

# 基于互信息的主成分分析用于声场景分类

范雪莉, 冯海泓, 原 猛

(中国科学院声学研究所东海研究站, 上海 200032)

**摘要:** 主成分分析是声场景分类中常用的特征选择方法。针对主成分分析的局限性, 提出一种基于互信息的主成分分析方法。这一方法引入类别信息, 用不同声场景条件下特征之间的互信息矩阵之和替代传统主成分分析中的协方差矩阵, 计算其特征向量与特征值, 特征向量表示由原始特征空间向新的主成分空间的转换系数, 特征值则用于计算主成分的累计贡献率并判断主成分维数。声场景分类实验结果表明, 该方法较之传统主成分分析方法降维效果更好, 辅以神经网络分类器, 计算得到的分类正确率更高。

**关键词:** 互信息; 主成分分析; 声场景分类; 特征选择

中图分类号: TP391.4

文献标识码: A

文章编号: 1000-3630(2013)-03-0222-06

DOI 编码: 10.3969/j.issn1000-3630.2013.03.010

## Mutual information based principal component analysis for acoustic environment classification

FAN Xue-li, FENG Hai-hong, YUAN Meng

(Shanghai Acoustics Laboratory, Chinese Academy of Science, Shanghai 200032, China)

**Abstract:** Principal Component Analysis (PCA) is a common method for feature selection. To improve selection, an algorithm of Principal Component Analysis based on Mutual Information (PCAMI) is presented. PCAMI introduces the category information, and uses the sum of mutual information matrices between features under different acoustic environments to replace the covariance matrix. The eigenvectors of the sum matrix represent the transformation coefficients and the eigenvalues of the sum matrix are used to calculate the accumulative contribution rate to determine the dimensionality of the principal components. The experiments of acoustic environment classification show that relative to the traditional PCA, PCAMI has better performance in dimensionality reduction and in classification accuracy under the assistance of neuron network.

**Key words:** mutual information; principal component analysis; acoustic environment classification; feature selection

## 0 引言

随着信号处理技术的发展, 信号处理策略越来越注重声场景的差异带来的影响, 表现为在不同的声场景下, 信号处理策略相异或策略相同, 但是参数的设置不同。自动识别当前声场景在很多领域得到了广泛的发展<sup>[1-6]</sup>, 如数字助听器自动程序切换技术<sup>[7,8]</sup>, 可以识别用户当前所处的环境, 自动调整后端信号处理策略或参数设置, 而无需用户手动切换。声场景分类属于模式识别<sup>[9]</sup>的范畴, 一般的处理方法是: 声音信号首先经过特征选择<sup>[10]</sup>, 选择出反映不同类别之间差异的有效特征, 之后经过分类

器<sup>[11]</sup>, 分类器结果表明当前输入信号属于哪个类别。Roberto Gil-Pita 等人利用量化后的 8 维特征和人工神经网络最高得到了 92.12% 的分类正确率<sup>[12]</sup>。Sourabh Ravindran 等人利用 104 维的特征和基于 ADABOOST 的分类器最高得到了 97.91% 的分类正确率<sup>[13]</sup>。Enrique Alexandre 等人利用 25 维特征和 Fisher 分类器最高得到了 95.8% 的分类正确率<sup>[14]</sup>。

在处理过程中, 特征选择是决定分类效果的关键问题之一。一方面, 特征的维数较高会带来“维数灾难”<sup>[15]</sup>, 使得分类效果恶化, 计算复杂度大幅度提升, 所需样本数量急剧增加; 另一方面, 特征之间的冗余会造成虚假信息, 或者特征集包含与分类无关的特征造成计算量的提高。因此如何更有效的进行特征选择是目前研究的一个热点。

主成分分析 (Principal Component Analysis, PCA)<sup>[16]</sup> 是目前应用较为广泛的特征选择方法之一。PCA 将原有特征进行线性加权, 得到各个主成分。权系数向量的方向称为主成分方向, 在主成分方向

