ＲＶＭを用いた多値文書分類のための効率的な二値判別器構成法に関する研究

情報数理応用研究　5210C009-8　小田井良輔
指導教員　後藤正幸

A Study of Multivalued Document Classification based on Combination of Binary RVM Classifiers

ODAI Ryosuke

1　はじめに

近年，情報化社会の到来により，World Wide Web，電子メール，電子図書館など，膨大なオンラインテキストが扱われるようになった．このような電子媒体のテキストデータを自動処理する技術の重要性は高まる一方であり，中でも高精度な文書自動分類技術が必要とされている．

文書の自動分類技術には様々な手法が提案されているが，特にカーネル法を用いた手法が高性能であると報告されている [1]．その代表的な手法として，Relevance Vector Machine (RVM) があげられ，優れた二値判別器として知られている [2]．RVM は Support Vector Machine (SVM) の特性の多くを引き継ぎながら確率モデルとして解釈できる点が最大の特徴である．

しかし，もともと RVM は二値判別器であり，これを多値判別問題に適用する場合，1 つの判別器で直接モデル化する方法が考えられるが，計算量の問題で実用的とは言えない [2]．この問題を回避するため，“1-vs-the rest”多値判別手法と呼ばれる方法が知られている [3]．これは，1 つのカテゴリとそれ以外を識別する二値判別器をカテゴリ数だけ用意する方法である．しかしながらこれは，簡単な方法である反面，1 カテゴリとそれ以外の間で学習データ数の偏りが生じるため，各判別器の性能悪化を招くという問題がある．一方，“1-vs-the rest”多値判別手法と同様に，複数の二値判別器の組み合わせで多値判別を実現する方法として，符号理論の枠組みを導入した ECOC 復号法に基づく多値判別法が提案されている [4]．本研究ではこの枠組みの立場で議論する．

上記の視点に基づいた研究として，RVM が確率モデルであることを利用し，複数の二値判別器の組み合わせと事後確率の計算により多値分類を行う手法が提案されている [5](以下，“half-vs-the rest ”多値判別手法)．この方法では，各二値判別器が分類するクラス間の学習データ数に偏りが生じないよう，なるべくカテゴリ数が半々に分かれるような構成法を与えており，分類精度の面で優れていることが示されている．これは，二値判別器の学習データの正例と負例に偏りがあると，分類精度が低下するため，なるべく偏りが少なくなるように判別器を構成しようとする方法である．しかし，全ての分け方に対して二値判別器を作るため，必要以上の判別器を用いている可能性がある．特にカテゴリ数が大きい場合，計算量が増大し，実用的とは言えない．

そこで本研究では，“half-vs-the rest ”多値判別手法のアイデアを保持したまま，計算量を抑える判別器構成法を提案する．提案法では ECOC の一種である BCH 符号を用いてカテゴリの大きさに応じた判別器構成を構成するアルゴリズムを構築すると共に，出力が確率値である RVM の特性を活かした判別法を用いる．提案手法を文書分類問題に適用し，従来手法より精度・計算量の面で性能が向上することを示す．

2　準備

2.1　多値判別問題

判別問題とはカテゴリラベルの付いた入力データを用いる学習を行い，新たに与えられた入力データ $\boldsymbol{x}$ に対応するカテゴリラベル $\mathcal{C} = \{C_1, C_2, \cdots, C_G\}$ を推定する問題のことである．ここで G はカテゴリ数を表し，多値判別問題とは $G \geq 3$ の場合を指す．

多値判別の手法としては，大きく分けて 2 通りのアプローチが存在するが，本研究では「正解カテゴリ (1)」と「不正解カテゴリ (0)」の二値に判別する二値判別器を複数組み合わせて多値判別器を構成する方法を対象として研究を行う．

2.2　Relevance Vector Machine

RVM [2] は Tipping によって提案された手法で，回帰および分類問題を解くために提案された疎なカーネルベースのベイズ流学習手法である．優れた分類性能を持つ Support Vector Machine (SVM) [6] の特性の多くを引き継ぎながら確率モデルとして解釈できる点が最大の特徴である．

