

文章编号:1005-0523(2009)06-0060-05

# 基于最小二乘支持向量机的乐器音乐分类

刘建辉,曾丽辉,许金凤,刘遵雄

(华东交通大学 信息工程学院,江西 南昌 330013)

**摘要:**提出使用最小二乘支持向量机 LS-SVM(Least Squares Support Vector Machines)算法进行乐器音乐分类,从而实现乐器的辨识。在对 LS-SVM 理论进行深入探讨的基础上,选择乐器音乐 clip 作为样本,进行特征提取,提取的特征包括频谱特征,短时自相关系数和 MFCC 等,然后用最小二乘支持向量机算法进行分类。对古琴、古筝、箜篌和琵琶音乐采取样本进行仿真实验,求得分类准确率和运行时间,同时使用逻辑回归(Logistic Regression)算法进行对比试验,其中最小二乘支持向量机和逻辑回归分类的准确率分别为 96.5%和 92.5%,且 LS-SVM 的运行时间比 Logist 的少。实验结果表明最小二乘支持向量机具有更为优越的分类性能和非线性处理能力,可以推广用于解决其它实际分类问题。

**关键词:**最小二乘支持向量机;乐器音乐;音乐特征

**中图分类号:**TP391.42

**文献标识码:**A

科学研究表明音乐特有的旋律与节奏能缓解人们的日常生活和工作的压力。但是在选择音乐方面,非专业人士很难根据乐曲来甄选出某个具体乐器的音乐,所以基于音乐内容的分类在人们的日常生活中具有重要的意义。在音乐分类的方面,前人已经进行了深入的研究,并取得了丰富成果。音乐分类系统主要由两部分组成:第一部分是特征提取<sup>[1]</sup>,实现样本的相关特征提取。第二部分是分类<sup>[1]</sup>,采用的分类算法包括高斯混合模型<sup>[2]</sup>、神经网络<sup>[3]</sup>、非负矩阵分解(Non-negative Matrix Factorization,即 NMF)<sup>[4]</sup>、支持向量机(Support Vector Machine,即 SVM)<sup>[5]</sup>、隐马尔科夫模型<sup>[6]</sup>和逻辑回归等,不同模型处理不同的分类问题表现出不同的分类性能,其中支持向量机和逻辑回归使用最为广泛。支持向量机是一种重要的机器学习算法,和传统的基于经验风险最小化准则的学习算法不同,支持向量机基于结构风险最小化准则工作,能在训练误差和分类器容量之间达到一个较好的平衡,因而具有更好的性能。1999年,Suykens J A K<sup>[7]</sup>提出了最小二乘支持向量机,主要是将 SVM 中的求解二次规划问题转化成求解一组线性关系式,从而演化成简单的矩阵逆运算。

本文将通过对古琴古筝演奏出来的乐曲进行特征提取频谱,运用 LS-SVM 算法进行分类,从而辨识演奏乐曲相应的乐器。同时使用逻辑回归算法进行仿真试验,求取其分类准确率,比较这两种算法的分类性能。

## 1 音乐信号的特征分析与提取

识别系统性能的好坏与音乐信号的特征参数选择有直接的关系,所以我们选取的特征应能准确合理地表现音乐的实质内容和内在特征,尽可能摒弃无用信息。在分析音乐信号性质时,我们将其切分成若干段来分析,在每一段内,信号具有短时平稳性,将这些段称之为分析帧。实际应用中的特征抽取包括 3 个步骤:原始音乐预处理,特征抽取和特征集的构造<sup>[8]</sup>。

在对音乐信号进行时域或频域分析前,要对音乐信号首先进行一定的预处理,主要方法有信号的预加重、分帧和加窗。本文采用的加窗算法是海明窗(Hamming Window)。本文中的音乐数据采样率取 22 050 Hz,精度为 16,帧长取 25 ms,clip 长度取 75 ms。在对音乐信号进行预处理之后,我们将对处理的音乐信号进行特征提取。

