题目：音乐流派自动分类系统的研究与实现

# 一．课题来源、项目名称

课题来源：

项目名称：基于内容的音乐流派自动分类系统的研究与实现

# 二.文献综述

## 2.1 课题背景和意义

随着互联网的迅猛发展，音乐的主要传播方式已经从过去的CD、唱片等迅速转变为互联网在线传播，音乐工业中一个深刻而显而易见的变化是，人们消费音乐的方式从购买物理媒介转变为消费音乐在线服务。音乐已经成为面向大众的最重要的在线服务之一，互联网上的音乐库规模已经极其庞大，各大在线曲库中不乏歌曲规模上百万者。随之而来的一个重要的问题和需求是如何检索、分类和组织这些音乐。将音乐按照流派分类和组织音乐是人们最自然也最重要的方式之一。

不同流派的音乐各有特色，不同的人对不同流派的音乐的喜好不同，人们希望可以方便快捷的检索和收藏不同流派的音乐。许多智能手机和桌面音乐播放器都提供了按genre分类的音乐列表方便用户检索和播放音乐，然而常见的情况是由于mp3文件的ID3信息错误或者缺失音乐播放器无法获得音乐文件正确的流派信息，这导致genre列表显示的音乐流派混乱或者甚至显示一些乱码。熟悉不同音乐流派的人只要听一首歌的片段就能对歌曲的流派做出判断，分辨一首歌是摇滚、流行、古典、民谣或是其他。基于内容的音乐流派自动分类系统试图代替人工完成这项任务，提供自动的音乐流派分类解决方案。

## 2.2音乐流派自动分类的研究现状

音乐最基本的要素是旋律和节奏。旋律、节奏以及和声等方面是不同音乐之间的主要区别，不同的音乐流派也因此产生。音乐流派又称作音乐风格，有十多种，这十多种有细分为100多种具体的流派，常见和主要的音乐流派中人们比较熟悉的有乡村音乐、蓝调、古典乐、摇滚乐、电子乐、爵士乐、舞曲、Hip Hop、流行乐、重金属、Alternative等等。

近年来，基于内容的音乐分类领域的研究工作取得了很大进展，国际上发表的有关论文数量大幅增长。音乐分类问题既有着各种模式分类问题所共有的特点，可以使用在其他领域中广泛应用的各类经典机器学习的算法；又有自己的特点，对特征提取、特征选择过程和分类器性能等有着特定的要求。音乐数据可抽取的数字特征数量巨大，往往可以达到上百维。目前音乐分类的研究框架所包含的两部分的重点内容是：特征提取和分类器的选择，已经有不少音乐分类领域的论文就这两个问题做了讨论和研究[1-4]。

早期基于相音频内容的容自动分类研究工作可以追溯到1996年，美国 Muscle Fish 公司的 Erling Wold[1] 等人首先提出了根据内容对音乐风格进行分类的方法(Muscle Fish)，在此方法中，从每一首样本(训练样本或测试样本)音乐中提取出一些时域或频域的统计特征，包括均值、方差及自相关系数等，它们代表了该首音乐的感知特征：响度(loudness)、带宽(bandwidth)、亮度(brightness)及音调(pitch)等等，这些特征用于后一阶段的分类器中进行分类。

Foote[2]从歌曲中提取 12 阶 MFCC (Mel Frequency Cepstrum Coefficient)和能量项构造特征向量以表示歌曲，使用一种树状结构的矢量量化器将特征空间分隔为数个不重叠的区域，计算待分类音乐特征与这些区域的距离(欧几里德距离或余弦距离)，使用 NN(Nearest Neighbor)规则作为分类器进行了音乐风格分类实验。

Li[3]选取 MFCC 和包括基音频率、子带能量等在内的感知参数的级联作为特征，并使用了新的模式分类方法 NFL(nearest feature line)作为分类器，他的实验表明这种分类方法优于 NN、NC 及 K-NN 等分类器，最终在 198 首待分类音乐中出现 40 首分类错误。

而Lin等[4]在Li的基础上使用小波变换的方法提取子带能量及基音频率等特征，这种方法得到的特征与其他方法相比更为精确，分类阶段采用一种由底向上的分类结构，并使用 SVM 作为分类器，利用 SVMs 优秀的泛化能力，取得了较好的分类效果。

