

# 基于连续小波和支持向量机分类音乐类型

颜景斌<sup>1</sup>, 伊戈尔·艾杜阿尔达维奇<sup>2</sup>

- (1. 哈尔滨理工大学 电气与电子工程学院, 黑龙江 哈尔滨 150040;
2. 白俄罗斯国立大学 无线电物理与电子系, 白俄罗斯 明斯克 220050)

**摘要:**音乐类型分类主要包括两个阶段:特征提取和分类。文中在研究小波变换理论基础上,采用连续小波分析方法提取音乐特征参数。支持向量机是专门针对有限样本情况下的一种分类方法。它是建立在统计学习理论的 VC 维理论和结构风险最小原理基础上,根据有限的样本信息在模型的复杂性和学习能力之间寻求最佳折衷,以期获得最好的推广能力。采用指数径向基函数(ERBF)内核,分类正确率可达 85%,比传统的混合高斯模型和 K 近邻分类器,分类性能分别提高了 21% 和 23%。实验结果表明,采用小波和支持向量机方法是一种相当有效的音乐类型分类方法。

**关键词:**音乐类型分类;小波;支持向量机;核函数

中图分类号:TN912.3

文献标识码:A

文章编号:1673-629X(2008)12-0019-03

## Musical Genre Classification Based on Continuous Wavelet and Support Vector Machines

YAN Jing-bin<sup>1</sup>, IGOR. Kheidorov<sup>2</sup>

- (1. Electric & Electronic Engineering College, Harbin University of Science and Technology, Harbin 150040, China;
2. Department of Radiophysics and Electronics, Belarusian State University, Minsk 220050, Belarus)

**Abstract:** Musical genre classification task falls into two major stages: feature extraction and classification. According to a research in wavelet theory, continuous wavelet analysis method is used to extract feature parameters of music. SVM is designed to classifying of limited samples. It is based on VC dimension and the ERM(expectation risk minimum)of statistical learning theory. According to information of limited samples, there is a trade-off existing between models complexities and learning capability to get best extending ability. Exponential radial basis function (ERBF) kernel function are used to classify the musical genre, 85% of classification are correct. In comparison with Gaussian mixture model (GMM) classifier and K nearest neighboring (KNN) classifier, the classification performances are improved by 21% and 23% respectively. Experimental results indicate that wavelet and SVM is useful method for musical genre classification.

**Key words:** musical genre classification; wavelet; support vector machines; kernel function

### 0 引言

音乐类型分类是多媒体应用中的重要组成部分。随着数据存储、压缩技术和互联网技术的飞速发展,音乐类型数据急剧增加。在实际应用中,所有的商业音乐数据库和 mp3 音乐下载站点首要的任务是将这些音乐收藏到不同音乐类型的数据库中。传统的人工信息检索技术已经不能满足对海量信息的检索分类。基于音乐类型分类技术目的就是为了解决上述问题,它可利用音乐本身的声学特征对其进行自动分类,取代手工的方法。同时,确定背景音乐类型也是检索视频

场景的一种有效方法。

从本质上讲,音乐类型分类是一个模式识别的问题,主要包括两个方面内容:特征提取和分类。很多研究者在这个领域做了大量的工作,采用了不同的音频特征和分类方法。文献[1]采用混合高斯模型,对 MPEG-7 格式的 13 种类型音乐分类。文献[2]使用 KNN 和 GMM 分类器和小波特征分类音乐类型,错误率分别为 38% 和 36%。文献[3]使用关于音乐特征的韵律特征和 GMM 实现分类。

传统参数虽然在实际中取得了较好的效果,但这些方法的鲁棒性、适应性和推广能力有限,特别是特征参数大都是采用短时平稳信号的分析方法取得的。小波理论是一种非平稳信号的分析方法,采用多分辨率的思想,非均匀划分时-频空间。小波变换是时频域分析中非常有效的工具,并有广泛应用<sup>[3-5]</sup>。SVM

收稿日期:2008-03-28

作者简介:颜景斌(1972-),男,黑龙江哈尔滨人,博士,副教授,从事语音识别的研究;伊戈尔·艾杜阿尔达维奇,博士,副教授,从事语音信号处理和多媒体信息处理的研究。

