文章编号:1001 - 9081(2009)10 - 2741 - 03

基于 Mean-Shift 的广播音频聚类算法

郑继明[1],俞 佳[2]
(1. 重庆邮电大学 应用数学研究所,重庆 400065; 2. 重庆邮电大学 计算机科学与技术学院,重庆 400065)
(yujiatoutou@163.com)

摘 要:针对大多数聚类算法依赖聚类数目这一先验知识的不足,提出一种基于均值漂移(Mean-Shift)的新广播音频聚类算法。对需聚类的音频段选取基于小波域的特征构造特征集合,通过主成分分析方法降低所提取特征中的冗余信息。在此基础上,采用 Mean-Shift 算法对音频信号进行初步聚类,然后利用快速近邻法对其聚类结果进行一次修正,最后合并仅含有单个样本类别的类进行二次修正。实验结果表明,该算法的聚类精度有一定的提高。

关键词:主成分分析;均值漂移算法;快速近邻法;二次修正;广播音频聚类
中图分类号: TP18;TP391 **文献标志码**:A

Audio clustering algorithm based on Mean-Shift in broadcasting

ZHENG Ji-ming[1], YU Jia[2]
(1. *Institute of Applied Mathematics, Chongqing University of Posts and Telecommunications, Chongqing* 400065, *China*;
2. *College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing* 400065, *China*)

Abstract: Since most clustering algorithms depend on the number of the prior knowledge, a novel broadcasting audio clustering algorithm based on Mean-Shift was proposed. Firstly, a Principal Component Analysis (PCA) transformation was used to reduce redundant feature set information, which based on wavelet domain. Mean-Shift thought was applied to cluster the audio as the primary clustering. Then it used the fast nearest neighbor algorithm to revise the result of the first time, and merged the class that only contained single sample to the second correction. The experiment results show that clustering accuracy has been improved to a certain extent.

Key words: Principal Component Analysis (PCA); Mean-Shift algorithm; fast nearest neighbor algorithm; two revisions; broadcast audio clustering

0 引言

近年来,广播新闻语料、电台语音等现实世界中的语音信号已经成为了研究重点。在现有的音频数据中,同一文件中不同的时间段往往对应着不同的说话人或者不同的录制语音的环境,因此有必要将其中具有不同特性的音频数据分成不同的段,然后将具有相同特性的音频段聚到一起,为提高音频检索系统的性能打下基础。

一个好的聚类算法应该能自动地把样本聚成应该有的类数,也就是说,不管待聚类的样本在空间的分布如何,算法都应该能准确地将其聚类。目前大多数聚类算法依赖于聚类数目这一先验知识[1],并且在特征空间的分析时加入了人为的假设条件,如:K 均值聚类算法需要指定聚类的类别数,并且只适合发现球状簇,面对延伸状的簇和大小差别很大的簇无能为力[2];模糊 C 均值聚类[3]则需要事先定义模糊隶属度函数。由于对于一个未知的音频流,了解其先验知识不太实际,因此本文采用 Mean-Shift 算法[4-6]进行聚类。Mean-Shift 算法不需要任何先验条件,数据集中的每一点都可作为初始点,分别执行 Mean-Shift 算法,收敛到同一个点算作一类,它能对任何维度、任何分布的采样点进行聚类。

图 1 是本文设计的音频聚类算法框图。由于聚类效果的好坏很大程度上取决于信号所在的特征空间,因此,在特征选取方面,本文将选取基于小波域的特征构造特征集合;同时为了消除特征数据的相关性,降低冗余信息,对于构造出来的特征集合进行 PCA 变换。在此基础上,采用 Mean-Shift 算法对音频信号进行初步聚类,然后利用快速近邻法对 Mean-Shift 算法的聚类结果进行一次修正,最后合并仅含有单个样本类别的类进行二次修正。实验结果表明,该算法的聚类性能较好。

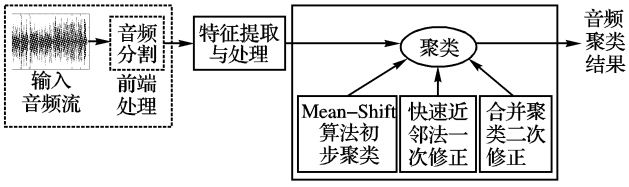


图 1 音频聚类算法

1 聚类特征选取与处理

本文选取了文献[7]分析的 1 维低帧能量比、1 维基因变化率和 12 维的 MFCC 特征值来构造 14 维的特征向量集作为聚类算法的输入。对于构造出来的特征集合再进行 PCA 变换,用于消除特征数据的相关性,降低所提取特征的冗余信息。

