

参赛队号#1129

## 第六届“认证杯”数学中国

### 数学建模网络挑战赛

#### 承 诺 书

我们仔细阅读了第六届“认证杯”数学中国数学建模网络挑战赛的竞赛规则。

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的，如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们将受到严肃处理。

我们允许数学中国网站([www.madio.net](http://www.madio.net))公布论文，以供网友之间学习交流，数学中国网站以非商业目的的论文交流不需要提前取得我们的同意。

**我们的参赛队号为：** 1129

**参赛队员（签名）：**

队员 1：万佳乐

队员 2：王志平

队员 3：何帅帅

**参赛队教练员（签名）：** 郭青宇

**参赛队伍组别：** 本科组

参赛队号#1129

## 第六届“认证杯”数学中国

### 数学建模网络挑战赛

#### 编号专用页

参赛队伍的参赛队号（请各个参赛队提前填写好）：

#1129

竞赛统一编号（由竞赛组委会送至评委团前编号）：

---

竞赛评阅编号（由竞赛评委团评阅前进行编号）：

参赛队号#1129

# 2013 年第六届“认证杯”数学中国 数学建模网络挑战赛

题 目 流行音乐发展简史关 键 词 神经网络;分类器;反文档频率;LSA;LDA

## 摘 要：

音乐风格划分是音乐推荐、欣赏的基础。本模型考虑各个音乐流派的声学、标签和主旋律特征。用 Audition 等软件处理音乐获取数据，在利用神经网络等方法进行数据处理，最终对音乐进行风格划分。

第一针对声学特征，先利用软件处理音乐，然后运用神经网络结构给出误差反传训练算法的分类器，最后对分类器进行训练和音乐风格划分。

第二针对标签特征，从音乐网站上利用 Audioscrobbler 抽取音乐标签，通过反文档频率和 LSA 方法对音乐——标签向量进行降维预处理，最后使用 LDA 模型处理降维后的向量，将标签属于每个流派的概率转换为对应的音乐属于每个流派的概率。

第三针对旋律特征，以 MIDI 乐曲为数据源提取出乐曲的旋律，然后对旋律数字化，通过经验公式对两首音乐进行旋律重合匹配，得出不同风格音乐的旋律。最后利用分类器进行音乐风格划分。

最终将基于三种特点所求得分类结果进行加权结合，得到音乐流派分类的最优解。

参赛队号： #1129参赛密码                     **(由组委会填写)**所选题目：     B     题

## Abstract:

How to divide the music into several categories is based on recommendation and appreciation of it. The model takes into account the acoustics, labels and the main theme of the characteristics of all kinds of the music schools. Processing the music with some softwares such as Audition to gain the data and then using methods such as neural networks for data processing, and ultimately the music style is divided.

First of the overall is in connection with the acoustic characteristics. We should process the music with software to gain the source data, and then based on neural network structure, given the error reversal training algorithm classifier, and finally training the classifier to achieve the purpose to divide the music.

Second of it is aimed at label feature. We extract music label with Audioscrobbler from music sites and process dimensionality reduction pretreatment to music labels vector through Anti document frequency and LSA. Finally we used the vectors which have finished Dimensionality reduction with LDA model and change each school probability of label into the probability of Corresponding music.

Third for melody features, we extract out the song's melody for source data with music in MIDI format and then digitize the melody and match the paragraphs of the same music section to two style music by empirical formula, comparing the size of the frequency same as music section, eventually divided into various types of music.

Finally we put the classification results based on the three characteristics into weighted combination and gain the optimal solution of the music schools.

**Key words:** neural network ; classifier ; anti-document frequency ; LSA; LDA

## 一、问题重述

随着互联网的发展,流行音乐的主要传播媒介从传统的电台和唱片逐渐过渡到网络下载和网络电台等。网络电台需要根据收听者的已知喜好,自动推荐并播放其它音乐。由于每个人喜好的音乐可能横跨若干种风格,区别甚大,需要分别对待。这就需要探讨如何区分音乐风格的问题。

在流行音乐中,传统的风格概念包括 Pop (流行)、Country (乡村)、Jazz (爵士)、Rock (摇滚)、R&B (节奏布鲁斯)、New Age (新世纪) 等若干大类,它们分别可以细分成许多小类,有些小类甚至可以做更进一步的细分。而每首歌曲只能靠人工赋予风格标签。这样的做法有许多不足:有的类别之间关系不清楚,造成混乱;有的类别过度粗略或精细;有的类别标签没有得到公认;有的音乐归属则存在争议或者难以划归。所以需要建立合理的数学模型,对流行音乐的风格给出一个自然、合理的分类方法,以便给网络电台的推荐功能和其它可能的用途提供支持。

首先对上述问题需要解决不同风格音乐特点的分析,然后建立模型给出一个自然、合理的分类方法。

## 二、问题分析

随着互联网的发展,网络电台需要根据收听者的喜好,自动推荐并播放其它音乐。而每首歌曲只能靠人工赋予风格标签。这样的做易造成混乱、过度粗略或精细、音乐归属则存在争议或者难以划归。因此,音乐流派的准确分类对于现代音乐信息检索系统来说是至关重要的。

人工标注的速度显然已经不能满足网络中音乐资源飞速增长的需求。为此,人们开始自动音乐流派分类的研究,之前的研究使得流派分类的准确率有了提高,但仅仅使用底层声学特征对音乐进行流派分类不能满足实际音乐信息检索系统<sup>[1]</sup>的需要。为此,需要找到新的研究方法来提高音乐流派分类的准确率。