George Tzanetakis[5] 在 2002 年开展的音乐流派自动分类研究成为了该领域的代表作之一，他介绍了音乐流派分类中常用的特征集合以及特征提取方法，他的实验中使用了分别代表音色纹理、旋律和音高的特征集，并对这些特征的性能做了比较性研究，得到了 61%的准确率，成为最有开创性的工作。实验的结果被广泛引用，许多类似研究都将这次实验的结果作为性能指标的参照。另外由于 George Tzanetakis 构建的 Gtzan 数据集也成为了众多音乐流派自动分类研究首选的训练数据集。

随后在 2003 年 Tao Li 和 George Tzanetakis[6] 共同发表的论文研究了影响分类性能的因素。这个实验研究了不同数字特征以及特征之间的组合对分类效果的影响，以及不同分类器对分类效果的影响。实验结果显示节奏相关的特征计算消耗最大，但对于改善流派分类效果的作用却并不明显；Full Beat Histogram 和 Full Pitch Histogram 在大幅增大特征向量维度的同时对分类效果的改善并不显著；使用 LDA 分类器得到的 71% 最佳分类效果比 2002 年使用 GMM 分类器得到的 61% 的分类效果有显著提升。

Tao Li 等人[7]在音乐风格分类实验中加入了 Daubechies 小波系数（DWCH）特征，对不同的特征组合和分类算法做了对比研究，结果显示 DWCH 与音色特征（MFCC 和 FFT）组合成的特征向量与多分类支持向量机构成的分类系统可以大致获得 80% 的分类精度，这比早前 Tzanetkis 的分类工作有了相当可观的性能提升。

E Benetos 等人[8]提出了一种基于特征张量的音乐自动分类方法。该研究中使用多线性技术（multilinear techniques）来处理音乐的特征提取问题，在多线性代数领域中，张量被认为是矩阵和向量的等价替代量。该研究中使用特征张量来表示音乐记录，这种表示方法更好的刻画了短时特性的时变属性。基于 NMF （non-negative matrix factorization）作者提出了一种名为 NTF （non-negative tensor factorization）的算法处理数据分类问题，应用该算法对包含 Blues、Classical 和Jazz等10个类别的音乐几率进行了分类实验并得到了不错的实验结果，值得一提的是该实验中 Rock 和Metal 两个分类的区分度很高，与许多音乐分类实验相比较而言是不错的提升。

Lin等[9]受到Li的研究的启发，使用了小波变换来提取特征；然后使用支持向量机作为分类器并采用自下往上的分类结构，这一算法获得了更高的分类精度。

Munoz等［10］将遗传模糊系统与 GMM 结合，对 GMM 进行了深度优化。

2009年Honglak Lee等［11］首次将深度学习算法：深度置信网络（DBN，Deep Belief Network ）应用于音乐分类，并与传统的音乐分类算法与神经网络算法进行了比较，取得了相对较好效果。

2012年，Google Brain 项目得到了广泛关注，它用并行计算机平台训练出了深度神经网络（DNN，Deep Neural Network），在语音处理和图像识别领域均获得了巨大的突破和成功。

除了以上经典算法，还存在很多其它有效的音乐分类算法。当前研究重点在于提取可以在最大程度上反映音乐本质属性的特征，和设计高性能的分类器以优化算法的分类结构。除此之二者之外，由于音乐分类问题中特征向量的维度常常是在几十甚至上百维，而且一个实际有意义音乐分类系统需要处理8个以上的分类，所以计算强大度是实现自动音乐分类中需要面对的另一个问题，对于本身计算复杂度大的分类算法尤其如是。随着分类需求的发展，当需要增加分类数目时所增加的计算复杂度也会继续上升。因此特征选择相关的研究和工作对于实现高效的音乐分类系统而言也是很有意义的，通过特征选择排除冗余和无关的特征以降低特征维度，减小计算强度，排除信号噪音，可以提高自动音乐分类系统的性能。

## 2.3 音乐信号的处理

在按帧进行语音分析，提取语音参数之前，有一些经常使用的、共同的短时分析技术必须预先进行，这些处理统称为预处理技术。图2.1是整个预处理阶段的流程图，整个流程包括了分帧、预加重、加窗和静音帧判别四个步骤。

### 2.3.1 分帧

分帧

预加重

加窗

静音判别

特征提取

抛弃

音乐信号

是

否

图2.1 音乐信号预处理流程图

为了保证了音乐信号在每一帧内保持短时平稳，在预处理阶段首先应对原音乐信号进行分帧处理，下面讨论帧长的选择问题。 采样周期 、窗口长度 N 和频率分辨率Δf 之间存在下列关系：