收稿日期: 2012-12-17; 修回日期: 2013-04-08

基金项目: 国家自然科学基金青年科学基金(11104316)、上海市自然科学基金(11ZR1446000)资助项目。

作者简介: 范雪莉(1984-), 女, 河南许昌人, 博士, 研究方向为信号与信息处理、模式识别。

通讯作者: 范雪莉, E-mail: fanxueli@mail.ioa.ac.cn

上数据的方差最大化。各个主成分不相关，主成分序号按照各自方向的方差大小降序排列，如主成分 1 表示在此方向上数据的方差最大，主成分 2 表示在此方向上数据的方差次大，依此类推。但是对于分类而言，主成分分析并没有用到类别的信息，而且方差并不能很好地表示类别之间的差异，方差最大的方向只能表示作为同一类别的数据在此方向上分布最离散，但是不能完全等同于多类数据在此方向上分离程度最大化。

基于信息的度量<sup>[17]</sup>能够以量化的形式描述随机变量之间的不确定程度，表示随机变量之间关系的强弱，并且这种度量不局限于线性关系。信息度量标准包括信息增益(Information Gain, IG)<sup>[18]</sup>或互信息(Mutual Information, MI)<sup>[19]</sup>等，可以计算各个类别条件下特征之间的信息。在特征选择时考虑到类别的作用，更加适合分类问题，同时可以度量特征之间相互依赖的程度，包括非线性相关度，较之主成分分析采用相关法有所改进，因此信息熵理论被引入到特征选择领域，逐步成为近年来一个研究热点<sup>[20]</sup>。

本文将互信息引入特征选择中，引入了类别信息与特征之间的互信息，用各个类别条件下特征之间的互信息之和取代特征的协方差矩阵，避免了方差最大化不能完全等同于类别数据分布最离散。通过声场景分类实验，利用神经网络分类器，在同样的主成分选择准则下，基于互信息的主成分分析特征选择算法所需特征的维数更少；在同样的特征维数下，其分类的准确率更高。

## 1 主成分分析特征选择

### 1.1 主成分分析

PCA 的主要目的是对一个维数较高、各维变量之间相互关联的数据集进行线性降维，并且降维后数据集尽量保留原始信息。降维后的低维空间称为主成分空间，降维后的新特征被称为主成分(Principal Component, PC)，且各个主成分之间是不相关的。PCA 的主要思路是在数据空间找到某一个方向，在此方向上数据的方差最大，此方向称为主成分方向 1，数据在此方向上的投影被称为主成分 1，在与此方向无关的其他方向上，找到数据的方差次大，则此方向称为主成分方向 2，数据在此方向上的投影被称为主成分 2，依此类推。

假设特征空间  $R^n$  上的样本数据  $x$ ，经过主成分分析后为主成分空间  $S^m$  内的数据  $y$ ，则

$$y = A'x \tag{1}$$

式中： $y$  的第  $k$  维元素为  $y_k$ ，也是第  $k$  个主成分； $A$  是转换矩阵， $A$  的第  $k$  列向量  $a_k$  表示第  $k$  个主成分方向：

$$y_k = a_k'x \tag{2}$$

根据方差最大化的要求，第一主成分  $y_1$  满足最大化此方向上的方差：

$$\text{var}[a_1'x] = a_1' \Sigma a_1 \tag{3}$$

式中， $\Sigma$  表示  $x$  的协方差矩阵，协方差矩阵是通过计算整个样本集合上所有数据任意两两特征的协方差得到的。由于满足此要求的向量  $a_1$  不唯一，因此需要添加一定限制，一般要求  $a_1'a_1 = 1$ 。利用拉格朗日因子法求解  $a_1$ ，最大化下式

$$a_1' \Sigma a_1 - \lambda(a_1'a_1 - 1) \tag{4}$$

可以推导出

$$(\Sigma - \lambda I_n)a_1 = 0 \tag{5}$$

其中  $I_n$  是  $n \times n$  的单位阵。

因此， $\lambda$  是此主成分方向  $a_1$  上的方差，且  $\lambda$  是  $\Sigma$  最大的特征值， $a_1$  是对应的特征向量。

$a_2$  的求解与  $a_1$  类似，再加上一个限制， $a_1'x$  与  $a_2'x$  无关，即

$$\begin{aligned} \text{cov}[a_1'x, a_2'x] &= a_1' \Sigma a_2 = a_2' \Sigma a_1 = \\ & a_2' \lambda a_1 = \lambda a_2' a_1 = 0 \end{aligned} \tag{6}$$

