Vol.2020-MPS-131 No.9
2020/12/17

多値分類問題における ECOC 法の最適性に関する一考察

雲居 玄道[1,a)]　八木 秀樹[2]　小林　学[1]　後藤 正幸[1]　平澤 茂一[1]

概要：与えられた二値判分類器を組合せて用いる多値分類器の構成法の 1 つに，符号理論の枠組みを導入した誤り訂正符号に基づく多値分類法（Error-Correcting Output Coding：ECOC 法）がある．この手法が実データに対して良い性能を示すことは実験的に知られているが，ECOC 法に対する分類精度について，理論的な最適性については明らかになっていない．そこで本研究では最大事後確率分類を可能とする二値分類器を仮定した場合，ECOC 法が最適な多値分類法になる十分条件を示す．この結果，同様の仮定のもとで n-vs-all 及び Exhaustive 符号が最適な多値分類法になることが示せる．これは種々の ECOC 法に対する最適性の議論の方向性の一つを示唆している．

キーワード：Multi-valued Classification，Error-Correcting Output Coding，最適性，最大事後確率

A Study on the Optimization of the ECOC Method for Multi-label Classification Problems

Gendo Kumoi[1,a)]　Hideki Yagi[2]　Manabu Kobayashi[1]　Masayuki Goto[1]　Shigeichi Hirasawa[1]

Abstract: One of the methods for constructing a multi-valued classifier that uses a combination of given two-valued classifiers is the Error-Correcting Output Coding (ECOC) method, which is based on error-correcting codes introducing a code theory framework. Although it is experimentally known that this method performs well on real data, the theoretical optimality of the classification accuracy for the ECOC method has not been clarified. In this study, we show sufficient conditions for the ECOC method to be an optimal multi-valued classification method under the assumption that binary classifiers achieve maximum posterior probability classification. As a result, we can show that n-vs-all and Exhaustive signs are the best multi-valued classification method under the same assumptions. This suggests one of the directions of the optimization debate for various ECOC methods.

Keywords: Multi-valued Classification, Error-Correcting Output Coding, Optimization, Maximum a Posteriori

1. はじめに

膨大なデータからの知識獲得のアプローチとして，分類問題がある．分類問題は，データからカテゴリを推定する問題である．この分類問題は，カテゴリ数 $M=2$ のとき二値分類問題，$M \geq 3$ は多値分類問題と呼ばれる．

1　早稲田大学
Waseda University, Shinjuku, Tokyo 169–8555, Japan.
2　電気通信大学
The University of Electro-Communications, Chofu, Tokyo 182–8585, Japan.
a)　moto-aries@ruri.waseda.jp

二値分類問題に対して Support Vector Machine (SVM) [1], [2]，AdaBoost [3]，Relevance Vector Machine (RVM) [4]，Regularized Least-Squares Classification (RLSC) [5], [6] など，精度の高い二値分類器がある．しかし，この二値分類器を多値分類器に拡張しようとする場合には，空間，時間計算量が膨大となる問題がある [2], [7]．その上，これらを解決しようとすると精度が低下する問題が生じ [8]，分類器の構築が難しい．

そこで，多値分類器構築のもう 1 つのアプローチとして，符号理論の枠組みを導入した誤り訂正符号に基づく多値分類法（Error-Correcting Output Coding Method：ECOC

法）[9] がある．ECOC 法は，二値分類器の組合せにより構成される多値分類器である．

二値判別器を組合せて多値分類器を構成する手法において，最もよく知られた方法として，“one-vs-all” (OVA) と呼ばれる方法 [10] がある．OVA は，1 つのカテゴリとそれ以外を識別する二値分類器を全通り組合せて構築する方法である．また，組合せた二値分類器の出力からカテゴリを推定する手法にも，様々な手法が存在する [11], [12], [13], [14]．ニューラルネットワークやディープラーニングでは，OVA が用いられていると見なすことができる．