是在统计学理论上发展起来的一种新的机器学习方法,它在小样本的条件下依然保持很好的推广能力,在结构风险最小化原则基础上,在特征空间建立最优的分类超平面,克服传统基于规则的分类算法的缺点。近年来 SVM 已广泛被应用于各领域,并在语音识别上也取得了很好的实验效果<sup>[6]</sup>。

因此,文中提出了一种基于小波特征和 SVM 的分类音乐类型的方法。

### 1 连续小波变换

在文献[3]中使用离散小波变换(DTW)系数提取特征的方法分类音乐类型。不同于 DTW,连续小波变换在任意的时间-尺度变换,具有更大的灵活性。音乐的连续小波变换如图 1,2 所示。图中白色部分为基频。

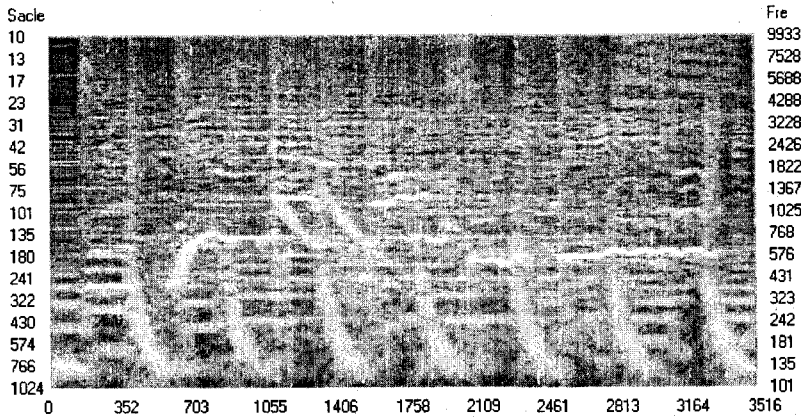


图 1 流行音乐的小波变换

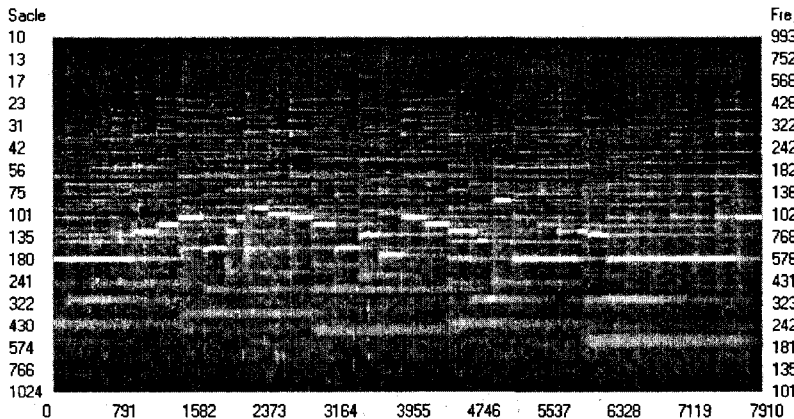


图 2 古典音乐的小波变换

设  $f(t)$  是平方可积函数[记作  $f(t) \in L^2(R)$ ],  $\psi(t)$  是基本小波或母小波函数,则

$$[W_{\psi}f](a, \tau) = \int x(t) \frac{1}{\sqrt{a}} \psi\left(\frac{t-\tau}{a}\right) dt \quad (1)$$

成为  $f(t)$  的小波变换。式中:

- $a$  - 尺度因子,其值大于零;
- $\tau$  - 时间位移,其值可正负。

计算机要求在有限的离散数据下工作,因此离散的小波变换公式如下:

$$[WD_{\psi}f](a, \tau) = \sum_{n=-\infty}^{+\infty} x[n] \frac{1}{\sqrt{a}} \psi\left(\frac{n-\tau}{a}\right) \quad (2)$$

#### 1.1 数据预处理

直接实现小波变换要求很大的计算量,这可能导致算法消耗大量内存。因此,对数据进行预处理,在不改变计算过程和精度的前提下实现数据计算量实质性的压缩。

可以预先设定不同尺度小波,需完成以下几步:

- (1)选择小波最大和最小尺度。
- (2)计算所有的小波尺度。
- (3)计算相应尺度下基小波离散点值。

小波尺度与频率成反比。选择对数规律改变尺度。尺度计算如公式(3)所示。

$$scale_{k \in [1, num]} = MaxWvLng \cdot \exp\left[-\frac{k}{num} \cdot \log\left(\frac{MaxWvLng}{MinWvLng}\right)\right] \quad (3)$$