所建模型需要考虑各个音乐流派的声学、标签和主旋律特征,用软件处理音乐获取数据,在利用神经网络等方法进行数据处理,最终对音乐进行风格划分。针对声学特征,先利用软件处理音乐,然后运用神经网络结构给出误差反传训练算法的分类器,最后对分类器进行训练和音乐风格划分。针对标签特征,从音乐网站上抽取音乐标签,通过反文档频率和 LSA 方法对音乐——标签向量进行降维预处理,最后使用 LDA 模型处理降维后的向量,将标签属于每个流派的概率转换为对应的音乐属于每个流派的概率。每首乐曲中,都会有几段相同的乐段,它是整个音乐的主旋律,也是一首曲子的灵魂,所以它也是一首音乐的特征之处,也就促成了音乐的分类,对旋律特征,以 MIDI 乐曲为数据来源提取出乐曲的旋律,然后对旋律数字化,通过经验公式对两首音乐进行旋律重合匹配,得出不同风格音乐的旋律。最后利用分类器进行音乐风格划分。

所建数学模型需要给出对流行音乐的风格给出一个自然、合理的分类方法,以便给网络电台的推荐功能和其它可能的用途提供支持。

### 三 模型假设

1. 假设我们研究音乐的特征时，除音高与音符的长短，其它对所研究的问题不产生干扰
2. 假设从 cn.last.fm 网站得到的标签信息完全正确
3. 假设我们在音乐库中得到的音乐具有代表性。
4. 假设音波其中的音乐信息，没有进行一定的损坏
5. 假设音符的长短对旋律的数值分析影响不大

### 四 变量与符号说明

变量符号	符号说明
$H_r$	相对音高
$L_r$	相对节拍
$H_{\min}$	最低音
$L_{\min}$	最短音
M	径向离散量
N	旋转角
$X_{Ki}$	K 层上 i 节点的输入
$Y_{Ki}$	K 层上 i 节点的输出
$W_{KLij}$	K 层上 i 节点和 L 层上 j 节点间的连接权重
$\theta_{Ki}$	K 层上 i 节点的阈值
$N_L$	L 层上的节点数
$\bar{Y}_{ij}$	第 j 层节点的输出期望

## 参赛队号#1129

$Y_{ii}$	第 i 层节点输出的真实值
$N_p$	输出层节点数
$E_p$	训练中各图形的积累误差
$x_i$	对应的每首歌
$T(x_i)$	标签集合
$Freq(t_j, x_i)$	标签 $t_j$ 被用来标注音乐 $x_i$ 的次数
$t$	音乐 $x_i$ 包含的标签数目
$\omega_{ts}$	标签在矩阵中的权重
$f_{ts}$	歌曲 s 中标签 t 出现的次数
$d_{fs}$	标签 t 的歌曲数目
$N$	歌曲的总数目
$\Sigma_k$	对角矩阵
$U_k$	矩阵中行向量对应原矩阵 A 的标签向量
$V_k$	矩阵中列向量对应原矩阵 A 的歌曲向量
$k$	降维后的维数
$D$	语料库
$M$	语料库中文档的总数
$\beta$	为一个矩阵（里面的元素代表某条件下生成某个单词的概率）
$pr(c_k   t_j)$	每个标签属于某个流派的概率
$f_i$	风格的趋向值
$m$	测试集的曲目数量
$n$	测试集中被正确识别的曲目数量

$p_A(c_k   x_i)$	特征 $i$ 的流派 $j$ 数值 $x_{ij}$ 占有流派总数值 $c_j$ 的概率
------------------	--

## 五、模型建立与求解

首先考虑各个音乐流派的声学特征、标签特征和主旋律特征。用 Audition 等软件处理音乐获取数据，接下来利用神经网络结构给出误差反传训练算法的分类器对音乐风格划分。从音乐网站上抽取音乐标签，通过反文档频率<sup>[2]</sup>和 LSA<sup>[3]</sup>方法对音乐——标签向量进行降维预处理，最后使用 LDA<sup>[4]</sup>模型处理降维后的向量，将标签属于每个流派的概率转换为对应的音乐属于每个流派的概率。对旋律特征以 MIDI<sup>[5]</sup>乐曲为数据源提取出乐曲的旋律，然后对旋律数字化，通过经验公式对两首音乐进行旋律重合匹配，得出不同风格音乐的旋律。最后利用分类器进行音乐风格划分。

### 5.1 声学特征

歌曲的声学特征包括节拍和音高，每一类歌曲的节拍和音高是相似的，所以节拍和音高可以用来划分音乐风格。但相同的歌曲以不同速度或不同人的演唱会影响其风格分类，绝对指标会在很多场合下容易产生很大的误差，因此，本模型选取相对音高  $H_r$  和相对节拍  $L_r$  两个不变量为歌曲的声学特征：

$$LH_r = H - H_{\min} \quad L_r = \frac{L}{L_{\min}}$$

其中  $H_{\min}$  为最低音， $L_{\min}$  为最短音节。

#### 5.1.1 数据的获取

本模型选取整首音乐作为处理对象，数据集共包括6个音乐流派即 pop（流行）、country（乡村）、Jazz（爵士）、Rock（摇滚）、R&B（节奏布鲁斯）、New Age（新世纪），其中每个流派由100首音乐组成。这些音乐是根据网站mp3.com列出的每个流派的经典专辑和经典曲目列表在Last.fm上免费下载获得的。然后利用软件Audition处理音乐获取音高的H值。图1中白点行为其中一首歌的H值。运用MATLAB对每首音乐的波进行数字化处理后会得到相应波长的两列矩阵，从而获得节拍L的值。



