

文章编号: 1005-0523(2010)04-0029-05

基于核 Logistic 回归的乐器音乐辨识

刘遵雄, 许金凤, 曾丽辉

(华东交通大学 信息工程学院, 江西 南昌 330013)

摘要: 基于统计学习的音频分类具有理论基础坚实, 实现机制简单等特点受到广泛关注并被很多音频分类系统所采用。本文对核 Logistic 回归算法(KLR)进行了深入分析, 提出基于 KLR 的音频分类器设计方法, 应用其解决同类型的乐器音乐分类问题。结合所采集的小提琴中提琴和大提琴的音乐信号样本进行了分类仿真试验, 并传统的 Logistic 回归(LR)和支持向量机(SVM)进行对比。结果表明, 核 Logistic 回归模型具有较为优越的分类性能和非线性处理能力。

关 键 词: 核 Logistic 回归; 音频分类; 特征提取

中图分类号: TP391.42

文献标识码: A

经典的 Logistic 回归(Logistic Regression, LR)是一种统计分析方法, 根据一个或多个连续性或属性型的自变量来分析和预测 0/1 二值型因变量的多元量化分析方法, 属于概率型非线性回归。与线性回归不同, LR 是一种非线性模型, 普遍采用的参数估计方法是最大似然估计法。可以证明, 在随机样本条件下, LR 模型的最大似然估计具有一致性、渐进性和渐进正态性^[1]。LR 模型所用假设简单, 不要求满足误差分布趋于正态分布的假设, 也不要求自变量符合正态分布的条件, 模型对识别变量的分布未作任何要求; 能用于因变量二值的判别并计算出其归属的概率, 而且可以给出判别结果在概率意义上的解释。目前, 该方法已广泛应用于经济、社会科学以及医学等诸多科学领域。

LR 通过二元取值, 直接利用样本所属类别的后验概率来进行分类^[2]。由于采用的是线性模型, 所以对概率估计的精度有限, 因此一些学者利用 SVM 中的核技巧手法将经典的 Logistic 回归(Logistic Regression, LR)推广到 RKHS(Reproducing Kernel Hilbert Space)空间从而得到非线性的核 Logistic 回归(Kernel Logistic Regression, KLR)^[3-4], 以提高概率估计的精度。

音频自动分类是解决音频结构化问题和提取音频内容语义的重要手段之一, 是当前基于内容的音频检索领域的研究热点。目前, 该领域的研究重点主要在以下两个方面: 音频特征分析与抽取以及分类器的设计。在音频分类的问题上, 国内外的研究人员进行了大量的研究。为了提高分类精度, 研究者们提出了不同的分类方法, 包括最近邻准则(NN), 支持向量机(SVM), 高斯混和模型(GMM), 神经网络, 隐马尔科夫模型(HMM)^[5-10]等。上述方法处理的分类问题比较单一, 通常只是对语音、音乐及环境音等进行分类, 在简单的分类中分类精度比较满意, 但是在相似度较高的音频信号例如同类型的乐器音乐等音频分类精度不高。

本文通过在 KLR 模型中采用不同的核函数、线性核函数、多项式核函数和径向基核函数, 建立多类 KLR 模型^[11]。通过对弦乐器中小提琴、中提琴、大提琴的音乐信号进行特征提取, 运用多类 KLR 模型进行分类试验, 从而实现对同类别乐器信号的辨识。使用传统的 LR 多分类模型和 SVM 进行对比试验, 比较算法之间的分类性能差别。

收稿日期: 2009-12-29

基金项目: 教育部人文社会科学研究项目(09YJA630036)

作者简介: 刘遵雄(1967—), 男, 博士, 副教授, 研究方向为机器学习与数据挖掘、智能信息处理等。

1 音频信号特征提取

在音频分类中,所选取的特征应该能充分刻画音频在时域和频域上的重要分类特性^[12],对环境的改变具有鲁棒性和一般性。

一般来说,音频特征提取是基于两种不同的时间长度:一种是基于音频帧(audio frame)的特征提取,时间为几十毫秒;另一种是基于音频段(audio clip)的特征提取,持续时间一般是几秒。本文通过将原始的音频信号切分成每 3 s 的音频段,然后将音频段通过加 23.21 s(512 个采样点)的汉宁窗(hanning)处理形成音频帧(audio frame),帧间重叠 25%。计算每一帧的傅里叶变换系数 $F(\omega)$ 和频域能量 E 。然后对每一音频帧进行特征提取,最后再根据所获得的帧层次上的特征来计算音频段层次上的特征,从而获得一个长度为 3 s 音频样本的特征数据集。