$scale_k$  - 第  $k$  个尺度;

MaxWvLng - 小波尺度最大值;

MinWvLng - 小波尺度最小值;

num - 尺度总数目(256)。

选择 Morlet 作为基小波。Morlet 小波满足连续小波的可允许条件,是一个具有解析表达式的小波。其解析表达式如公式(4)。对于不同尺度下的 Morlet 小波函数如图 3 所示。

$$\psi(x) = Ce^{-x^2/2} \cos(5x) \quad (4)$$

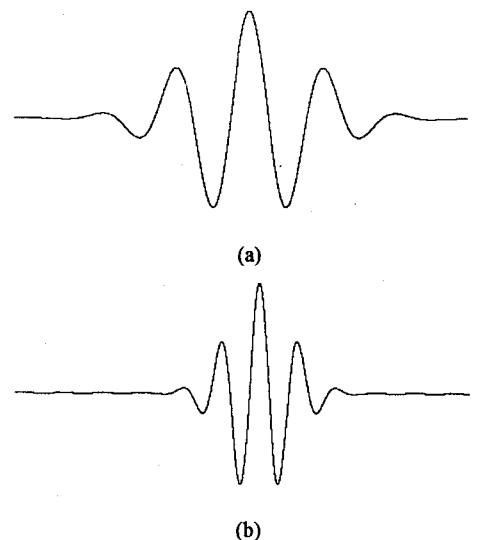


图 3 Morlet 小波不同尺度下的形状

#### 1.2 计算连续小波变换

计算连续小波变换如公式(5)所示。

$$CWT(pos, k) =$$

$$\sqrt{\text{rscale}[k]} \cdot \sum_{i=0}^{\text{MaxWvLng}} \left( \text{wave}[\text{pos} - \frac{\text{MaxWvLng}}{2} + i] \cdot \text{wvlt}[k][i] \right) \quad (5)$$

pos - 时间轴采样点的位置;

k - 小波编号;

rscale[k] - 系数,与小波尺度成反比;

wave[] - 选择分析信号值的数据集;

wvlt[k][i] - 第 k 个小波的第 i 点采样值。

由公式(5)计算得到二维时频域的小波系数。连续小波变换后某时刻的小波系数如图 4 所示。

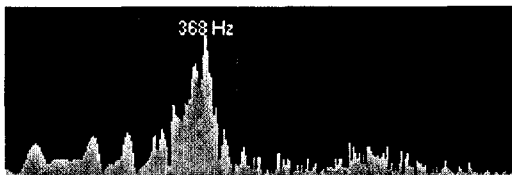


图 4 流行音乐某时刻小波系数

### 1.3 提取特征向量

将所有的时间频率区域沿着时间轴分割成子段,沿着频率轴分割成子带。然后将每个分割块单元中小波系数取平均值,并将这些值排成序列作为特征向量,如图 5 所示。

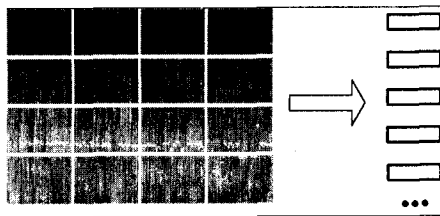


图 5 小波提取特征向量示意图

## 2 支持向量机分类

SVM 是 Vapnik 在统计学理论上提出的小样本情况下机器学习的方法,它证明优于普通的分类器(高斯混合模型, K 近邻, 隐马尔可夫)。

SVM 的基本理论如下。

假设训练数据集为

$$\{x_i, y_i\}, i = 1, \dots, n$$

$$x_i \in R^d, y_i \in \{-1, 1\} \quad (6)$$

$x_i$  表示由  $d$  维特征组成的第  $i$  个样本,  $y_i$  表示第  $i$  个样本所属的类别。支持向量机训练分类器的目的就是要通过这些训练数据得到一个期望错误最小的分类决策函数  $f(x) = \text{sgn}(\omega \cdot x - b)$ , 即在特征空间中寻找最大间隔(超平面与每类数据点间的距离)的超平面  $H$ :

$$(\omega \cdot x) - b = 0 \quad (7)$$

对于不可分的情况, 引入松弛变量  $\xi$  和错误惩罚力度因子  $C$ 。训练错误最小和建立最大间隔的超平面

的优化问题, 可由拉格朗日函数鞍点对偶式给出。

$$L_d = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (8)$$

约束条件为  $\sum_{i=1}^n \alpha_i y_i = 0; 0 \leq \alpha_i \leq C; i = 1, \dots, n$ 。