一般に ECOC 法では二値分類器の構成及びカテゴリの推定手法の組合せにより分類精度が大きく変化することが知られている．その上で，二値分類器に RLSC を用いた場合 [10] や二値分類器の出力に基づく重み付けを用いた場合 [15] に，二値分類器の組合せと分類精度について実験的に明らかにしている研究がある．一方で，OVA を含め，ECOC 法に対する分類精度について，理論的な最適性については明らかになっていない．

そこで本研究では最大事後確率分類を可能とする二値分類器を仮定した場合，ECOC 法が最適な多値分類法になる十分条件を示す．この結果，同様の仮定のもとで “n-vs-all” (nVA) 及び “Exhaustive 符号”などが最適な多値分類法になることが示せる．これは種々の ECOC 法に対する最適性の議論の方向性の 1 つを示唆している．

また，これらの結果を踏まえて人工データ及び実データを用いて本研究で与える十分条件に基づく分類を行い，知見の有効性を示す．

2. 多値分類問題

分類問題は，与えられた入力データ $\boldsymbol{x}$ に対応するカテゴリ c_i ($i \in C \coloneqq \{1, 2, \ldots, M\}$) を推定する問題である．$M$ はカテゴリ数を表し，$M \geq 3$ の場合，この問題は多値分類問題と呼ばれる．

各カテゴリの事後確率が推定できると仮定した場合，事後確率を

$$P_i \coloneqq P(c_i|\boldsymbol{x}) \tag{1}$$

と定義する．このとき，最大事後確率におけるカテゴリ $c_{\hat{y}}$ は，

$$\hat{y}_{\mathrm{MAP}} = \arg\max_{i \in C} P_i \tag{2}$$

と推定される．

最大事後確率分類は，分類誤り率を最小にすることができるという意味で最適である．

3. Error-Correcting Output Coding Method：ECOC 法

ECOC 法は，符号理論の枠組みを導入した誤り訂正符号に基づく多値分類法 [9] である．この手法では，事前 [9], [10], [14] または逐次的 [16] に符号語表と呼ばれる二値分類器の構成法が与えられる．この符号語表に基づき構成された各二値分類器の出力をもとにカテゴリが推定される．

3.1 符号語表

長さ N の符号語を M 個，行として並べた $M \times N$ 行列 W を符号語表と呼ぶ．行列 W，行はカテゴリを表し，列は二値分類器の構成を意味している．このとき，行列 W の (i, j) 成分を w_{ij}，カテゴリを表す行ベクトルを $\boldsymbol{w}_i$，二値分類器の構成である列ベクトルを $\boldsymbol{w}^j$ と表す．w_{ij} がとる値の集合には，{0,1} の 2 元を仮定する場合や，学習に用いないという意味で，*を導入し，{0,*,1} の 3 元を仮定する場合 [17] などがある．本研究では $w_{ij} \in \{0, 1\}$ が 2 元のケースを仮定する．

符号語表 W の j 列目 f_j ($j \in F \coloneqq \{1, 2, \ldots, N\}$) は，$w_{ij}$ が 0 となるカテゴリと w_{ij} が 1 となるカテゴリに分ける二値分類器となる．

3.2 カテゴリ推定

二値分類器は，N 個存在し，各二値分類器の出力を $f_j(\boldsymbol{x})$ と表す．Dietterich ら [9] は，$f_j(\boldsymbol{x}) \in \{0, 1\}$ のとき，

$$g(c_i|\boldsymbol{x}) = \sum_{j=1}^{N} w_{ij} \oplus f_j(\boldsymbol{x}) \tag{3}$$

によるハミング距離に基づく硬判定法を用いている．ここで，$\oplus$ は排他的論理和を表す．この二値分類器の出力は，分類器ごとに定められたしきい値をもとにカテゴリを判定した結果である．そのため，判定前の値を用いて軟判定を行うことが可能となる．

そこで，Allwein ら [11] は，分類器に，$w_{ij} = 1$ の時に正，$w_{ij} = 0$ のときに負をとる SVM を用いた．いま，出力が $f_j(\boldsymbol{x}) \in \mathbb{R}$ のとき，

$$g(c_i|\boldsymbol{x}) = \sum_{j=1}^{N} \left(1 - (2w_{ij} - 1) f_j(\boldsymbol{x})\right)_+ \tag{4}$$