## 参赛队号#1129

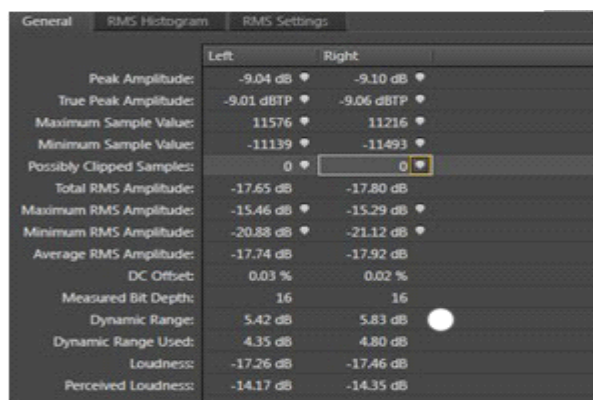
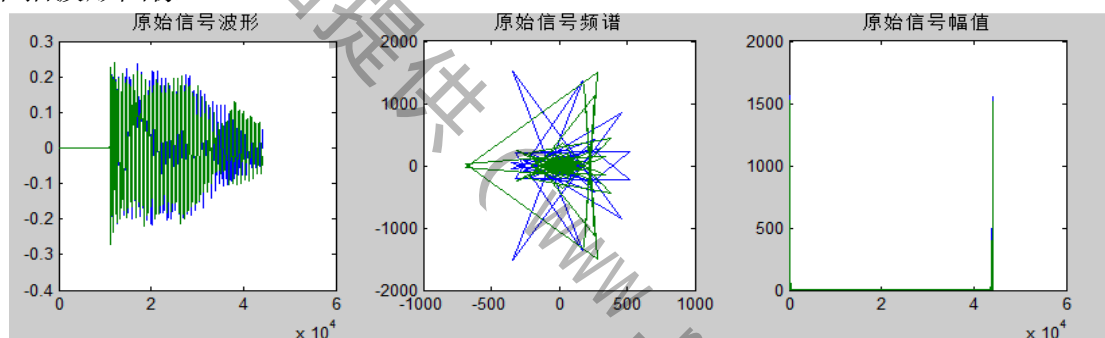


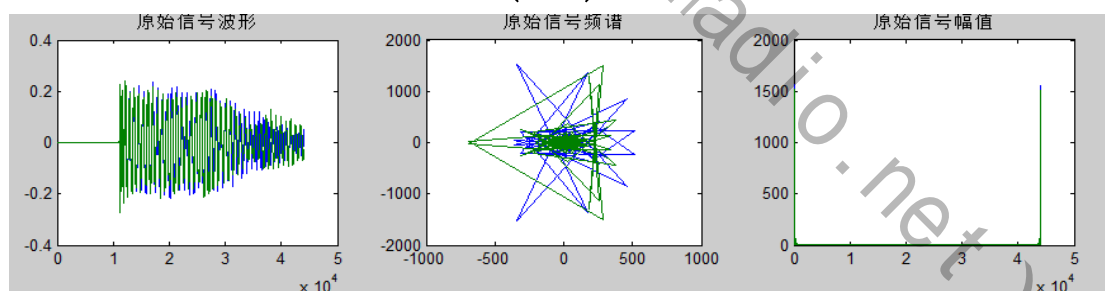
图1 H值

## 5.1.2数据的处理

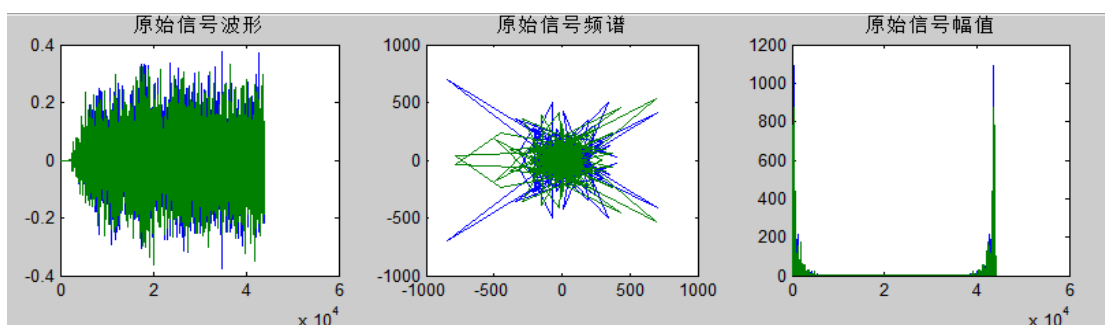
将mp3格式的音乐利用软件Audition转换成wav格式，同时用MATLAB将音频波图数字化，可以获得每一种音乐风格的音节数据，可以感性的观察相关差别，以下为各类音乐的节拍波形图像：



(blue)

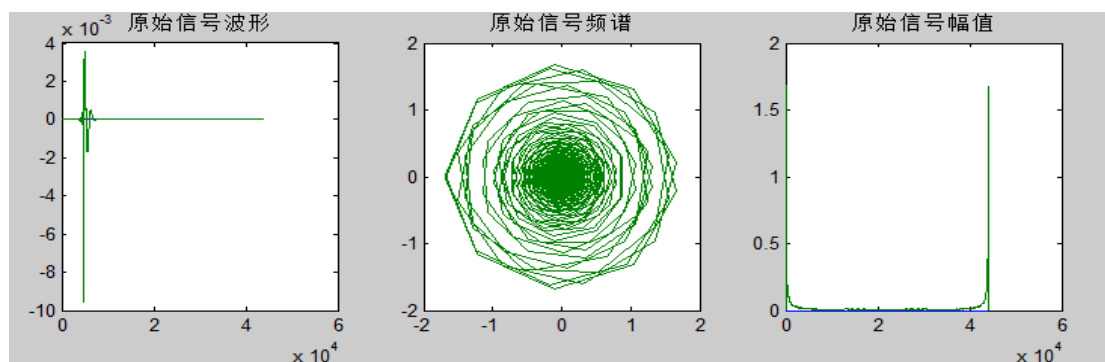


(classical)

