

文章编号: 1001-0920(2013)05-0671-06

基于纠错输出编码的多类概率建模

张玉玺^{1,2}, 王晓丹¹, 周进登¹, 姚旭¹

(1. 空军工程大学 防空反导学院, 西安 710051; 2. 93424部队, 北京 102101)

摘要: 目前模式识别领域中缺乏有效的多类概率建模方法, 对此提出利用纠错输出编码作为多类概率建模框架, 将二元纠错输出编码研究的概率输出问题转化为线性超定方程的求解问题, 通过线性最小二乘法来求解并获取多类后验概率的结果; 而对于三元纠错输出编码的等价非线性超定方程组, 提出一种迭代法则来求解多类概率输出. 实验中通过与3种经典方法相比较可以发现, 新方法求取的概率输出具有更好的分布形态, 并且该方法具有较好的分类性能.

关键词: 模式识别; 多类分类; 概率建模; 纠错输出编码

中图分类号: TP391

文献标志码: A

Probability modeling of multi-class based on error correcting output codes

ZHANG Yu-xi^{1,2}, WANG Xiao-dan¹, ZHOU Jin-deng¹, YAO Xu¹

(1. Air and Missile Defence College, Air Force Engineering University, Xi'an 710051, China; 2. Air Force 93424, Beijing 102101, China. Correspondent: ZHANG Yu-xi, E-mail: zhyuxibj@126.com)

Abstract: For the lack of effective method on multi-class' probabilities modeling, a posterior probability estimating approach based on error correcting output codes(ECOC) is presented. In this approach, probability estimation is translated into solving overdetermined linear equations with the binary ECOC as a modeling framework. Moreover, when the modeling framework is changed to ternary ECOC, the equivalence problem is overdetermined nonlinear equations. Therefore, an iterative algorithm is proposed to solve this problem for obtaining probabilities. Compared with three classic approaches, the proposed approach has better classification ability and probability distribution of the posterior probability in experiments.

Key words: pattern recognition; multi-class classification; probability modeling; error-correcting output codes

0 引言

在模式识别领域中, 经典二类分类的输出为硬输出, 同二类分类问题一样, 多类分类的结果往往也是硬输出, 所能得到的信息就是测试样本最后归属到具体的哪一类. 在实际应用中, 有时更需要得到样本可能属于每一类的度量值(即概率), 这样结果将包含更丰富的信息. 尤其当多类分类的结果并不是最终结果, 而是另一个子系统的输入, 这样概率输出的意义更显而易见. 比如在语音识别系统中, 声音的识别结果被作为一个隐马尔科夫模型的输入而用于词汇识别. 概率估计在机器学习领域的研究已有很多学者在进行, 文献[1-3]最早在此方面做了很多工作, 近年来随着多类分类问题的研究深入, 特别是纠错输出编码(ECOC)的大量应用, 基于ECOC的概率输出也逐渐引起众多研究者的关注. 在此方面的工作有, Reza

等^[4]提出通过构造等距纠错输出编码(Equi-distane ECOC)得到能同时最小化多类后验概率和二类后验概率损失的改进最小二乘法, 但该方法的前提是必须构造等距编码矩阵, 若编码矩阵不符合条件则不能得到此结论, 因此减弱了其应用推广的能力. Friedman提出的Max-Wins方法(即投票法), 是基于成对分类编码方法(也称“一对一”编码矩阵)计算多类别中每个两类组合形成的二分器得到的结果, 并将组合各类别对应的所有二分器输出之和作为该类的最终输出概率. 该方法的不足表现在对各两类分类器的结果进行组合时, 各分类器的重要性被同等看待, 即在合并每两个分类器得到的后验概率时, 并没有考虑其他类别出现的概率, 因此该方法得到的是一种近似的后验概率. 此外, 该方法的编码矩阵仅限定为成对编码阵, 不能推广到其他类型的纠错输出编码.

收稿日期: 2011-12-02; 修回日期: 2012-02-14.

基金项目: 国家自然科学基金项目(60975026, 61273275).

作者简介: 张玉玺(1976—), 男, 博士生, 从事智能信息处理、机器学习的研究; 王晓丹(1966—), 女, 教授, 博士生导师, 从事智能信息处理、机器学习等研究.

