

# Hadamard 纠错码结合支持向量机在多分类问题中的应用

尹安容, 谢 湘, 匡镜明  
(北京理工大学电子工程系, 北京 100081)

**摘 要:** 多分类问题一直是模式识别领域的一个热点, 本文提出了将 Hadamard 纠错码同二元分类器相结合的方法来解决此问题, 相对于其它类型的纠错码多分类器法, 该方法的实现简单快捷, 且更容易构造出性能优越的纠错码本. 本文将 Hadamard 纠错码和支持向量机相结合, 应用于说话人辨认这样一个多分类问题中, 并同传统的“1 对余”的多类推广方式进行了比较. 实验结果表明在多分类任务中, Hadamard 纠错码对于不同的类别都表现出了很强的分类能力, 且性能优于“1 对余”法, 对于类间码字的不同分配方式也具有良好的鲁棒性.

**关键词:** 模式识别; 说话人辨认; Hadamard 矩阵; 支持向量机; 纠错码

中图分类号: TN912.34 文献标识码: A 文章编号: 0372-2112 (2008) 01-0122-05

## Application of Hadamard ECOC in Multi-Class Problems Based on SVM

YIN An rong, XIE Xiang, KUANG Jing ming

(Department of Electronics Engineering, Beijing Institute of Technology, Beijing 100081, China)

**Abstract:** In this paper, we proposed applying Hadamard Error Correcting Output Code to extend binary classifier to multi class problems. Compared with other ECOC approaches, Hadamard ECOC is easy to construct and suitable to any number of classes. We combine it with binary support vector machine (SVM) to solve the multi class problem of speaker identification. Compared with the traditional “1 against rest” method, the experiment result shows that Hadamard ECOC has much better and more stable performance to any number of classes for the multi class problem and is robust with respect to the assignment of distributed representations to particular classes.

**Key words:** pattern recognition; speaker identification; Hadamard matrix; support vector machine; error correcting output code

### 1 引言

在模式识别领域中, 如何有效地解决多分类问题是模式分类研究中的重要课题, 现有的方法包括直接运用多类分类器(如决策树)、为每个类单独建立一个二元分类器(即“1 对余”法)<sup>[1]</sup>、根据分布式输出码来构造多个二元分类器<sup>[2]</sup>等等, 其中最值得注意的是 1995 年 Dietterich 和 Bakiri 所提出的纠错码(Error Correcting Output Code)法, 他们利用纠错码来将二元分类器推广到多分类问题中, 并结合决策树和神经网络算法针对手写字母识别等多分类任务进行了实验, 在同其它几种分类方法的比较中, 纠错码法表现出了优越的性能<sup>[3]</sup>. 但是, 他们并没有给出一种可以通用的纠错码, 而只是针对不同的类别范围给出不同的纠错码生成办法, 故当类别数改变

时, 纠错码的生成算法也要相应地变化, 此外, 他们所采用纠错码算法也比较复杂, 这都会给实际应用带来困难.

针对这种情况, 本文提出了一种通用的纠错码——Hadamard 纠错码, 它的生成简单快捷且适用于任意的类别数, 在本文进行的多分类实验——说话人辨认中, 它有效地将二元分类器扩展到了多类问题中.

本文的实验中所采用的二元分类器为支持向量机<sup>[4]</sup>(support vector machine). 支持向量机因其强大的区分能力而在模式识别领域被广泛运用, 将其推广到多分类问题中的方法包括“1 对余”、“1 对 1”等<sup>[5]</sup>.