とマージンによる軟判定を用いている．ここで，$(z)_+ = \max\{z, 0\}$ である．つまり，式 (4) は SVM の出力 $|f_j(\boldsymbol{x})|$ の大きさが推定の信頼度を表すと想定した推定法である．

これらに対して，Passerini ら [13] は，分類器の出力が $f_j(\boldsymbol{x}) \in [0, 1]$ となる二値分類器を仮定している．特に，$f_j(\boldsymbol{x})$ は，$w_{ij} = 1$ となる事後確率を表すものとし，

$$g(c_i|\boldsymbol{x}) = \prod_{j=1}^{N} f_j(\boldsymbol{x})^{w_{ij}} (1 - f_j(\boldsymbol{x}))^{1 - w_{ij}} \tag{5}$$

とおく．このとき，入力データ $\boldsymbol{x}$ のカテゴリ $c_{\hat{y}}$ は，

$$\hat{y} = \arg\max_{i} g(c_i|\boldsymbol{x}), \tag{6}$$

と推定される．

4. 問題設定

4.1 符号語表の性質

本研究では，符号語表 W において，2 元のケースを対象としている．このとき，ECOC 法においては一般性を失うことなく，W の性質として以下を仮定して良い．

仮定 1. $w_{ij} \in \{0, 1\}$.

仮定 2. カテゴリごとの符号語は全て異なる．

仮定 3. 同一の二値分類器が 2 つ以上存在しない．

二値分類器は，カテゴリの組合せが同じである場合，同一のものとなる．よって，仮定 3 は，W の列において，0,1 の並びが同じものや，0,1 を反転すると同じ列になるものが存在しないことを意味する．いま f_j と f_l の二値分類器間のハミング距離 d_H を，

$$d_H(j,l) = \sum_{i=1}^{M} w_{ij} \oplus w_{il} \tag{7}$$

とするとき，仮定 3 から符号語表 W は $j \neq l$ となる任意の $j, l \in F$ について，

$$\begin{aligned} d_H(j,l) &\neq 0, \\ d_H(j,l) &\neq M \end{aligned} \tag{8}$$

を満たす．仮定 1, 3 のもと，二値分類器 f_j $(j \in F)$ で同一のカテゴリとみなされるカテゴリの組からなる集合族 $\mathcal{T}_j$ を以下のように定義する．

定義 1. 各 $j \in F$ について

$$\mathcal{T}_j \coloneqq \{\{i \in C | w_{ij} = 1\}, \{i \in C | w_{ij} = 0\}\}. \tag{9}$$

仮定 1 より各二値分類器において，w_{ij} は 0 または 1 の値をとる． 加えて，仮定 3 より，同一の二値分類器は存在しないため，以下の定理が成り立つ．

定理 1. $\mathcal{T}_j$ の要素は全て異なる．すなわち，$j \neq j'$ を満たす任意の $j, j' \in F$ について，集合 $T \in \mathcal{T}_j, T' \in \mathcal{T}_{j'}$ ならば，$T \neq T'$ となる．

このとき，符号語表 W に対応する $\mathcal{S}$ を以下のように定義する．

定義 2.

$$\mathcal{S} \coloneqq \bigcup_{j \in F} \mathcal{T}_j. \tag{10}$$

各 $j \in F$ について，必ず $|\mathcal{T}_j| = 2$ が成り立つ．これより，$|\mathcal{S}|$ に関する以下の定理が成り立つ．ただし，$|\cdot|$ は要素数を表す．

定理 2. 仮定 1, 3 を満たす W に対して，以下が成り立つ．

$$\begin{aligned} |\mathcal{S}| &= \sum_{j=1}^{N} |\mathcal{T}_j| \\ &= 2N. \end{aligned}$$