即  $a_2'a_1 = 0$

利用拉格朗日因子法得到  $a_2$  为  $\Sigma$  次大的特征值对应的特征向量。以此类推，各个主成分方向为降序排列的  $\Sigma$  的特征值对应的特征向量。

因此

$$A' \Sigma A = \Lambda \tag{7}$$

$A$  由  $\Sigma$  的特征向量为列向量组成的矩阵， $\Lambda$  是由  $\Sigma$  的特征值组成的对角阵。主成分维数  $m$  由前  $m$  个主成分对应特征值之和所占总体特征值之和的比重确定，一般选择 85%~95%。

### 1.2 主成分分析的弊端

方差最大化准则只能表示作为同一类别的数据在此方向上分布最离散，但是不能完全等同于多类数据在此方向上离散程度最大化。应用于分类时，尤其对于类别的分布方向与所有数据分布方向不同的情况下，主成分分析忽略类别间的差异，将所有数据默认为一类，仅关注所有数据的分布特点，得到的主成分方向不能反应类别差异。如下列情况：

假设数据集  $X$  是一个两维( $f_1, f_2$ )的数据集，共 200 个数据，分为两类，每类 100 个数据都遵循高斯分布，各自的均值和协方差矩阵分别为  $[1,3], [1,0;$

0,50]; [5,3], [1,0; 0,50]。将所有数据进行主成分分析,得到主成分 1 的贡献率为 91.2%,图 1 显示了原始数据以及主成分 1 的方向,为清楚地显示主成分 1 的方向,将其方向不变,长度变为原来的 10 倍。

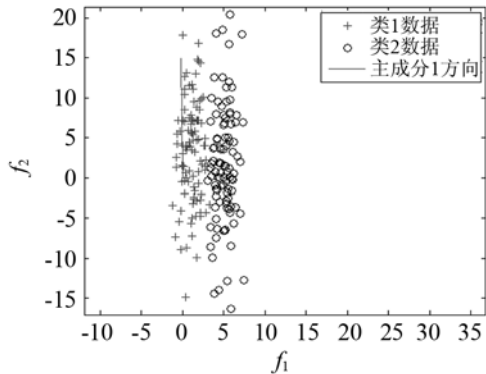


图 1 高斯分布下 PCA 得到的 PC1 方向  
Fig.1 The PC1 direction from PCA of Gaussian data

从图 1 可以看出,类 1 数据和类 2 数据在主成分 1 方向的方差是较大的,计算得到主成分 1 的贡献率达到了 90%以上,但是两类数据在主成分 1 方向的投影是混叠在一起的,完全不能分出两类。虽然数据方差在主成分 1 的方向最大,但是由于方差最大化的方向与类别差异最大化的方向不同,因此得到的主成分 1 对识别两类数据来说并不是一个好的特征。传统的主成分分析并不考虑类别这一因素,而是把所有类别的数据混为一谈,这样做不仅忽略了类别信息,而且方差最大化原则只适合特定情况,即方差最大的方向同时也是类别差异最大的方向,若两者不相同,特征选择的效果将大大下降,甚至如例子所显示的,主成分完全不能分开不同类别的数据。

## 2 基于互信息的主成分分析特征选择算法

### 2.1 基于互信息的主成分分析法

互信息<sup>[21]</sup>是对消息之间或消息集合之间相互提供信息量多少的度量。对于两个随机变量  $F_1$  和  $F_2$ ,互信息可以衡量两者之间相互依赖的程度,表示两个变量间共同拥有信息的含量,并且这种度量不局限于线性相关的两个随机变量,对于非线性相关也可以进行衡量。若  $F_1$  和  $F_2$  各自的边缘概率分布及联合概率分布分别为  $p(f_1)$ ,  $p(f_2)$  和  $p(f_1, f_2)$ ,则它们之间的互信息  $I(F_1; F_2)$  定义为

$$I(F_1; F_2) = \sum_{f_1} \sum_{f_2} p(f_1, f_2) \log \frac{p(f_1, f_2)}{p(f_1)p(f_2)} \quad (8)$$

