

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

受付日 2020年11月19日, 再受付日 2021年1月7日,
採録日 2021年1月20日

概要: 与えられた二値判分類器を組み合わ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 および 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1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 YAGI² MANABU KOBAYASHI¹ MASAYUKI GOTO¹ SHIGEICHI HIRASAWA¹

Received: November 19, 2020, Revised: January 7, 2021,
Accepted: January 20, 2021

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.

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]. そのうえ, これらを解決しようとする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,

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

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、二値分類器の組み合わせにより構成される多値分類器である。

二値判別器を組み合わせる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, 1VA) 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み合わせる構築する方法である。また、組み合わ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およびカテゴリの推定手法の組み合わ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。そのうえで、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、二値分類器の組み合わ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) および “Exhaustive 符号” [9] 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、これらの結果をふまえて人工データおよ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 \mathbf{x} に対応するカテゴリ c_i ($i \in C := \{1, 2, \dots, M\}$) 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

$$P_i := P(c_i | \mathbf{x}) \tag{1}$$

と定義する。このとき、最大事後確率におけるカテゴリ $c_{i_{\text{MAP}}}$ は、

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

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 の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} 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 元の場合を仮定する。

符号語表 W の j 列目により構成される二値分類器を f_j ($j \in F := \{1, 2, \dots, N\}$) とする。このとき f_j 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。 w_{ij} が 2 元の場合の代表例として、以下に、 $M = 3$ のときの、OVA の例を示す。

例 1 (one-vs-all (OVA)).

$$W = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

このとき、二値分類器 f_1 は、カテゴリ c_1 とカテゴリ c_2, c_3 を分類す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、入力データ \mathbf{x} に対する各二値分類器の出力を $f_j(\mathbf{x})$ と表す*1。Dietterich ら [9] は、 $f_j(\mathbf{x}) \in \{0, 1\}$ のとき、

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

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、 $w_{ij} = 1$ のときに正、 $w_{ij} = 0$ のときに負をとる SVM を用いた。いま、出力が $f_j(\mathbf{x}) \in \mathbb{R}$ のとき、

*1 f_j に対して、入力データ \mathbf{x} に対する出力を行う識別関数を $f_j(\mathbf{x})$ と書くことにする。

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

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 $|f_j(\mathbf{x})|$ 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 $f_j(\mathbf{x}) \in [0, 1]$ となる二値分類器を仮定している．特に、 $f_j(\mathbf{x})$ は、 $w_{ij} = 1$ となる事後確率を表すものとし、

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

とおく．このとき、入力データ \mathbf{x} のカテゴリ c_i は、

$$\hat{i} = \arg \max_i g(c_i|\mathbf{x}), \quad (6)$$

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 元の場合を対象としている．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よい．

仮定 1. $w_{ij} \in \{0, 1\}$ ．

仮定 2. カテゴリごとの符号語はすべて異なる．

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、

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

とすると、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、

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

を満たす．一方、例 2 に示すように、二値分類器の構成を示す列の 0, 1 の並びが同じもの（ハミング距離が 0）、および、0, 1 が反転したもの（ハミング距離が M ）は、二値分類器として同一のものである．

例 2. (同一の二値分類器)

$$W = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

ここで $d_H(1, 2) = 0$ より、 f_1 と f_2 は同一の二値分類器となり、 $d_H(1, 3) = 4 = M$ より、 f_1 と f_3 は同一の二値分類器となる．

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見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

$$\mathcal{T}_j := \{\{i \in C | w_{ij} = 1\}, \{i \in C | w_{ij} = 0\}\}. \quad (9)$$

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 の要素はすべて異なる．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 S を以下のように定義する．

定義 2.

$$S := \bigcup_{j \in F} \mathcal{T}_j. \quad (10)$$

各 $j \in F$ について、必ず $|\mathcal{T}_j| = 2$ が成り立つ．これより、 $|S|$ 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．

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

また、 W と 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 S は符号構成を一意に定めるという意味で等価である．