また，W と $\mathcal{S}$ は，相互に変換可能であり，以下の定理が成り立つ．

定理 3. 仮定 1, 3 を満たす W に対して，W と $\mathcal{S}$ は符号構成を一意に定めるという意味で等価である．

4.2 対象とするカテゴリ推定法

本研究では多値分類問題に対して，仮定 1–3 を満たす符号語表 W を対象とする．各二値分類器の事後確率を以下のように定義する．

定義 3. 各 $j \in F$ に対し，

$$f_j(\boldsymbol{x}) \coloneqq \sum_{i=1}^{M} w_{ij} P_i. \tag{11}$$

この定義のもと，本研究では，以下を仮定する．

仮定 4. 式 (1) を直接計算することは難しい．一方，2 値分類器は最適で，2 値カテゴリに対する事後確率を求めることが可能である．すなわち，式 (11) の $f_j(\boldsymbol{x})$ を計算可能とする．

この仮定のもとでは，二値分類器の出力は，Passerini ら [13] らの ECOC 法と同様に $f_j(\boldsymbol{x}) \in [0, 1]$ となる．定義 2, 3 から，式 (5) は，

$$\begin{aligned} g(c_i|\boldsymbol{x}) &= \prod_{j=1}^{N} f_j(\boldsymbol{x})^{w_{ij}} (1 - f_j(\boldsymbol{x}))^{1-w_{ij}} \\ &= \prod_{j=1}^{N} \left(\sum_{i'=1}^{M} w_{i'j} P_{i'} \right)^{w_{ij}} \left(\sum_{i'=1}^{M} (1 - w_{i'j}) P_{i'} \right)^{1-w_{ij}} \\ &= \prod_{S \in \mathcal{S}_i} \sum_{s \in S} P_s \end{aligned} \tag{12}$$

となる．ことから，ECOC 法におけるカテゴリ推定を以下のように定義する．

定義 4. 各 $i \in C$ に対し，

$$g(c_i|\boldsymbol{x}) \coloneqq \prod_{S \in \mathcal{S}_i} \sum_{s \in S} P_s. \tag{13}$$

定義 5.

$$ECOC(\boldsymbol{x}) \coloneqq \arg\max_{i \in C} g(c_i|\boldsymbol{x}). \tag{14}$$

また，事後確率最大法による最適な多値分類法を以下のように定義する．

定義 6.

$$MAP(\boldsymbol{x}) \coloneqq \arg\max_{i \in C} P_i. \tag{15}$$

本研究では，定義 5, 6 が一致し，ECOC 法が最適な多値分類法になる符号語表の与え方の十分条件を示すことが目的である．

5. ECOC 法が MAP となる十分条件

本節では，仮定 4 のもとで定義 5, 6 による分類結果が一致し，ECOC 法が最適な多値分類法になる符号語表の与え

方の十分条件を示す．

はじめに，W に対応するカテゴリの組の集合族 $\mathcal{S}$ について以下を定義する．

定義 7 (部分集合族 $\mathcal{S}_i$)**.** $\mathcal{S}$ に対して，$i \in C$ を含む集合のみを取り出した集合族 $\mathcal{S}_i$ を

$$\mathcal{S}_i \coloneqq \{S \in \mathcal{S} | i \in S\}$$

とする．

定義 8 (部分集合族 $\mathcal{S}_{ik}$)**.** $\mathcal{S}$ に対して，$i, k \in C$ を両方含む集合のみを取り出した集合族 $\mathcal{S}_{ik}$ を

$$\begin{aligned}\mathcal{S}_{ik} &\coloneqq \{S \in \mathcal{S} | i, k \in S\} \\ &= \mathcal{S}_i \cap \mathcal{S}_k\end{aligned}$$

とする．

定義 9 (集合族の集合要素置換)**.** いま集合 $A \subseteq C$ の要素 i を k に置換する操作を $\pi_{ik}(A)$ と表し，以下のように定義する．

$$\pi_{ik}(A) \coloneqq \begin{cases} A \cup \{k\} \backslash \{i\}, & i \in A \wedge k \notin A, \\ A, & \text{otherwise.} \end{cases}$$

