基于 MFCC和 HMM的音乐分类方法研究

张 燕[1,2], 唐振民[2], 李燕萍[2], 邹 益[2]

(1. 金陵科技学院 信息技术学院, 江苏 南京 210006; 2. 南京理工大学 计算机学院, 江苏 南京 210094)

[摘要] 采用基于 Mel倒谱系数特征的隐马尔可夫模型对音乐进行分类. 对音乐通过有监督的学习方式进行聚类, 分类时将测试样本归入似然值最大的类别, 对同一音频抽取若干样本, 对样本识别结果采用投票法判定该音频的音乐类别, 使分类的准确率得到进一步的提高. 仿真实验对 4种分类器在有干扰和无干扰的环境下的分类性能进行了比较, 实验结果表明该方法具有更好的抗干扰能力和正确率.

[关键词] Mel倒谱系数, 音乐分类, 隐马尔可夫模型

[中图分类号] TN 912. 34 [文献标识码] A [文章编号] 1672-1292(2008)04-0112-03

Research of Music Classification Based on MFCC Feature and HMM Model

Zhang Yan[1,2], Tang Zhenmin[2], Li Yanping[2], Zou Yi[2]

(1. College of Information, Jinling Institution of Technology, Nanjing 210006, China;
2. College of Computer Science, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract: In this paper, we use hidden Markov Model based on Mel-frequency cepstrum coefficients to classify the music. Classification divides the test samples into categories according to the largest likelihood value. We draw several samples of the same music frequency, identify the results of the samples using the voting method, and thus determine the category of the audio to further improve classification accuracy. We make a simulation experiment to compare the performance of four different classifications in the environments of disturbance and nodistabance. The results show that HMM classification has more advantages on performance and is less sensitive to disturbance.

Key words: Mel frequency cepstrum coefficients; music classification; hidden Markov model

基于内容的音频信息检索技术(CBAIR)研究如何利用音频的幅度、频谱等物理特征, 响度、音高、音色等听觉特征, 词字、旋律等语义特征实现基于内容的音频信息检索[1]. 随着音频数据量的快速增长, 对于种类繁多的音乐数据, 人们要求有快速高效的方法对它们进行分类管理(根据不同风格或演唱者等), 这需要有效的自动分类技术对音频数据进行整理, 以便于检索和相关的分析处理. 音频分类技术是音频检索以及其它音频处理的重要辅助手段.

目前大部分的音频音乐分类算法都包含两个阶段: 特征提取和分类阶段. 许多音乐特征可用于实现这一算法, 包括时域的短时能量、短时过零率、频域的带宽、谱质心等, 还有基于听觉感受的 MFCC等. 而分类算法可利用模式分类中现存的大量高效算法, 例如高斯混合模型、神经网络、支持向量机、隐马尔可夫模型等[2-4]. 本文提出了基于 Mel倒谱系数特征的隐马尔可夫模型对音乐进行分类. 在音乐特征提取方面, 以感知特征和 Mel倒谱系数组成特征向量; 在音乐分类方面, 以隐马尔可夫模型作为分类器, 对同一音频抽取若干样本, 对样本识别结果采用投票法判定该音频的音乐类别. 仿真实验结果表明该方法具有更好的抗干扰能力和正确率.

1 Mel倒谱系数

Mel倒谱系数(MFCC)反映了人耳的音高听觉特性, 而且计算量不大, 广泛应用于语音处理领域. 研究

收稿日期: 2008-06-18
基金项目: 江苏省教育技术研究"十一五"规划重点课题(2007-L-4704)资助项目.
通讯联系人: 张 燕, 副教授, 博士研究生, 研究方向: 模式识别和音频信号处理. E-mail: zy@jit.edu.cn

结果表明 MFCC 系数可以用作音频分类特征, 并且可以提高音频分类的精度. MFCC 特征的计算过程为:

对每一帧信号作 DFT变换计算幅度频谱, 然后将幅度频谱用 Mel尺度变换到 Mel域, 经过等带宽的 Mel滤波器组滤波之后, 将滤波器的输出能量进行叠加.

$$e[j] = \log\left(\sum_{k=0}^{N-1} w_j[k] \times |s[k]|\right), \quad j = 1, 2, \cdots, P, \tag{1}$$