(証明) 付録 A.1 を参照．

ここで、 W と S の関係を例 3 の符号語表をもとに示す．

例 3 (2-vs-all, 2VA)．

$$W = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

この W において、 f_j に対応する集合族 \mathcal{T}_j は、

$$\begin{aligned} \mathcal{T}_1 &= \{\{1, 2\}, \{3, 4\}\}, \\ \mathcal{T}_2 &= \{\{1, 3\}, \{2, 4\}\}, \\ \mathcal{T}_3 &= \{\{1, 4\}, \{2, 3\}\} \end{aligned}$$

となる．よって、

$$S = \{\{1, 2\}, \{3, 4\}, \{1, 3\}, \{2, 4\}, \{1, 4\}, \{2, 3\}\}$$

となる. また, $\mathcal{T}_1 = \{\{1, 2\}, \{3, 4\}\}$ に対応する列ベクトル \mathbf{w}^1 は, $\mathbf{w}^1 = (1, 1, 0, 0)^\top$ または, $\mathbf{w}^1 = (0, 0, 1, 1)^\top$ となる. どちらでも同一の二値分類器であるので, S が与えられれば, 同じ二値分類の組合せという意味で等価な W が得られることが分かる.

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,

$$f_j(\mathbf{x}) := \sum_{i=1}^M w_{ij} P_i. \quad (11)$$

つまり, OVA について考えると以下のようになる.

例 4. 例 1 で与えた W に対し,

$$\begin{aligned} f_1(\mathbf{x}) &= P_1, \\ 1 - f_1(\mathbf{x}) &= 1 - P_1 \\ &= P_2 + P_3. \end{aligned}$$

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, 二値分類器は最適で, 二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 $f_j(\mathbf{x})$ 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 $f_j(\mathbf{x}) \in [0, 1]$ となる. 定義 2, 3 から, 式 (5) は,

$$\begin{aligned} g(c_i|\mathbf{x}) &= \prod_{j=1}^N f_j(\mathbf{x})^{w_{ij}} (1 - f_j(\mathbf{x}))^{1-w_{ij}} \\ &= \prod_{j=1}^N \left(\sum_{i'=1}^M w_{i'j} P_{i'} \right)^{w_{ij}} \left(\sum_{i'=1}^M (1 - w_{i'j}) P_{i'} \right)^{1-w_{ij}} \end{aligned} \quad (12)$$

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.

$$ECOC(\mathbf{x}) := \arg \max_{i \in C} g(c_i|\mathbf{x}). \quad (13)$$

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 5.

$$MAP(\mathbf{x}) := \arg \max_{i \in C} P_i. \quad (14)$$

本研究では, 定義 4, 5 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

本章では, 仮定 4 のもとで定義 4, 5 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 S について以下を定義する.

定義 6 (部分集合族 S_i). S に対して, $i \in C$ を含む集合のみを取り出した集合族 S_i を

$$S_i := \{S \in S | i \in S\}$$

とする.

定義 7 (部分集合族 S_{ik}). 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 S_{ik} を

$$\begin{aligned} S_{ik} &:= \{S \in S | i, k \in S\} \\ &= S_i \cap S_k \end{aligned}$$

とする.

定義 8 (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.

$$\pi_{ik}(A) := \begin{cases} A \cup \{k\} \setminus \{i\}, & (i \in A) \cap (k \notin A), \\ A, & \text{otherwise.} \end{cases}$$

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 $(\mathcal{A})_{i \rightarrow k}$ と表し, 以下のように定義する.

$$(\mathcal{A})_{i \rightarrow k} := \{\pi_{ik}(A) | A \in \mathcal{A}\}.$$

以下に例を示す.

例 5 (集合族の集合要素置換). 集合族 $\mathcal{A} = \{\{1, 2\}, \{1, 4\}, \{2, 3\}\}$ とするとき,

$$(\mathcal{A})_{1 \rightarrow 4} = \{\{4, 2\}, \{1, 4\}, \{2, 3\}\}$$

となる.

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

$$(S_i \setminus S_{ik})_{i \rightarrow k} = S_k \setminus S_{ik}$$

が成り立つとき, W を対称性を持つ符号語表と呼ぶ.