このとき，集合族 $\mathcal{A}$ に対して，集合族内の集合 A の要素 i を k に置換する操作をして得られる集合族を $(\mathcal{A})_{i \to k}$ と表し，以下のように定義する．

$$(\mathcal{A})_{i \to k} \coloneqq \{\pi_{ik}(A) | A \in \mathcal{A}\}.$$

これらの定義に基づき，符号語表の性質として，以下の対称性を定義する．

定義 10 (対称性)**.** 符号語表 W に対する集合族 $\mathcal{S}$ に対し，任意の $i \neq k$ となる $i, k \in C$ について，

$$(\mathcal{S}_i \backslash \mathcal{S}_{ik})_{i \to k} = \mathcal{S}_k \backslash \mathcal{S}_{ik}$$

が成り立つとき，W を対称性をもつ符号語表と呼ぶ．

対称性をもつ符号語表 W が仮定 2 を満たすとき，任意の $i \neq k$ について $\mathcal{S}_i \backslash \mathcal{S}_{ik} \neq \phi$ が成り立つ．

いま，定義 4, 5 より以下の定理が成り立つ．

定理 4 (十分条件)**.** 対称性をもつ W を用いるとき，任意の入力データ $\boldsymbol{x}$ に対し，$ECOC(\boldsymbol{x}) = MAP(\boldsymbol{x})$ が成り立つ．

6. 対称性をもつ符号語表

OVA とは，M 個のカテゴリを 1 対 $M-1$ に分ける全ての組合せをもつ符号語表のことである．同様に，n-vs-all (nVA) を n 対 $M-n$ に分ける全ての組合せをもつ符号語表とする[*1]．

定義より，nVA の符号語表に対する $\mathcal{S}$ は，任意の $S \in \mathcal{S}$ に対し $|S| = n$ または $|S| = M - n$ となる．これは，S が $C = \{1, 2, \ldots, M\}$ から n 個および $M-n$ 個の要素を抜き出した全ての組合せをもつことを意味している．これより，以下の定理が成り立つ．

定理 5. nVA の符号語表は，$i \neq k$ となる任意の $i, k \in C$ について $(\mathcal{S}_i \backslash \mathcal{S}_{ik})_{i \to k} = \mathcal{S}_k \backslash \mathcal{S}_{ik}$ を満たす．すなわち対称性をもつ．

また，定理 5 より，対称性をもつ符号語表の組合せで与えられる W について，以下の定理が成り立つ．

定理 6. 符号語表 W が nVA の組合せで構成されるとき，W は対称性をもつ．

Exhaustive 符号 [9] とは，カテゴリ数 M に対して，全ての二値分類器の組合せをもつ符号語表のことである．よって，以下の系が成り立つ．

系 1. Exhaustive 符号は，$1 \leq n \leq \lfloor M/2 \rfloor$ を満たす nVA の組合せにより構成される．そのため，符号語表は対称性をもつ．

7. 評価実験

本研究で最適な多値分類法である対称性をもつ符号語表について，人工データ，実データを用いて非対称性と分類誤り率について評価実験を行う．

7.1 実験データ

7.1.1 人工データ

カテゴリ数 M に対して，式 (16) に従う M 次元正規乱数を用いて実験を行う．

$$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_i, \Sigma) = \frac{1}{(\sqrt{2\pi})^M \sqrt{|\Sigma|}} \times \exp\left(-\frac{(\boldsymbol{x} - \boldsymbol{\mu}_i)^\top \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_i)}{2}\right). \quad (16)$$

ここで，$\boldsymbol{\mu}_i = (\mu_{ik} : 1 \leq k \leq M)$ は，カテゴリ c_i ごとの平均ベクトル．分散共分散行列 $\Sigma = (\sigma_{ik}^2 : 1 \leq i, k \leq M)$ は，以下のように与えられる．

$$\mu_{ik} = \begin{cases} 1, & i = k, \\ 0, & \text{otherwise.} \end{cases} \quad (17)$$

$$\sigma_{ik} = \begin{cases} 0.5, & i = k, \\ 0, & \text{otherwise.} \end{cases} \quad (18)$$