当变量  $F_1$  和  $F_2$  完全无关或相互独立时,互信息最小,结果为 0,这意味着两个变量之间不存在重叠的信息;反之,两者的相互依赖程度越高,互信息的值就越大,所包含的相同信息也越多。

鉴于 PCA 忽略类别信息以及采用方差最大化的原则带来的不良影响,提出一种基于互信息的主成分分析(Principle Component Analysis based on Mutual Information, PCAMI)特征选择算法。PCAMI 加入类别信息,并选择用各个类别条件下特征之间的互信息矩阵之和,而非原始算法中的协方差矩阵,进行主成分的计算。将式(7)改为

$$B^T \Psi_{I_F} B = A \quad (9)$$

式中,  $B$  为 PCAMI 的转换矩阵,其列向量为  $\Psi_{I_F}$  的特征向量,  $A$  为  $\Psi_{I_F}$  特征值  $(\mu_1, \mu_2, \dots, \mu_n)$  为对角线元素的对角阵,  $n$  为数据集的原始维数,  $\Psi_{I_F}$  表示各个类别条件下特征之间的互信息矩阵之和。

$$\Psi_{I_F} = \sum_{c=1}^l MI(F|C) \quad (10)$$

式中,  $l$  表示类别总数,  $MI(F|C)$  表示类别  $C$  条件下特征  $F$  两两之间的互信息矩阵,计算时利用数据集集合中的所有数据。 $MI(F|C)$  第  $i$  行第  $j$  列的元素为  $I(f_i; f_j | C)$ ,表示类别  $C$  条件下特征  $f_i$  和  $f_j$  之间的互信息。 $\Psi_{I_F}$  中对角线元素表示特征的自信息,也就是信息熵,非对角线元素表示特征之间的互信息。无论是互信息或者是信息熵都是实数,当两个特征之间不相关时,互信息为 0,否则为正数,因此  $\Psi_{I_F}$  为非负实数阵。互信息满足对称性,因此  $\Psi_{I_F}$  为非负实数对称阵。所以  $\Psi_{I_F}$  的特征值为实数,  $A$  为实数对角阵,对应的特征向量两两正交,矩阵  $B$  为正交阵。

PCAMI 的主成分  $z$  为

$$z = B^T x \quad (11)$$

第  $k$  维主成分  $z_k = \beta_k^T x$ ,且主成分两两正交。

接下来判断主成分的维数  $m$ 。依照 PCA 第  $k$  维主成分的贡献率  $\theta_k$  和第  $k$  维主成分的累计贡献率  $\zeta_k$  的定义<sup>[16]</sup>,PCAMI 中  $\theta_k$  与  $\zeta_k$  沿用 PCA 中的定义:  $\theta_k$  为单一主成分占总体主成分的比率,  $\zeta_k$  为前  $k$  个主成分贡献率之和

$$\theta_k = \frac{\mu_k}{\sum_{k=1}^n \mu_k}, \quad \zeta_k = \sum_{i=1}^k \theta_i \quad (12)$$

选择  $\zeta$  为 85%~95% 的前  $m$  个主成分作为新的特征。

### 2.2 计算互信息

互信息的计算最重要的部分是概率的估计,本

文选择直方图法计算边缘概率和联合概率密度。直方图法凭借数据集本身得到样本的近似分布，因此也称作经验估值，在统计学中广泛应用。给定样本集  $X$  的  $n$  个观察值  $x_1, \dots, x_n$ ，其中相同值可能重复出现多次。令  $c(x)$  为观察值  $x$  的计数函数，表示  $x_k$  在  $n$  次观察中所出现的次数，那么  $x_k$  的近似概率  $p(x)$  为它在观察中出现次数与总次数的比率，即

$$p(x) = \frac{c(x)}{n} \quad (13)$$

倘若  $x$  是连续变量，那么可通过离散化方法将  $x$  的取值范围划分为  $k$  个不相交区间  $(x_i, x_{i+1}]$ ,  $i=0, \dots, k-1$ 。则  $c(x)$  为落在各个区间内的  $x$  的次数，之后再根据式(13)计算变量  $x$  的每个取值范围所相应的概率  $p(x)$ 。对于两个(或多个)随机变量  $X$  和  $Y$ ，它们的联合概率分布可通过相同方式获得，不同之处在于计数函数  $c(x,y)$  表示  $x$  和  $y$  同时出现的次数。