可见，采样周期一定时，Δf 随窗口宽度 N 的增加而减小，即频率分辨率相应得到提高，但同时时间分辨率降低；如果窗口取短，频率分辨率下降，而时间分辨率提高，因而两者是矛盾的。应该根据不同的需要选择合适的窗口长度。例如对于时域分析来讲，如果 N 很大，则它等效于很窄的低通滤波器，语音信号通过时，反映波形细节的高频部分被阻碍，短时能量随时间变化很小，不能真实地反映语音信号的幅度变化，不能得到平滑的能量函数。因此窗口的长度选择应合适。

### 2.3.2 预加重

由于语音信号的平均功率谱受声门激励和口鼻辐射影响，高频端大约在80OHz 以上按 6dB/倍频程跌落，所以求音乐信号频谱时，频率越高相应的成分越小，高频部分的频谱比低频部分的难求，为此要在预处理中进行预加重处理。预加重的目的是提高高频部分，使信号的频谱变得平坦，保持在低频到高频的整个频带中，能用同样的信噪比求频谱，以便于频谱分析或声道参数分析。预加重可在语音信号数字化时在反混叠滤波器之前进行，这样不仅可以进行预加重，而且可以压缩信号的动态范围，有效地提高信噪比。但预加重一般是在语音信号数字化之后，在参数分析之前在计算机里用具有 6dB/倍频程的提升高频特性的预加重数字滤波器来实现，它一般是一阶的数字滤波器：

μ值接近于 1。

有时要恢复原信号，需要从做过预加重的信号频谱来求实际的频谱时，要对测量值进行去加重处理，即加上 6d B/倍频程的下降的频率特性来还原成原来的特性。

### 2.3.3 加窗

进行预加重数字滤波处理后，接下来就要进行加窗分帧处理。一般每秒的帧数约为 33-100 帧，视实际情况而定。分帧虽然可以采用连续分段的方法，但一般要采用交叠分段的方法，这是为了使帧与帧之间平滑过渡，保持其连续性。前一帧和后一帧的交叠部分称为帧移。帧移与帧长的比值一般取为 0-1/2。分帧是用可移动的有限长度窗口进行加权的方法来实现，这就是用一定的窗函数 w(n)来乘 s(n)，从而形成加窗语音信号 。

在语音信号数字处理中常用的窗函数是矩形窗和汉明窗。窗函数 w(n)的选择(形状和长度)对于短时分析参数的特性影响很大。虽然不同的短时分析方法(时域、频域、倒谱域分析)以及求取不同的语音特征参数可能对窗函数的要求不尽一样，但一般来讲，一个好的窗函数的标准是:在时域因为是语音波形乘以窗函数，所以要减小时间窗两端的坡度，使窗口边缘两端不引起急剧变化而平滑过渡到零，这样可以使截取出的语音波形缓慢降为零，减小语音帧的截断效应;在频域要有较宽的 3d B 带宽以及较小的边带最大值。汉明窗的主瓣宽度比矩形窗大一倍，即带宽约增加一倍，同时其带外衰减也比矩形窗大一倍多。矩形窗的谱平滑性能较好，但损失了高频成分，使波形细节丢失;而汉明窗则相反，从这一方面来看，汉明窗比矩形窗更为合适。

### 2.3.4静音帧判别

最后，若音乐信号中存在静音帧（信号幅度为零或很小），则从这些帧中提取出的参数将不具代表性，而且会恶化最终提取的音乐特征参数，因此为了排除一首音乐中静音帧对提取特征的影响，首先应计算每一帧音乐信号的短时能量，若该帧能量低于一阈值时，该帧被判为静音帧，对其不作任何后继处理而抛弃。阈值的选取对最终音乐分类的精确度有较大的影响，因此最终阈值的选取可由实验中具体数据决定，本论文对阈值的选取在 10-20 之间。

经过上面四个步骤的处理过程，音乐信号就已经被分割成一帧一帧的加过窗函数的短时信号，然后再把每一个短时音乐信号帧看成平稳的随机信号，利用数字信号处理技术来提取音乐特征参数。