これらのパラメータに従い，表 1 のデータを正規乱数より発生させる．

表 1 人工データ
Table 1 Artificial data

カテゴリ数 M	8
データ数	2,000 件/カテゴリ

[*1] W は同一の二値分類器を除き，列重みが n または $M-n$ の全通りからなる符号語表である．

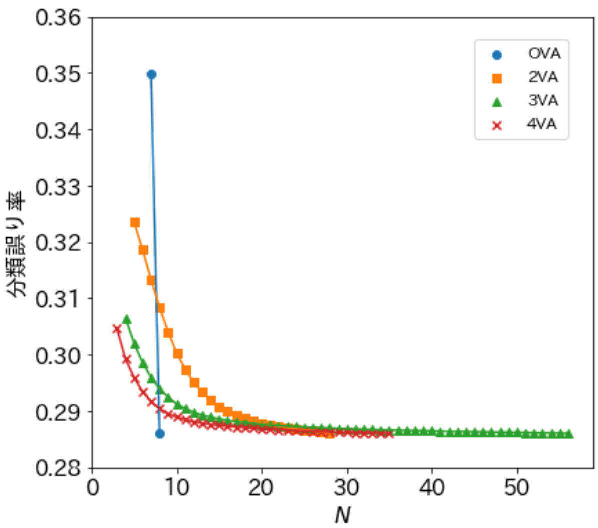

図 1　人工データ
Fig. 1　Artificial data

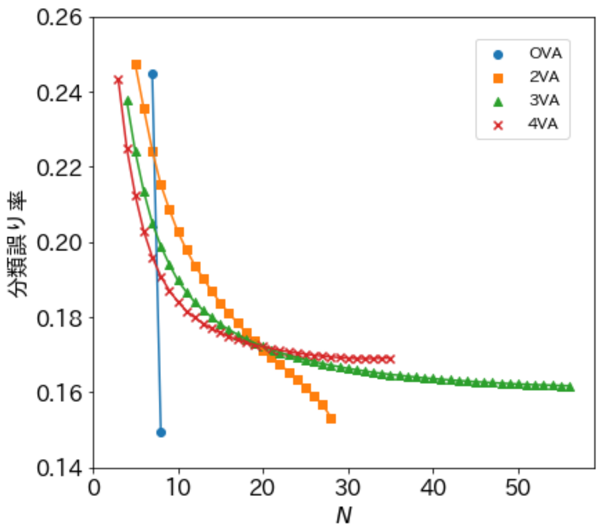

図 2　実データ（SVM）
Fig. 2　Real data (SVM)

7.1.2　実データ

実データにはベンチマークデータとして，2015 年の読売新聞 [18] の記事を用いる．このデータは，文書分類問題のベンチマークデータとして広く用いられるものである．読売新聞は，経済，スポーツなどの記事の掲載面レベルでのカテゴリがラベル付けされる．10 カテゴリの大分類からデータ数の大きい順にデータを抽出し，カテゴリ数を $M = 8$ とした．

2015 年の読売新聞は，表 2 のデータセットに基づき，10 セットから 1 セットを入力データとする実験を 10 パターン行い平均をとる 10 分割ローテーションによって評価する．

7.2　非対称性と分類誤り率

本論文で対称性をもつことが，最適な分類となることを示した．本実験では，この対称性がくずれるにつれて，どのように分類性能が変化するかを評価する．これにより，非対称性と分類誤り率の関係が明らかにする．

7.2.1　実験設定

対称性をもつ符号語表として，OVA，2VA，3VA，4VA を比較する．このとき，nVA から列（二値分類器）を削除すると対称性が保たれなくなる．そこで，仮定 2 を満たす範囲で nVA からランダムに列を削除し，誤り率を計算し比較する．このとき，削除する列数ごとに 10,000 回ランダムに繰り返しその誤り率の平均値で評価する．