次に RVM の分類モデルを説明する．入力ベクトルを $\boldsymbol{x}$，カテゴリ集合を $\mathcal{C} = \{C_1, C_2\}$，$N$ 個のトレーニング文書セットを $\{\boldsymbol{x}_n, t_n\}_{n=1}^N$ とする．ただし，$t_n \in \mathcal{C}$ である．このとき $\boldsymbol{x}$ が判別されるカテゴリ $C = C_1$ となる確率をロジスティック回帰関数を使って以下の式で表す．

$$p(C = C_1|\boldsymbol{x}) = \frac{1}{1 + \exp(-f_{RVM}(\boldsymbol{x}))}, \tag{1}$$

$$f_{RVM}(\boldsymbol{x}) = \sum_{i=1}^{N} w_i K(\boldsymbol{x}, \boldsymbol{x}_i). \tag{2}$$

ただし，$w_i \sim N(0, \alpha_i^{-1})$ である．$K(.,.)$ はカーネル関数であり，入力された 2 つのデータ点を高次元空間上に写像し，内積を計算したものである．

1 つの RVM で多値判別を行う際には，G 個の線形モデルを組み合わせる確率的な方法を用いる．α_i^{-1} は 2 クラスの場合と同じように計算する．この方法は一貫性があるという点では有利ではあるが，学習にかかる計算量が 2 クラス RVM の G^3 倍となってしまう点が不利である [6].

3　従来手法

前述の通り，複数の二値判別器を組み合わせ，多値判別器を構成する方法は既に多くの有効な手法が提案されている．本章では，これらの従来手法について述べる．

3.1　“1-vs-the rest ”多値判別手法

二値判別器を複数用いて，多値判別を行う方法の代表的手法として，“1-vs-the rest ”多値判別手法があげられる [3].

“1-vs-the rest ”多値判別手法では，全ての $i = 1, 2, \cdots, G$ に対して判別対象カテゴリ C_i とそれ以外のカテゴリに分け

る “1-vs-the rest ”判別器を作る．入力 $\boldsymbol{x}$ に対する各々の判別器の出力を $R_{C_1}, R_{C_2}, \cdots, R_{C_G}$ とすると，

$$\hat{C} = \arg\max_{C_i} R_{C_i}, \tag{3}$$

とするカテゴリ $\hat{C}$ に判別する．

3.2 “half-vs-the rest ”多値判別手法

“1-vs-the rest ”多値判別手法の問題点として，

- 1 つでも判別器の性能が低いと，全体の分類性能が悪くなってしまう，
- 1 対多判別器では，1 カテゴリの学習データ数がその他の多カテゴリの学習データ数に比べ少なくなってしまうため，良い判別器を構成できない可能性が高い，

という 2 つの問題点が挙げられる．

この問題を解決する手法として提案されたのが “half-vs-the rest ”多値判別手法である [5]．二値判別器構成として，「正解カテゴリに属する学習データ」と「正解カテゴリに属さない学習データ」の比が等しいときが最も正解率が高いため [5]，カテゴリを $\lceil G/2 \rceil$ 個と $\lfloor G/2 \rfloor$ 個に分けるような判別器を全ての組み合わせについて作る．ただし，$\lceil x \rceil$ は x 以上の最小の整数，$\lfloor x \rfloor$ は x 以下の最大の整数である．作成する判別器の個数 g は $G \geq 4$ のとき，

$$g = \begin{cases} \frac{G!}{(G/2)!^2 \times 2}, & G \text{ が偶数の場合,} \\ \frac{(G+1)!}{((G+1)/2)!^2 \times 2}, & G \text{ が奇数の場合,} \end{cases} \tag{4}$$

で与えられる．

判別器構成は $G \times g$ 行列を $\boldsymbol{W}$ で表すことができる．行列 $\boldsymbol{W}$ の各行を g 次元ベクトル $\boldsymbol{W}_{C_i}$ とし，カテゴリ C_i を意味する符号語とすることで，1 つのカテゴリ C_i に対応させる $(i = 1, 2, \cdots, G)$．符号語 $\boldsymbol{W}_{C_i}$ は g 次元 $0-1$ ベクトルで構成され，$\boldsymbol{W}$ の各列は判別器の分け方を意味し，それぞれ対応する $\{0, 1\}$ を二値判別する．$G = 6$ の時の判別器構成と符号語を図 1 に与える．

$$\begin{array}{c} \\ W_{C_1} \\ W_{C_2} \\ W_{C_3} \\ W_{C_4} \\ W_{C_5} \\ W_{C_6} \end{array} \begin{array}{c} \begin{array}{cccccccccc} f_1 & f_2 & f_3 & f_4 & f_5 & f_6 & f_7 & f_8 & f_9 & f_{10} \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix} \end{array}$$