### 2 支持向量机

支持向量机(SVM)是一种基于结构风险最小化原

则的机器学习方法,早在二十世纪六十年代就由 V. Vapnik 提出<sup>[4]</sup>,但对它研究热潮的兴起却是近十年的事。由于其强大的区分能力,它在很多模式识别领域如人脸识别、说话人识别等被广泛研究和应用。

SVM 是一种二元分类器,构造 SVM 的过程就是通过解一个二次规划问题,来找到能分开两类训练数据的最优超平面的过程。所谓最优超平面,是指此分类面不仅能正确地分开两类数据,且要使两类间的间隔最大。当引入新的样本  $x$  时,则通过下面的决策函数来对其类别进行判别:

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^n y_i \partial_i K(x, x_i) + b\right) \quad (1)$$

其中,  $\partial_i$  为训练得到的拉格朗日因子,  $x_i$  为支持向量,  $y_i$  是  $x_i$  对应的类别标注,  $y_i \in \{0, 1\}$ ,  $b$  为分类阈值,  $K(\cdot, \cdot)$  为核函数,核函数也是 SVM 的核心,通过它 SVM 将线性不可分的样本映射到高维空间,从而使得这些样本在高维空间线性可分。

比较常见的核函数有 sigmoid 函数、径向基函数(radial base function)、多项式核函数(polynomial kernel)等,在本次实验中的 SVM 采用的核函数为径向基函数:

$$K(x, x_i) = \exp\{-r \|x - x_i\|^2\} \quad r > 0 \quad (2)$$

其中  $r$  为核参数。正如引言所提过的,将支持向量机这种二元分类器推广到多类问题中的经典方法有“1 对余”、“1 对 1”等<sup>[5]</sup>。“1 对余”是指在训练时,将某个类的数据样本作为一类,而其余所有类的数据样本作为一类,相当于为每个类都构造了一个支持向量机。在本文的实验中,也将“1 对余”的方式应用于了说话人辨任务中,以和 Hadamard 纠错码法进行对比。

### 3 纠错码

纠错码(ECOC)是 Bose 和 Ray Chaudhuri 在 1960 年提出的一种分布式输出码<sup>[6]</sup>,1963 年, Duda, Machanik 和 Singleton 将其应用于了机器学习领域中<sup>[7]</sup>,1995 年, Dietterich 和 Bakiri 提出了利用纠错码来将二元分类器扩展到多类的分类问题中<sup>[3]</sup>:若类别数为  $k$ ,则为每个类别分配一个长度为  $L$  的二元纠错码序列  $W_i$ ,形成一个  $k$  行  $L$  列的码本(见表 1)。在训练时,对于第  $j$  列,将该列中码“0”对应的所有类别的数据归为一类,码“1”对应的所有类别的数据归为一类,以此来构造一个二元分类器——这样  $k$  类的分类问题就转化成了  $L$  个两类的分类问题。在测试时,  $L$  个分类器都对新样本进行判别,得到各分类器的输出——一个二元序列  $B = \{b_2, b_2, \dots, b_L\}$ ,然后计算此序列和各类别的码字  $\{W_i\}$  间的汉明距离(Hamming distance),最小距离对应的那个码字所代表的类就是最终的判别结果,即:

$$W = \arg \min d(B, W_i) = \sum_{j=1}^L b_j - W_{i,j} \quad (3)$$

对于任一分类问题而言,都需要经过样本采集、特征提取和训练识别算法实施这些环节,如果样本不足,或者特征不完整,或者算法存在缺陷,都会对分类的正确性造成影响,使分类产生误差,而 ECOC 最大的优点就在于它能够有效地对这些误差进行修正,对于一组最小汉明距离为  $d$  的纠错码来说,它至少能修正  $\lfloor (d-1)/2 \rfloor$  位误差,即,即便有  $\lfloor (d-1)/2 \rfloor$  个分类器出错,系统还是能给出正确的判别结果。

当应用纠错码来将二元分类器扩展到多类问题中时,对纠错码的选择要基于以下两点<sup>[3]</sup>:

(1) 行分离,即尽量使行与行之间的汉明距离最大,以便使纠错码有更强的纠错能力;

(2) 列分离,即尽量使列与其它各列及各列的补集的汉明距离最大,以此来削弱分类器间的相关性。当列间距离相似或相同时,对应的分类器也会相似或相同,那当出错时它们就会同时产生相似或相同的错误,而 ECOC 纠错能力有限,如果同时有很多错误,ECOC 将无法对其修正;之所以还要强调列的补集,是因为对于很多二元分类器而言,根据某个码字训练得到的分类器和根据此码字的补集训练得到的分类器是一样的(支持向量机即是如此)。