Hastie等^[5]利用成对分类编码方法并基于Kullback-Leibler距离得到概率输出(Pairwise Coupling),我国台湾学者Wu等^[6]对该方法进行了改进,得到3种等价问题.然而,此类方法同样是将编码矩阵限定为成对编码阵,从而存在泛化能力弱的缺点.

针对上述文献在研究基于ECOC获得多类概率输出时的不足,本文首先分析了Hastie等提出的Pairwise Coupling获取概率方法,并将该方法扩展到任意输出编码矩阵,进而分别研究针对二元和三元输出编码的多类概率建模方法.

1 相关概念介绍

1.1 纠错输出编码(ECOC)

ECOC框架即用一种二元或三元的编码矩阵实现多类类别分解和基分类器集成.在其编码矩阵中,二数码用 $M = \{-1, +1\}^{K \times L}$ 表示,三数码用 $M = \{-1, 0, +1\}^{K \times L}$ 表示.其中:“-1”代表一类,“+1”代表另一类;“0”表示该码字位所对应的类在其列所形成的二类划分中被忽略(即不参与由该列所产生的基分类器的训练)^[7-8]; K 为类别数; L 为编码长度.图1给出了4种常见的ECOC分类系统示意图,以编码矩阵来区分,分别是:“一对多”编码阵(one-versus-all)^[7,9],“一对一”编码阵(one-versus-one)^[5],密集随机阵(dense random)^[10]和稀疏随机阵(sparse random)^[10].

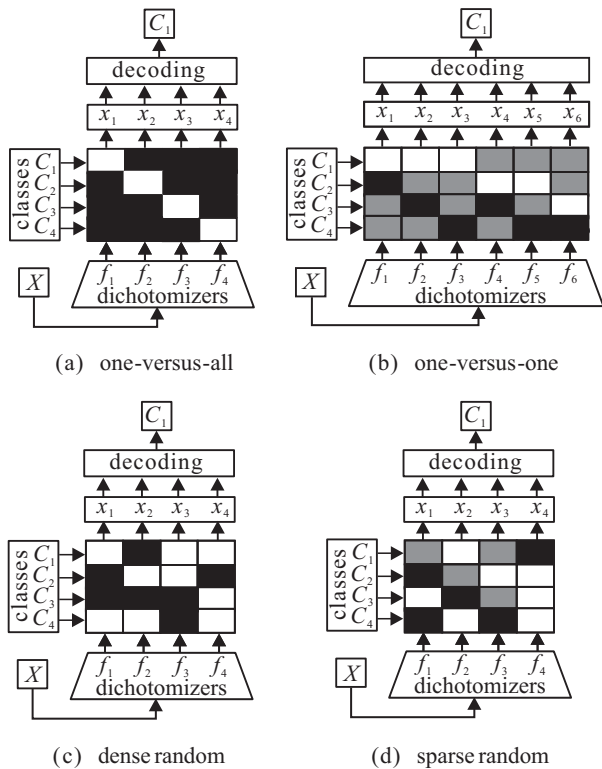


图1 4种常见的ECOC

图1中所有编码阵的每一行代表某一类 C_i ($i = 1, 2, 3, 4$)的码字,每一列代表样本的一种二类划分,码

元“1”、“-1”和“0”分别用白色、黑色和灰色表示.

在训练阶段,每一个基分类器 f_i ($i = 1, 2, \dots, 6$)的训练样本先根据其在编码阵对应的列重新划分,将多类转化成二类,然后分别训练得到与该列对应的二类分类器.例如,在图1(d)中对基分类器 f_3 进行训练时,白色对应的 C_2 为一类,黑色对应的 C_4 为另一类,而灰色对应的 C_1 和 C_3 不参与该基分类器的训练.由此依次训练得到4个二类分类器 $\{f_1, f_2, f_3, f_4\}$.在测试阶段,给定一个测试样本 X ,同时利用这4个二类分类器对其进行分类,结果为一码字向量 (x_1, x_2, x_3, x_4) (其中 $x_i \in \{-1, +1\}$),最后根据某种解码规则(即融合策略)对其进行解码,即可得最终分类结果.