本文选择的音频帧(frame)上的特征:

(1) 率中心(Frequency Centroid):在一帧中音频亮度(Brightness)是用频率中心来度量的,是声音信号频

谱能量分布的中心点。定义为: $\omega_c = \int_0^{\pi} \omega |F(\omega)|^2 d\omega / \int_0^{\pi} |F(\omega)|^2 d\omega$ 。

式中: ω 为音频信号频率; $F(\omega)$ 为傅里叶变换系数。

(2) 宽(Bandwidth):带宽用于描述信号在频域中所占的有效宽度,是衡量音频频域范围的指标。定义

为: $B = \sqrt{\int_0^{\pi} (\omega - \omega_c)^2 |F(\omega)|^2 d\omega / \int_0^{\pi} |F(\omega)|^2 d\omega}$ 。

(3) 关键子带能量比(Critical Sub-band Power Ratios):在频域中划分子带区间,选取部分关键子带区间。

关键子带能量比定义为每个关键子带的能量对数值与整个音频帧的能量对数值之比。

基于音频段(audio clip)上的特征:

(1) 子带能量比均值定义为片段(clip)中每个关键子带能量比的均值。

(2) 带宽均值与方差定义为 clip 中各个帧的带宽均值。

(3) 频谱中心均值和方差定义为 clip 中亮度均值和方差。

2 核 Logistic 回归(KLR)问题

给定 N 个独立同分布的观测样本 $\{(x_i, y_i)\}_{i=1}^N$, 对于二元分类问题目标值 $y_i \in \{0, 1\}$, 样本 X 隶属于类别 1 的后验概率 $P(y=1|X)$, 属于类别 0 的后验概率 $P(y=0|X)=1-P(y=1|X)$ 。定义判别函数为

$$f(X) = \log \frac{P(y=1|X)}{P(y=0|X)} = \log \frac{P(X, \beta)}{1 - P(X, \beta)} \quad (1)$$

式中: β 为样本 X 的权重向量。

若已知 $f(X)$, 则可以求得给定样本隶属于类别 1 的后验概率为

$$P(y=1|X) = P(X, \beta) = \frac{1}{1 + \exp(-f(X))} \quad (2)$$

在 Logistic 回归中,我们用一个线性函数 $f(X) = \beta^T X + \beta_0$ 去估计 $f(X)$ 。

利用核技巧(kernel trick)可将这样一个广义线性模型扩展到非线性模型而得到核逻辑回归。令 $K(x, y)$ 为任意满足 Mercer 条件的核函数,则 K 唯一定义了一个从样本空间 R^n 到由 K 所定义的 RHKS 空间的映射: $\Phi: X \rightarrow \Phi(X)$ 。每个权重向量都位于所有 $\Phi(X_i)$ 所组成的空间^[13]: $\beta = \sum_{i=1}^N \alpha_i \Phi(X_i)$

引入核矩阵 K , 使得 $K_{ij} = [\Phi(X_i) \Phi(X_j)] = K(X_i X_j)$, 则可得到:

$$f(X) = \beta^T \Phi(X) = \left[\sum_{i=1}^N \alpha_i \Phi(X_i) \right]^T \Phi(X) = \sum_{i=1}^N \alpha_i K(X_i, X) \quad (3)$$

由此得到后验概率

$$P(y \mid X, \alpha) = \frac{1}{1 + \exp \left[- \sum_{i=1}^N \alpha_i K(X_i, X) \right]}$$

(4)

在核 Logistic 回归中,样本概率估计的负对数似然可以写成:

$$\ell(\alpha) = \sum_{i=1}^N -y_i K_i \alpha + \log(1 + \exp(K_i \alpha))$$

(5)

加入正则化项后(为避免训练数据的过拟合,加入惩罚项以防止估计参数 β 出现较大的波动),得到最小化下面的目标泛函:

$$\ell(\alpha)^* = \ell(\alpha) + \frac{\lambda}{2} \|\alpha\|_2^2$$

(6)

λ 为平衡回归函数光滑性与损失函数的系数(正则化项的参数)。求解回归参数是一个无约束的凸优化问题,利用 Newton-Raphson 规则解等式,得到回归参数 α^{new} :

$$\alpha^{\text{new}} = (K + \lambda W^{-1})^{-1} K W \tilde{z}$$

(7)

$$\tilde{z} = (K \alpha^{\text{old}} + W^{-1}(y - P))$$

(8)

P 是拟合概率向量,它的第 i 个元素是 $P(\alpha^{\text{old}}, X_i)$, W 是 $N \times N$ 权重矩阵,对角线上元素为 $P(\alpha^{\text{old}}, X_i)(1 - P(\alpha^{\text{old}}, X_i))$ 。