PCA 是模式识别分析中最常用的一种线性映射方法[8-9],该方法是根据样本点在多模式空间的位置分布,以样

收稿日期:2009 - 04 - 16;**修回日期**:2009 - 06 - 22。 **基金项目**:重庆市教育委员会科学技术研究项目(KJ080524)。
作者简介:郑继明(1963 -),男,重庆人,副教授,主要研究方向:小波分析、多媒体技术; 俞佳(1984 -),女,安徽宣城人,硕士研究生,主要研究方向:小波分析、基于内容的音频检索。

本点在空间中变化最大方向，即方差最大的方向，作为判别矢量来实现。它能够对原始数据进行简化，有效地找出原始数据中最主要的元素和结构，去除噪音和冗余，将复杂的原始数据进行简化。

给定一组 m 维输入向量 $\boldsymbol{x}_t; t = 1,\cdots,l$ 且 $\sum_{t=1}^{l} x_t = 0$，通过式(1)PCA 将每一个 m 维向量 $\boldsymbol{x}_t$ 线性地变换为一新的 s 维向量 $\boldsymbol{y}_t(s \leqslant m)$：

$$\boldsymbol{y}_t = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}_t \tag{1}$$

其中：$\boldsymbol{U}$ 表示一个 $m \times m$ 的正交矩阵，它的第 i 列 $\boldsymbol{u}_i$ 是样本协方差矩阵 $\boldsymbol{C} = \frac{1}{l}\sum_{t=1}^{l}\boldsymbol{x}_t\boldsymbol{x}_t^{\mathrm{T}}$ 的第 i 个特征向量。PCA 需首先求解式(2) 的特征值：

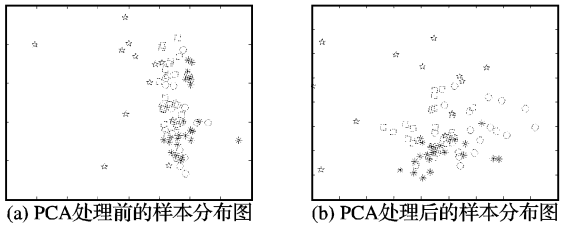
$$\lambda_i\boldsymbol{u}_i = \boldsymbol{C}\boldsymbol{u}_i;\ i = 1,\cdots,m \tag{2}$$

其中：λ_i 表示样本协方差矩阵 $\boldsymbol{C}$ 的第 i 个特征值，$\boldsymbol{u}_i$ 是相应的特征向量。协方差矩阵 $\boldsymbol{C}$ 的特征值按降序排列 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_m \geqslant 0$，相应的特征向量构成矩阵 $\boldsymbol{U} = [\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_m]$。计算出 $\boldsymbol{u}_i$ 后，向量 $\boldsymbol{y}_t$ 可以通过式(3) 对向量 $\boldsymbol{x}_t$ 进行正交变换得到：

$$\boldsymbol{y}_t = \boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{x}_t;\ i = 1,\cdots,m \tag{3}$$

通常只选择有序特征向量最前面的几个向量表示原输入，这样 $\boldsymbol{y}_t$ 中主分量的个数将减少，从而可以达到降维的效果。但是，在本文中，要进行 PCA 变换的特征向量集已经经过小波变换并且是由挑选出的特征构成，进行 PCA 变换是为了消除彼此之间的相关性，因此本文选择了所有的有序特征向量作为聚类的输入。

本文将实验样本在 PCA 处理前后的分布作了比较，如图 2 所示，图中的四个类型分别代表实验中设计的聚类类别，即男播音、女播音、外景采访和音乐。从图 2 可以看出，经过 PCA 变换后的同类样本分布较处理前的分布更聚集一些，不同类的样本分布更散开一些，满足聚类的要求，在下面的实验部分中将具体给出聚类的效果比较。