基于以上两个原则, Dietterich 和 Bakiri 针对不同的类别范围给出了 4 种设计纠错码的方法:详尽码(Exhaustive codes) ( $3 \leq k \leq 7$ )、在详尽码中取列( $8 \leq k \leq 11$ )、随机登山算法(Randomized Hill Climbing) ( $k \geq 11$ )和 BCH 码( $k \geq 11$ ),它们的详细介绍见文献<sup>[3]</sup>。但这种得到纠错码的方法存在三个缺点:

(1) 没有一个对于任意类别都通用的算法,当类别改变时,就必须相应地更换算法;

(2) 算法实现复杂,以 BCH 码为例,本来它的生成就比较麻烦,在生成码阵之后,因为其行数始终为 2 的幂,且存在互补列,所以还要采用一定的算法对行、列进行截取和挑选,往往要通过多次尝试才能找到合适的码。

(3) 得到的纠错码码长很长(如详尽码,其类别数  $k$  和码长  $L$  的关系为  $L = 2^k - 1$ ,故 7 类问题所需的纠错码码长为 63),相应的需要构建的二元分类器的数目也就比较庞大,这给参数存储和实时运行都带来了难度,故系统性能虽提高了,但代价过大。

本文采用的 Hadamard 纠错码,则有效地解决了这三个问题。

### 4 Hadamard 纠错码

Hadamard 纠错码是由 Hadamard 矩阵得到的,2 阶的

Hadamard 矩阵如下所示:

$$H_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad (4)$$

Hadamard 矩阵的阶数按  $N = 2^j (j = 1, 2, \dots)$  的规律变化, 高阶的 Hadamard 矩阵可以由低阶 Hadamard 矩阵递推得到<sup>[8]</sup>:

$$H_N = \begin{pmatrix} H_{N/2} & H_{N/2} \\ H_{N/2} & -H_{N/2} \end{pmatrix} \quad (5)$$

其中  $-H_{N/2}$  表示对  $H_{N/2}$  中的元素取补, 即 1 变为 0, 0 变为 1. 可见 Hadamard 矩阵的任意不同的两行或两列都正交(注: 本来 Hadamard 矩阵中的元素应为“-1”和“1”, 此处为与前文的码字统一, 故用“0”代替了“-1”), 且对于  $N$  阶的 Hadamard 阵来说, 其各行间及各列间距离均为  $N/2$ , 有着良好的区分性, 它的这些特性使得它在编码理论、通信理论、数字信号处理等领域得到了广泛应用. 此外, 我们也可以注意到, 它的行、列间的分离度满足多分类问题中对纠错码的要求, 但是, 它无法被直接运用于多分类问题中, 因为根据其特性, 存在有两点不足:

- (1) 其第一列全零. 全零列意味着此列实际无分类器, 因此它的存在没有任何作用, 应该被删除.
- (2) 对类别数有限制. 因为 Hadamard 矩阵为一个  $2^j \times 2^j$  的方阵, 所以它只能对  $2^j$  类进行编码, 如果要对任意类别都提供纠错码, 就必须对码阵进行一些取舍;

经过一番综合考虑后, 我们得到 Hadamard 纠错码的生成过程如下(设类别数为  $k$ ):

- (1) 根据式(5)生成 Hadamard 矩阵. 如果  $2^{j-1} < k \leq 2^j$ , 则生成阶数为  $2^j$  的 Hadamard 矩阵;
- (2) 删除第一列全零列, 得到  $2^j \times 2^j - 1$  的矩阵;
- (3) 对第 2 步得到的方阵取其前  $k$  行, 则得到所需的  $k \times 2^j - 1$  的纠错码阵.