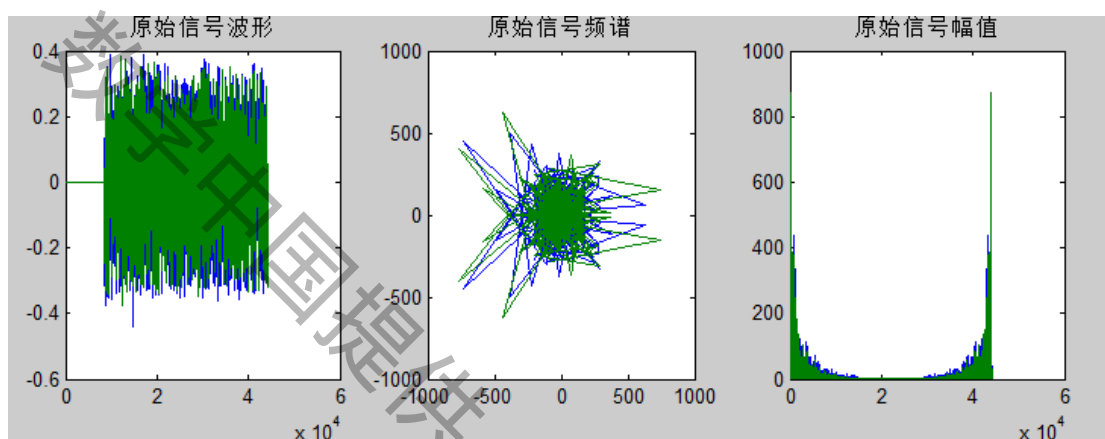


(country)

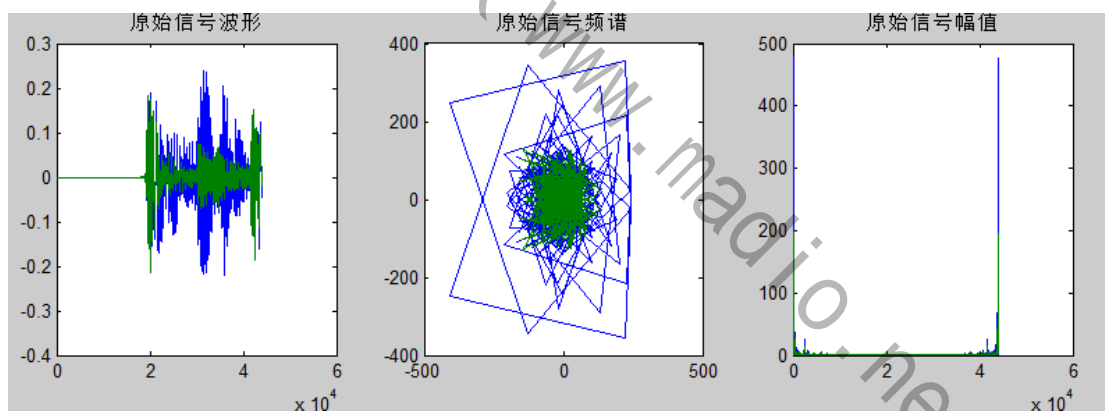
## 参赛队号#1129



(jazz)



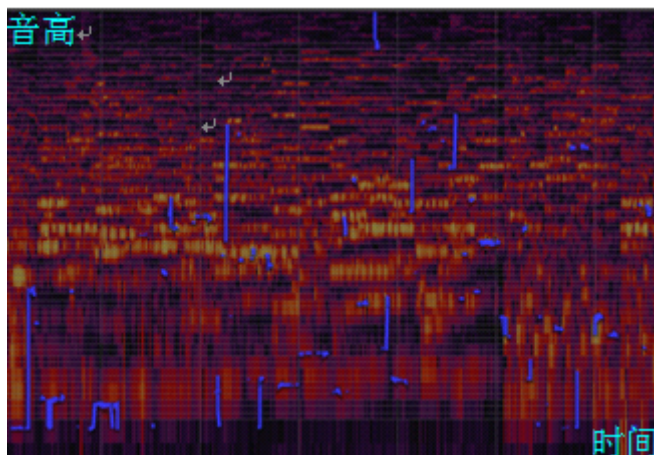
(New Age)



(rock)

观察数字化图像后，由于图形本身的随意性和不确定性，需要先利用其自相关函数值作为神经网络<sup>[6]</sup>的输入值，以获取不同风格音乐的特征值。

利用上文获取的音高数据，在Audition软件上形成不同风格音乐唯一的二维特征图，下图2为其中一种风格音乐的特征图：



图（2）音高二维特征图

对于一种风格音乐的二维特征值图有：

$$f(x, y) = \begin{cases} 1, (x, y) \in I \\ 0, (x, y) \notin I \end{cases}$$

定义：

$$g(u, v) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) f(x+u, y+v) dx dy$$

其中  $g(u, v)$  是  $f$  的自相关函数。几何意义上讲， $g(u, v)$  表示而为特征图的中频区域，得标准自相关函数：

$$G_0(u, v) = \frac{g(u, v)}{A} = \frac{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) f(x+u, y+v) dx dy}{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy}$$