(a) PCA处理前的样本分布图　(b) PCA处理后的样本分布图

图 2　PCA 处理前后样本分布图比较

2　基于 Mean-Shift 的广播音频算法

利用前面中提取出的特征向量，采用 Mean-Shift 算法对音频信号进行初步聚类，然后利用快速近邻法对 Mean-Shift 的结果进行一次修正，最后合并含有单个元素类别的类进行二次修正。

2.1　Mean-Shift 算法

Mean-Shift 算法[8]是从密度函数梯度的非参数估计中推导获得，而非参数估计则是从样本集出发对密度函数进行估计，它不需要任何先验知识，对任意形状的分布都有效。其中最常用的是核密度估计，它根据核函数 $k(\boldsymbol{x})$ 对样本集进行计算得到密度函数。给定在 d 维空间 $\mathbf{R}^d$ 的 n 个样本数据 $\boldsymbol{x}_i, i = 1,\cdots,n$，多维变量的核函数估计可写成：

$$\hat{f}_{h,k}(\boldsymbol{x}) = \frac{c_{k,d}}{nh^d}\sum_{i=1}^{n}|\boldsymbol{H}|^{-1/2}k\left(\left\|\frac{\boldsymbol{x}-\boldsymbol{x}_i}{\boldsymbol{H}}\right\|^2\right) \tag{4}$$

其中：$c_{k,d}$ 为归一化常数；$k(\boldsymbol{x})$ 为核函数；$\boldsymbol{H}$ 为 $d \times d$ 维的带宽矩阵，完整的参数表示 $\boldsymbol{H}$ 会增加估计的复杂性，在实际中，$\boldsymbol{H}$ 一般被限定为一个对角矩阵 $\boldsymbol{H} = \mathrm{diag}[h_1^2,\cdots,h_d^2]$，甚至更简单地被取为正比于单位矩阵，即 $\boldsymbol{H} = h^2\boldsymbol{I}$，$\boldsymbol{I}$ 为单位矩阵。由于后一形式只需要确定一个系数 h，在 Mean-Shift 中常常被采用，在本文的后面部分也采用这种形式，则核密度估计为：

$$\hat{f}_{h,k}(\boldsymbol{x}) = \frac{c_{k,d}}{nh^d}\sum_{i=1}^{n}k\left(\left\|\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right\|^2\right) \tag{5}$$

由文献[5] 得到核密度估计的梯度如式(6) 所示，其中定义 $g(\boldsymbol{x}) = -k'(\boldsymbol{x})$。

$$\nabla\hat{f}_{h,k}(\boldsymbol{x}) = \frac{2c_{k,d}}{nh^{d+2}}\left[\sum_{i=1}^{n}g\left(\left\|\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right\|^2\right)\right]\times\left[\frac{\sum_{i=1}^{n}\boldsymbol{x}_i g\left(\left\|\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n}g\left(\left\|\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right\|^2\right)} - \boldsymbol{x}\right] \tag{6}$$

Mean-Shift 向量为：

$$\nabla\boldsymbol{m}_{h,g}(\boldsymbol{x}) = \frac{\sum_{i=1}^{n}\boldsymbol{x}_i g\left(\left\|\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n}g\left(\left\|\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right\|^2\right)} - \boldsymbol{x}$$

由 Mean-Shift 向量得到 Mean-Shift 迭代公式：

$$y_{t+1} = y_t + \nabla m_{h,g}(y_t) \tag{7}$$

式(7) 经过一些转换得到 Mean-Shift 算法的迭代式：

$$y_{t+1} = \frac{\sum_{i=1}^{n}\boldsymbol{x}_i g\left(\left\|\frac{\boldsymbol{y}_t-\boldsymbol{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n}g\left(\left\|\frac{\boldsymbol{y}_t-\boldsymbol{x}_i}{h}\right\|^2\right)} \tag{8}$$

Mean-Shift 算法具体步骤描述如下：

1) 在特征空间中任意选择初始搜索区域圆，半径为带宽 h，设置结束条件 ε；

2) 用式(8) 计算 y_{t+1} 的值；

3) 判断是否满足 $|y_{t+1} - y_t| < \varepsilon$，如果满足，则退出；否则，用 y_{t+1} 代替 y_t 转 2)。

在上述的均值漂移算法的带宽矩阵 h 和核函数 $k(x)$ 始终保持不变。本文中所使用的核函数为高斯函数：

$$k_N(x) = \exp\left(-\frac{1}{2}\|x\|^2\right) \tag{9}$$

因为它估计得更准确，收敛路径更平滑。

2.2　快速最近邻算法一次修正

Mean-Shift 算法能够将大多数的音频片段聚为一类，但是会遗漏一些较为“偏远” 的但是应该属于一类的音频片段，因此，本文采用了快速最近邻算法对 Mean-Shift 算法的聚类结果进行一次修正。