1.2 Pairwise Coupling 概率建模法

假设 K 类样本数据中, n_{ij} 为第 i 类和第 j 类($i \neq j$)的训练样本之和, $r_{ij} = \text{Pr ob}(i|i \text{ or } j)$ 为此两类对应基分类器的输出,由文献[5]可知

$$\mu_{ij} = E(r_{ij}) = \frac{p_i}{p_i + p_j}, \tag{1}$$

p_i 和 p_j 分别为测试样本对应第 i 类和第 j 类的后验概率, μ_{ij} 为基分类器概率输出的期望值.模型如图2所示, C_i ($i = 1, 2, \dots, K$)表示类别, f_i ($i = 1, 2, \dots, L$)表示基分类器.

	r_{ij}						
	f_1	f_2	\dots	f_l	\dots	f_{l-1}	f_l
c_1	1	1	\dots	0	\dots	0	0
c_2	-1	0	\dots	0	\dots	0	0
\vdots	\vdots	\vdots		\vdots	\ddots	\vdots	\vdots
c_i	0	0	0	1	\dots	0	0
c_j	0	0	0	-1	\dots	0	0
\vdots	\vdots	\vdots	\vdots	\vdots		\vdots	\vdots
c_{k-1}	0	0	0	0	\dots	0	1
c_k	0	0	0	0	\dots	-1	-1

图2 基于“一对一”编码矩阵的概率输出模型示意图

在每一个基分类器中,参与训练的正类和负类各对应模式中的一类,即 $\text{class}^+ = \{c_i\}$, $\text{class}^- = \{c_j\}$,故基分类器的输出(概率输出,下文所提到的基分类器输出在不做特殊说明时都为概率输出)可以看成是正类所对应类别在负类所对应类别条件下的概率.式(1)等价于

$$\log \mu_{ij} = \log(p_i) - \log(p_i + p_j). \tag{2}$$

若要精确获得样本 x 对每一类的后验概率估计值,则须找到使 $\hat{\mu}_{ij}$ 和 r_{ij} 无限接近的 \hat{p}_i .然而,由式(1)组成的方程组为

$$\begin{cases} \mu_{12} = \frac{p_1}{p_1 + p_2}, \\ \mu_{13} = \frac{p_1}{p_1 + p_3}, \\ \vdots \\ \mu_{ij} = \frac{p_i}{p_i + p_j}, \\ \vdots \\ \mu_{(K-1)K} = \frac{p_{(K-1)}}{p_{(K-1)} + p_K}, \end{cases} \quad (3)$$

其中 $i < j$. 由于 $\sum_{i=1}^K p_i = 1$, 可知方程组有 $K - 1$ 个未知变量且有 $K(K - 1)/2$ 个方程与之对应, 这里研究多类问题, 故 $K > 2$. 因此, 式(3)为超定方程组, 而且为非线性超定方程组, 直接求解难度较大. 为此, Pairwise Coupling 方法通过构建 μ_{ij} 和 r_{ij} 之间的 Kullback-Leibler 距离作为目标函数, 即

$$l(x) = \sum_{i < j} n_{ij} \left[r_{ij}(x) \log \frac{r_{ij}(x)}{\mu_{ij}(x)} + (1 - r_{ij}(x)) \log \frac{1 - r_{ij}(x)}{1 - \mu_{ij}(x)} \right], \quad (4)$$

将非线性不定方程组的求解转化为求使式(4)值最小的各类后验概率估计值 $\hat{p}_i(x)$ 的最优化问题, 进而得到样本对各类的后验概率. 下面将重点介绍如何将此类方法扩展到纠错输出编码中.

2 基于纠错输出编码的概率建模

由 1.2 节可以知道, Pairwise Coupling 概率建模法是利用基分类器中正、负类别分别对应模式域中的某一类, 其输出可以看成是此两类中其中一类在另一类条件下的概率, 而该条件概率的计算可看成是由样本对两类后验概率的比值关系. 推广到纠错输出编码中, 由于基分类器中的正、负两类不再只包含模式域中的两类, 而是可能由一个或多个类别组成, 下面将讨论如何将 Pairwise Coupling 建模方法加以改进, 使其适应一般类型的纠错输出编码.