対称性を持つ符号語表 W が仮定 2 を満たすとき, 任意の $i \neq k$ について $S_i \setminus S_{ik} \neq \phi$ が成り立つ.

いま, 式 (12), 定義 4 より以下の定理が成り立つ.

定理 4 (十分条件). 対称性を持つ W を用いるとき, 任意の入力データ \mathbf{x} に対し, $ECOC(\mathbf{x}) = MAP(\mathbf{x})$ が成り立つ. (証明) 付録 A.2 を参照.

6. 対称性を持つ符号語表

OVA とは, M 個のカテゴリを 1 対 $M-1$ に分けるすべての組合せを持つ符号語表のことである. 同様に, n -vs-all (nVA) を n 対 $M-n$ に分けるすべての組合せを持つ符号語表とする*2.

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 $C = \{1, 2, \dots, M\}$ から n 個および $M-n$ 個の要素を抜き出したすべての組合せを持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 $(\mathcal{S}_i \setminus \mathcal{S}_{ik})_{i \rightarrow k} = \mathcal{S}_k \setminus \mathcal{S}_{ik}$ を満たす. すなわち対称性を持つ.

(証明) 付録 A.3 を参照.

例により nVA が対称性を持つことを説明する.

例 6 (対称性を持つ符号語表). $M = 3$ のときの OVA は例 1 により与えられる. このとき, c_1, c_2 に対する $\mathcal{S}_1, \mathcal{S}_2$ は,

$$\begin{aligned} \mathcal{S}_1 &= \{\{1\}, \{1, 3\}, \{1, 2\}\}, \\ \mathcal{S}_2 &= \{\{2, 3\}, \{2\}, \{1, 2\}\} \end{aligned}$$

となる. $\mathcal{S}_{1,2} = \{1, 2\}$ であるため,

$$\begin{aligned} \mathcal{S}_1 \setminus \mathcal{S}_{1,2} &= \{\{1\}, \{1, 3\}\}, \\ (\mathcal{S}_1 \setminus \mathcal{S}_{1,2})_{1 \rightarrow 2} &= \{\{2\}, \{2, 3\}\} \\ &= \mathcal{S}_2 \setminus \mathcal{S}_{1,2} \end{aligned}$$

となる. 任意の $i, k \in C$ に対し同様の関係が成り立つため, OVA は対称性を持つ.

また, $M = 4$ のときの 2VA は例 3 により与えられる. このとき, c_1, c_2 に対する $\mathcal{S}_1, \mathcal{S}_2$ は,

$$\begin{aligned} \mathcal{S}_1 &= \{\{1, 2\}, \{1, 3\}, \{1, 4\}\}, \\ \mathcal{S}_2 &= \{\{1, 2\}, \{2, 4\}, \{2, 3\}\} \end{aligned}$$

となる. このとき, $\mathcal{S}_{1,2} = \{1, 2\}$ であるため,

$$\begin{aligned} \mathcal{S}_1 \setminus \mathcal{S}_{1,2} &= \{\{1, 3\}, \{1, 4\}\}, \\ (\mathcal{S}_1 \setminus \mathcal{S}_{1,2})_{1 \rightarrow 2} &= \{\{2, 3\}, \{2, 4\}\} \\ &= \mathcal{S}_2 \setminus \mathcal{S}_{1,2} \end{aligned}$$

となる. 同様の関係が任意の $i, k \in C$ に対し成り立つため, 2VA も対称性を持つことが確認できる.

また, 定理 5 より, 対称性を持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,

*2 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.

W は対称性を持つ.

Exhaustive 符号 [9] とは, カテゴリ数 M に対して, すべての二値分類器の組合せを持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持つ.