Mean-Shift 算法的聚类输出包括：每一类的中心矢量、属于同一类的所有原始音频片段的序号以及每一个音频片段属于的类别号。

快速最近邻算法的步骤如下所示。

1) 从第一类开始，每个类根据自己的中心矢量找到与自己距离最近的那个类，形成最近邻对。本文中采用的距离公式是均方误差欧氏距离，其定义[10] 为：

$$d_2(X,Y) = \frac{1}{K}\sum_{i=1}^{K}(x_i - y_i)^2 \tag{10}$$

其中,$d_2(X,Y)$ 的下标 2 表示平方误差,K 为特征向量维数。

2) 假设类 C_i 的最近邻为 C_j,对于所有以 C_i 和 C_j 为最近邻的类,找出这些最近邻对中的最小距离:

$$d_{\min}(C_i,C_j) = \min_{(C_n,C_m)\in\phi} d(C_n,C_m) \tag{11}$$

其中 $\phi = \{(C_n,C_m) \mid C_n = C_i \text{ or } C_m = C_j\}$。

3) 将所有满足以下条件的最近邻类对聚成新类:

$$\begin{aligned} &d(C_n,C_m) < \beta d_{\min}(C_i,C_j);(C_n,C_m)\in\phi \\ &d(C_n,C_m) < DT \end{aligned} \tag{12}$$

其中 β 为比例因子,本文取的是 1.5;DT 为最大可以容忍的类间距离,本文取为 0.000 2。

2.3　合并聚类二次修正

由于音频样本多种多样,经过上面两个步骤后,有可能会存在一个类别中只包含单个音频片段的情形。为了便于管理聚类结果,本文把聚类类别中只含有单个音频片段的类合并为一类,作为聚类结果输出的最后一类。

3　实验测试

实验中使用的音频数据来自于美国之音(VOA)的广播新闻和 CD 音乐,数据的采样频率选择 11.025 kHz,精度为 16 位。该数据集由 80 段长度不等的音频数据组成,时间长度在 15 ~80 s,其内容包括男女播音员的标准语音、外景采访人员和被采访人员语音、演讲现场音频、电话录音以及音乐等,设计被聚为男播音、女播音、外景采访和音乐四类。

聚类的正确率采用式(13)来衡量:

$$\text{聚类的正确率} = \frac{\text{聚类正确的音频段}}{\text{所有的音频总数}} \tag{13}$$

在音频聚类算法中,初始点的选取对聚类效果有一定的影响。由于本文提出的音频聚类算法随机选取初始点进行迭代,为了评估本聚类算法的性能,将数据集进行五次四交叉验证,设置初始搜索区域圆半径即带宽 h 为 3.0,实验结果如表 1 所示。

表 1　本文算法四交叉验证聚类精度正确率　%

实验	第一组	第二组	第三组	第四组	平均
实验一	84	86	90	80	85
实验二	82	84	80	78	81
实验三	80	76	90	74	80
实验四	84	84	76	76	80
实验五	80	82	80	78	80
平均	82.0	82.4	83.2	77.2	81.2

Mean-Shift 算法虽然不需要任何先验知识,但带宽矩阵是其中一个重要参数,它不但决定了参与迭代的采样点数量,而且还会影响算法的收敛速度和准确性。目前主要有两种带宽矩阵计算方法:固定带宽法和自适应计算法。

本文采用的是前者,并在相同条件下变化带宽,以达到较优聚类效果。由于第一组实验样本包括的各类样本数目比较平均,因此采用第一组的实验样本,同样进行五次验证,实验结果如表 2 所示。

表 2　影响聚类效果的带宽变化

带宽值	聚类类别数目	聚类正确率/%					
		实验一	实验二	实验三	实验四	实验五	平均
2.8	4	78	78	82	78	80	79.20
2.9	4	76	84	82	76	82	80.00
3	4	84	82	80	84	80	82.00
3.1	3	84	86	86	86	86	85.60
3.2	2	88	82	88	88	88	86.80