2.1 二元纠错输出编码概率建模

如 1.1 节所述, 在二元纠错输出编码中各码元值的取值范围为 $\{-1, +1\}$, 各基分类器所包含的类别为模式全部类别, 基分类器中正类和负类将至少有 1 个包含的类别数大于 2, 如图 3 所示. 此时基分类器的输出不再是两类之间的类条件概率, 而是 2 个子类划分集之间的类条件概率, 其中正子类集为编码矩阵中第 l 列编码为“+1”所对应的类别集合 $\text{class}_l^+ = \{c_i | m_{il} = +1, i = 1, 2, \dots, K\}$, 负子类集为编码为“-1”所对应的类别集合 $\text{class}_l^- = \{c_i | m_{il} = -1, i = 1, 2, \dots, K\}$. 因此式(1)可改写为

		f_1	f_2	\dots	f_l	\dots	f_{L-1}	f_L
					\downarrow			
c_1		1	1	\dots	1	\dots	-1	1
c_2		-1	1	\dots	1	\dots	-1	1
\vdots		\vdots	\vdots	\ddots	\vdots	\ddots	\vdots	\vdots
c_i		1	-1	-1	1	\dots	1	-1
c_j		-1	-1	1	-1	\dots	1	1
\vdots		\vdots	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
c_{K-1}		1	1	1	-1	\dots	0	1
c_K		-1	1	-1	-1	\dots	-1	-1

图 3 基于二元编码矩阵的概率输出模型示意图

$$\mu_l = p \left(\bigcup_{c_i \in \text{class}_l^+} c_i \mid \bigcup_{c_i \in \text{class}_l^+ \cup \text{class}_l^-} c_i \right) = E(r_l) = \sum_{c_i \in \text{class}_l^+} p_i / \left(\sum_{c_i \in \text{class}_l^+} p_i + \sum_{c_j \in \text{class}_l^-} p_j \right). \quad (5)$$

注意到式(5)右边的分母可以简化为

$$\sum_{c_i \in \text{class}_l^+} p_i + \sum_{c_j \in \text{class}_l^-} p_j = 1,$$

故式(5)可以简化为

$$\mu_l = E(r_l) = \sum_{c_i \in \text{class}_l^+} p_i. \quad (6)$$

假设各基分类器的输出组成向量为 $\mathbf{r} = (r_1, r_2, \dots, r_L)$, $\boldsymbol{\mu} = E(\mathbf{r})$, $\mathbf{p} = (p_1, p_2, \dots, p_K)$, 则式(6)可以写成

$$M^T \mathbf{p} = \boldsymbol{\mu}. \quad (7)$$

可以看出, 方程(7)为一线性超定方程组, 方程组的系数矩阵 M^T 为编码矩阵. 为求解该方程组, 可采用线性最小二乘问题广义逆法进行求解, 从而获得多类问题的输出概率建模.

2.2 三元纠错输出编码概率建模

在二元纠错输出编码中, 由于码元值“0”的引入, 各基分类器所包含的类别不一定是模式的全部类别, 如图 4 所示. 因 $\sum_{c_i \in \text{class}_l^+} p_i + \sum_{c_j \in \text{class}_l^-} p_j \neq 1$, 所以由式(5)得不到相对应的式(6), 故该超定方程组为非线性超定方程组. 为避免直接求解该方程所带来的计算复杂度, 参考 Pairwise Coupling 概率建模法, 通过构建一个目标函数, 将方程组的求解转化为最优化问题的求解, 为此先构建以 μ_l 和 r_l 之间的 Kullback-Leibler 距离作为目标函数, 即

$$l(x) = \sum_{l=1}^L n_l \left[r_l(x) \log \frac{r_l(x)}{\mu_l(x)} + (1 - r_l(x)) \log \frac{1 - r_l(x)}{1 - \mu_l(x)} \right], \quad (8)$$

其中 n_l 为编码矩阵中第 l 列所对应的各类别(类别标签为“+1”和“-1”)样本数总和, 该目标函数的梯度方

程为

$$\sum_{l=1}^L n_l r_l(x) = \sum_{l=1}^L n_l \mu_l(x). \quad (9)$$

由于

$$\mu_l = \sum_{c_i \in \text{class}_i^+} p_i / \left(\sum_{c_i \in \text{class}_i^+} p_i + \sum_{c_j \in \text{class}_i^-} p_j \right), \quad (10)$$

且满足 $\sum p_i = 1$, 参考 Hastie-Tibshirani 法则提出如下利用迭代过程求解步骤:

Step 1: 随机初始化 $\hat{p}_i(x)$ 值, 并利用式 (1) 计算对应的 $\hat{r}_l(x)$.