$G_0(u, v)$  同时满足评议不变量和比例不变量的特点，但不具备旋转不变量的特征，为此改用极坐标  $(r, \theta)$  系统，并加以离散化：

$$\text{令：} \begin{cases} u = a\sqrt{A}\cos\theta \\ v = a\sqrt{A}\sin\theta \end{cases},$$

$$\text{得到：} G_0(a, \theta) = G_0(a\sqrt{A}\cos\theta, a\sqrt{A}\sin\theta) =$$

$$\frac{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) f(x+a\sqrt{A}\cos\theta, y+a\sqrt{A}\sin\theta) dx dy}{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy}$$

$$G_0(a, \theta + \pi) = G_0(a, \theta)$$

其中  $G_0(r, \theta)$  是以  $\pi$  为周期的周期函数， $\theta$  可取  $\theta \in [0, \theta]$ 。

$$\text{令} \begin{cases} a = a, i = 1, 2, \dots, M \\ \theta = \theta, \frac{(j-1)\pi}{N}, j = 1, 2, \dots, N \end{cases}$$

M 为径向离散量, N 为旋转角。可以得到可以被看做是音乐而为特征图的离散的  $G_0(r, \theta)$ , 音乐二维特征图的特征向量为:

$$F_c = [G(a_1, \theta_1), G(a_2, \theta_2), \dots, G(a_M, \theta_{N_1})]^T$$

基于上式, 根据音乐的二维特征图样本建立一个特征向量, 将全部样本按照 4 : 1 的比例随机地划分为训练集和测试集, 分别用于网络训练和效果测试的工作。

### 5.1.3 分类器

我们找了有关资料, 发现了一种分类器。其是基于音乐特征聚类分析的优势且在普通 BP 网的基础上发展的一种新型、有两个隐含层和非毗邻层连接的前馈神经网络结构。

新设计的非毗邻层连接神经网络的拓扑结构。通常的带有毗邻层连接的 BP 网络可以等价为一个串联连接的生物回路模型, 新设计的非毗邻层连接神经网络则可以等价为一个串联或并联连接的生物回路模型<sup>[7]</sup>, 可见下图 3:

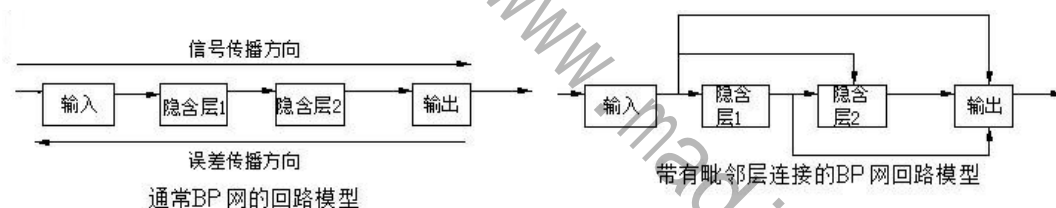


图 (3) 生物回路模型

由于这一新型网络的结构变化, 求和函数和 BP 算法的矫正参数将发生变化。设  $X_{Ki}$  代表 K 层上 i 节点的输入,  $Y_{Ki}$  为 K 层上 i 节点的输出,  $W_{KLij}$  代表 K 层上 i 节点和 L 层 j 节点间的连接权重,  $\theta_{Ki}$  为 K 层上 i 节点的阈值,  $N_L$  为 L 层上的节点数, 则有:

$$Y_{Ki} = X_{Ki}, \text{ for } k=1$$

$$Y_{Ki} = f\left(\sum_{L=1}^{K-1} \sum_{j=1}^{N_L} W_{KLij} Y_{Lj} + \theta_{Ki}\right), \text{ for } K > 1$$

$$f(x) = \frac{1}{1 + e^{-(x)}}$$

## 参赛队号#1129

误差反转计算和连接权重计算如下：

$$E_p = \frac{1}{2} \sum_{j=1}^{N_p} (\overline{Y_{tj}} - Y_{ti})^2$$

其中  $\overline{Y_{tj}}$  为第  $j$  层节点的输出的期望， $Y_{ti}$  为输出的真实值， $N_p$  为输出层节点数， $E_p$  为训练中各图形的积累误差。令：

$$\text{net} Y_{Kj} = \sum_{L=1}^{K-1} \sum_{i=1}^{N_L} W_{KLij} Y_{Li}$$

上式为  $K$  层节点的输入和函数，对于  $K$  层  $i$  节点，误差为：

$$e_{Kj} = -\frac{\partial E_p}{\partial \text{net} Y_{Kj}} = -\frac{\partial E_p}{\partial Y_{Kj}} \cdot \frac{\partial Y_{Kj}}{\partial \text{net} Y_{Kj}} = -\frac{\partial E_p}{\partial Y_{Kj}} \cdot f(\text{net} Y_{Kj})$$

$$\frac{\partial E_p}{\partial Y_{Kj}} = \sum_{L=K+1}^P \sum_{i=1}^{N_L} \frac{\partial E_p}{\partial \text{net} Y_{Lj}} \cdot \frac{\partial \text{net} Y_{Lj}}{\partial Y_{Kj}}$$

如果  $K=P$ ，对于输出层节点得：

$$e_{Pj} = (\overline{Y_{Pj}} - Y_{Pj})(\text{net} Y_{Pj})$$

如果  $K < P$ ，对于隐含层及该点得：

$$e_{Kj} = f(\text{net} Y_{Kj}) \sum_{L=K+1}^P \sum_{i=1}^{N_L} e_{Li} W_{LKij}$$