从表 2 中可以看出,随着带宽的增加,平均聚类正确率单调递增,但是它的聚类类别数目是单调递减的。表中的 4 类是按照期望聚为男播音、女播音、外景采访和音乐四类;3 类是将男播音和女播音聚为一类,即聚为标准语音、外景采访和音乐;2 类是将男女播音聚为一类,音乐和外景采访聚为一类。由此可知当带宽为 3.1 或者 3.2 时,聚类正确率虽然较高,但是只将音频样本聚为 2 类,不符合我们的期望结果。经过综合比较,本文中选用的带宽值为 3.0。

Mean-Shift 算法能够广泛应用于聚类,但是应用在音频聚类上的较少。本文提出的广播音频聚类算法是基于 Mean-Shift 算法的改进音频算法,因此,为了评估本聚类算法的性能,将其与单一 Mean-Shift 聚类算法进行比较,同样将数据集进行五次四交叉验证,实验结果如表 3 所示。

对比表 1 和表 3 可以得出,本文提出的改进音频算法的平均聚类正确率较单一 Mean-Shift 聚类算法提高了 2% 左右。其中,第一组样本集的平均聚类正确率提高近 5%,其他三组都在 2% 左右。原因可能是由于在带宽比较实验中,选择的是第一组数据作为样本集,因此,实验中用到的带宽 3.0 较适合第一组的样本分布,聚类正确率相对要高一些。第四组的聚类精度在两次实验中都不高,分别只有 77.2% 和75.6%,可能是由于该组的音频样本中音乐样本的数目较多,包括有不同的音乐类型,速度平缓的和激昂的,因此影响了整体的聚类正确率。

表 3　Mean-Shift 算法四交叉验证聚类正确率　%

实验	第一组	第二组	第三组	第四组	平均
实验一	82	84	88	78	83.0
实验二	82	84	78	78	80.5
实验三	72	76	88	72	77.0
实验四	76	82	76	74	77.0
实验五	76	82	80	76	78.5
平均	77.6	81.6	82.0	75.6	79.2

由于聚类效果的好坏很大程度上取决于信号所在的特征空间,为了评估本文提出的特征集的有效性,本文将进行 PCA 变换的特征集和未进行 PCA 变换的特征集的聚类效果进行了比较。同样将数据集进行五次四交叉验证,实验结果如表 4 所示。

(下转第 2750 页)

推广性较强。但是要准确识别出编码蛋白质基因需要进一步的研究。一方面,仅从GC含量和Z曲线特征上识别编码ORF还有一定的局限性,可能还需要借助其他的特征;另一方面,非线性支持向量机的训练速度极大地受到训练集规模的影响。对于超大规模的数据集,如何高效地进行训练和测试也是一个需要研究的重要问题。

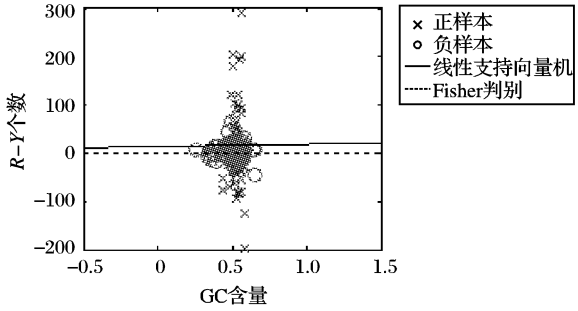


图4 Fisher判别和线性支持向量机

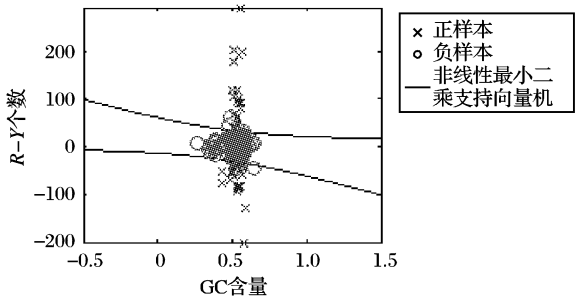


图5 非线性最小二乘支持向量机

参考文献:

[1] 郭峰彪. 原核生物蛋白质编码区识别及基因组序列分析[D]. 天津: 天津大学理学院, 2005.