## 2.4 用于分类的音乐特征

特征提取是音乐分类的关键部分，它的主要功能是从音乐中提取出表征音乐特征的特征向量，即把音乐信号中具有辨识性的成分提取出来[13][14][15][16]。音乐信号特征应该能够最大程度的表征音乐的属性，并且具有不变性和可区分性。Scaringella[12]将音频特征从听觉属性的角度分为基于音色、节奏和音调的三大类。人们对声音音质的感知主要由音调、音色和节奏来度量。音调是指人耳对声音的调子高低的主观感受，它与声音的频率成正比；音色是指人们对声音音质的感觉，在音乐信号中除了最低的频率之外的一切频率统称为泛音，泛音的组合成就了特定的音色；节奏定义为音乐的规律程度，通常用单位时间的节拍数来表征。音调、音色和节奏是人耳对音乐信号的主观感受，它们均可抽象成为表征它们的特征向量，其中音色抽象而成的大多为短时特征，而音调和节奏抽象而成的大多为长时特征，除了直接抽象而成，长时特征也可以由短时特征组合而成。

反映音色的特征

频谱中心值（Spectral Centroid）: 频谱中心是从能量分布中推导出的一个重要特征，它是一个声音频谱能量分布的中心点。频谱中心也被称为亮度。频谱中心在刻画音乐乐器音色方面是一个非常重要的感知参数。

衰减截止频率（Spectral Rolloff Point）：衰减截止频率是谱形状的另一个重要测度，它反映了信号能量在低频上的 n 分布。

频谱流量（Spectral Flux）：频谱流量定义为相邻两帧之间频谱分布的变化量，它体现了信号的动态特性。

过零率（Zero Crossings）： 过零率指在一个短时帧内，离散采样信号值由正到负和由负到正变化的次数。这个量能够反映信号在短时帧内的平均频率。音乐在 ZCR 上具有较低的可变性。

Mel 倒谱系数（MFCC：Mel-Frequency Cepstrum Coefficient）： Mel 倒谱系数（MFCC）是建立在傅立叶和倒谱分析基础上的，对短时音频帧中的[K/M]个采样点进行傅立叶变换，得到这个短时音频帧在每个频率上的能量大小。

线性预测系数（LPC：Linear Prediction Coefficient） ：如果用有限个参数的线性模型来近似表示音频序列{si}，这些参数就可以成为描述序列的重要特征，称为线性预测系数。在该线性预测模型下，对下一个样本的预测可以表示之前样本的加权和。

最强频率值（Strongest Frequency Via FFTMaximum）：该值是信号进行快速傅立叶频谱变换后计算得到的最强值。

2.反映节奏的特征

通常，节奏特征对于流派分类同样起着重要的作用，不同流派的音乐节奏差异较大。

最强节拍均值（Strongest Beat Overall Average）：反映了音乐中每秒的节拍数，是节拍直方图中的最强节拍。而节拍直方图是通过离散小波变换识别基频峰值获得的。

节拍力度和均值（Beat Sum Overall Average）

最强节拍力度均值（Strength Of Strongest Beat Overall Average）

3.反映响度的特征

能量均方根（Root Mean Square）：RMS能很好的反映信号的强度，是信号在一帧内采样数据的均方值。

低能量帧比例（Fo LE:Fractionof Low Energy Frames）：低能帧比例为能量小于RMS的帧数占总帧数的比例。

4．DWCH（Daubechies Wavelet Coefficient Histograms）特征

基于 Daubechies 小波系数直方图的前三阶参数：均值、方差、偏度和子带能量。

## 2.5 特征选择

典型的分类研究中包括预处理、特征提取和分类器训练和测试阶段，为了减低分类计算的代价、提高分类准确率本研究中在特征提取之后加入一个特征选择的过程。

经典的特征选择的定义为从 N 个特征集合中选出 M 个特征的子集，并满足条件 M≤N，它可以看作一个优化问题，其关键是建立一种评价标准来区分哪些特征组合有助于分类，哪些特征组合存在冗余性、部分或者完全无关。不同的评价函数可能会给出不同的结果。