对于任意数目的类别都可以用上面的方法得到相应的 Hadamard 纠错码阵, 且此纠错码生成法有如下特点:

- (1) 简单快捷. Hadamard 矩阵很容易生成, 再对其做简单的截取就可以得到所需的码本, 不用再作挑选.
- (2) 纠错码阵的码长为  $2^j - 1$ , 行间汉明距离均为  $2^{j-1}$ , 且列间不会出现相同和互补的情况, 保证了行间分离度和列间分离度.
- (3) 码字长度较短(对于  $k$  类情况, 码长  $L$  满足  $k - 1 \leq L < 2k - 1$ ), 这样构造的分类器数目也减少很多(如 7 类问题只需 7 个二元分类器), 降低了存储量提高了运行效率.

表 1 给出了一个表示 10 类分类问题的 15 位 Hadamard 纠错码, 其行间汉明距离均为 8.

表 1 表示 10 个类别的 15 位 Hadamard 纠错码

类	码 字														
	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$	$f_8$	$f_9$	$f_{10}$	$f_{11}$	$f_{12}$	$f_{13}$	$f_{14}$	$f_{15}$
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
2	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
3	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0
4	0	0	0	1	1	1	1	0	0	0	1	1	1	1	1
5	1	0	1	1	0	1	0	0	1	0	1	1	0	1	0
6	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0
7	1	1	0	1	0	0	1	0	1	1	0	1	0	0	1
8	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
9	1	0	1	0	1	0	1	1	0	1	0	1	0	1	0

## 5 纠错码与支持向量机的结合

当具体结合支持向量机时, 则上面所介绍的纠错码码本中的每一列就对应一个支持向量机. 训练时每列中码字“0”对应的那些类别的训练样本作为一类, “1”对应的那些类别的训练样本作为一类, 以此来构造一个 SVM.

测试时, 输入测试样本(如一段语音)的特征参数序列  $X = \{x_2, x_3, \dots, x_T\}$  到各支持向量机中, 支持向量机会对每帧参数  $x_i$  都进行判别, 并给出两个值: 决策函数的判决结果“0”或“1”和决策函数取“sgn”之前的值. 如何根据各个支持向量机的输出来得到对于  $X$  的最终判决结果, 下面给出了两种方法:

- (1) 令各支持向量机的最终输出直接就为“0”或“1”, 判决依据可以是该段测试数据中“0”或“1”类别的帧数和总帧数的比率大小; 得到各支持向量机的输出后, 将此输出的二元序列和各类的码字进行比较, 汉明距离最小的码字所对应的类为最后判别结果, 此即式(3)所表示的方法, 这是一种硬判决.

- (2) 令各支持向量机的最终输出在某种程度上表征此测试样本属于“0”类或“1”类的概率, 如可以取该测试数据中“0”或“1”类别的帧数和总帧数之比; 得到各支持向量机的输出  $P = \{p_1, p_2, \dots, p_L\}$  后, 再根据下面的式子计算此输出序列和各类的码字间的距离:

$$d(P, W_i) = \sum_{j=1}^L |p_j - W_{ij}| \quad (6)$$

若  $p_i$  表征的是测试样本属于“0”类别的概率, 则距离最大的码字所对应的类为最后判决结果; 若  $p_i$  表征的是测试样本属于“1”类别的概率, 则距离最小的码字所对应的类为最后判决结果. 这是一种软判决.

在本次任务中采用的是第二种方法, 但在表征概率时用的不是帧数, 而是式(1)中决策函数取“sgn”之前的值的绝对值  $g(x)$  的和,  $g(x)$  代表了测试样点到分界面的距离, 衡量了此次判决的置信度;

$$g(x) = \left| \sum_{i=1}^n y_i \delta K(x, x_i) + b \right| \quad (7)$$

若  $p_i$  表征的是第  $l$  个支持向量机所给出的测试样本属于“0” (“1”) 类的概率, 且第  $l$  个支持向量机的决策函数所给出的判决结果中, 被判为“0” (“1”) 类的那些帧的集合为  $s \subseteq \{1, 2, \dots, T\}$ , 则有:

$$P_l = \sum_s g(x_i) \setminus \sum_{i=1}^T g(x_i) \quad (8)$$

因为对于各帧来说, 支持向量机的决策函数(式(2))所给出的判决值 (“0” 或 “1”) 没有考虑  $g(x)$  的大小 (即没有考虑对样点的判决的置信度), 所以式(8)比类别帧数之比能更本质地反映测试样本属于“0” 或 “1” 类的概率。

