, 最終的な多値分類は各二値分類器の出力を組み合わせるため, 各分類器の出力をどのように統合し, より誤分類による損失を減少させるかについても検討の余地がある.

一方, 様々な分類手法に対して精度の高い分類器を学習する手法として, 分類誤り数損失を平滑な関数で近似しその損失の最小化を行う MCE 学習法が提案されている.

そこで本研究では、MCE 学習の考えを導入し、各誤分類に対して異なる損失を設定しその損失を最小化するパラメータを求めることで、ECOC 多値分類手法においてコスト考慮型学習を実現する手法を提案する。誤分類による損失をより小さくする分類器を学習するために、ECOC 多値分類手法を 2 段階で捉え、MCE 学習の考えを各段階で用いてパラメータを学習する。具体的には、学習段階では各二値分類器の学習パラメータを MCE 学習によりコストを考慮して学習し、予測段階で用いる各カテゴリに対する各二値分類器の出力の重みについても、MCE 学習により最適化する。

4.2 学習アルゴリズム

コスト考慮型学習では、各誤分類によって異なる損失を設定する。そのため、カテゴリ c_q に属するデータの カテゴリ c_r への誤分類を $M_{q,r}$ と定義し、はじめに各誤分類に対して損失関数のパラメータを設定する。続いて、その損失に基づいた学習データの経験的平均損失を最小化する L 個の二値分類器を学習する。さらに、二値分類器の組み合わせの際に、各分類器の出力に対して適切な重みを与えることで、分類器の学習の枠組みにとらわれずに、経験的平均損失をより小さくすることを考える。

4.2.1 損失関数の設定

多値分類におけるコスト考慮型学習のために、各誤分類の損失関数に関するパラメータの設定を行う。まず、各誤分類に対して共通に平滑化分類誤り数損失のパラメータ β を設定する。合わせて、各誤分類に異なる損失を考慮するため、カテゴリ c_q に属するデータをカテゴリ c_r へ誤分類したときの平滑化分類誤り数損失を $\delta_{q,r}$ ($q, r = 1, \dots, K$) 倍することを考え、 $\delta_{q,r}$ を設定する。

以降、式 (14) の各学習データ \mathbf{x}_i の誤分類尺度 $d(\mathbf{x}_i)$ をもとに、各カテゴリに属する各学習データ \mathbf{x}_i がどのカテゴリに誤分類され易いかに応じて、設定した β , $\delta_{q,r}$ を用いて平滑化分類誤り数損失を式 (15) により求める。

$$d(\mathbf{x}_i) = -F_{y_i}(\mathbf{x}_i) + F_{I_i}(\mathbf{x}_i) \quad (14)$$

$$l(d(\mathbf{x}_i)) = \frac{\delta_{y_i, I_i}}{1 + \exp(-\beta d(\mathbf{x}_i))} \quad (15)$$

I_i は、学習データ \mathbf{x}_i に対して、正解カテゴリ以外でもっとも属する度合いが大きいカテゴリを表す。

$$I_i = \arg \max_{c_k, c_k \neq y_i} F_{c_k}(\mathbf{x}_i) \quad (16)$$

各誤分類 $M_{q,r}$ に対して損失を $\delta_{q,r}$ 倍し MCE 学習を行うことで、 $\delta_{q,r}$ の大きい誤分類をより考慮しながらパラメータを学習することを考える。

経験的平均損失の最小化のために、各二値分類器の学習においては学習パラメータ、各分類器の組み合わせにおいては各分類器の出力に対する重みの最適化を考える。

4.2.2 各分類器の学習 (MCE)

ECOC 多値分類手法において、 L 個の SVM を作成する。ここで、分類器 f_l における学習データ \mathbf{x}_i に対するラグランジュ未定乗数を $\alpha_{i,l}$ と定義する。SVM による分類器 f_l の出力値は $\alpha_{i,l}$ を用いて、式 (17) で表される。

$$D_l(\mathbf{x}) = \sum_{i=1}^N (\alpha_{i,l} t_i \mathbf{x}_i^T \mathbf{x}) + b \quad (17)$$

いま、事前に設定した β , $\delta_{q,r}$ を用いて、最適なパラメータ $\alpha_{i,l}$ を求めることを考える。式 (8) の尺度により、各学習データ \mathbf{x}_i に対し $F_{c_k}(\mathbf{x}_i)$ を求め、式 (14) から誤分類尺度 $d(\mathbf{x}_i)$ を求める。

また、本研究では、各分類器に対して $\sum_{i=1}^N \alpha_{i,l}$ の値は変化させずに、式 (18) で $\omega_{i,l}$ を定義し、各分類器において $\sum_{i=1}^N \omega_{i,l} = 1$ という条件を満たしながら経験的平均損失を最小化する $\omega_{i,l}$ を求めることを考える。

$$\omega_{i,l} = \frac{\alpha_{i,l}}{\sum_{i=1}^N \alpha_{i,l}} \quad (18)$$

以上の設定を用いて、以下のステップで最適なパラメータ $\omega_{i,l}$ を求める。

Step1) SVM の結果として得られた $\alpha_{i,l}$ から $\omega_{i,l}$ を式 (18) を用いて算出し、これを初期値とする。