3 试验结果与性能分析

3.1 特征数据集构造

本文从标准乐器录音数据库^[14]中获取三类乐器原始音乐信号样本:小提琴、中提琴和大提琴。所有原始音频样本均为单声道,采样频率为 44.1 kHz,精度为 16 位,AIFF 格式。利用音频格式转换软件转化为 WAV 格式后,将原始音频样本在时域上分割成每 3 s 的片段(clip),每个类别获得 100 个音频片段,共 300 个音频信号样本。再对每段加 23.21(512 个采样点)的汉宁窗(hanning)形成帧,帧间重叠 25%。

选取 clip 中每一音频帧中的 22 个关键子带,获得各个子带的子带能量比,然后基于帧层次上计算音频段中 22 个关键子带能量比均值和标准差共 44 维特征;根据每一帧的频率和带宽,然后计算音频段上的频率均值和标准差,带宽均值和标准差共 4 维向量,最终构建 48 维的分类特征数据集。

最终实验特征数据集组成如表 1 所示。

表 1 音频实验数据集组成

音频类别		训练样本尺寸/测试样本尺寸						片段总数
小提琴	20/80	30/70	40/60	50/50	60/40	70/30	80/20	100
大提琴	20/80	30/70	40/60	50/50	60/40	70/30	80/20	100
中提琴	20/80	30/70	40/60	50/50	60/40	70/30	80/20	100

3.2 实验结果与性能分析

本文在 Matlab 平台下,使用核 Logistic 回归模型对上述特征数据集进行分类仿真实验,为评价本文算法提出的有效性使用传统的 Logistic 回归和支持向量机 SVM 完成相同的音频分类任务。其中核 Logistic 回归和 SVM 中核函数分别选择线性核、多项式核和 RBF 核进行实验。

通过大量实验统计,在核 Logistic 回归和 SVM 中多项式核函数参数均取 $d=2$;logistic 回归中 RBF 核函数的参数取 $\sigma=2$,正则化项参数 $\lambda=1E-5$;SVM 中 RBF 核函数的参数 $\sigma=2$,惩罚因子 $\lambda=1E-2$ 。

本文采用分类准确度评价分类器的性能。其定义如下:分类准确率=分类正确的样本数/样本总数。实验结果如表 2 所示。

表 2 不同分类器下的分类准确率 %

训练样本尺寸	LR	KLR			SVM		
		线性核	多项式核	RBF 核	线性核	多项式核	RBF 核
20	89.17	95.00	95.00	91.25	95.42	88.33	89.17
30	88.10	90.95	91.90	88.57	92.86	84.76	88.10
40	91.11	92.78	92.78	91.11	94.34	78.33	90.00
50	93.33	94.67	94.67	94.67	94.00	80.00	94.67
60	94.17	92.50	92.50	95.83	91.67	74.17	94.17
70	96.67	96.67	98.89	96.67	92.22	86.67	97.78
80	96.67	95.00	95.00	96.67	90.00	85.00	96.67

从实验结果来看,应用核 Logistic 回归算法建立的分类模型分类性能良好,准确度均在 90%以上。分析表 1 中数据,可以得到如下结论:(1) 在训练样本尺寸相同时,KLR 算法的分类准确率要高于支持向量机(SVM)和传统的 LR 回归。(2) 三类算法在训练样本尺寸增加的情况下,分类准确度随之提高。(3) 总体看来,KLR 和 SVM 中选取 RBF 核函数时,分类准确率和稳定性要高于选择线性核函数和多项式核函数。

三类算法的准确度的曲线如图 1 所示,其中 KLR 和 SVM 中的核函数均选择 RBF 核函数。

在训练样本尺寸不同的情况下,KLR 算法的分类性能始终优于或等于 LR 算法,因为在 KLR 算法将广义线性模型 LR 扩展为 RKHS 空间的非线性模型,较好的解决了在特征空间中相似度较大音频样本线性不可分的问题。KLR 算法分类性能与 SVM 相比,在每个类别的训练样本尺寸为 50 时,二者的分类性能相等,只有在训练样本尺寸为 70 时 SVM 的分类准确度要高于 KLR,KLR 通过学习能给出样本所属类别在概率意义上的解释,更具实际意义。