为减少误差函数  $E_p$ ，在最初连接权重的基础上，可以采用下降梯度法调整连接  $W_{KLij}$ 。

从上述讨论中可以看出，在带有毗邻层连接的 BP 网中，和函数极大的增加，同时  $f(x)$  迅速趋于稳定，学习过程被加速了。尽管这一网络中乘法和加法操作有所增加，但  $f(x)$  的计算量极大地减少，这一模型也更接近于生物的神经网络结构。实验显示了理论分析的正确性。与普通的 BP 网相比，这一带有毗邻层连接的 BP 网达到了令人满意的收敛速度和识别率。

#### 5.1.4 结果和讨论

采用乡村音乐、爵士、摇滚、布鲁斯、新世纪、古典音乐 6 种音乐流派，通过实验验证上述神经网络结构的可行性和优越性，按照下图的神经网络结构方案，分类器以每种流派各 50 首乐曲训练，另外 50 首乐曲样本进行测试。实验结果见下表 1：

## 参赛队号#1129

作品类型	布鲁斯	古典音乐	乡村音乐	爵士	新世纪	摇滚
正确识别率(%)	88.71	75.86	80.05	81.29	77.57	86.71

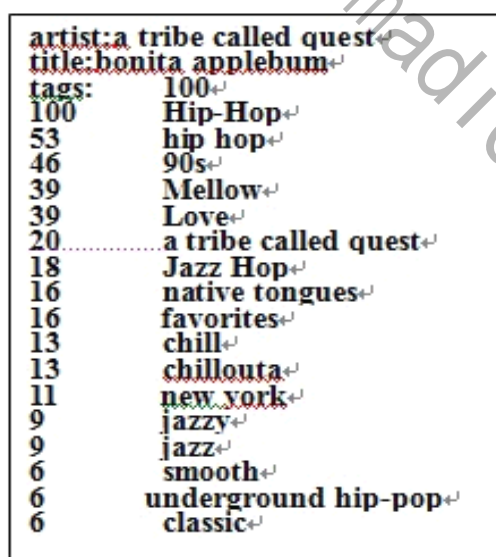
表（1）基于声学特点风格划分的实结果

## 5.2 标签特征

Geleijnse<sup>[8]</sup>等人提出 (<http://cn.last.fm>) 网站上的标签是进行音乐分类研究的可靠特征。每一个标签前面都有一个表示该标签使用频率的数字，模型中就把这些数字作为该标签在歌曲中出现的次数。

## 5.2.1 数据值获取

为了从音乐网站 Last.fm 上获取与上述音乐对应的标签，首先使用 jAudioTagger Library 从音乐文件包含的 ID3 标签中抽取音乐的标题和艺术家的姓名。然后以此作为输入值利用网站 Last.fm 提供的 API 接口 (Track.GetTopTags) 下载和获取音乐对应的标签。Last.fm 网站规定每首歌曲的标签数量范围在 1 到 100 之间，而且除了标签以外还可以得到每个标签被使用的频率。用上文下载的音乐中选取其中一首为例，歌曲的标题和艺术家的姓名分别是 a tribe called quest 和 bonita applebum。见下图 4，共有 100 个标签被用来标注这首歌曲，其中使用次数最多的标签就是 Hip-Hop，其频率为 100 次。



图（4）通过 Audioscrobbler 获取的标签

## 5.2.2 数据值处理

先利用反文档频率的方法<sup>[9]</sup>和LSA方法对取得的音乐标签进行预处理降维，定义集合  $\chi = \{x_1, x_2, \dots, x_n\}$  为歌曲的集合，对于每首歌曲  $x_i$  而言均有与之对应的标签集合  $T(x_i)$ 。定义集合  $C = \{c_1, c_2, \dots, c_k\}$  为音乐的流派，使得对  $\forall x_i \in \chi$  都有  $C(x_i) \in c$ 。基于标签的音乐流派分类就是要求得函数  $F$  从而建立由  $\chi \rightarrow c$  的映射。函数  $F$  可由下面的公式定义：

$$c_k = \arg \max_{c_k} \sum_{j=1}^l [Freq(t_j, x_i) \times pr(c_k | t_j)]$$

其中， $Freq(t_j, x_i)$  表示标签  $t_j$  被用来标注音乐  $x_i$  的次数， $t$  表示音乐  $x_i$  包含的标签的数目。

对得到的标签集  $T$  和歌曲集  $S$ ，我们建立一个  $M \times N$  的标签-歌曲矩阵  $A$ ， $A = \{a_{i,j}\}$ ，每个标签词对应着矩阵  $A$  的每一行，每一个歌曲则对应着  $A$  的每一列，其中  $a_{i,j}$  为非负值，表示第  $j$  个标签词在第  $i$  首歌曲中的权重。对于标签在歌曲中的权重值计算，使用词频  $\chi$  反文档频率的方法来表示标签和歌曲之间的关系，每首歌曲的权重值的计算如下：

$$\omega_{ts} = \begin{cases} (1 + \log_2 tf_{ts}) \log_2 \frac{N}{d_{fs}}, & tf_{ts} > 0, \\ 0, & tf_{ts} \leq 0, \end{cases}$$

其中  $\omega_{ts}$  为该标签在矩阵中的权重， $f_{ts}$  为歌曲  $s$  中标签  $t$  出现的次数， $d_{fs}$  为包含标签  $t$  的歌曲的数目， $N$  为歌曲的总数目。