Step 2: 重复以下步骤直到 $\hat{p}_i(x)$ 值收敛:

Step 2.1: For each $i = 1, 2, \dots, K$,

$$\hat{p}_i(x) \leftarrow \hat{p}_i(x) \left(\frac{\sum_{l=1:L, \text{s.t. } M(i,l)=1} n_l r_l(x)}{\sum_{l=1:L, \text{s.t. } M(i,l)=1} n_l \hat{\mu}_l(x)} + \frac{\sum_{l=1:L, \text{s.t. } M(i,l)=-1} n_l (1 - r_l(x))}{\sum_{l=1:L, \text{s.t. } M(i,l)=-1} n_l (1 - \hat{\mu}_l(x))} \right);$$

Step 2.2: 归一化所得 $\hat{p}_i(x)$, $i = 1, 2, \dots, K$;

Step 2.3: 根据式 (10) 重新计算对应的 $\hat{\mu}_l(x)$.

此外需要注意的是, 要得到各类后验概率估计值的最优解, 该迭代过程必须收敛. 文献 [5] 对此已有证明, 本文不再详述. 至此, 便可以获得测试样本 x 对各类后验概率的输出.

	f_1	f_2	...	f_i	...	f_{L-1}	f_L
c_1	1	1	...	1	...	0	1
c_2	-1	0	...	0	...	-1	0
\vdots	\vdots	\vdots	\ddots	\vdots	\ddots	\vdots	\vdots
c_i	0	-1	0	1	...	0	-1
c_j	-1	0	1	-1	...	1	1
\vdots	\vdots	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
c_{K-1}	1	0	1	-1	...	1	0
c_K	0	1	-1	0	...	0	-1

图 4 基于三元编码矩阵的概率输出模型示意图

3 实验分析

为了评价本文提出的基于纠错输出编码多类概率建模方法, 实验中将分别与经典的 Equidistance-Least Square、Max-Wins 方法和 Pairwise Coupling 方法进行比较. 首先利用已知概率分布的 5 类高斯分布数据集, 比较 4 种概率建模方法所得的样本后验概率与真实后验概率的差异度, 对比分析不同建模方法的优劣; 接着, 利用 5 种典型飞机缩比模型的实测数据

转换为一维高分辨距离像 (HRRP) 用于目标识别, 比较在编码矩阵长度相同时基于 4 种不同概率建模方式分类器分类性能的好坏.

3.1 实验数据

实验中采用的第 1 种数据为 5 类二维正态分布数据集, 各类别数据的先验概率相同, 其概率密度函数如下所示 (各类别分布参数如表 1 所示):

$$p(x|\text{class}_k) = \frac{1}{2\pi\sigma_k^2} \exp \left[\frac{\|x - \mu_k\|^2}{-2\sigma_k^2} \right], \quad k = 1, 2, \dots, 5. \quad (11)$$

表 1 5 类人工数据集各类别分布参数

Class	Prior Probabilities	Mean Vectors	Covariance Matrices
C_1	$P(C_1) = \frac{1}{5}$	$\mu_1 = (0, 0)^T$	$\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
C_2	$P(C_2) = \frac{1}{5}$	$\mu_2 = (3, 0)^T$	$\Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$
C_3	$P(C_3) = \frac{1}{5}$	$\mu_3 = (0, 5)^T$	$\Sigma_3 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$
C_4	$P(C_4) = \frac{1}{5}$	$\mu_4 = (7, 0)^T$	$\Sigma_4 = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}$
C_5	$P(C_5) = \frac{1}{5}$	$\mu_5 = (0, 9)^T$	$\Sigma_5 = \begin{bmatrix} 8 & 0 \\ 0 & 8 \end{bmatrix}$

使用表 1 所示参数, 根据贝叶斯法则可以得到各类别的后验概率. 因此, 由贝叶斯分类器可以得到各类别的决策分界面, 且通过计算可以得到该 5 类人工数据集的贝叶斯分类错误率为 25.76%.