**收稿日期:**2009-08-04

**基金项目:**国家自然科学基金项目(60963012);江西省教育厅科学研究项目(GJJ09507)

**作者简介:**刘建辉(1963-),江西宜春人,讲师,主要研究领域为计算机应用及多媒体技术。

使用表 1 中的公式计算每个帧的以下特征值,包括频率中心  $F$ 、带宽  $B$ 、频谱扩展  $\sigma$ 、频谱谱峰  $\gamma_2$ 、频谱中心  $S$ 、频谱偏斜  $\gamma_1$ 、频谱斜率  $m$ 、频谱滚降  $f_c$ 、自相关函数  $R_n(k)$  和 Mell 倒谱系数  $M$ 。

表 1 特征公式表

名称	公式
$\sigma$	$\sigma^2 = \int (f - s_c)^2 p(f) df$
$\gamma_2$	$m_4 = \int (f - s_c)^4 \cdot p(f) df, \gamma_2 = \frac{m_4}{\sigma^4}$
$s_c$	$s_c = \int f \cdot p(f) df$
$\gamma_1$	$m_3 = \int (f - s_c)^3 \cdot p(f) df, \gamma_1 = \frac{m_3}{\sigma^3}$
$R_n(k)$	$R_n(k) = \sum_{i=0}^{N-1-k} [ [ S_w(n+i) ] S_w(n+i+k) ]$
$F$	$F = \int_0^\pi \omega  F(\omega) ^2 d\omega / \int_0^\pi  F(\omega) ^2 d\omega$
$B$	$B = \sqrt{\int_0^\pi (\omega - F)^2  F(\omega) ^2 d\omega / \int_0^\pi  F(\omega) ^2 d\omega}$
$m$	$m = \frac{1}{\sum_f A(f)} \frac{N \sum_f f \cdot A(f) - \sum_f \sum_f A(f)}{N \sum_f f^2 - (\sum_f f)^2}$
$f_c$	$\sum_0^{f_c} A^2(f) = 0.95 \sum_0^{f_{ny}} A^2(f)$
$M$	$C_n = \sum_{k=1}^M \ln x'(k) \cos [\pi(k-0.5)n/M], n=1,2,\dots,l$

表 1 中的  $A(f) = |F[x(t)]|, p(f) = A(f) / \sum_f A(f)$ 。本文取自相关系数 12 个,取 MFCC 的前 13 维。计算出相对应特征值,求取上述特征值的均值作为 clip 片的特征值。在特征抽取的基础上构造出 35 维的音乐分类特征集合。由于不同音乐特征的值相差特别大,这里对特征集合进行相应的归一化处理

$$x_i' = (x_i - \mu_i) / \sigma_i \tag{1}$$

式中,  $\mu_i$  为均值,  $\sigma_i$  为方差。

由于 MFCC 归一化处理后维间数值差别过小,试验结果不具有理想性,所以对 MFCC 不做归一化处理。计算一个 clip 中的各帧的 13 维 MFCC 系数,然后在 clip 内对各维取平均值,作为该片的 MFCC 特征值。

## 2 最小二乘支持向量机

最小二乘支持向量机是支持向量机的改进,与标准 SVM 模型比较,该方法优势在于:

- (1) 用等式约束代替标准 SVM 算法中的不等式约束;
- (2) 将求解二次规划问题转化为直接求解线性方程组。

LS-SVM 分类器使用二次惩罚函数,而不像 SVM 通常采用  $\epsilon$  不敏感函数。设训练集为  $S = \{(x_i, y_i), x_i \in R^n, y_i \in \{-1, +1\}\}_{i=1}^l$ , 其中  $x_i$  是第  $i$  个输入模式,  $y_i$  是相应于  $x_i$  的目标,  $l$  为样本数目。二分类问题的目标是确定决策函数  $y(x) = \text{sign}(f(x))$ , 函数  $f(x)$  具有如下的形式