実験データには，人工データおよび実データを用いる．また，実データの二値分類器には，出力が $f_j(\boldsymbol{x}) \in [0, 1]$ となる Logistic Regression [19]，Gaussian SVM [20] を用いる．

図 1，2，3 に実験結果を示す．

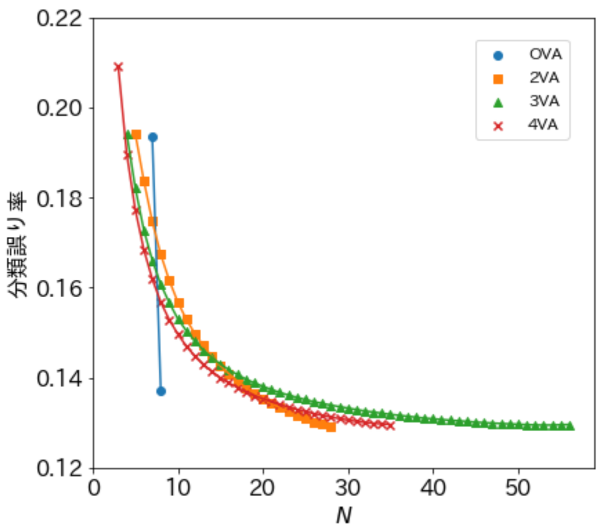

図 3　実データ（ロジスティクス回帰）
Fig. 3　Real data (Logistics regression)

7.2.2　考察

OVA では 1 列削除したもののみが仮定 2 を満たし，2 列以上削除するとカテゴリ間で同じ符号語が存在する．そのため，OVA は 1 列削除のみを実験した．図 1，2，3 より，OVA においては，1 列削除した場合，大きな分類誤り率の上昇が見られた．これは，$\boldsymbol{w}_i$ が全 0 ベクトルまたは全 1 ベクトルとなる符号語をもつカテゴリの誤り率が上昇したためと考えられる．

人工データは，事後確率が計算可能なため，最大事後確率分類を可能とする二値分類器となっている．図 1 より，削除する列数を増やすごとに，2VA，3VA，4VA では徐々に誤り率が上昇していくことがわかる．しかし，その上昇は緩やかであるため，本研究の対称性の条件を緩和できる可能性を示している．一方で，実データでは，図 2，3 よ

表 2　ベンチマークデータ (2015 年読売新聞)
Table 2　Benchmark data (Yomiuri Newspaper, 2015)

カテゴリ（数）	政治，経済，スポーツ，社会，文化，生活，犯罪事件，科学 (8)
文書の特徴ベクトル（次元）	形態素解析による名詞・動詞抽出：頻度 30 以上（7,432）
実験データ数	合計 12,000 件
データセット	150 件/カテゴリ ×10 セット，合計 12,000 件

表 3　分類誤り率（人工データ）
Table 3　Classification error rate (Artificial data)

手法	MAP	ECOC				
		OVA	2VA	3VA	4VA	Exhaustive 符号
誤り率	0.286	0.286	0.286	0.286	0.286	0.286
符号長 (N)		8	28	56	35	127

り，削除する列数を増やすごとに，急激の誤り率の上昇が見られた．このことから，本研究の対称性は，分類誤り率を低下させる条件として，非常に重要であることがわかる．

また，図 1，2，3 について，$N = 10$ の誤り率は，4VA，3VA，2VA の順に高くなる．このことから，nVA において，n が大きいほど，誤り率に対して頑健な分類器となっていることがわかる．

8.　まとめと今後の課題

本研究では最大事後確率分類を可能とする二値分類器を用いた場合，ECOC 法が最適な多値分類法になる十分条件を示した．この十分条件により，従来よりよい性能を示すとされてきた，OVA や Exhaustive 符号が最大事後確率分類となることを論理的に示した．

加えて，評価実験により対称性がくずれるにつれて生じる分類性能の変化を示し，非対称性と分類誤り率の関係を示した．これにより，ECOC 法による分類誤り率に対して，対称性をもつことの重要性を示すことができた．