Step2) 各学習データに対し、式 (14) の誤分類尺度から式 (15) により平滑化分類誤り数損失を求め、その平均を経験的平均損失とする。

Step3) Step2 で求めた経験的平均損失を $\omega_{i,l}$ に関して偏微分する。式 (13) に基づき、 $\sum_{i=1}^N \omega_{i,l} = 1$ を満たすよう、パラメータの更新を行う。

Step4) Step3 で求めたパラメータを用いて経験的平均損失を求め、収束していなければ、Step2 へ戻る。収束していれば、その時点におけるパラメータ $\omega_{i,l}$ を学習パラメータとする。

4.2.3 各分類器の出力に対する重みの学習 (MCE)

ECOC 多値分類手法では、複数の二値分類器を組み合わせることで多値分類を実現する。このため、各カテゴリに対する各分類器の出力に適切な重みを与えることで、さらに経験的平均損失を最小化することを考える。

いま、カテゴリ c_k に対する分類器 f_l の出力の重み $a_{c_k,l}$ を用いて、カテゴリ c_k の識別関数を式 (19) で定義する。

$$F_{c_k}(\mathbf{x}) = \sum_{l=1}^L a_{c_k,l} D_l(\mathbf{x}) g_{c_k,l} \quad (19)$$

事前に設定した β , $\delta_{q,r}$ を用いて最適な重み $a_{c_k,l}$ を求めることを考える。各学習データ \mathbf{x}_i に対し、式 (19) の識別関数から $F_{c_k}(\mathbf{x}_i)$ を求め、式 (14) により誤分類尺度を求める。以上の設定を用いて、以下のステップで最適なパラメータ $a_{c_k,l}$ を求める。

Step1) 各分類器の各カテゴリに対する重みを等しく設定するために各 c_k , 各 l に対して共通に $a_{c_k,l} = 1$ を初期値とする。

Step2) 各学習データに対し、式 (14) の誤分類尺度から式 (15) で平滑化分類誤り数損失を求め、その平均を経験的平均損失とする。

Step3) Step2 で求めた経験的平均損失を $a_{c_k,l}$ に関して偏微分する。式 (13) に基づき、 $\sum_{l=1}^L a_{c_k,l} = L$ を満たすようにパラメータの更新を行う。

Step4) Step3 で求めたパラメータを用いて経験的平均損失を求め、収束していなければ、Step2 へ戻る。収束していれば、その時点における $a_{c_k,l}$ を各分類器の各カテゴリに対する重みとする。

4.3 提案手法による分類

MCE 学習により求めたパラメータを用いて、式 (19) の識別関数にテストデータ \mathbf{x} を代入した値を求め、その値が最大となるカテゴリへ分類を行う。

5 実験

提案手法の有効性を検証するため、ベンチマークデータセットに対して分類実験を行い、提案手法の評価を行った。

5.1 実験条件

本実験では、ベンチマークデータセットとして用いられる UCI 機械学習レポジトリから 3 種類を使用した。データセットの概要を表 1 に示す。データセットの 75% を学習データ、残りの 25% をテストデータにランダムに分割し、同様の分割を行い、各データセットに対して 10 回の分類実験を繰り返し、平均により評価する。ECOC 多値分類手法の分類器構成には、Exhaustive 符号 [1] を用いた。

また、本実験では提案手法におけるパラメータの設定による挙動を確認するために、各データセットにおいて、通常の ECOC SVM による分類で最も誤分類しやすい組み合わせの損失に関するパラメータを変化させ結果を示す。具体的には、最多誤分類が $M_{q,r}$ であるデータセットに対し、 $\delta_{q,r}$ のみを 1 から 9 まで 2 刻みで変化させ、それ以外の δ を 1 と設定した際の分類結果の推移を示す。

指標としては、 $\delta_{q,r}$ を変化させる最多誤分類 $M_{q,r}$ の誤分類数とその対となる $M_{r,q}$ の誤分類数、その他の誤分類の総数の 10 回平均を用い、 $\delta_{q,r}$ を変化させたときの推移から手法の有効性を検証する。また、シグモイド関数のパラメータは $\beta = 10$ とした。

表 1. データセット概要

データセット名	K	d	データ数	最多誤分類
Iris	3	4	150	$M_{2,1}$
Wine	3	13	178	$M_{2,3}$
Vehicle	4	18	846	$M_{4,2}$