实验中采用的第 2 类数据集为 5 种典型飞机目标缩比模型在实验室的实测数据. 原始数据为各目标的 X 波段水平极化幅相数据: 频段为 8~12 GHz, 共 101 个频率点, 均匀间隔; 方位角度范围为 $0^\circ \sim 160^\circ$ (采样间隔 0.5°), 横滚角和俯仰角都为 0° . 实测数据经过变换可以得到目标相应姿态角的 HRRP 数据集, 变换过程中选取原始数据的前 64 个频率点, 故 HRRP 数据维数为 64 维.

3.2 实验设计

为验证本文基于两种不同类型纠错输出编码的概率建模方法优劣, 实验中通过与 3 种经典的概率建模方法在 5 类人工数据集进行比较, 观察各模型概率输出与真实概率输出的差异. 为避免由于基分类器数量不同带来的影响, 基分类器的数量以 Pairwise Coupling 模型所包含的基分类器数量为标准, 在具体操作中二元纠错输出编码选择 Dense Random 编码, 三元纠错输出编码选择 Sparse Random 编码. 在选择两种随机编码方法时, 将分别从已产生的 2000 个密集及稀疏随机编码阵集 (对应各码元概率分别为 $p(-1) = 0.5$, $p(+1) = 0.5$; $p(-1) = 1/3$, $p(0) = 1/3$, $p(+1) = 1/3$) 中随机选择所需要的编码阵. 接着, 将

本文所提出的方法和3种经典概率建模方法应用于5种不同飞机的HRRP目标识别中,用以检验各方法在实际应用中的效果. 实验中将分别选取4个不同角度范围(0° ~ 30°, 60° ~ 90°, 120° ~ 150°和60° ~ 150°)的HRRP数据集加以识别. 为简化实验的操作性,选择支持向量机(SVM)为基分类器,采用十重交叉验证来估计分类错误率,并将利用双边估计t检验法来计算置信水平为0.95的分类错误率置信区间作为最终结果,计算公式如下:

$$\frac{|\bar{x} - \mu|}{\sigma/\sqrt{n}} \geq t_{0.025}(n-1). \quad (12)$$

其中: μ, σ 分别表示十重交叉验证的均值和标准差, $t_{0.025}(9) = 2.2622$. 同时为了对实验结果进行统计分析,采用Nemenyi检验法对各编码矩阵分类效果之间的差异显著性进行检验. 实验中所用基分类器均来自PRTool (<http://www.prtools.org>) 工具箱,实验机器配置为1G内存,2.30G CPU,算法基于Matlab7.7 (R2008b)实现.

3.3 实验结果与分析

图5为各概率建模方法所获得概率与真实概率的对角线图,在每一个子图中,横坐标为估计概率,纵

坐标为真实概率,且每个子图代表一类数据的概率估计差异,分布在对角线的样本越多说明此建模方法在该类样本数据的概率建模越准确. 图5中从上至下依次为Equidistance-Least Square方法、Max-Wins方法、Pairwise Coupling方法以及本文提出的基于Dense Random建模法和基于Sparse Random建模法,从左至右依次为第1类到第5类数据样本的估计概率与真实概率对角线图. 从图5可看出,基于纠错输出编码概率建模法相比其余3种经典概率建模法能更好地估计样本的概率,表现为各样本的概率比值分布都围绕在对角线附近. 可以得出这样的结论:在基分类器数量相同或相近时,基于纠错输出编码(包括二进制码和三元码)能获得更接近真实概率的估计.

为进一步验证基于各概率建模所获得概率的分类性能,表2~表5为选择4种不同方位角范围数据集,基于各概率建模法的分类错误率,基分类器为SVM. 从表中结果可以看出,基于纠错输出编码矩阵的概率模型的分类错误率明显低于其他模型方法,从而说明识别多类HRRP目标时,基于纠错输出编码概率建模法设计的分类器具有更好的分类效果. 由于