我们对 100 首歌曲进行标签挖掘，得到 7082 个标签，将其表示成一个矩阵。但我们会发现得到的矩阵是一个松散矩阵，即存在一词多义和多词一义。运用LSA提取并化这些语义结构，消除同义词、多义词的影响，从而达到降维的目的。最后使用LSA对降维后的矩阵进行奇异值分解，提取出前  $k$  个最大的奇异值以及其对应的奇异矢量，构成一个  $k$  秩近似矩阵来近似表示原标签-歌曲矩阵。原始矩阵  $A$  经过奇异值分解可以表示成 3 个矩阵的乘积：

$$A \approx A_K = U \sum_k V_k^T$$

其中  $U_K^T U_K = V_K^T V_K = I_K$ ， $U_K$  和  $V_K$  的列分别被列为矩阵  $A_K$  的左右奇异向量， $\sum_k$  是对角矩阵，对角元素被称为的奇异值。 $U_K$  矩阵中的行向量对应原矩阵  $A$  的标

## 参赛队号#1129

签向量,  $V_K$  矩阵中的行向量则对应原矩阵  $A$  的歌曲向量, 标签与标签以及歌曲与歌曲之间的关系是通过整行之间的相关关系来获得的。矩阵的奇异值按照递减排列, 取前  $k$  个最大的奇异值构成了  $\sum_K$ ,  $k$  是降维后的维数。查的资料得出当  $k$  取得 100—300 就可以取得良好的结果, 所以我们取  $k=200$ 。

然后利用 LDA 方法对上述得到的矩阵进行建模。LDA 是文本建模的一种有效方法, 它属于生成模型, 它基于 bag of word 的前提。在本文的方法中将标签类比于文档中的词, 而将每首歌曲对应的标签向量类比于语料库中的文档。对于本文的 LDA 模型, 仅仅需要对主题的词汇分配, 也就是对变量进行抽样。对应的概率模型为:

$$p(\theta, z, w | \alpha, \beta) = p(\theta | \alpha) \prod_{n=1}^N p(z_n | \theta) p(w_n | z_n, \beta)$$

对上式计算边缘概率得到:

$$p(D | \alpha, \beta) = \prod_{d=1}^M \int p(\theta_d | \alpha) \left( \prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn} | \theta_d) p(w_{dn} | z_{dn}, \beta) \right) d\theta_d$$

其中,  $D$  代表一个语料库,  $M$  代表语料库中文档的总数, 而  $\beta$  是一个矩阵, 里面的元素代表了某个主题条件下生成某个单词的概率。而使用 LDA 对由标签所组成的语料库进行建模是 LDA 生成文档的逆过程, 训练集中的标签数据就构成了由多个文档所组成的语料库, 通过上述 LSA 方法的预处理, 可以求得矩阵  $\beta$ , 也就可以得到每个标签属于某个流派的概率, 用  $pr(c_k | t_j)$  来表示。

### 5.2.3 实验结果

将数据按照 4:1 的比例随机地划分为训练集和测试集, 按上述方法生成相应的特征向量, 对训练集使用上面提到的特征选择算法进行特征选择。最后使用分类器进行分类。结果如下表 2 所示:

音乐类型	布鲁斯	古典音乐	乡村音乐	爵士	新世纪	摇滚
正确识别率 (%)	88.57	90.71	87.86	90.86	89.29	88.75

表 (2) 通过标签获取的结果

### 5.3 旋律特征

每类乐曲中, 都会有几段相同的乐段, 它是整个音乐的主旋律, 是一类音乐的特征之处, 也就促成了音乐的分类, 我们识别这样的特征时需要考虑的是, 可能有的曲子中





## 参赛队号#1129

感，于是我们对各首曲子进行了一定简化，以便于匹配更多的片断。将相邻各音符之间音长后者比前者呈增大的现状标为p，而将减少的现状标为m，等音长的标为e，这样上图中的旋律可表示为：

$$(1/4, p), (1/4, p), (1/4, p), (1/4, p), (1, p), (1/4, p), (1/4, p), (1/4, p), \\ (1/4, p), (1, p), (1/4, m), (1/4, m), (1/4, p), (1/4, p), (2, e)$$

## 5.3.2 乐曲的数据挖掘

划分乐曲的风格是在两两音乐中进行比较，设其中的两种音乐分别为a和b对a和b乐曲相同乐段A和B取并集： $H = A \cup B$ 。其中的每个模式为：

$$K_i (0 < i \leq n, n = |H|)$$

计算其风格的趋向值 $f_i$ 。设 $r_{Ai}$ ， $r_{Bi}$ 分别是模式 $K_i$ 分别在a，b两类风格讨论中出现的频率。下面是一个经验公式，它可以保证在两种风格的讨论中出现频率相差较大的相同乐段，得到绝对值较高的分值：

$$f_i = \frac{r_{Ai}}{r_{Bi} + h} - k \frac{r_{Bi}}{r_{Ai} + h}$$

公式中的k也是一个经验常数，通常不为1，从而导致给分公式的不对称。其原因在于，不同风格乐曲对各自风格所拥有相同乐段的出现频率是不一样的，这是由特定风格乐曲本身的性质决定的，常数h则是一个正常数。

如果 $f_i$ 的绝对值小于阈值y，将模式 $K_i$ 从H中删除，这样做是为了剔除那些对识别风格没有什么作用的相同乐段，使得H精简化。

将H中各模式的风格趋向值排列组成向量a：

$$a = (f_1, \dots, f_m)' \quad (0 \leq i \leq m, m = |H|)$$

扫描待测风格的乐曲s，统计其中属于集合H的各个模式 $K_i$ 的出现次数 $t_i$ 。按照上文中同样的顺序将乐曲S中属于H的各相同乐段的出现次数排列组成列向量 $\beta$ ：

$$\beta = (t_1, \dots, t_m)'$$

待测乐曲s的风格分值： $F = (\alpha, \beta), (\alpha, \beta) = \sum_{i=1}^m f_i t_i$

当 $F > 0$ 时表示两乐曲属于一风格，反之则不属于一类。

## 5.3.3 结果与讨论

识别的正确率的计算公式是：

$$Ac = \frac{n}{m};$$

## 参赛队号#1129

其中 $m$ 代表测试集的曲目数量， $n$ 代表测试集中被正确识别的曲目数量。

为了验证乐曲的旋律片段、节奏各自在决定乐曲风格时所起的作用，我们选了6种2流行乐曲风格的其中15种组合的3种，从旋律片段以及节奏2方面的特征进行独立测试，比较它们指导风格识别的正确率。其结果如下表3所示：