数据的直方宽度，即等宽区间  $(x_i, x_{i+1}]$  的宽度，遵循 Scott 原则<sup>[22]</sup>：

$$h = 3.5\sigma n^{-1/3} \quad (14)$$

式中： $h$  是直方图直方宽度； $\sigma$  是数据的标准方差， $n$  是数据数量。

### 3 实验分析

实验一：比较 1.2 节的数据经过两种方法后的结果。

数据集如 1.2 节所述，数据集方差最大的方向与两类数据差异最大的方向不一致(图 2)。

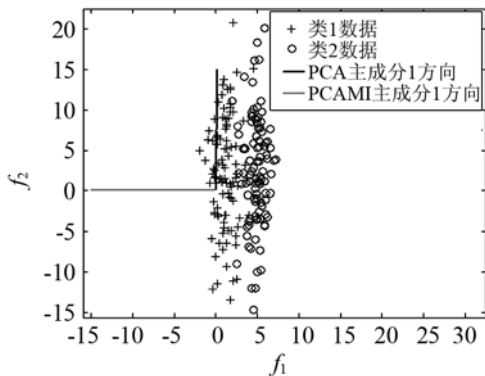


图 2 高斯分布下 PCA 和 PCAMI 得到的 PC1 方向  
Fig.2 The PC1 direction from PCA and PCAMI of Gaussian data

从图 2 可以看出，PCAMI 得到的 PC1 的方向基本上是平行于  $f_1$  轴的，在此方向上两类数据的投影可以较好地地区分。从两类数据的分布来看，协方差矩阵分别为  $[1,0;0,50]$  和  $[1,0;0,50]$ ， $f_2$  方向上数据的方差远大于  $f_1$  方向，即每一类数据在  $f_1$  方向上分布较集中，均值分别为  $[1,3]$  和  $[5,3]$ ，中心不重合，

且连线平行于  $f_1$  轴，因此两类数据在  $f_1$  方向上紧紧围绕在均值中心处，可以较好地分辨出来，与 PCAMI 的实验结果一致。而 PCA 将两类数据混合在一起，默认为一类，寻找方差最大的方向，得到的 PC1 方向如前所述，并不能反映两类数据的差异，而将两类数据的投影混叠在一起，不能分辨。

实验二：应用于声场景分类时，比较 PCA 和 PCAMI 特征提取算法得到的主成分信息量和累积贡献率。

选择 NOISEUS, NOISEX-92 等声音数据库，包括纯净语音、语音在交谈噪声中、语音在工厂噪声中、噪声和音乐 5 类声场景，一共 500 个声音文件，每类 100 个。采用常用的 28 维的特征，包括倒谱及其方差和倒数、谱重心、包络调制深度、过零率及其方差、基频等。

表 1 为利用所有声音文件进行 PCA 和 PCAMI 分析后，各个主成分对应的贡献率以及累积贡献率。

表 1 PCA 和 PCAMI 得出的主成分贡献率以及累积贡献率  
Table 1 Contribution rate and accumulative contribution rate of principal components from PCA and PCAMI

	PCA		PCAMI	
	特征值 $\mu$	主成分累积贡献率 $\zeta/\%$	特征值 $\mu$	主成分累积贡献率 $\zeta/\%$
PC1	12.78	45.64	16.92	61.53
PC2	5.72	66.05	5.69	82.23
PC3	2.73	75.80	1.17	86.50
PC4	1.18	80.02	0.95	89.94
PC5	0.98	83.51	0.60	92.13
PC6	0.65	85.82	0.46	93.80
PC7	0.63	88.06	0.32	94.98
PC8	0.52	89.90	0.28	95.99
PC9	0.49	91.65	0.15	96.52
PC10	0.46	93.31	0.12	96.98
PC11	0.32	94.46	0.11	97.37
PC12	0.32	95.60	0.10	97.74
PC13	0.26	96.54	0.09	98.07
PC14	0.16	97.13	0.08	98.36
PC15	0.14	97.64	0.07	98.61
PC16	0.12	98.07	0.06	98.84
PC17	0.11	98.48	0.05	99.03
PC18	0.09	98.79	0.05	99.22
PC19	0.08	99.08	0.04	99.37
PC20	0.06	99.28	0.04	99.53
PC21	0.04	99.44	0.03	99.65
PC22	0.04	99.58	0.02	99.72
PC23	0.03	99.71	0.02	99.80
PC24	0.03	99.80	0.02	99.87
PC25	0.02	99.88	0.02	99.94
PC26	0.02	99.95	0.01	99.97
PC27	0.01	100.00	0.01	99.99
PC28	0.00	100.00	0.00	100.00