根据评价函数与分类器的关系，特征选择方法分成筛选器（Filter）[17][18]和封装器（Wrapper）[19][20]两种。Filter 的评价函数与分类器无关，而封装器采用分类器的错误概率作为评价函数。Filter 的评价函数又可以细分为距离测度、信息测度、相关性测度和一致性测度。Filter 方法仅仅使用数据集评价每个特征(子集)的相关性，不考虑后来的学习算法。ReliefF[21]就是这类算法的代表。信息理论的方法也被用于评价特征：相关特征和类标签的相互信息值应该是高的。非参数的方法也可以用来计算包含连续值特征的相互信息：如果一个参数在给定其他特征的条件下，条件独立于类标签，则该参数被认为是无关特征。Wappers方法使用学习算法来评价每个特征(子集)的质量。具体地说，一个学习算法(例如最邻近分类器、决策树、朴素贝叶斯方法)运行在一个特征子集，该特征子集用分类准确率的估计值来评价。同 Filters 方法相比，Wappers 方法通常计算量大，但是可以获得更高的准确率。

数据集的大小可以从两方面衡量：特征的数目 n 和样本的数目 p，n 和 p 可能很大，而 n 的庞大经常会引起维数灾难（Curse of Dimensionality）等问题。特征空间的维数不宜过高，这已经是模式识别领域中一条经验性的“公理”[22]。为了准确的估计参数或者对未知样本进行预测，训练样本的个数是随着特征数目指数增长的，这就是所谓的维数灾难(Curse of Dimensionality)。许多学习算法的性能受到无关特征和冗余特征的影响，比如最近邻法，需要计算样本间的距离，这样不管是冗余特征还是无关特征，都会对距离的计算产生影响，势必会降低算法性能，而且研究者发现最近邻分类器容易陷入维数灾难中，尤其是当存在很多无关特征的情况下[23]。另外，决策树和神经网络的性能也同样受到无关或者冗余特征的影响[24]。

## 2.5.1 Wrapper 特征选择方法

封装式 Wrapper 特征选择算法最早由 John 等在 1994 年提出[25]。该算法的核心思想是：和学习算法无关的过滤式特征评价会和后续的分类算法产生较大的偏差，而学习算法基于所选特征子集的性能是更好的特征评价标准。不同学习算法偏好不同的特征子集，既然特征选择后的特征子集最终将用于后续的学习算法，那么该学习算法的性能就是最好的评估标准。因此在Wrapper 特征选择中将学习算法的性能作为特征选择的评估标准。

由于采用学习算法的性能作为特征评估标准，Wrapper 特征选择算法比过滤式特征选择算法准确率高，但算法效率较低。Wrapper 方法的另一个缺点是过适应问题，但该问题主要发生在训练数据规模较小的情况，而且研究者实验发现 Wrapper 算法的过适应并不多见。高维问题的不断出现对特征选择研究提出了严峻的挑战，这不仅是在音乐分类领域中存在的突出问题，在其他领域中也非常突出。 在 Wrapper 类算法中，SFS、SBS 和 GA（Genetic Algrithm）是比较经典且有效的方法：

顺序前进法(SFS)

这是最简单的自下而上搜索方法。首先计算每个特征单独进行分类的准确率值，并选择其中准确率最大的特性，作为入选特征。然后每次从未入选的特征中选择一个特征，使得它与已入选的特征组合在一起时所得的准确率为最大，直到特征数增至 d 个为止，d 为要求选中的特征个数。顺序前进法与单独特征最优化组合相比，一般说来，由于考虑了特征之间的相关性，在选择特征时计算与比较了组合特征的判据值，要比单独特征最优化组合好些。其主要缺点是，一旦某一特征被选入，即使由于后加入的特征使它变为多余，也无法再把它剔除。

顺序后退法(SBS)

这是一种与 SFS 相反的方法，是自上而下的方法。做法也很简单，从现有的特征组中每次减去一个不同的特征并计算其准确率，找出这些准确率中之最大值，如此重复下去直到特征个数达到预定数值 d 为止。与 SFS 相比，此法计算判据值是在高维特征空间进行的，因此计算量比较大。

遗传算法(GA)

遗传算法最早由 Holland[26]提出，它是一种进化学习方法，通过迭代搜索技术来实现。与传统的优化过程不同，遗传算法从多个初始解开始，经过演化，同时到达一个近似的最优解。这一过程是通过模拟自然界生物的选择和繁衍而进行随机的信息交换来完成的。待解决的问题通过编码由多个基因构成的染色体来描述。每个染色体都有一个适应值，由适应函数评价得到。这个适应函数是用来判断染色体的好坏及其生存概率的，适应值越高，则该染色体越有可能被选择进行繁殖。在遗传算法中，初始种群由较少的个体组成，然后反复地根据适应值繁衍出优化的后代。