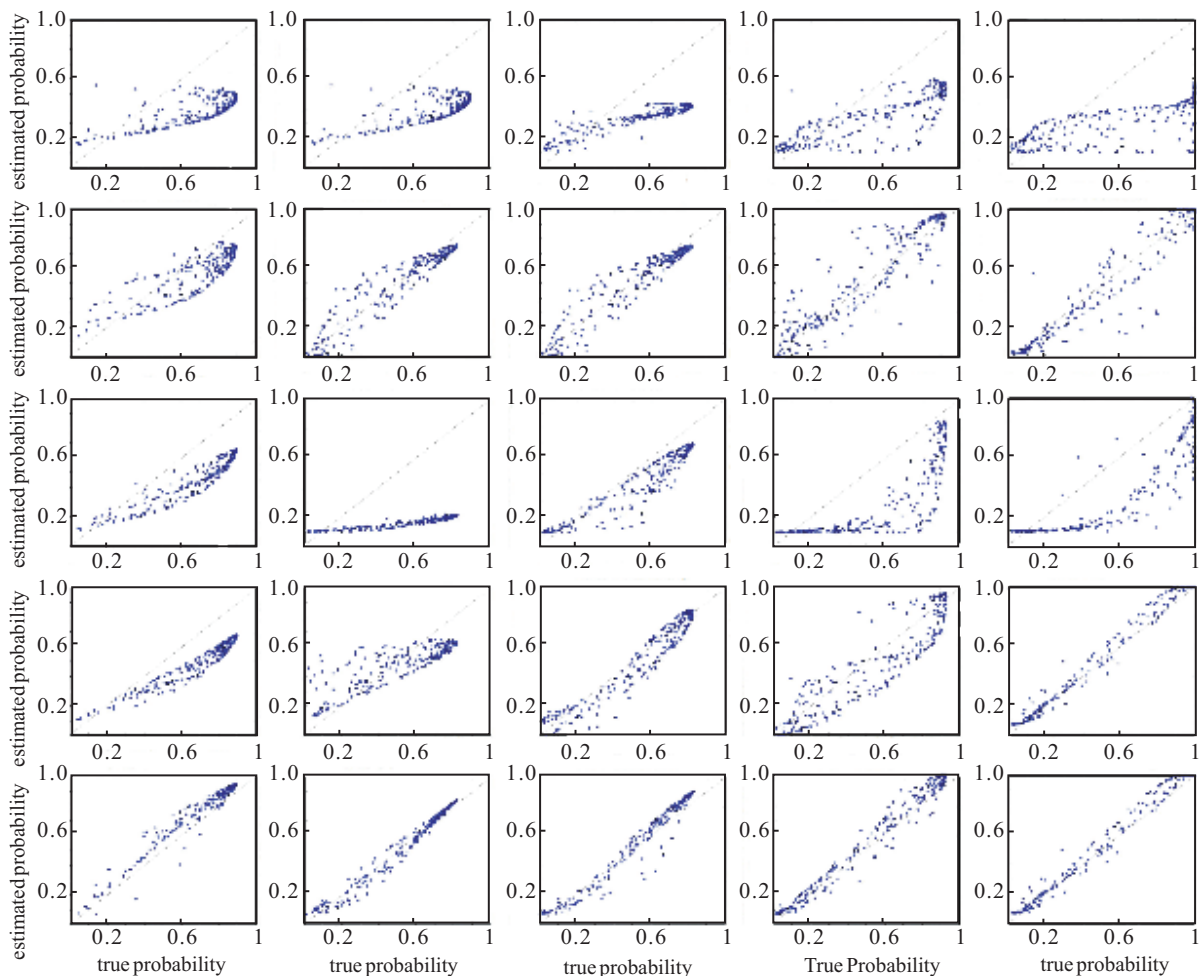


图5 Equidistance-Least Square方法、Max-Wins方法、Pairwise Coupling方法以及纠错输出编码概率建模法在5类高斯分布样本数据概率估计与真实概率的对角线图

表 2 角度范围为 $0^\circ \sim 30^\circ$ 时 HRRP 分类错误率 %

Target	Equidistance-Least Square	Max-Wins	Pairwise Coupling	Dense Random	Sparse Random
A	11.89±2.41	13.41±2.45	9.21±1.24	6.95±1.09	5.19±0.92
B	9.79±1.69	16.62±2.87	10.80±2.20	6.47±1.02	7.99±1.87
C	16.70±2.05	16.46±2.53	12.81±1.41	8.55±1.79	8.35±0.81
D	9.36±1.95	10.24±1.72	8.66±2.09	4.91±1.45	4.75±0.65
E	13.72±3.14	15.24±2.36	10.43±2.25	5.56±0.88	5.35±1.28