7. 評価実験

本研究で最適な多値分類法である対称性を持つ符号語表について, 人工データ, 実データを用いて証明の確認および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 (15) に従う M 次元正規乱数を用いて実験を行う.

$$\begin{aligned} \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_i, \Sigma) &= \frac{1}{(\sqrt{2\pi})^M \sqrt{|\Sigma|}} \\ &\times \exp\left(-\frac{(\mathbf{x} - \boldsymbol{\mu}_i)^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)}{2}\right). \end{aligned} \quad (15)$$

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.

$$\mu_{ik} = \begin{cases} 1, & i = k, \\ 0, & \text{otherwise.} \end{cases} \quad (16)$$

$$\sigma_{ik} = \begin{cases} 0.5, & i = k, \\ 0, & \text{otherwise.} \end{cases} \quad (17)$$

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.

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

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

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

ン行い平均をとる 10 分割交差検証法によって評価する.

7.2 定理 4, 6 および系 1 の確認

人工データを生成させ, 仮定 4 を満たす ECOC 法が最適な多値分類法の十分条件を満たすとき, 分類性能が MAP 法と一致することを実験により確認する.

7.2.1 実験設定

定義 5 の MAP および対称性を持つ符号語表として, OVA, 2VA, 3VA, 4VA, Exhaustive 符号の誤り率を比較する.

7.2.2 実験結果および考察

実験結果を表 3 に示す. $M = 8$ の場合, nVA の符号語表は, $1 \leq n \leq 4$ の値域で定義が可能である. 加えて, OVA, 2VA, 3VA, 4VA を組み合わせた Exhaustive 符号も対称性を持つ. これらすべての誤り率が MAP と一致することが確認できる.

7.3 非対称性と分類誤り率

本論文で対称性を持つことが, 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.

7.3.1 実験設定

対称性を持つ符号語表として, 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 $f_j(\mathbf{x}) \in [0, 1]$ となる Logistic Regression [19], Gaussian SVM [20] を用いる.

図 1, 図 2, 図 3 に実験結果を示す.

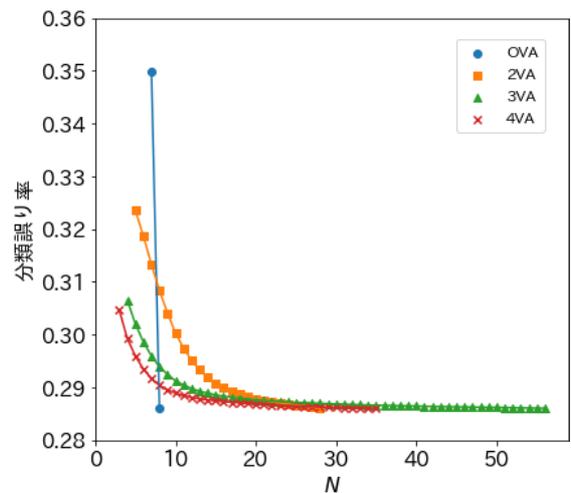


図 1 人工データ

Fig. 1 Artificial data.

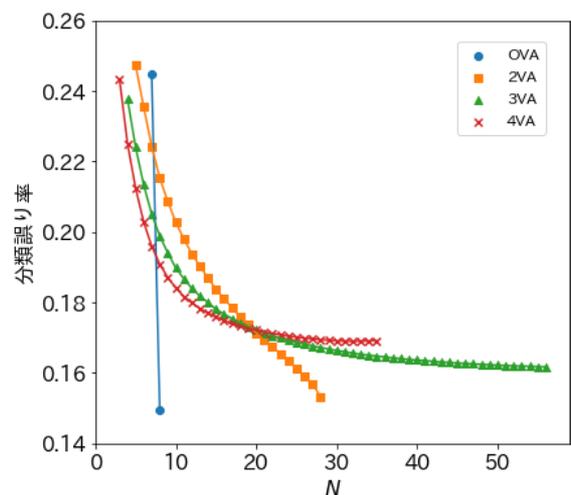


図 2 実データ (SVM)

Fig. 2 Real data (SVM).