## 2.5.2 Filter 特征选择

过滤式（Filter）特征选择的评估标准独立于学习算法，直接由数据集求得。过滤式特征选择的评估依赖于数据集本身，通常是选择和目标函数相关度大的特征或者特征子集，一般认为相关度较大的特征或者特征子集会对应得到后续学习算法较高的准确率。过滤式特征选择的评估方法很多，如类间距离、信息增益、关联度(Correlation)以及不一致度等等。过滤式特征选择因为通常运行效率较高而适用于大规模数据集，但寻找和类别相关的特征/特征子集和选择可最优化分类准确率的特征/特征子集是两个不同的任务。直接用分类器准确率作为特征子集评估标准的特征选择算法就是下面要介绍的 Wrapper 式特征选择。

Relief 是公认的性能较好的特征评估方法，其特征评价借鉴了最近邻学习算法的思想，其核心思想是：一个好的特征应该使最近邻的同类样本之间特征值相同或相近，而使最近邻的不同类样本之间值不同或者差别很大。Relief 系列算法的优点是：运行效率高，对数据类型没有限制，对特征间的关系不敏感（相对很多特征评估方法的特征间独立的假设）。Relief 算法的缺点是：和很多特征评估算法（如信息增益，Gini-Index 等）一样，Relief 算法不能去除冗余特征，算法会赋予所有和类别相关性高的特征较高的权值，而不管该特征是否和其余特征冗余。

## 2.6参考文献

1. Meng A, Shawe-Taylor J. An Investigation of Feature Models for Music Genre Classification Using the Support Vector Classifier.[C]// Ismir 2005, International Conference on Music Information Retrieval, London, Uk, 11-15 September 2005, Proceedings. 2005:604-609.
2. Norowi N M, Doraisamy S, Rahmat R W O K. Factors Affecting Automatic Genre Classification: An Investigation Incorporating Non-Western Musical Forms.[C]// Ismir 2005, International Conference on Music Information Retrieval, London, Uk, 11-15 September 2005, Proceedings. 2005:13--20.
3. Mckay C. jAudio: Towards a standardized extensible audio music feature extraction system[J]. 2005.
4. Li M, Sleep R. Genre Classification via an LZ78-Based String Kernel.[C]// Ismir 2005, International Conference on Music Information Retrieval, London, Uk, 11-15 September 2005, Proceedings. 2005:252--259.
5. Tzanetakis G, Cook P. Musical genre classification of audio signals[J]. Speech & Audio Processing IEEE Transactions on, 2001, 10(5):293-302.
6. Li T, Tzanetakis G. Factors in automatic musical genre classification of audio signals[C]// Ismir 2001, International Symposium on Music Information Retrieval, Indiana University, Bloomington, Indiana, Usa, October 15-17, 2001, Proceedings. 2001:293-302.
7. Li T, Ogihara M. Towards Intelligent Music Information Retrieval[J]. IEEE Transactions on Multimedia, 2006, 8(3):564-574.
8. Benetos E, Kotropoulos C. A tensor-based approach for automatic music genre classification[J]. 2008:1-4.
9. Lin C C, Chen S H, Truong T K, et al. Audio Classification and Categorization Based on Wavelets and Support Vector Machine[J]. IEEE Transactions on Speech & Audio Processing, 2005, 13(5):644-651.
10. Molla M K I, Hirose K. On the effectiveness of MFCCs and their statistical distribution properties in speaker identification[C]// IEEE, 2004:462-7.
11. Lee H, Pham P T, Yan L, et al. Unsupervised feature learning for audio classification using convolutional deep belief networks[C]// Advances in Neural Information Processing Systems 22:, Conference on Neural Information Processing Systems 2009. Proceedings of A Meeting Held 7-10 December 2009, Vancouver, British Columbia, Canada. 2009:1096-1104.
12. Scaringella N, Zoia G, Mlynek D. Automatic Genre Classification of Music Content: a Survey[J]. IEEE Signal Processing Magazine, 2006, 23(2):133-141.
13. 韩纪庆. 语音信号处理[M]. 清华大学出版社, 2004.
14. 赵力. 语音信号处理[M]. 机械工业出版社, 2003.
15. Xu C, Maddage M C, Shao X, et al. Musical genre classification using support vector machines[J]. 2003, 5.
16. Zhang T, Kuo C C J. Audio content analysis for online audiovisual data segmentation and classification[J]. IEEE Transactions on Speech & Audio Processing, 2001, 9(4):441-457.
17. Huang Y, Tseng S S, Wu G, et al. A two-phase feature selection method using both filter and wrapper[C]// IEEE Smc 99 Conference IEEE International Conference on Systems. IEEE, 1999:132 - 136.
18. 白亮, 老松杨, 陈剑赟,等. 音频自动分类中的特征分析和抽取[J]. 小型微型计算机系统, 2005, 26(11):2029-2034.
19. Kira K, Rendell L A. The feature selection problem: traditional methods and a new algorithm[C]// National Conference on Artificial Intelligence. San Jose, Ca, July. 1992:129-134.
20. Cardie C. Using Decision Trees to Improve Case-Based Learning[C]// Proc. Int. Workshop on Machine Learning. 1993.
21. Jain A K, Duin R P W, Mao J. Statistical Pattern Recognition: A Review[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2000, 22(1):4-37.
22. Jerome H. Friedman. On Bias, Variance, 0/1—Loss, and the Curse-of-Dimensionality[J]. Data Mining & Knowledge Discovery, 1997, 1(1):55-77.
23. Langley P. Selection of Relevant Features in Machine Learning[J]. Proceedings of the Aaai Fall Symposium on Relevance, 1997:140--144.
24. John G H, Kohavi R, Pfleger K. Irrelevant Features and the Subset Selection Problem[J]. Machine Learning Proceedings, 1994:121-129.
25. Holland J H. Adoption in Natural and Artificial System[M]// Adaptation in natural and artificial systems. MIT Press, 1975:126–137.