図 1．$G = 6$ の時の判別器構成

$\boldsymbol{f} = \{f_1, f_2, \cdots, f_g\}$ がそれぞれ判別器に対応し，例えば，図 1 の f_5 は $\{C_1, C_3, C_4\}$ と $\{C_2, C_5, C_6\}$ を判別する．つまり，$\boldsymbol{W}$ の各行は各カテゴリに対応する符号語となり，各列は各カテゴリをどのように分けるかを $\{0, 1\}$ で表した各判別器を意味する．

一方，判別法については，入力 $\boldsymbol{x}$ に対して，カテゴリ C_i の符号語 $\boldsymbol{W}_{C_i}$ の l 番目の判別器の値 $W_{C_i l}$ が 0 ならば $1 - R_l$，1 ならば R_l を g 個の判別器の出力をかけあわせ，以下の式で判別する．

$$\hat{C} = \arg\max_{C_i} \prod_{l=1}^{g} R_l^{W_{C_i l}} (1 - R_l)^{1 - W_{C_i l}}, \tag{5}$$

とするカテゴリ $\hat{C}$ に判別する．これはカテゴリの事前確率が等確率のとき，C_i に対応した符号語の事後確率を計算し，それを最大にするカテゴリに判別することと等価である．

3.3 ECOC 復号法に基づく多値判別法

誤り訂正符号 (ECOC) とは $\{0, 1\}$ の二値で表現される情報系列に対し，より組織的に機械で処理しやすい形で冗長性を付加し，信頼性の向上を図る技術であり，多少雑音が混入しても元の情報に訂正することができる符号を指す．Dietterich ら [4] はこの手法を多値判別に援用し，多値判別問題を複数の二値判別問題に分解するための枠組みを与えた．

Dietterich らによる判別器構成法は Exhaustive Codes を用いるものであり [4]，判別器を $g = 2^{G-1} - 1$ 個作成する．これは，$\boldsymbol{W}_{C_1}$ は全て 1 で構成し，$\boldsymbol{W}_{C_2}$ は 2^{G-2} 個の 0 に続き $2^{G-2} - 1$ 個の 1 で構成する．$\boldsymbol{W}_{C_3}$ は 2^{G-3} 個の 0，2^{G-3} 個の 1，2^{G-3} 個の 0 に続き $2^{G-3} - 1$ 個の 1 で構成し，$\boldsymbol{W}_{C_i}$ は 2^{G-i} 個の 0 と 1 を交互に並べて構成することで $\boldsymbol{W}$ を作成する．$G = 5$ の場合の判別器構成を図 2 に与える．

$$\begin{array}{c} \\ W_{C_1} \\ W_{C_2} \\ W_{C_3} \\ W_{C_4} \\ W_{C_5} \end{array} \begin{array}{c} \begin{array}{ccccccccccccccc} f_1 & f_2 & f_3 & f_4 & f_5 & f_6 & f_7 & f_8 & f_9 & f_{10} & f_{11} & f_{12} & f_{13} & f_{14} & f_{15} \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix} \end{array}$$

図 2．$G = 5$ の時の判別器構成

上記の Exhaustive Codes のように符号語をカテゴリ数個作成することで，各判別器を構成することができ，符号語の符号長は判別器数となる．

判別方法は，符号語 $\boldsymbol{W}_{C_i}$ と入力 $\boldsymbol{x}$ に対する g 個の二値判別器の $\{0, 1\}$ の硬判定出力を $\boldsymbol{b} = (b_1, b_2, \cdots, b_g)$，両者のハミング距離を $H(\boldsymbol{W}_{C_i}, \boldsymbol{b})$ とし，

$$\hat{C} = \arg\min_{C_i} H(\boldsymbol{W}_{C_i}, \boldsymbol{b}), \tag{6}$$

とするカテゴリ $\hat{C}$ に判別する．

その特徴は各カテゴリに対応する符号語間のハミング距離を大きくすることで，分類精度をあげる点にある．各カテゴリ間の距離が大きいということは，少量の判別器が誤判別をしたとしても，誤りを訂正できる能力を有するということである．この手法は g 個の二値判別器の精度が同等のとき，性能が良いとされている．