$$f(x) = w^T \varphi(x) + \delta \tag{2}$$

式中,  $\varphi(\cdot)$  为输入空间到特征空间的映射, 系数向量  $w$  和偏差项  $\delta$  为待求量。在权  $w$  空间(原始空间)中构造 LS-SVM 分类器可以描述为求解等式约束  $y_i [\omega^T \varphi(x_i) + \delta] = 1 - e_i, (i = 1, 2, \dots, l)$  下的最小化问题, 即

$$\min_{\omega, b, e} J_p(\omega, e) = \frac{1}{2} \omega^T \omega + \gamma \cdot \frac{1}{2} \sum_{i=1}^l e_i^2 \tag{3}$$

式中, 前项表示最小化模型 VC 维, 后项表示最小化训练误差,  $\gamma$  是惩罚系数, 表示模型函数光滑和最小经验误差的折中,  $e$  表示单位向量。  $\varphi(\cdot): R^n \rightarrow R^h$  为核空间映射函数, 权矢量  $\omega \in R^h$ 、误差变量  $e_i \in R, \delta$  是

偏差量。在对偶空间求解此优化问题。定义相应的拉格朗日函数式

$$L(\omega, \delta, e, \beta) = J_p(\omega, e) - \sum_{i=1}^l \beta_i [y_i \omega^T \varphi(x_i) + y_i \delta + e_i - 1] \quad (4)$$

式中,拉格朗日乘子  $a_i \in R$  求解优化问题如下,即

$$\begin{cases} \frac{\partial L}{\partial \omega} = 0 \Rightarrow \omega = \sum_{i=1}^l \beta_i y_i \varphi(x_i) \\ \frac{\partial L}{\partial \delta} = 0 \Rightarrow \sum_{i=1}^l y_i \beta_i = 0 \\ \frac{\partial L}{\partial e_i} = 0 \Rightarrow \beta_i = \gamma e_i \\ \frac{\partial L}{\partial \beta_i} = 0 \Rightarrow y_i (\omega^T \varphi(x_i) + \delta) - 1 + e_i = 0 \end{cases} \quad i = 1, 2, \dots, l \quad (5)$$

上述式子可以转化为下面的矩阵方程

$$\begin{bmatrix} 0 & 1 \\ 1 & K + \gamma^{-1} \mathbf{1} \end{bmatrix} \begin{bmatrix} \delta \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ d \end{bmatrix} \quad (6)$$

式中,  $d = [d_1, d_2, \dots, d_l]^T$ ;  $\beta = [\beta_1, \beta_2, \dots, \beta_l]^T$ ;  $\mathbf{1} = [1, 1, \dots, 1]^T$ ;  $K(x_i, x_j) = \varphi(x_i) \varphi(x_j)^T$ 。这样,从而分类函数式可从下式得到

$$y = \sum_{i=1}^l \beta_i K(x, x_i) + \delta \quad (7)$$

LS-SVM 仅需要确定核函数的形状参数和惩罚系数  $\gamma$ ,而不需要选取不敏感损失函数的  $\epsilon$  值,这不仅简化了计算,而且便于使用。本文使用 LS-SVM 的核函数是 RBF 核。

### 3 实验结果与分析

本文利用网上资源,下载了用古琴、古筝、箜篌和琵琶独奏的乐曲,利用转化软件,将 mp3 格式的乐曲转化成 wav 格式。乐曲的采样频率为 22 050 Hz,精度 16 位,存储格式为 wav 格式。采集的每种乐曲样本数为 4 种,分别为 500,800,1 100 和 1 400 个,每个样本的特征数为 35 维,在 MATLAB 平台下进行仿真实验。

首先比较古琴和古筝独奏曲在频谱特征值上的不同。古琴和古筝乐曲的频谱特征值,实验数据见表 2。从表 2 中可知,古筝乐曲的频谱特征值普遍要比古琴乐曲的频谱特征值要大,说明古筝发出来的声音音域宽广、音量宏大、音色含蓄而深沉。