5.2 実験結果と考察

各データセットの結果を図 3, 図 4, 図 5 に示す。

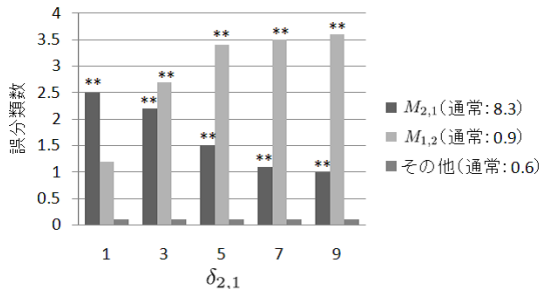


図 3. Iris 実験結果

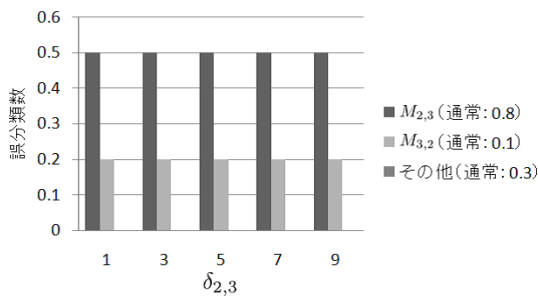


図 4. Wine 実験結果

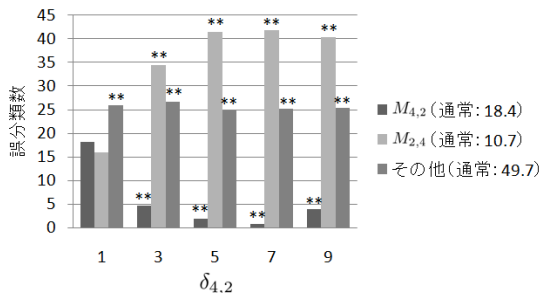


図 5. Vehicle 実験結果

各図において、括弧内に記載した数字は、コスト考慮型学習を考えず通常の ECOC SVM を行った際の誤分類数の 10 回平均である。また、各誤分類に対して、通常の ECOC SVM と提案手法で有意に差があるか t 検定を行った。各図において、**は 1% 有意を示す。

図 3, 図 5 より、データセット Iris, Vehicle において、それぞれ対応する $\delta_{q,r}$ を大きく設定することで、 $M_{q,r}$ の誤分類数が減少し、同時に $M_{r,q}$ の誤分類数が増加するような多値分類が行われ易くなると考えられる。これは、 $M_{q,r}$ の誤分類数を減少させるようにパラメータの学習を行うことで、 c_q と c_r の 2 カテゴリ間で誤分類されやすいデータがカテゴリ c_q に分類されやすくなるためであると考えられる。さらに、 $\delta_{q,r}$ を変化させた場合の $M_{q,r}$ と $M_{r,q}$ 以外のその他の誤分類数は、通常の ECOC SVM の誤分類数と比較して改善され、さらに $\delta_{q,r}$ の変化による影響はあまり見られなかった。このことから、カテゴリ集合を分類するような分類器を組み合わせる ECOC 多値分類手法において、ある特定の誤分類にのみ着目し、他のカテゴリの分類に悪影響を与えずにパラメータを学習することができたと考えられる。

また、図 4 より、データセット Wine に関しては、 $\delta_{2,3}$ を変化させた際の分類結果に差異が見られなかった。これは、データセット Wine が通常の ECOC SVM においてすでに誤分類数が少なく、提案手法による改善の余地が大きくなかったためであると考えられる。

6 まとめと今後の課題

本研究では、ECOC 多値分類手法に対するコスト考慮型学習の導入を考え、より一般的なコスト考慮型学習のための手法を提案した。平滑化分類誤り数損失を定義しその最小化を目指す MCE 学習の考えを採用し、分類器の学習段階では各二値分類器の学習パラメータを、分類器の組み合わせの段階では各分類器の各カテゴリへの出力に対する重みの最適化を行った。検証実験の結果、提案手法を用いることにより、設定した誤分類による損失に基づいた分類結果が得られることを示すことができた。

今後の課題として、適切なパラメータの設定法や、SVM 以外の二値分類の手法を用いた際の有効性の検証が挙げられる。また、本実験ではある特定の誤分類に対しより大きな損失を与えることを考えたが、より複雑に誤分類による損失を設定する場合の有効性の検証も必要である。

参考文献

- [1] T. G. Dietterich, G. Bakiri, "Solving Multiclass Learning Problems via Error-Correcting Output Codes," *Journal of Artificial Intelligence Research*, vol. 2, pp. 263–286, 1995.
- [2] Z. H. Zhou, X. Y. Liu, "On Multi-Class Cost-Sensitive Learning," *Computational Intelligence* vol. 26, no. 3 pp. 232–257, 2010.
- [3] C. Elkan, "The Foundations of Cost-Sensitive Learning," *International Joint Conference on Artificial Intelligence*, vol. 17, no. 1, pp. 973–978, 2001.
- [4] H. He, E. A. Garcia, "Learning from Imbalanced Data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.
- [5] B. H. Juang, S. Katagiri, "Discriminative Learning for Minimum Error Classification," *IEEE Transactions on Signal Processing*, vol. 40, pp. 3043–3054, 1992.
- [6] C. Cortes, V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.