[2] 张春霆. 人与其他生物基因组若干重要问题的生物信息学研究[J]. 自然科学进展, 2004, 14(12): 1367 - 1374.

[3] BESEMER J, LOMSADZE A, BORODOVSKY M. GeneMarkS: A self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions [J]. Nucleic Acids Research, 2001, 29(12): 2607 - 2618.

[4] DELCHER A L, HARMON D, KASIF S, *et al*. Improved microbial gene identification with GLIMMER[J]. Nucleic Acids Research, 1999, 27: 4636 - 4641.

[5] YADA T, TOTOKI Y, TAKAGI T, *et al*. A novel bacterial gene-finding system with improved accuracy in locating start codons[J]. DNA Research, 2001, 8: 97 - 106.

[6] 史良. 国内外基因计算机识别的研究方法及进展[J]. 北京生物医学工程, 2004, 23(1): 73 - 74.

[7] SUYKENS J A K, VANDEWALLE J. Least squares support vector machine classifiers[J]. Neural Processing Letters, 1999, 9(3): 293 - 300.

[8] 闻芳. 基于支持向量机(SVM)的剪接位点识别[J]. 生物物理学报, 1999, 15(4): 733 - 738.

(上接第2743页)

对比表1和表4可以得出,音频特征向量经过PCA变换后,聚类正确率提高了近8%,说明通过PCA变换有效地去除了类别之间特征数据的相关性,降低了所提取特征中的冗余信息。

表4 未进行PCA变换的四交叉验证聚类正确率 %

实验	第一组	第二组	第三组	第四组	平均
实验一	72	62	72	84	72.5
实验二	74	62	72	84	73.0
实验三	74	72	72	84	75.5
实验四	76	68	72	84	75.0
实验五	68	62	72	84	71.5
平均	72.8	65.2	72.0	84.0	73.5

4 结语

本文提出的基于广播音频聚类算法,利用小波变换和PCA变换对音频片段进行特征提取和处理,采用Mean-Shift算法对音频信号进行初步聚类,然后利用快速近邻法对Mean-Shift的结果进行一次修正,最后合并仅含有单个样本类别的类进行二次修正。该音频算法不需要任何先验条件,而且执行速度快。仿真实验表明,该算法较之单一Mean-Shift算法和未进行PCA处理的特征集的聚类,效果有一定的提高,但对自然的音频流中存在着少数的时间间隔较短的讨论、采访、背景音较大等现象处理得不是很好,因此下一步可以从提高算法的综合性能方面进行改进,以便使算法的应用更加稳定灵活。

参考文献:

[1] (加)韩家炜, (加)坎伯. 数据挖掘概念与技术[M]. 范明, 孟小峰, 译. 北京: 机械工业出版社, 2001: 223 - 262.

[2] 毛韶阳, 李肯立. 优化K-means初始聚类中心研究[J]. 计算机工程与应用, 2007, 43(22): 179 - 181, 219.

[3] 王元珍, 王健, 李晨阳. 一种改进的模糊聚类算法[J]. 华中科技大学学报: 自然科学版, 2005, 33(2): 92 - 94.

[4] FUKUNAGA K, HOSTETLER L. The estimation of the gradient of a density function, with applications in pattern recognition[J]. IEEE Transactions on Information Theory, 1975, 21(1): 32 - 40.

[5] COMANICIU D, MEER P. Mean shift: A robust approach toward feature space analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(5): 603 - 619.

[6] CHENG Y Z. Mean shift, mode seeking, and clustering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1995, 17(8): 790 - 799.

[7] 郑继明, 俞佳. 基于小波变换和支持向量机的音频分类[J]. 计算机工程与应用, 2009, 45(11): 158 - 161.

[8] PARK C H, PARK H, PARDALOS P. A comparative study of linear and nonlinear feature extraction methods[C]// ICDM: Proceedings of the 4th IEEE International Conference on Data Mining. Washington, DC: IEEE Computer Society, 2004: 495 - 498.

[9] KIM H C, KIM D, BANG S Y. A PCA mixture model with an efficient model selection method[C]// Proceedings of International Joint Conference on Neural Networks. Washington, DC: IEEE Press, 2001: 430 - 435.

[10] 赵力. 语音信号处理[M]. 北京: 机械工业出版社, 2003: 84 - 86.