4 提案手法

4.1 従来手法の問題点と本研究への展開

本研究では，各カテゴリに所属する学習データの数が全て等しく，各カテゴリの事前確率が全て等しいという問題設定とする．

従来手法はカテゴリ数の増加に伴い，次のような問題点が挙げられる．

- “1-vs-the rest ”多値判別手法：各判別器の精度の低下．
- “half-vs-the rest ”多値判別手法：判別器数の飛躍的な増加による計算量の増加．

前者は「正解カテゴリに属する学習データ」と「正解カテゴリに属さない学習データ」でのデータ数の偏りが大きくなるために発生するものであり，後者は例えばカテゴリ数

$G=10$ の時 $g=127$，$G=16$ の時 $g=6435$ と判別器数が増大してしまい，実用的ではないという問題がある．

そこで，本研究では “half-vs-the rest ”多値判別手法の精度を保ったまま，カテゴリ数が増えたときの計算量を抑える方法として誤り訂正符号の一種である BCH 符号の適用を考える．BCH 符号は符号長や誤り訂正能力をパラメータとして与えることで自由に設定でき，目的に応じてカスタマイズされた符号を設計できるという特徴がある．これを利用し，カテゴリ数に応じた判別器数を設定することで，判別器数の増加を抑えつつハミング距離が大きな符号語集合を作成可能である．しかしながら，BCH 符号は符号語間の距離をとることはできるが，構成された二値判別器の学習データの偏りが考慮されていない．そこで符号語集合の中から，各判別器の学習データの偏りを小さくするような符号語を選択することで，各判別器の精度が良いものを使用する判別器構成を構築するアルゴリズムを提案する．

4.2 判別器構成法

4.2.1 概要と BCH 符号

判別器構成方法は BCH 符号を用いた符号語作成と各判別器の学習データバランスを考慮した符号語選択の 2 つのステップで行う．まず，符号語作成のステップで BCH 符号を用いて符号語集合を作成することで各符号語間のハミング距離と計算量を考慮する．さらに，符号語集合はカテゴリ数より多い符号語で構成されているため，符号語集合からカテゴリ数の符号語を選ばなければならない．そこで，符号語選択のステップで各判別器の学習データの偏りを考慮する判別器を構成する．

次に BCH 符号について説明する．符号長 n, 情報ビット数 k, 誤り訂正可能ビット数 t で定義される BCH 符号を，(n,k,t)BCH 符号と表す．一般に BCH 符号はあらかじめパラメータが定められたものが用いられ，その中から条件に合った BCH 符号を選ぶ．

4.2.2 符号語作成

性能のよい判別器構成にするためには，各符号語間の最小ハミング距離を大きくすることが必要である．本研究では以下の条件の (n,k,t)BCH 符号を選択する．BCH 符号で作成される符号語数は $D=2^k$ である．作成する符号語数はカテゴリ数より大きくなければならないため，定められた n に対して，$2^k \geq G$ である k の下で，最大の t を持つ (n,k,t)BCH 符号を用いて符号語集合を作成する．符号長 n の符号語を $\boldsymbol{d}_i=(d_{i1},d_{i2},\cdots,d_{in})$，符号語の集合を $\Delta=\{\boldsymbol{d}_1,\boldsymbol{d}_2,\cdots,\boldsymbol{d}_D\}$ とし，符号語集合の最小ハミング距離は $2t+1$ となる．

4.2.3 符号語選択

二値判別器を構成する際に，学習データの偏りを考慮しながら，各カテゴリを表す符号語同士の距離を大きくとることは難しい．したがって，すでに距離について保証されている符号語集合から 1 つずつ符号語を削っていくことで，学習データの偏りが小さくなるように G 個の符号語を選択する方法を提案する．

各判別器の学習データの偏りを考慮するために，符号語集合の重みベクトルを定義する．重みベクトルとは符号語集合の各列の 1 の数を足し合わせたベクトルである．最終的に符号語が G 個になった時に，重みベクトルの各要素をおよそ $G/2$ にしつつ，ハミング距離を大きくするための判別器数を確保したい．すなわち，最終的に残った G 個の符号語集合の重みベクトルの要素がばらつかず，均等になるように符号語を 1 つずつ削除していく．そこで，重みベクトルの値が大きいところでは 1，小さいところでは 0 をとる符号語を (n,k,t)BCH 符号で作成した符号語集合から削除することで，重みベクトル内の偏りを小さくする．そのために，目標ベクトルという重みベクトルの要素のうち，その値が (最大値-1) 以上のものは 1，他は 0 であるベクトルを定義し，目標ベクトルとハミング距離が小さい符号語を削ることで各列の重みベクトルを均等に減らしながら，符号語候補を減らしていく．これを符号語候補がカテゴリ数と等しくなるまで繰り返すことで，最終的に学習データの偏りを小さくした判別器構成を達成することが可能となる．