## 6 实验结果及分析

### 6.1 实验数据

本次实验所用的数据均来自于北京理工大学现代通信实验室所录制的电话语音库 (采集于公用电话网), 语音采样率为 8kHz, 16bits 量化。一共有 15 个人 (8 男 7 女) 的语音, 每人 40 句话, 每句话皆为一个长度为 3s 左右短句。

实验采用的特征参数为 12 维的梅尔倒谱 (MFCC), 每 20ms 语音一帧, 帧移 10ms。

### 6.2 实验结果及分析

为了对 Hadamard 纠错码的性能进行较为全面的研究, 本文组织了两组实验。在第一组实验中, 我们分别应用 Hadamard 纠错码和“1 对余”法来将支持向量机扩展到多分类问题中, 并对它们的性能进行了比较; 在第二组实验中, 我们探讨了类间码字的不同分配情况对识别性能的影响。

#### 6.2.1 实验一

本文对文本无关的说话人辨认问题进行了研究, 进行了 7、8、9、10、11、12、13、14、15 个说话人共 9 组实验。为了充分利用数据, 本文采用了交错验证的方法, 即, 对于每类问题, 都将每个人总的的数据分成两组 (即 20 句话一组), 一组训练另一组测试, 交错进行。若类别数为  $k$ , 则对应的总的测试次数为  $M = k \times 2 \times 20$ , 故每个人的总的训练数据长度约为 1 分钟, 每段测试数据长度约为 3s。在性别比例方面, 人数为偶数时, 则男女个数相等; 人数为奇数时, 则男性数目比女性数目多 1。

本文采用的二元分类器为支持向量机, 多类扩展方法为“1 对余”和 Hadamard 纠错码法。识别结果 (识别率) 见图 1:

从图中可看出:

(1) 在同类别情况下, Hadamard 纠错码得到的识别性能总是优于“1 对余”的推广方式, 类别数越大时, 优

势越明显——当类别数为 15 时, Hadamard 纠错码法的识别率为 94.83%, 而“1 对余”法则只有 52.83%, 差距为 40%; 实际上, “1 对余”法也可以看成是纠错码的一种特例——它在构造 SVM 时用的码矩阵是一个单位阵, 行间汉明距离为 2, 不具备任何纠错能力, 故识别性能也弱于 Hadamard 纠错码法。

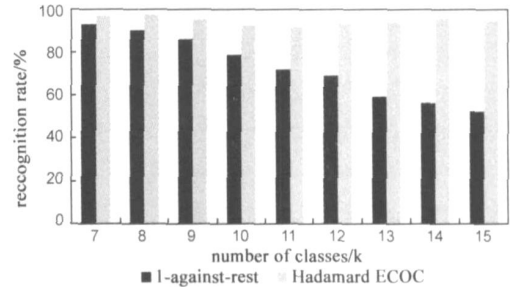


图 1 “1 对余”法和 Hadamard 码法的识别结果

(2) 随着类别数的增加, “1 对余”方式的识别性能迅速下降, 这是因为人数增多时, 这种方式所建立的模型间的区分度也随之下降了。相比之下, Hadamard 纠错码法的性能就稳定得多, 在较大类别数的情况下也能维持较好的性能。

因为本实验的训练数据每人只有 2 分钟, 并不充分, 对于“1 对余”方式而言, 在类别数增大后, 训练数据的不充分性就体现得更明显, 而 Hadamard 纠错码法的高行列分离度则在此时则表现出了优良的性能。另外值得一提的是, 解决训练数据不充分问题, 常用的还有经典的通用背景模型 (UBM) 法<sup>[9]</sup>。