7.3.2 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~図 3 より,

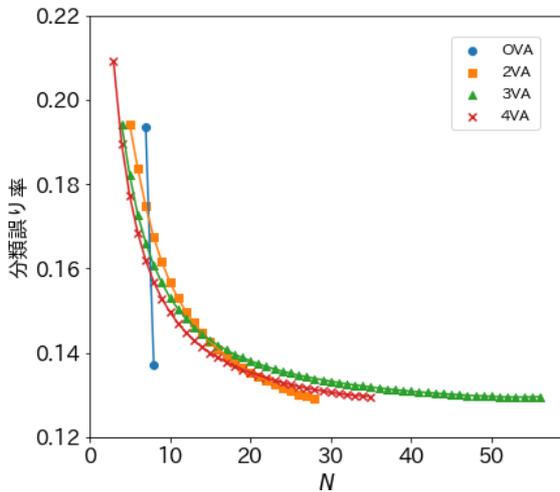


図 3 実データ (ロジスティック回帰)
Fig. 3 Real data (Logistic regression).

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、符号語が全 0 ベクトルとなるカテゴリができてしまうことにより誤り率が上昇したためである。

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分かる。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, 図 3 より、削除する列数を増やすごとに、急激な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分かる。

また、図 1~図 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分かる。

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、従来より良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。そして、評価実験によりこの結果の正当性を確認した。

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、対称性を持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。

謝辞 本研究の一部は、JSPS 科研費 JP16K06340, JP17K00020, JP18K11585, JP19H01721, JP19K04914, JP20K04462 の助成による。

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., et al.: 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, 石井直宏: 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 Output 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).

付 録

A.1 定理 3 の証明

W に対応する S をもとに、符号語表 W' に戻すことを考える。このとき、 W と S が等価とは、 W と W' の二値分類の組合せが等しいことを意味する。

証明. 二値分類器 f_j ($j \in F$) に対応する T_j が与えられたとする。集合 $T \in T_j$ であるとき、仮定 1 よりカテゴリ番号集合 C から T を除いたものが、もう 1 つの T_j の要素である。つまり、

$$T_j = \{T, C \setminus T\}$$

となる。 T_j に対応する二値分類器 f'_j は、2 通り考えられる。2 通りの分類器を表す列ベクトルを w'^j, w''^j とすると、それぞれの第 i 成分は

$$w'^j = \begin{cases} 0, & i \in T, \\ 1, & i \in C \setminus T \end{cases}$$

$$w''^j = \begin{cases} 1, & i \in T, \\ 0, & i \in C \setminus T \end{cases}$$

となる。このとき、 w^j と w'^j のハミング距離は M となり、仮定 3 より、同一の二値分類器となる。また、 f_j に対応する列ベクトル w^j は、 w'^j または w''^j に一致する。これより、 f_j と f'_j は同一の二値分類器である。

さらに、定理 1 より、 f_j と f'_j は同一の二値分類器であれば、 W に対応する S から符号語表 W' に戻した場合、 W と W' は二値分類の組合せが等しいという意味で等価となる。□

A.2 定理 4 の証明

証明. まず式 (12) は、 S_i を用いて表すと、

$$g(c_i|\mathbf{x}) = \prod_{S \in S_i} \sum_{s \in S} P_s \quad (\text{A.1})$$

と表せる。これを、 S_{ik} によって分割すると

$$g(c_i|\mathbf{x}) = \prod_{S \in S_{ik}} \sum_{s \in S} P_s \times \prod_{S \in S_i \setminus S_{ik}} \sum_{s \in S} P_s, \quad (\text{A.2})$$

と表せる。このとき、右辺の第 2 項に注目する。

いま、 $P_i > P_k$ ($i, k \in C, i \neq k$) と仮定する。 $S \in S_i \setminus S_{ik}$ に対し $k \notin S$ が成り立つ。また、仮定 2 より、 $S_i \setminus S_{ik} \neq \emptyset$ である。これより、

$$\begin{aligned} \prod_{S \in S_i \setminus S_{ik}} \sum_{s \in S} P_s &= \prod_{S \in S_i \setminus S_{ik}} \left(P_i + \sum_{s \in S \setminus \{i\}} P_s \right) \\ &> \prod_{S \in S_i \setminus S_{ik}} \left(P_k + \sum_{s \in S \setminus \{i\}} P_s \right) \\ &= \prod_{S \in S_i \setminus S_{ik}} \sum_{s \in S \cup \{k\} \setminus \{i\}} P_s \\ &= \prod_{S \in S_i \setminus S_{ik}} \sum_{s \in \pi_{ik}(S)} P_s \\ &= \prod_{S \in (S_i \setminus S_{ik})_{i \rightarrow k}} \sum_{s \in S} P_s. \end{aligned} \quad (\text{A.3})$$