符号語は以下のアルゴリズムによって選択する．

Step0) $D=2^k \geq G$ を満たす BCH 符号を用いて作成した符号語集合を $\Delta=\{\boldsymbol{d}_1,\boldsymbol{d}_2,\cdots,\boldsymbol{d}_D\}$ とする．

Step1) 符号語集合の重みベクトル $\boldsymbol{Q}=(q_1,q_2,\cdots,q_n)$ を以下の式で計算する．

$$q_s=\sum_{h=1}^{D} d_{hs} \tag{7}$$

Step2) 重みベクトルから目標ベクトル $\boldsymbol{V}=(v_1,v_2,\cdots,v_n)$ を以下の式で作成する．

$$v_s=\begin{cases} 1, & q_s \geq \max\limits_i q_i - 1, \\ 0, & q_s < \max\limits_i q_i - 1, \end{cases} \tag{8}$$

Step3) 目標ベクトルと各符号語とのハミング距離を $H(\boldsymbol{d}_i,\boldsymbol{V})$ とし，

$$\hat{\boldsymbol{d}}=\arg\min_{\boldsymbol{d}_i} H(\boldsymbol{d}_i,\boldsymbol{V}), \tag{9}$$

とする符号語 $\hat{\boldsymbol{d}}$ を符号語集合 Δ から削除し, $D=D-1$ とする．

Step4) $D>G$ であれば Step1 に戻り，$D=G$ になるまで繰り返す．

Step5) G 個の符号語を各カテゴリに対応させ，同一の分け方の判別器を削除する．

このアルゴリズムを行うことで，各カテゴリ間の最小ハミング距離が t 以上であることを保証しつつ，判別器構成に含まれる各判別器のデータの偏りが小さい判別器構成が得られる．

判別方法は 3.2 節で記した “half-vs-the rest ”多値判別手法と同様に，(5) 式を用いる．

5 実験による評価

提案手法の有効性を検討するため，新聞記事データを用いて分類実験を行い，分類精度の評価を行なった．

5.1 実験方法

実験には，読売新聞 2000 年 10 カテゴリ (政治・経済・スポーツ・社会・文化・生活・犯罪事件・科学・国際・皇室) を使用した．すべての記事は 1 カテゴリのみに属し，カテゴリの重複はない．データから各カテゴリ 150 記事をランダム

に選び，それを各カテゴリ学習データ 100 個，テストデータ 50 個にランダムに分ける．

カテゴリ数が大きな場合での提案手法の有効性を示すため，カテゴリ数を変えた際の性能変化について実験を行う．後者は，カテゴリ数を $G=8$(政治 · 経済 · スポーツ · 社会 · 文化 · 生活 · 犯罪事件 · 科学) と $G=10$(全カテゴリ) の 2 パターンで実験を行う．

提案手法で使用する BCH 符号は，符号長 n が提案手法における最大の判別器数となる．“half-vs-the rest ”多値判別手法よりも少ない判別器構成でも性能が劣化しないことを示すため，各々のカテゴリで “half-vs-the rest ”多値判別手法の判別器数よりも n が小さい範囲で設定し，$G=8$ では $n=31$ の $(31,6,7)$BCH 符号，$G=10$ では $n=63$ の $(63,7,15)$BCH 符号を使用した．

従来手法として，3.1 節で記した “1-vs-the rest ”多値判別手法, 3.2 節で記した “half-vs-the rest ”多値判別手法と 3.3 節で記した ECOC 法を用いた．

さらに，提案手法の符号語選択の有効性を示すために，$(31,6,7)$BCH 符号は $D=64$ 個，$(63,7,15)$BCH 符号は $D=128$ 個の符号語集合から，$G=8,10$ それぞれに対して，符号語をランダム選択し，分類を行った平均と比較し検討する．