#### 6.2.2 实验二

在实验一中, 码字是随机分配给各个类别 (说话人) 的, 那么, 类间码字分配的不同是否对识别性能会有影响? 即, Hadamard 纠错码是否对于码字的任意分配具有鲁棒性呢? 本文对此问题也进行了探讨, 以 7 类分类问题为例, 按照五种不同的码字分配方式来构建 SVM, 并进行了测试, 结果如下:

表 2 五种 Hadamard 纠错码码字的不同分配方式得到的识别结果

1 对余 (%)	7-bit Hadamard 纠错码的五种分配方式 (%)				
	a	b	c	d	e
93.21	97.14	96.43	97.86	97.14	96.43

从表 2 中可见, Hadamard 纠错码对于类间码字的不同分配具有较好的鲁棒性, 码字分配方式的改变, 并没有引起识别率的明显变化。

从前述实验中可知, 当应用于多分类问题中时, 对于不同的类别数, Hadamard 码都能提供稳定而优良的识别性能, 且对于类间码字的随机分配也具有良好的鲁棒性。

但值得注意的是, 对于  $k$  类问题, Hadamard 纠错码

法需要训练  $2^k-1$  个二元分类器,而“1 对余”则只用训练  $k$  个二元分类器,可见在大多数情况下( $k$  不是 2 的幂时),采用纠错码法的计算量要大,需要被存储的参数也多,所以在一些实时性要求较高的场合,这种方法仍需进一步改进.

## 7 结论

本文提出了利用 Hadamard 纠错码将二元分类器扩展到多分类问题中的办法,并将 Hadamard 纠错码和支持向量机相结合应用于说话人辨认任务中,对 7~15 人共 9 类的多分类问题进行了实验.此外,还将识别结果和经典的“1 对余”的支持向量机的多类推广方式进行了比较,实验结果表明 Hadamard 纠错码能成功地将支持向量机推广到多分类问题中,有效地提高识别系统的性能,且对于码字的随机分配也具有良好的鲁棒性.但它总体的计算量较大,影响了其综合性能,如何提高这种纠错码多分类器的效率,既是令之更具实用性的关键,也将是我们进一步的改进目标.

## 参考文献:

[1] Nilsson, N J. Learning Machines[M]. New York: McGraw

Hill, 1965.

- [2] Sejnowski, T J, Rosenberg, C R. Parallel networks that learn to pronounce English text[J]. Journal of Complex System, 1987, 1(1): 145-168.
- [3] T G Dietterich, G Bakiri. Solving multiclass learning problems via error correcting output codes[J]. Journal of Artificial Intelligence Research 1995, 2: 263-286.
- [4] V Vapnik. 统计学习理论的本质[M]. 北京: 清华大学出版社, 2000.
- [5] 刘志刚, 李德仁, 等. 支持向量机在多类分类问题中的推广[J]. 计算机工程与应用, 2004, 40(7): 10-13.
- [6] Bose R C, Ray Chaudhuri, D K. On a class of error correcting binary group codes[J]. Information and Control, 1960, 3(1): 68-79.
- [7] Duda, R O, Machanik, J w, Singleton, R C. Function modeling experiments[R]. Tech. rep. 3605, Stanford Research Institute, 1963.
- [8] 张彦仲, 沈乃汉. 快速傅里叶变换及沃尔什变换[M]. 北京: 航空工业出版社, 1989.
- [9] Douglas A Reynolds. Speaker identification and verification using Gaussian mixture speaker models[J]. Speech Communication, 1995, 17(1-2): 91-108.

## 作者简介:



尹安容 女, 1981 年 11 月出生于湖南岳阳. 2003 于北京理工大学电子工程系获取学士学位, 之后攻读通信与信息系统的硕士学位, 从事语音识别方面的有关研究.  
E-mail: anrongcat@hotmail.com



谢湘 男, 1976 年 1 月出生于河南淮阳, 北京理工大学副教授. 2002 年毕业于北京理工大学, 获博士学位. 主要研究领域为语音信号处理、无线通信技术. 累计发表学术论文 40 余篇, 译著 1 部. E-mail: xiexiang@bit.edu.cn