二次规划问题的解为  $\alpha^* = (\alpha_1^0, \alpha_2^0, \dots, \alpha_n^0)$ 。

到此为止, 目标函数就变为在约束条件下最大化式(10),  $\alpha_i^0 > 0$  所对应的训练样本, 称之为支持向量。

最后得到的分类决策函数为

$$f(x) = \text{sgn} \left( \sum_{\text{支持向量}} \alpha_i^0 y_i (x_i \cdot x) - b \right) \quad (9)$$

对于线性不可分的情况, SVM 通过选择合适的非线性变换将输入向量映射到高维特征空间, 在高维空间建立线性超平面。设存在映射  $\phi: X \rightarrow Z, X \in R^d, Z \in R^k, k \geq d$ , 它可以把输入样本  $x_i \in X$  映射到  $k$  维特征空间  $Z$ 。若函数  $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$ , 则称  $K$  为内核函数。代入式(8)得

$$L_d = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i \cdot x_j) \quad (10)$$

在高维空间  $Z$  中, 最终的分类决策函数是

$$f(x) = \text{sgn} \left( \sum_{\text{支持向量}} \alpha_i^0 y_i k(x, x_i) - b \right) \quad (11)$$

常用的内核函数有多项式内核函数, 高斯径向基函数 RBF, 指数径向基函数 ERBF, 线性内核函数。实验表明, 采用高斯径向基函数效果最好。其函数形式如公式(12)所示。

$$K(x, y) = \exp \left( - \frac{\|x - y\|^2}{2\delta^2} \right) \quad (12)$$

SVM 训练算法采用序列最小优化 (Sequential Minimal Optimization, SMO) 算法<sup>[7]</sup>。该算法的优点是针对 2 个样本的二次规划问题可以有解析解的形式, 从而避免了多样本情况下的数据解不稳定及耗时问题, 同时也不需要大的矩阵存储空间。其工作集的选择采用启发式, 通过两个嵌套的循环来寻找待优化的样本变量, 循环优化, 直到全部样本都满足最优条件。

## 3 实验结果

研究选用 4 种音乐类型: 古典, 爵士, 流行和摇摆。实验数据库包含 400 首音乐录音, 16 位 PCM 格式, 每首截取 30 秒长度片段, 采样速率 22.050kHz。在 400 个录音中出自同不同作品。每种音乐类型 100 个录音。数据库录音来自 CD, mp3, 数据库和无线电。

每类 50 个样本作为训练集, 50 个样本作为测试集。由于小波系数值较大, 提取特征时需要作归一化处理, 提高识别精度。SVM 核函数选择径向基函数。循环迭代参数  $C$  和  $\xi$ , 从实验中得到精度最高的  $C$  和

(下转第 24 页)

### 3 实验结果与分析

为了对所提出的 LIC 并行算法的有效性进行验证,在 20 个节点的曙光 TC4000L 集群及 Linux 操作系统的实验环境和 MPI 并行编程环境下对渤海流场矢量数据进行了实验,积分长度为 100,输出图像大小为 562×376(如图 1 所示),并行实验结果如表 1 所示。

表 1 并行实验结果

节点数	计算时间(秒)	并行加速比	并行效率(%)
1	267.5	1.00	100
4	88.9	3.01	75
8	46.2	5.79	72
16	24.6	10.88	68
20	22.1	12.12	61