其中 $e[j]$ 表示第 j 个滤波器的对数能量输出; $w_j[k]$ 表示第 j 个三角滤波器的第 k 个点对应的权值; $|s[k]|$ 表示变换到 Mel尺度上的 DFT频谱幅值; P 是滤波器的个数, 一般为 24个. 将滤波器的对数能量进行离散余弦变换, 可以得到如下的倒谱域 MFCC 系数:

$$x_i = \sqrt{\frac{2}{P}} \sum_{j=1}^{P} \left(e[j] \times \cos\left(\frac{i\pi}{P}(j - 0.5)\right)\right), \quad i = 1, 2, \cdots, L, \tag{2}$$

其中, L 是 MFCC 系数的维数, 一般 $L \leqslant P$, 本文取 12维.

音乐信号的 MFCC特征参数主要反映音乐信号的静态特征, 音乐信号的动态特征可以通过这些静态特征的差分谱来描述, 结合一阶差分和二阶差分作为动态特征. 这些动态信息和静态信息形成互补, 能够很大程度上提高系统的识别性能.

2　隐马尔可夫模型

隐马尔可夫模型本质上是一种双重随机过程有限状态自动机[5], 可以用三元数组来表示: $\lambda = (A, B, \pi)$, 其中, A 是状态 S_i 到 S_j 的转换概率矩阵; B 是状态的观察输出概率密度; π 是状态的初始分布概率. HMM 需要研究的 3个基本问题是: (1) 已知 HMM 模型 λ 的各参数, 求某一观察序列 O 在该模型下的极大似然, 即 $P(O|\lambda)$, $O = o_1, o_2, \cdots o_T$; T 为观察序列长度; (2) 在给定的 HMM 模型 λ 的条件下, 求观察序列 O 最有可能历经的状态序列 S; (3) 在已知样本集合的条件下, 如何根据样本集合训练模型并获得模型参数. 问题 (a) 可以由前向 (Forward) 或者后向 (Backward) 算法解决. 问题 (b) 是典型的状态空间搜索问题, 经典的算法有基于动态规划的 Viterbi算法、Beam Search和 A* 算法. 问题 (c) 是统计学习过程, 其学习算法有 Baum-Welch算法、梯度算法等.

3　音频分类系统结构

系统结构首先对音乐文件进行预处理, 分割成音乐帧、加窗、端点检测, 而后进行特征提取, 提取出感知特征 Mel倒谱系数 (MFCC) 的特征序列作为特征向量, 通过基于隐马尔可夫模型 (HMM) 的分类器, 对已知类别的音频数据样本进行训练聚类, 对于未知类别的音频数据样本进行分类, 得出分类结果. 分类过程中, 当有待分类样本需要识别时, 利用已经建立的 HMM 参数来计算每套参数产生该音乐序列 O 的似然值 $P(O/\lambda_i)$, 将新样本归入似然值最大的类别中, 并给出分类结果.

4　实验数据与分析

实验中将音乐分为流行音乐 (*P*)、民歌 (*F*)、古典音乐 (*C*)、戏曲 (*O*) 和语音 (*S*) 5个类别, 所有音乐皆由 Internet下载, 其中流行音乐 (115首)、民歌 (85首)、古典音乐 (78首)、戏曲 (61首) 和语音 (92首), 提取的 MFCC特征维数 $L = 12$ 实验中抽取样本的长度为 5 s, 采样率 11.025 kHz, 采样精度为 16 bit. 训练过程中在每类中随机选取 40首音乐归入训练集, 其余音乐归入测试集. 在此基础上进行了几项实验, 每项实验中训练集和测试集的选择都是随机的, 进行 10次分类并取均值作为最终的分类结果. (*P*: 流行音乐, *F*: 民歌, *C*: 古典音乐, *O*: 戏曲, *S*: 语音)

表 1　不同分类器对不同音乐类型的分类性能

Table 1　The classification performance of different classifier and music

音乐类别	正确率 /%			
	NC	K-NN	PNN	HMM
流行音乐	80.00	81.33	80.00	82.67
民歌	68.89	73.33	71.11	75.56
古典音乐	76.32	78.95	81.58	78.95
戏曲	80.95	80.95	80.95	85.71
语音	90.38	92.31	92.31	90.38