5.2 実験結果

表 1 はカテゴリ数が変化する時の各手法の判別器数の表である．

表 1． 各手法のカテゴリ変化における判別器数

カテゴリ数	1-vs-rest	half-vs-rest	ECOC	提案
8	8	35	127	16
10	10	127	511	32

本実験において提案手法は $(31,6,7)$BCH 符号と $(63,7,15)$BCH 符号を用いて符号語集合を作成し，その符号語を選択した．その際に，同一の判別器を削除し，最終的に $G=8$ の時は 16 個，$G=10$ の時は 32 個の判別器構成となった．

“1-vs-the rest ”多値判別法, ECOC 法，“half-vs-the rest ” 多値判別手法，提案手法の正解率の実験結果を図 3 に示す．

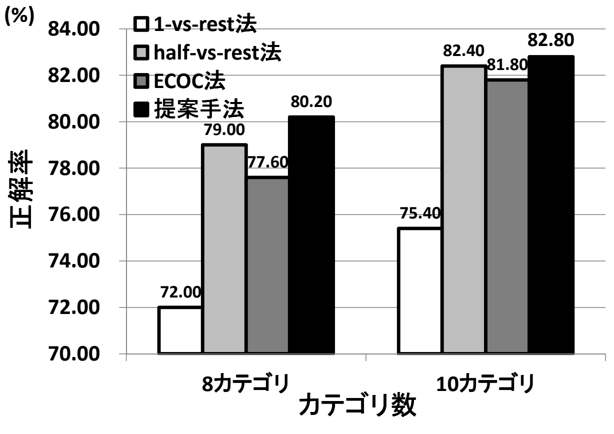

図 3．カテゴリ数と分類精度 (読売，2000 年)

図 3 からカテゴリ数が変化しても提案手法は “half-vs-the rest ”多値判別手法より高い精度である．提案手法は判別器数が “half-vs-the rest ”多値判別手法や ECOC 法よりはるかに少ない構成にも関わらず，同等以上の性能を示した．

表 2 は提案手法の符号語選択方法とランダムに選択した方法との分類精度の比較の結果である．

表 2．提案手法とランダム選択による分類精度

BCH サイズ	(31,6,7)		(63,7,15)	
カテゴリ数	8	10	8	10
ランダム選択	77.80	80.20	80.00	80.60
提案手法	80.20	81.40	81.80	82.80

表 2 からランダム選択よりも提案手法が分類精度で勝っていることより，提案手法の符号語選択手法の有効性が示せた．

5.3 考察

本研究で実施した実験において，提案手法は判別器数が少ないにも関わらず，“half-vs-the rest ”多値判別手法や ECOC 法より優れた精度を得た．つまり，提案手法ではデータの偏りとハミング距離の両方を考慮することにより，効率的な判別器構成が可能となったと考えられる．すなわち，提案手法では BCH 符号を使用しているため，条件に応じ，その符号パラメータを変更することで判別器数を任意に決めることができるという利点がある．

表 2 から，提案手法を用いた場合，ランダム選択より判別器数が少ないにも関わらず，高い分類精度が得られた．その理由として，ランダムに選択するとデータに偏りがある判別器が判別器構成に含まれてしまう可能性があり，判別器の精度が落ちてしまったと考えられる．以上より，符号語の選択法が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，カテゴリ数が大きな文書分類問題を対象とし，二値判別器の効率的な構成法を誤り訂正に用いられる手法を基に提案し，評価実験によって有効性を示した．

今後の課題として，文書分類以外の様々な多値分類問題での性能評価が挙げられる．また，各カテゴリによってデータの偏りがある条件における有効性の証明も必要である．

参考文献

[1] C. Silva and B. Ribeiro, “Scaling Text Classification with Relevance Vector Machines,” *IEEE International Conference on Systems, Man, and Cybernetics*, pp. 4186–4191, Oct. 2006.

[2] M.E. Tipping, “Sparse Bayesian Learning and the Relevance Vector Machine,” *Journal of Machine Learning Research*, pp. 211–244, Jun. 2001.

[3] 池田思朗, “2 クラス判別器の組み合せによる多クラス判別統計モデルとパラメータ推定,” 統計数理研究所, 特集「統計的機械学習」, vol.2, pp. 157–166, 2010.

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

[5] 小田井良輔, 雲居玄道, 三川健太, 後藤正幸, “二値判別器の組み合わせによるＲＶＭ多値文書分類手法に関する一考察,” 第 10 回情報科学技術フォーラム, pp. 425–428, Sep. 2011.

[6] C.M. Bishop, “Pattern Recognition And Machine Learning,” Springer, 2008.