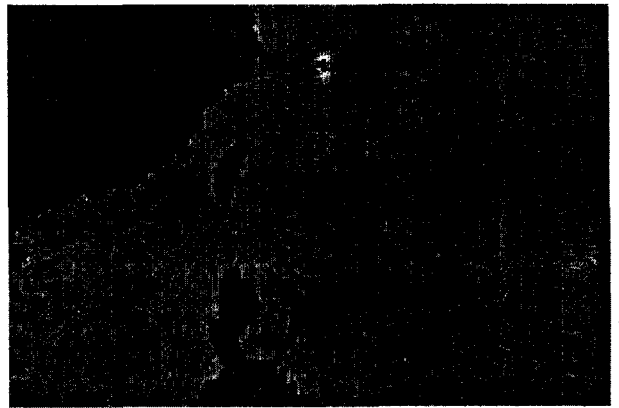


图 1 渤海流场的 LIC 纹理图

### 4 结束语

在矢量场可视化领域,LIC 算法具有很大的优越性,但其计算量过大,导致耗时过多,制约了它的实际应用,并行计算是解决该问题的有效途径。提出了 LIC 算法的并行实现算法,并运用到海洋流场可视化当中,实验结果表明,该算法设计合理,具有较高的并行度和加速比,使 LIC 算法的交互式探索成为可能,具有一定的实际意义。

#### 参考文献:

- [1] Van Wijk J J. Spot noise: texture synthesis for data visualization[J]. Computer Graphics, 1991, 25(4): 309-318.
- [2] Cabral B, Leedom L. Imaging vector fields using line integral convolution[J]. Computer Graphics, 1993, 27(4): 263-270.
- [3] 余永胜,顾耀林. 一个基于 LIC 的矢量场可视化算法的改进[J]. 工程图学报, 2006(5): 34-37.
- [4] 陈国良. 并行计算-结构算法编程(修订版)[M]. 北京: 高等教育出版社, 2003.
- [5] 都志辉. 高性能计算并行编程技术-MPI 并行程序设计[M]. 北京: 清华大学出版社, 2001.

(上接第 21 页)

ξ, 当存在多个相同的最高测试率时,取支持向量少的那组参数作为最优参数。窗口长度 4096 个采样点。对每种音乐类型提取特征向量,使用不同的子带和子段的数目进行测试。实验中使用 6 种组合类型形式用 SVM 分类。实验结果见表 1。整个实验在 Visual C++ 6.0 平台下开发完成。

表 1 不同音乐类型间组合的分类错误率

子段数	子带数	错误率%					
		古典/流行	古典/摇摆	古典/爵士	爵士/流行	爵士/摇摆	摇摆/流行
1	16	16.3	6.7	22.3	26.7	14.1	19
1	32	16.9	6.7	20	26.3	11.9	14.1
2	64	17.5	7.0	21.0	26.4	10.3	16.2
4	64	19.6	7.2	19.6	26.6	9.4	17.2

### 4 结束语

音乐类型分类在多媒体应用中有重要的应用价值。文中采用 Morlet 小波变换作为音乐特征,SVM 分类音乐类型,分类正确率可达 85%,尤其对古典/摇摆这对类型分类表现最好。这表明该方法合理,分类性能较好。

#### 参考文献:

- [1] Burred J, Lerch A. A Hierarchical approach to automatic musical genre classification[C]//Int. Conf. on Digital Audio Effects (DAFx-03). London, UK: [s. n.], 2003: 308-311.
- [2] Li T, Oginara M, Li Q. A comparative study on content-based music genre classification[C]//in Proc. of the 26th annual int. ACM SIGIR conf. on Research and development in information retrieval. [s. l.]: ACM Press, 2003: 282-289.
- [3] Tzanetakis G, Cook P. Musical genre classification of audio signals[J]. IEEE Trans. on Speech and Audio Processing, 2002, 10(5): 293-302.
- [4] Mallat S G. A Theory for Multiresolution Signal Decomposition: the Wavelet Representation[J]. IEEE Trans Pattern Analysis and Machine Intelligence, 1989, 11(7): 674-693.
- [5] Downie T R, Silverman B W. The discrete multiple wavelet transform and thresholding methods[J]. IEEE Trans on Signal Processing, 1998, 46(9): 2558-2561.
- [6] 王欢良,韩纪庆,张磊. 基于支持向量机的变异语音分类[J]. 哈尔滨工业大学学报, 2003, 35(4): 389-393.
- [7] Platt J C. Fast training of SVMs using sequential minimal optimization[C]//Scholkopf B, Burges C J C, Smola A J. Advances in Kernel Methods - Support Vector Learning. Cambridge, MA: MIT Press, 1998.