表 2 古琴乐曲,古筝乐曲频谱特征值对比

表 2 古琴乐曲,古筝乐曲频谱特征值对比					Hz
频谱特征	乐器	最小值	最大值	标准差	均值
频谱中心	古琴	38.125 480	151.784 356	23.984 641	88.309 812
	古筝	25.787 530	213.158 617	40.777 140	104.434 910
频谱扩展	古琴	76.498 823	223.535 516	32.361 050	165.638 700
	古筝	83.235 461	228.937 712	35.935 390	171.333 900
频谱偏斜	古琴	2.272 616	7.841 646	0.964 699	3.563 798
	古筝	1.264732	8.492119	1.357 509	3.360 953
频谱谱峰	古琴	7.458 539	75.064 587	11.013 893	17.579 891
	古筝	3.725 267	80.384 371	13.654 432	16.719 941
频谱斜率	古琴	$-2.5 \times 10^{-7}$	$-1.9 \times 10^{-7}$	$1.24 \times 10^{-8}$	$-2.2 \times 10^{-7}$
	古筝	$-1.7 \times 10^{-7}$	$-1 \times 10^{-7}$	$1.41 \times 10^{-8}$	$-1.4 \times 10^{-7}$
频谱滚降	古琴	516.796 932	3 380.713 015	773.764 448	1 094.979 017
	古筝	516.796 932	7 978.052 017	1 548.243 056	1 977.037 015

从样本的频谱特征中可以看出,古琴乐曲和古筝乐曲在频谱滚降处有非常大的区别,所以可以说明从

频谱滚降上进行分类具有其有效性。

将频率中心均值、频率中心标准差、带宽均值、带宽标准差为一组,用  $F$  表示;将 6 个频谱特征为一组,用  $S$  表示;用自相关系数用  $A$  表示;倒谱系数 MFCC 用  $M$  表示;用 LS-SVM 和逻辑回归(Logistic)算法进行分类试验,对采样数据进行 10 次仿真实验,计算分类准确率均值和标准差,实验结果见表 3。

表 3 LS-SVM 和 Logist 的分类正确率的对比结果

算法	正确率	$F$	$F&A$	$F&A&M$	$S$	$S&A$	$S&A&M$	$F&S&A&M$
LS-SVM	正确率均值	0.512 1	0.650 2	0.910 0	0.723 0	0.881 0	0.924 0	0.965 0
	正确率标准值	0.010 2	0.011 1	0.006 0	0.012 0	0.043 3	0.009 0	0.013 0
Logist	正确率均值	0.512 1	0.611 2	0.889 0	0.596 0	0.717 0	0.913 0	0.925 0
	正确率标准值	0.010 9	0.012 3	0.009 1	0.017 0	0.057 2	0.014 0	0.014 3

从表 3 中可知,当维数达到 35 维时,两种算法的分类效果相比于少于 35 维时要好,且 LS-SVM 的分类性能优于 Logist 方法。计算两种分类算法的运行时间,实验结果见表 4。

表 4 运行时间的对比

乐器	样本数 时间	运行平均时间/ms	
		Logist	LS-SVM 改进
古琴	500	100.52	75.27
	800	120.12	77.54
	1 100	158.32	80.72
	1 400	198.48	78.59
古筝	500	100.42	73.72
	800	120.24	79.78
	1 100	158.42	85.72
	1 400	199.10	80.27
箏篌	500	100.50	83.04
	800	120.30	90.50
	1 100	158.30	95.03
	1 400	198.40	93.27
琵琶	500	100.35	93.03
	800	120.27	94.54
	1 100	158.43	96.03
	1 400	198.60	95.03

从表 4 中可以看出,随着样本数的增加,LS-SVM 的运行时间依然比 Logist 方法要少。

从表 3 和表 4 中可以得出,LS-SVM 分类的准确率均值要比 Logist 高,而分类准确率标准差要小,Logist 数据分类准确率波动较大。随着特征数的增加,分类正确率也随着上升。当用 35 个特征数的时,LS-SVM 分类的准确率均值达到 96.4%,比 Logist 分类的准确率均值要高出 4 个百分点。从分类的正确率上和运行时间上可以看出,在解决音乐分类问题上 LS-SVM 的分类性能优于 Logist 方法。