从表 1 可以看出, 相同维数情况下, PCAMI 的主成分累积贡献率要高于 PCA 的主成分累积贡献率。如 PCA 的主成分 1 的累积贡献率为 45.64%, 而 PCAMI 的主成分 1 的累积贡献率为 61.53%, 并且这种优势持续保持。在选择主成分维数时, 若以主成分累积率 85%为限度, 采用 PCA 方法需要 6 个主成分, 但是采用 PCAMI 方法时只需要 3 个主成分; 若以主成分累积率 90%为限度选择主成分的维数时, 采用 PCA 方法需要 9 个主成分, 但是采用 PCAMI 方法时只需要 6 维, 因此大大降低了数据维度。

以上结果说明, 采用 PCAMI 方法较之 PCA 方法降低了数据的维数, 但是降维后的主成分是否能够取得较好的分类效果呢? 以下的实验将进行验证分析。

实验三: 分析 PCA 和 PCAMI 得到的主成分分类的效果。

以实验二中 PCA 和 PCAMI 降维后的主成分作为新的特征, 选择主成分贡献率为 85%、90%、95% 三种情况下的主成分维度, 以神经网络作为分类器验证声音分类的效果。选择多层感知器(Multi-Layer Perceptron, MLP)网络结构, 网络共分三层, 输入层, 中间层和输出层。输入层的神经元个数与 PCA 或 MIPCA 选择的维度相同, 中间层的神经元个数固定为 5, 输出层的神经元个数与所选类别数一致。输入层的传递函数为 logsig 函数, 中间层的传递函数为 purelin 函数。训练方法为后向传播(Back-Propagation, BP)方法。5 类样本每类各 100 例, 总数量 500 例, 其中, 随机抽取 60%用于训练 MLP 网络, 剩余 40%用于测试分类正确率, 保证训练集与测试集不存在任何交集。利用训练数据集计算主成分, 得到的主成分系数用于测试集, 评价方法的分类正确率。

从表 1 可得到主成分贡献率首次大于 85%时, PCAMI 得到 3 维主成分, PCA 得到 6 维主成分; 主成分贡献率首次大于 90%时, PCAMI 得到 5 维主成分, PCA 得到 9 维主成分; 主成分贡献率首次大于 95%时, PCAMI 得到 8 维主成分, PCA 得到 12 维主成分。为同 PCA 的 6 维主成分进行比对, 选择 PCAMI 的 6 维主成分, 而非 5 维, 此时 PCAMI 6 维主成分的贡献率为 93.80%。将以上主成分作为新的特征输入神经网络进行分类, 结果见表 2。

从表 1 可以看出, PCA 得到的 6 维主成分累积贡献率为 85.82%, MIPCA 得到的 3 维主成分累积贡献率为 86.50%, 两者相差不到 1%, 而表 2 中, 以 PCA 6 维数据进行分类的正确率为 79.5%, 以

表 2 分类正确率  
Table 2 Classification accuracy

PCA 维数	分类正确率/%	PCAMI 维数	分类正确率/%
6	79.5	3	79.5
9	89.5	6	90.5
12	92.5	8	93.0

PCAMI 3 维数据进行分类的正确率也为 79.5%, 两者相同, 其他横向比较结果也相差无几。因此以相同的主成分贡献率为原则计算得到的主成分维数作为新的特征, 分类正确率相似。若是维度相同, 如都是 6 维, PCA 6 维主成分的累积贡献率从表 1 可以看到为 85.82%, PCAMI 6 维主成分的累积贡献率为 93.80%, 也就是说, 以累积贡献率为例, 后者高出前者约 8 个百分点, 从表 2 可以看出, 前者的分类正确率为 79.5%, 后者为 90.5%, 后者高出前者 11 个百分点。因此, 累积贡献率越高, 分类正确率也越高。