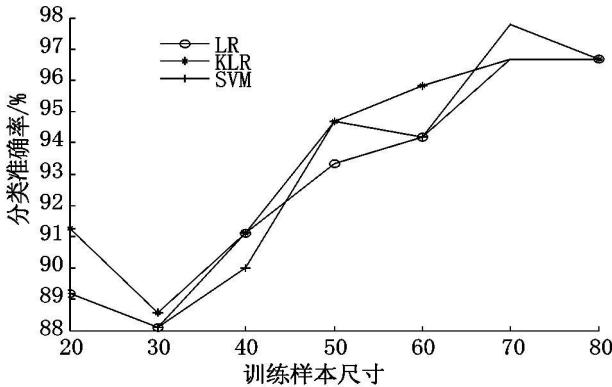


图 1 三种分类算法的准确度曲线图

4 结束语

本文使用核 Logistic 回归(KLR)算法进行音频分类。利用对样本数据的后验验概率估计判定样本所属的类别,并对三类弦乐器音频信号:小提琴、大提琴、中提琴进行分类实验,实验结果表明,KLR 算法分类的误差明显小于传统的 LR 算法和 SVM 算法,取得了较为满意的分类准确度。

在 KLR 中,由于其解不存在稀疏性,意味着计算新样本后验概率时需要所有的训练样本参与运算,计算量较大,这样就限制了 KLR 在大规模数据集中的应用。下一步的工作是对 KLR 解的稀疏性问题进行研究。

参考文献:

[1] 王济川, 郭志. Logistic 回归模型方法与应用[M]. 北京: 高等教育出版社, 2001.

[2] HASTIE T, TIBSHIRANI R, FRIEDMAN J. The Elements of Statistical Learning: Data Mining, Inference, And Prediction [M]. Berlin: Springer Verlag, 2002.

[3] JAAKKOLA T S, HAUSSLER D. Probabilistic Kernel Regression Models[C]//Proceedings of the Conference on AI and Statistics. San Francisco. USA: Morgan Kaufmann, 1999: 99-108.

[4] OTH V. Probabilistic Discriminative Kernel Classifiers for Multi-class Problems, Lecture Notes in Computer Science[R]. London, UK: Springer Verlag, 2001: 246-253.

[5] WOLD E, BLUM T, KEISLAR D, et al. Content-based classification, search and retrieval of audio[J]. IEEE Multimedia Magazine, 1996, 3(3): 27-36.

- [6] 白亮,老松杨,陈剑赟,等.基于支持向量机的音频分类与分割[J].计算机科学,2005,32(4):87-90.
- [7] MUBARAK O M, AMBIKAIRAJAH E, EPPS J. Novel Features for Effective Speech and Music Discrimination[C]//Proc of the IEEE Int'l Conf on Engineering of Intelligent Systems. 2006:22-23.
- [8] 卢坚,陈毅松,孙正兴,等.基于隐马尔可夫模型的音频自动分类[J].软件学报,2002,13(8):1 593-1 597.
- [9] 孟永辉,蒋冬梅,付中华,等.一种新颖的语言/音乐分割与分类方法[J].计算机工程与科学,2009,31(4):106-109.
- [10] MARQUES J, MORENO P J. A Study of Musical Instrument Classification Using Gaussian Mixture Models and Support Vector Machines[R]. COMPAG:Cambridge Research Laboratory, 1999.
- [11] HASTIE T, TIBSHIRANI R. Classification by pairwise coupling[J]. Ann Statist, 1998, 26(2):451-471.
- [12] AUCOUTURIER J, PACHET F, SANDLER M. "The way it sounds":timbre models for analysis and retrieval of music signals [J]. IEEE Transactions on Multimedia, 2005, 7(6):1-8.
- [13] 李涛,王俊普,吴秀清,唐金辉.后验概率估计及其应用:基于核 Logistic 回归的方法[J].模式识别与人工智能,2006,19(16):589-695.
- [14] FRITTS L. Musical Instrument Samples[DB/OL]. [2009-12-26], <http://theremin.music.uiowa.edu/>.

Musical Instrument Audio Identification Based on Kernel Logistic Regression

Liu Zunxiong, Xu Jinfeng, Zeng Lihui

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

Abstract: Audio classification based on statistical learning has attracted widespread attention and been widely used in audio classification system, because of better theoretical foundation and simple implementation mechanism. Based on exploration on theory about kernel logistic regression (KLR), a novel approach for audio classifier is put forward with the help of KLR in this paper. It is used to handle music from the same type of musical instrument. Music signals are collected with violin, viola and cello, and all the signals are preprocessed to extract features. The processed samples are used in experiments, the classification performances are compared with traditional LR and Support Vector Machines (SVM). Simulation results show that KLR performs better on classification accuracy and non-linear processing ability.

Key words: kernel logistic regression; audio classification; feature extraction

(责任编辑 刘棉玲)