これより、式 (A.2) の右辺の第 2 項について、

$$\prod_{S \in S_i \setminus S_{ik}} \sum_{s \in S} P_s > \prod_{S \in (S_i \setminus S_{ik})_{i \rightarrow k}} \sum_{s \in S} P_s \quad (\text{A.4})$$

が成り立つ。

ここで、符号語表 W が定義 9 の対称性を持つことから、 $i \neq k$ を満たす任意の $i, k \in C$ について $(S_i \setminus S_{ik})_{i \rightarrow k} = S_k \setminus S_{ik}$ が成り立つ。したがって式 (A.2), (A.4) から

$$\begin{aligned} g(c_i|\mathbf{x}) &> \prod_{S \in S_{ik}} \sum_{s \in S} P_s \times \prod_{S \in (S_i \setminus S_{ik})_{i \rightarrow k}} \sum_{s \in S} P_s \\ &= \prod_{S \in S_{ik}} \sum_{s \in S} P_s \times \prod_{S \in S_k \setminus S_{ik}} \sum_{s \in S} P_s \\ &= \prod_{S \in S_k} \sum_{s \in S} P_s \\ &= g(c_k|\mathbf{x}) \end{aligned} \quad (\text{A.5})$$

となる。いま、 $i \neq k$ を満たす任意の $i, k \in C$ について $P_i > P_k$ ならば $g(c_i|\mathbf{x}) > g(c_k|\mathbf{x})$ が成り立つことから、 $MAP(\mathbf{x}) = ECOC(\mathbf{x})$ となる。□

A.3 定理 5 の証明

証明. nVA は、 n 対 $M - n$ に分けるすべての組合せを持つ符号語表である。このことと、定義 8 より、任意の $i, k \in C$ について、

$$(S_i)_{i \rightarrow k} \subseteq S_k \quad (\text{A.6})$$

が成り立つ。 $(T_j)_i$ を、 T_j に対して $i \in C$ を含む集合のみを取り出した集合族とする。このことと、定義 1 より、任意の $i \in C$ について、

$$|(T_j)_i| = 1 \quad (\text{A.7})$$

となる。よって、定義 2 より、任意の $i \in C$ について、

$$|S_i| = N \quad (\text{A.8})$$

となる。よって、任意の $i, k \in C$ について、

$$|\mathcal{S}_i| = |\mathcal{S}_k| \quad (\text{A.9})$$

が成り立つ。定義 8 の操作は、要素数を変化させないことから、

$$|\mathcal{S}_i| = |(\mathcal{S}_i)_{i \rightarrow k}| \quad (\text{A.10})$$

が成り立つ。式 (A.9), (A.10) より、

$$|(\mathcal{S}_i)_{i \rightarrow k}| = |\mathcal{S}_k| \quad (\text{A.11})$$

が成り立つ。よって、式 (A.6), (A.11) より、

$$(\mathcal{S}_i)_{i \rightarrow k} = \mathcal{S}_k \quad (\text{A.12})$$

となる。定義 6, 7 より、

$$\mathcal{S}_i = \mathcal{S}_{ik} \cup \mathcal{S}_i \setminus \mathcal{S}_{ik} \quad (\text{A.13})$$