以累计贡献率首次大于 85%作为降维后选择主成分维度的最低限, 则 PCA 方法需要 6 维, PCAMI 方法需要 3 维。以 PCA 得到的 6~28 维主成分和 PCAMI 得到的 3~28 维主成分依次输入到分类器中, 得到如图 3 所示的主成分与分类正确率的关系图。从图中可以看出, 两种方法的分类正确率与主成分维度成正比, 且维度相同时, PCAMI 方法的分类正确率均高于 PCA 方法, 或者说, 在相同分类正确率下, PCAMI 所需的主成分更少, 即降维效果优于 PCA。PCA 方法在 21 维时分类正确率达到饱和, PCAMI 方法在 16 维时达到饱和, 此时两者的分类正确率均为 98%。

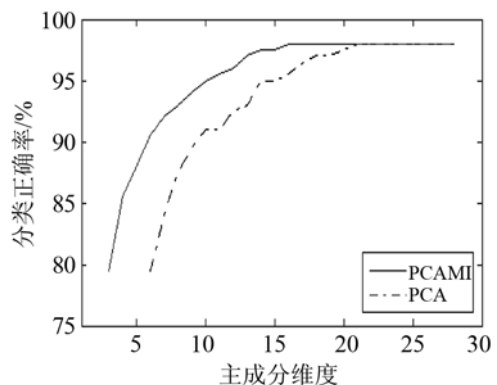


图 3 主成分维度与分类正确率的关系  
Fig.3 The relationship between the dimensionality of PCs and classification accuracy

综前所述, 累积贡献率不仅反映了数据的保留信息, 也在一定程度上反映了主成分对正确分类的影响, 也就是说以累积贡献率为准则对主成分维度进行选择是可行的, 除此以外, 在相似的分类正确率下, 如正确率为 90%左右, 需要 9 维 PCA 的主

成分, 相对仅需要 6 维 PCAMI 的主成分, 其他的类似, 因此采用 PCAMI 需要的特征维数少于采用 PCA 的维数, 也就是说 PCAMI 的降维效果优于 PCA, 则后端分类器的计算量会大大减少, 且降低了维数灾难的风险。

## 4 结论

本文将互信息引入特征选择中, 在进行主成分分析时, 用各个类别条件下特征之间的互信息之和取代特征的协方差矩阵, 避免了方差最大化不能完全等同于类别数据分布最离散, 引入了类别信息与特征之间的互信息。通过声场景分类实验, 利用神经网络为分类器, 分析了基于互信息的主成分分析特征选择算法的性能。

实验表明: (1) 传统 PCA 利用方差最大化原则不能很好地用于分类问题, PCAMI 将类别信息引入, 较好地反映了不同类别的差异; (2) 应用于声场景分类时, 以相同的主成分贡献率为准则判断 PCs 维数, PCAMI 特征选择算法较之 PCA 得到更少的主成分维度; 相同的主成分维度下, PCAMI 中 PCs 的贡献率更多; (3) 应用于声场景分类时, 以神经网络为分类器, 同样维数的主成分, 采用 PCAMI 较之 PCA 特征选择算法得到更高的分类正确率; 相同主成分贡献率情况下, PCAMI 特征选择算法需主成分的维数更少。