4.1　不同分类器的分类性能比较

以感知特征 MFCC及其一、二阶差分作为分类特征向量, 本文使用 3种经典分类器和 HMM 分类器对特征向量进行分类. 对 4种分类器的分类性能进行比较, 如表 1所示:

由于语音内在的与一般音乐不同的时、频特征,因此语音的分类正确率超过了 90%.又由于戏曲中固有的特点,特征较其它音乐类型明显,所以使戏曲的分类正确率较高.而流行音乐、民歌和古典音乐由于在音乐类型上的相似性和重叠性,所以这 3类比较容易造成误分类,其中流行音乐的误识率稍低一些.从上述表中可以看出,HMM 在识别流行音乐、民歌、戏曲方面比其余的分类器正确率高, PNN 在识别古典音乐、语音方面表现较好,而 HMM 与其余的分类器性能相当,可见 HMM 在 4种分类器中还是有一定的性能优势,这主要是因为 HMM 对时间统计特性的较好表征.

使用 HMM 进行音乐分类的详细结果如表 2所示(表格中的数字表示纵向的音乐类别被分类到横向音乐类别的音乐数目),其中语音和戏曲的分类正确率很高,而有较多的流行音乐、民歌和古典音乐之间比较容易混淆.如流行音乐被误分类为民歌(8首)、古典音乐(5首),民歌被误分类为流行音乐(6首)、古典音乐(4首),古典音乐被误分类为流行音乐(3首)、民歌(3首),这是因为它们之间本来就固有的相似性,例如某些流行音乐和民歌在旋律和音调方面具有相当大的类似性,因此提取的特征向量也很接近,分类器无法对其进行准确的分类.

表 2　HMM 分类详细结果

Table 2　The classification results of HMM

	P	*F*	*C*	*O*	*S*
P	62	8	5	0	0
F	6	34	4	1	0
C	3	3	30	2	0
O	0	1	2	18	0
S	2	3	0	0	47

表 3　加入 4个干扰样本后的分类结果

Table 3　The classification results when adding four disturbance samples

	P	*F*	*C*	*O*	*S*
P	58	13	4	0	0
F	9	31	4	1	0
C	4	4	29	1	0
O	0	1	3	17	0
S	2	3	0	0	47

4.2　干扰对分类的影响

最后,对系统进行了抗干扰实验.以 HMM 为分类器,在流行音乐的训练样本集中加入了少量民歌样本(取 40个训练样本中有 4个干扰样本),经过训练后进行测试.加入 4个干扰样本后对分类结果的影响如表 3所示,从表中可以看出,由于流行音乐训练样本加入了民歌样本,使训练的类别模型参数产生了改变,使流行音乐和民歌之间的误识率有所增加.

5　结语

本文提出了基于 Mel倒谱系数特征的隐马尔可夫模型对音乐进行分类.对同一音频抽取若干样本,对样本识别结果采用投票法判定该音频的音乐类别.对相关理论进行了实验验证,开发了音乐自动分类的实验系统.将音乐文件分为 5类,在分类中采用了投票法来对分类结果进行判定,提高了分类的准确率和稳定性.实验对比了 4种不同分类器的性能,最后对有干扰的模型进行分类实验,实验结果表明本文方法具有更好的抗干扰能力和正确率.

[参考文献](References)

[1] Foote J. An overview of audio information retrieval[J]. Multimedia Systems, 1999, 7(1): 2-10.

[2] Foote J. Content-based retrieval of music and audio[J]. Multimedia Storage and Archiving System II, 1997, 32(29): 138-147.

[3] Li S Z. Content-based classification and retrieval of audio using the nearest feature line method[J]. IEEE Trans on Speech Audio Processing, 2000, 8(5): 619-625.

[4] Lu Guojun, Templar H. A technique towards automatic audio classification and retrieval[C] // Proceedings of the 4th International Conference on Signal Processing. Beijing: IEEE Xplore, 1998: 1 142-1 145.

[5] 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, 8(13): 1 593-1 597.
Lu Jian, Chen Yisong, Sun Zhengxing, et al. Automatic audio classification by using hidden Markov model[J]. Journal of Sofeware, 2002, 8(13): 1 593-1 597. (in Chinese)

[责任编辑: 刘　健]