本研究で仮定したカテゴリ推定法以外にも，様々な推定法が存在する．今後は，他の推定法に対して，本論文で示した十分条件が成り立つかを検討する必要がある．

参考文献

[1] Vapnik, V. N.: *The Nature of Statistical Learning Theory*, Springer-Verlag, Berlin, Heidelberg (1995).
[2] Vapnik, V. N. and Vapnik, V. A.: *Statistical Learning Theory*, Wiley (1998).
[3] Freund, Y. and Schapire, R. E.: A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, *J. Comput. System Sci.*, Vol. 55, No. 1, pp. 119–139 (1997).
[4] Tipping, M. E.: Sparse Bayesian Learning and the Relevance Vector Machine, *J. Mach. Learn. Res.*, Vol. 1, No. Jun, pp. 211–244 (2001).
[5] Rifkin, R. M.: Everything Old Is New Again : a Fresh Look at Historical Approaches in Machine Learning, PhD Thesis, Massachusetts Institute of Technology (2002).
[6] Rifkin, R., Yeo, G. and Poggio, T.: Regularized Least-Squares Classification, *Nato Science Series Sub Series III* (2003).
[7] Weston, J., Watkins, C. and Others: Support Vector Machines for Multi-Class Pattern Recognition, *Esann*, Vol. 99, pp. 219–224 (1999).
[8] Crammer, K. and Singer, Y.: On the Algorithmic Implementation of Multiclass Kernel-based Vector Machines, *J. Mach. Learn. Res.*, Vol. 2, No. Dec, pp. 265–292 (2001).
[9] Dietterich, T. G. and Bakiri, G.: Solving Multiclass Learning Problems via Error-Correcting Output Codes, *J. Artif. Intell. Res.*, Vol. 2, pp. 263–286 (1994).
[10] Rifkin, R. and Klautau, A.: In Defense of One-Vs-All Classification, *J. Mach. Learn. Res.*, Vol. 5, No. Jan, pp. 101–141 (2004).
[11] Allwein, E. L., Schapire, R. E. and Singer, Y.: Reducing Multiclass to Binary: A Unifying Approach for Margin Classifiers, *J. Mach. Learn. Res.*, Vol. 1, No. Dec, pp. 113–141 (2000).
[12] 山口暢彦：WLS-ECOC における事後確率の推定誤差を用いたエラー訂正符号の生成法，電子情報通信学会論文誌 D，Vol. 89, No. 2, pp. 371–380 (2006).
[13] Passerini, A., Pontil, M. and Frasconi, P.: New Results on Error Correcting Output Codes of Kernel Machines, *IEEE Trans. Neural Netw.*, Vol. 15, No. 1, pp. 45–54 (2004).
[14] Escalera, S., Pujol, O. and Radeva, P.: Error-Correcting Ouput Codes Library, *J. Mach. Learn. Res.*, Vol. 11, pp. 661–664 (2010).
[15] 山口暢彦，石井直宏：ECOC における事後確率の推定誤差の上界，電子情報通信学会技術研究報告. NC, ニューロコンピューティング，Vol. 102, No. 730, pp. 149–154 (2003).
[16] Rocha, A. and Goldenstein, S. K.: Multiclass from Binary: Expanding One-Versus-All, One-Versus-One and ECOC-Based Approaches, *IEEE Trans Neural Netw Learn Syst*, Vol. 25, No. 2, pp. 289–302 (2014).
[17] Nilsson, N. J.: Learning Machines. (1965).
[18] 読売新聞社：2015 年　読売新聞記事データ集（邦文），日外アソシエーツ (2016).
[19] Yu, H.-F., Huang, F.-L. and Lin, C.-J.: Dual Coordinate Descent Methods for Logistic Regression and Maximum Entropy Models, *Machine Learning*, Vol. 85, No. 1-2, pp. 41–75 (2011).
[20] Platt, J. et al.: Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods, *Advances in Large Margin Classifiers*, Vol. 10, No. 3, pp. 61–74 (1999).