が成り立つ。定義 8 より、

$$(\mathcal{S}_{ik})_{i \rightarrow k} = \mathcal{S}_{ik} \quad (\text{A.14})$$

となる。以上より、

$$\begin{aligned} (\mathcal{S}_i)_{i \rightarrow k} &= \mathcal{S}_k \\ \Leftrightarrow (\mathcal{S}_{ik} \cup \mathcal{S}_i \setminus \mathcal{S}_{ik})_{i \rightarrow k} &= \mathcal{S}_{ik} \cup \mathcal{S}_k \setminus \mathcal{S}_{ik} \\ \Leftrightarrow (\mathcal{S}_{ik})_{i \rightarrow k} \cup (\mathcal{S}_i \setminus \mathcal{S}_{ik})_{i \rightarrow k} &= \mathcal{S}_{ik} \cup \mathcal{S}_k \setminus \mathcal{S}_{ik} \\ \Leftrightarrow \mathcal{S}_{ik} \cup (\mathcal{S}_i \setminus \mathcal{S}_{ik})_{i \rightarrow k} &= \mathcal{S}_{ik} \cup \mathcal{S}_k \setminus \mathcal{S}_{ik} \\ \Leftrightarrow (\mathcal{S}_i \setminus \mathcal{S}_{ik})_{i \rightarrow k} &= \mathcal{S}_k \setminus \mathcal{S}_{ik} \end{aligned}$$

となる。□



雲居 玄道 (正会員)

2008 年早稲田大学創造理工学部経営システム工学科卒業, 2008 年早稲田大学理工学術院総合研究所嘱託研究員, 2015 年浄土真宗本願寺派総合研究所研究助手, 2017 年早稲田大学大学院創造理工学研究科博士後期課程, 2019 年早稲田大学創造理工学部経営システム工学科助手, 現在に至る。情報数理応用・テキストマイニングの研究に従事。IEEE, 経営情報学会, 日本気象学会各会員。



八木 秀樹

2001 年早稲田大学創造理工学部経営システム工学科卒業, 2005 年同大学大学院理工学研究科博士後期課程修了, 2005 年早稲田大学メディアネットワークセンター助手, 2007 年同助教, 2008 年電気通信大学先端領域研究センター特任助教, 2008, 2010 年プリンストン大学客員研究員, 2012 年電気通信大学大学院情報理工学研究科准教授, 現在に至る。博士 (工学)。情報理論, 符号理論の研究に従事。IEEE, 電子情報通信学会, 信号処理学会各会員。



小林 学 (正会員)

1994 年早稲田大学創造理工学部経営システム工学科卒業, 2000 年同大学大学院理工学研究科博士後期課程修了, 1998 年早稲田大学理工学部助手, 2002 年湘南工科大学教員, 2018 年早稲田大学データ科学総合研究教育センター教授, 現在に至る。博士 (工学)。機械学習理論, 情報理論の研究に従事。IEEE, 電子情報通信学会等会員。



後藤 正幸 (正会員)

1994 年武蔵工業大学大学院修士課程修了, 2000 年早稲田大学大学院理工学研究科博士課程修了, 博士 (工学), 1997 年早稲田大学理工学研究科助手, 2000 年東京大学大学院工学研究科助手, 2002 年武蔵工業大学環境情報学部助教授, 2008 年早稲田大学創造理工学部経営システム工学科准教授, 2011 年同大学教授, 現在に至る。情報数理応用とデータサイエンスの研究に従事。著書に、『入門パターン認識と機械学習』コロナ社 (2014), 『ビジネス統計～統計基礎とエクセル分析』オデッセイコミュニケーションズ (2015) 等。IEEE, INFORMS, 電子情報通信学会, 人工知能学会, 日本経営工学会等会員。



平澤 茂一 (正会員)

1961年早稲田大学第一理工学部数学科卒業。1963年同大学電気通信学科卒業。同年三菱電機(株)入社。1981年早稲田大学理工学部工業経営学科(現、経営システム工学科)教授、2009年早稲田大学名誉教授、早稲田大学理

工学院総合研究所研究員、現在に至る。情報理論とその応用、ならびに計算機応用システム等の研究に従事。工学博士(大阪大学)。1979年UCLA計算機科学科客員研究員。1985年ハンガリー科学アカデミー、1986年イタリアトリエステ大学客員研究員、2002年UCLA訪問教員。1993年電子情報通信学会小林記念特別賞、業績賞受賞。IEEE Life Fellow, 電子情報通信学会フェロー、経営情報学会各会員。