#### 4 结论与展望

针对收集的古琴古筝弹奏乐曲进行特征提取以及分类试验,实现对古琴古筝乐器的辨识。实验结果表明,LS-SVM 分类性能比 Logist 分类方法要好,说明 LS-SVM 分类具有良好的非线性数据处理能力,可以推广用于其它实际分类问题的处理。音乐信号的特征还有待进行深入研究,结合多种算法进行乐器辨识。

#### 参考文献:

- [1] XU Chang-heng, Maddage, N C, XI Shao. Automatic music classification and summarization[J]. IEEE Transactions on Speech and Au-

- dio Processing, 2005, 13(3): 441 – 450.
- [2] Marques J, Moreno P J. A study of musical instrument classification using gaussian mixture models and support vector machines[DB/OL]. <http://www.hpl.hp.com/techreports/Compaq-DEC/CRL-99-4.pdf>. 1999-06-10.
- [3] Charlie K. D. Adaptive music classification using neural network architectures[J]. The Artificial Neural Networks in Engineering Conference, 2000, 34(16): 917 – 921.
- [4] Benetos E, Kotti M, Kotropoulos C. Musical instrument classification using non-negative matrix factorization algorithms[J]. IEEE International Symposium, 2006, 15(3): 1 – 4.
- [5] Usama Fayyad. A tutorial on support vector machines for pattern recognition[J]. Data Mining and Knowledge Discovery, 1998, 12(7): 121 – 167.
- [6] Lee J, Chun J. Musical instruments recognition using hidden markov model[C]. Computing Literature, South Korea: Conference Record of the Thirty-Sixth Asilomar Conference, 2002. 196 – 199.
- [7] Suykens J A K, Vandewalle J. Least squares support vector machine classifiers[J]. Neural Processing Letters, 1999, 9(3): 293 – 300.
- [8] 白 亮. 音频分类与分割技术研究[D]. 长沙: 国防科学技术大学, 2004.

## Classification of Music Instruments Based on LS – SVM

LIU Jian-hui, ZENG Li-hui, LIU Zun-xiong

(School of Information Engineering, East China Jiaotong University, Nanchang 330013, China)

**Abstract:** The algorithm of least squares support vector machines (LS – SVM) is proposed for classification of music instruments to recognize them. The corresponding music clips are chosen as experimental samples, and feature selection including spectrum features, short-time auto-correlation coefficients, MFCC, etc., is implemented, and then LS – SVM is used in the categorization. The simulating experiments of four traditional musical instruments are conducted to obtain classifying accuracy and running time. Meanwhile, logistic regression is also used in experiments as comparison. Accuracy rate of LS – SVM and logistic regression are respectively 96.5% and 92.5%. Running time of LS – SVM is less than that of Logist. The results demonstrate that LS – SVM possesses better performance, and can be generalized to cope with other practical classification problems.

**Key words:** LS – SVM; musical instruments; music feature

(责任编辑: 王建华)

(上接第 47 页)

## A Design of Intelligent Trajectory Car Based on DSP

ZHANG Yong-gao<sup>1</sup>, LI Yin-ling<sup>1</sup>, NIE Shui-guo<sup>2</sup>

(1. School of Electrical and Electronic Engineering, East China Jiaotong University, Nanchang 330013, China; 2. Patent Examination Cooperation Center of SIPO, Beijing 100190, China)

**Abstract:** The intelligent car adopting DSP TMS320LF2407A as its kernel is designed. The paper introduces the hardware and software design of the intelligent car in detail. The intelligent car uses infrared sensor to identify the path. The movement of the car is driven by the motor. The speed and direction of the car is controlled by PWM of the DSP.

**Key words:** intelligent car; DSP TMS320LF2407A; PWM control

(责任编辑: 刘棉玲)