表 3 角度范围为 $60^\circ \sim 90^\circ$ 时 HRRP 分类错误率 %

Target	Equidistance -Least Square	Max-Wins	Pairwise Coupling	Dense Random	Sparse Random
A	14.40±2.79	18.29±2.76	8.65±1.70	5.39±1.15	4.05±0.44
B	11.08±3.32	19.93±3.16	11.31±1.63	6.05±2.74	8.72±1.21
C	13.63±1.66	16.74±1.39	7.60±1.37	4.14±1.56	3.99±0.32
D	12.36±2.19	13.24±1.72	9.73±1.41	5.56±0.72	5.12±0.63
E	10.70±2.83	19.63±0.98	10.46±2.65	8.43±1.17	7.31±1.91

表 4 角度范围为 $120^\circ \sim 150^\circ$ 时 HRRP 分类错误率 %

Target	Equidistance-Least Square	Max-Wins	Pairwise Coupling	Dense Random	Sparse Random
A	11.89±2.17	12.41±2.33	9.13±2.04	8.17±1.53	6.04±0.83
B	12.79±1.73	17.62±1.92	11.78±1.18	5.39±0.92	4.71±1.05
C	9.70±1.30	19.24±2.31	8.96±0.68	6.73±1.41	4.20±0.69
D	11.70±2.01	15.24±0.89	11.04±1.57	6.34±1.09	5.61±1.85
E	13.70±2.51	17.24±2.02	9.70±1.12	8.92±2.03	7.04±0.99

表 5 角度范围为 $60^\circ \sim 150^\circ$ 时 HRRP 分类错误率 %

Target	Equidistance-Least Square	Max-Wins	Pairwise Coupling	Dense Random	Sparse Random
A	25.89±6.41	28.41±3.75	22.21±4.24	18.97±3.59	15.19±2.02
B	23.79±3.69	29.62±2.87	21.80±2.20	16.47±1.07	19.91±1.87
C	23.70±3.05	25.24±2.66	23.11±2.41	17.56±3.73	19.85±1.81
D	21.70±2.60	24.24±2.03	21.46±3.54	19.45±1.51	16.92±2.32
E	28.70±3.19	30.24±2.99	25.09±3.86	20.53±1.94	22.28±2.43

HRRP 具有姿态敏感性, 当角度范围较大时, 将影响分类器的识别效果, 因此, 表 5 所有方法的分类错误率都高于表 2~表 4 对应的结果. 表 2~表 4 中基于 Dense Random 和 Sparse Random 建模法的分类错误率均低于 10%, 说明基于这两种方法设计的分类器对不同方位的 HRRP 目标具有稳定的识别性能.

4 结 论

多类概率输出建模是模式识别领域的难点, 现有方法大都存在模型不准确、求解难度大以及推广能力弱的缺点. 本文利用纠错输出编码作为建模框架, 在深入分析基于“一对一”编码方法即 Pairwise Coupling 方法的基础之上, 将其扩展到基于纠错输出编码的多类概率建模方法中, 进而得到基于二元纠错输出编码的线性超定方程组, 通过线性最小二乘法解此方程组从而获得多类后验概率. 虽然在对基于三元纠错输出编码的概率建模过程中也可获得类似的超定方程组, 但考虑到该超定方程组为非线性的, 直接求解存在很大困难, 因此参照 Pairwise Coupling 法将其转化为一个最优化问题, 并利用迭代法来获取对应的概率输出. 利用 5 类高斯分布数据集和 HRRP 数据

集分别进行概率准确率分布和基于此概率模型的分类型性能评估, 验证发现本文所提出的基于纠错输出编码概率获取法在不增加额外的基分类器情况下能准确地对各样本后验概率建模, 同时基于此概率建模能明显地提高分类性能.

参考文献(References)

- [1] Quinlan J R. Decision trees as probabilistic classifiers[C]. Proc of the 4th Int Workshop on Machine Learning. Tottori, 1987:31-37.
- [2] Izenman A J. Recent developments in nonparametric density estimation[J]. J of the American Statistical Association, 1991, 86(413): 205-224.
- [3] Richard D M, Lippmann R P. Neural network classifiers estimate Bayesian a posteriori probabilities[J]. Neural Computation, 1991, 4(3): 461-483.
- [4] Ghaderi R, Windeatt T. Least squares and estimation measures via error correcting output code[R]. The 2nd Int Workshop on Multiple Classifier Systems, Lecture Notes in Computer Science, Springer-Verlag, 2001.

(下转第 687 页)