## 参 考 文 献

- [1] Ghoraani B, Krishnan S. Time-frequency matrix feature extraction and classification of environmental audio signals[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2011, **19**(7): 2197-2209.
- [2] Vesa Peltonen, Juha Tuomi, Anssi Klapuri, Jyri Huopaniemi, and Timo Sorsa, Computational Auditory Scene Recognition[C]// *Proc. of ICASSP Florida, USA, May 2002*.
- [3] LU Lie, ZHANG Hongjiang, JIANG Hao. Content analysis for audio classification and segmentation[J]. *IEEE Transactions on Speech and Audio Processing*, 2002, **10**(7): 504-516.
- [4] MA Ling, Ben Milner, Dan Smith. Acoustic environment classification[J]. *ACM Transactions on Speech and Language Processing*, 2006, **3**(2): 1-22.
- [5] Choi, Jae-Hun, Chang, Joon-Hyuk. On using acoustic environment classification for statistical model-based speech enhancement[J]. *Speech Communication*, 2012, **54**(3): 477-490.
- [6] Lucas Cuadra, Roberto Gil-Pita, Enrique Alexandre, et al. Joint design of Gaussianized spectrum-based features and least-square linear classifier for automatic acoustic environment classification in hearing aids[J]. *Signal Processing*, 2010, **90**(8): 2628-2632.
- [7] 蔡银华, 王冬鹤, 闻雨婷, 等. 助听器自动程序切换技术的发展——从自动导声到动态声景自适应[J]. *中国听力语言康复科学杂志*, 2008, **30**(5): 62-63.  
CAI Yinhua, WANG Donghe, WEN Yuting, et al. Research and development of automatic program——From AutoPilot to SoundFlow[J]. *Chinese Scientific Journal of Hearing and Speech Rehabilitation*, 2008, **30**(5):62-63
- [8] Enrique Alexandre, Lorena Alvarez-Perez, Roberto Gil-Pita, et al. On the design of automatic sound classification systems for digital hearing aids[C]// *AES 126th Convention, Munich, Germany, 2009 May 7-10*.
- [9] Bishop, Christopher. *Pattern Recognition and Machine Learning*[M]. Berlin: Springer, 2006.
- [10] 边肇祺, 张学工. *模式识别*[M]. 第二版. 北京: 清华大学出版社, 2000.  
BIAN Zhaoqi, ZHANG Xuegong. *Pattern Recognition*[M]. Second Edition. Beijing: Tsinghua University Press, 2000.
- [11] Yongkoo Han, Kisung Park, Young-Koo Lee. Confident wrapper-type semi-supervised feature selection using an ensemble classifier[C]// *2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC)*, 2011, 8-10 Aug. 2011: 4581-4586.
- [12] Roberto G P, Enrique A. Analysis of the effects of finite precision in neural network-based sound classifiers for digital hearing aids[J]. *EURASIP Journal on Advances in Signal Processing*, 2009, **2009**(1): 2-2.
- [13] Ravindran S, Anderson D V. Audio classification and scene recognition and for hearing aids[C]// *IEEE International Symposium on Circuits and Systems*, 2005: 860-863.
- [14] Alexandre E, Cuadra L, Álvarez-Perez L, et al. Automatic sound classification for improving speech intelligibility in hearing aids using a layered structure[J]. *Intelligent Data Engineering and Automated Learning-IDEAL*, 2006, **4224**: 306-313.
- [15] Bellman R. *Adaptive Control Processes: A Guided Tour*[M]. Princeton: Princeton University Press, 1961.
- [16] Jolliffe I T. *Principal Component Analysis*[M]. Springer-Verlag, 1986.
- [17] Memisevic Roland, Sigal Leonid, Fleet David J. Shared kernel information embedding for discriminative inference[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34**(4): 778-790.
- [18] YU Lei, LIU Huan. Efficient feature selection via analysis of relevance and redundancy[J]. *J of Machine Learning Research*, 2004, **5**(1): 1205-1224.
- [19] Battiti R. Using mutual information for selecting features in supervised neural net learning[J]. *IEEE Trans on Neural Networks*, 1994, **5**(4): 537-550.
- [20] 姚旭, 王晓丹, 张玉玺, 等. 特征选择方法综述[J]. *控制与决策*, 2012, **27**(2): 161-166.  
YAO Xu, WANG Xiaodan, ZHANG Yuxi, et al. Summary of Feature Selection Algorithms[J]. *Control and Decision*, 2012, **27**(2): 161-166.
- [21] 田宝玉, 杨洁, 贺志强, 等. *信息论基础*[M]. 第一版. 北京: 人民邮电出版社, 2008.  
TIAN Baoyu, YANG Jie, HE Zhiqiang, et al. *Elements of Information Theory*[M]. First Edition. Beijing: Posts and Telecom Press, 2008.
- [22] Scott D W. On optimal and data-based histograms[J]. *Biometrika*, 1979, **66**(1): 605-610.