# 三.研究计划

## 3.1本选题研究的主要内容和目标

研究的目标是建立和实现一套自动音乐流派分类系统，为了具有实际应用的意义，预设的分类包括 rock、classical、pop、blues、jazz、metal、alternative、country、electronic、folk、reggae、hiphop等12类。

通过机器学习的方法构建分类系统。同时为了保证系统可用性，分类准确率是本课题首要最求的性能指标。其次良好的分类速度也是本课题最求的第二个性能指标。

本课题从一下应用场景出发来验证分类系统最终的可用性：

为 Android 移动设备上的本地音乐库提供自动分类服务：Android 手机处于充电和 WiFi 连接条件时在云端为其本地音乐库提供音乐流派分类服务，并返回结果。

## 3.2技术方案

本研究的主要工作主要概括为两大部分内容：

构建高性能的自动音乐流派分类系统；

这部分的工作主要包括三个方面：特征提取、特征选择和分类算法改进。通过特征选择排除无关和冗余特征，减轻分类计算代价，提高分类准确率。通过分类算法的改进或多种算法的整合提高分类准确率。

特征选择过程中将会涉及到的技术有ReliefF、SFS等；分类器调优过程中将要涉及的技术有 Logistics Regression、k-Nearest Neighbor、Linear Discriminant、Support Vector Machine、Hidden Markov Model、Gaussian Mixture Model等。

为移动设备提供本地音乐库流派分类的服务和接口。

## 3.3 主要问题和难点

1.测试数据的补齐。音乐自动分类研究中常用的测试数据集如Gtzan只有10个genre的数据，需要补齐alternative和eletronic等满足本课题需求的数据集。

2.计算工作量大。由于目标分类数目多，特征维度大，为了实现一个高性能的自动音乐流派分类系统无论是特征选择还是分类器调优过程都许多对多种算法和算法的组合进行计算强度大的计算和比较。

3.Music Genre Classification 这样的多分类问题中从现有研究的结果所展示的混淆矩阵来看各种分类算法对不同音乐流派的分类能力是不一样的，整合多种分类算法构建一套对各个音乐流派的分类准确率都足够高的分类系统具有挑战性。

## 3.4 研究进度计划

初步的进度计划安排如下：

|  |  |
| --- | --- |
| 2016.10-2016.11 | 查阅文献资料 |
| 2016.12-2017.03 | 初步实验特征选择和分类器 |
| 2017.04-2017.05 | 对初步成果进行验证 |
| 2017.06-2017.08 | 在基础研究成果之上深入研究 |
| 2017.09-2017.10 | 对更多分类算法进行整合实验和测试 |
| 2017.11-2017.12 | 对研究成果进行整合 |
| 2018.01-2018.03 | 撰写论文 |