	乡村--爵士	新世纪--摇滚	古典--节奏布鲁斯
旋律片段	76.85%	85.36%	87.23%
节奏	52.68%	61.55%	64.12%

表（3） 旋律片段、节奏

如上图表中所示，我们使用了2类：旋律片断、节奏。从它们指导乐曲风格识别的效果可以看出：旋律片断的效果较好，节奏的较差一些，对上述旋律片断、节奏进行加权处理。最终计算出每一种风格的正确识别率都会得到10个值，我们最后将采用10个值的平均值，将得到总共6个正确识别率。其结果如下表4所示：

作品类型	布鲁斯	古典音乐	乡村音乐	爵士	新世纪	摇滚
正确识别率%	83.23	77.67	84.86	80.53	74.74	88.53

表（4） 基于音乐旋律获取的结果

## 5.4 三种特征的加权结合

以  $p_A(c_i | x_{ij})$  表示使用特征  $i$  中求得的歌曲流派  $j$  数值  $x_{ij}$  占该特征所有流派总数值  $c_i$  的概率。则最终利用加权结合的方法求得的歌曲  $x_{ij}$  属于流派  $c_i$  的概率  $p(c_i | x_{ij})$  可由下面的公式计算得到：

$$p(c_i | x_{ij}) = x_{ij} / c_i$$

对上文计算结果进行统计整理得出下表 5：

布鲁斯	古典音乐	乡村音乐	爵士	新世纪	摇滚
88.71	75.86	80.05	81.29	77.57	86.71
88.57	80.71	87.86	90.86	89.29	88.75
83.23	77.67	84.86	80.53	74.74	88.53
260.51	234.24	252.77	252.68	241.6	263.99

表（5） 不同分类方法下的分类准确率

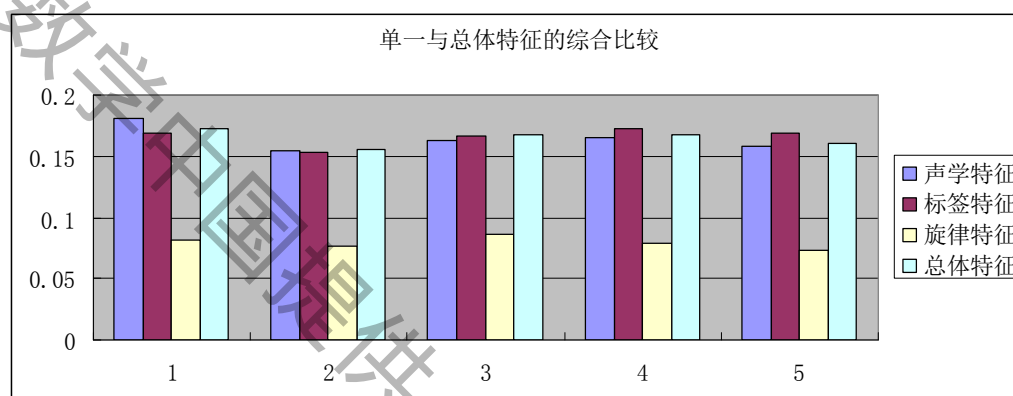
## 参赛队号#1129

我们将三个特征的结果进行权重的计算，其结果如下表 6 所示：

0.180971	0.154756	0.163304	0.165834	0.158245	0.176891
0.168371	0.153429	0.167022	0.172725	0.16974	0.168713
0.081901	0.07643	0.086457	0.079244	0.073546	0.087116
0.173006	0.15556	0.167865	0.167806	0.160447	0.175317

表（6）权重计算

我们通过转置，利用 Matlab 软件最终得到了分组条图 6：



图（6） 综合结果与单一结果的对比

从图 6 中可以很明显看出这样求得的综合分类准确率要高于使用单一方法求得的准确率。

## 六、模型评价与改进

### 1、模型评价

1. 为了对 6 种音乐风格进行分析和获得音乐风格的划分标准。首先考虑各个音乐流派的声学、标签和主旋律特征。本模型考虑各个音乐流派的声学、标签和主旋律特征。针对声学特征，先利用 Audition 软件处理音乐，然后运用神经网络结构给出误差反传训练算法的分类器，最后对分类器进行训练和音乐风格划分。针对标签特征，从音乐网站上抽取音乐标签，通过反文档频率和 LSA 方法对音乐——标签向量进行降维预处理，最后使用 LDA 模型处理降维后的向量，将标签属于每个流派的概率转换为对应的音乐属于每个流派的概率。最终将基于三种特点所求得的分类结果进行加权结合，得到音乐流派分类的最优解，可以发现分类的准确率较高，结果令人满意。

2. 针对音乐的声学特征描述，发展了前馈神经网络结构，给出了误差反传训练算法的分类器，并进行了实际操作。结果表明，前馈神经网络结构有优越的识别性能和极快

## 参赛队号#1129

的收敛速度，可以较好地满足音乐作品风格流派的实际的需要。

## 2、改进方向

1. 为了进一步提高分类的准确率，应该尝试使用音乐网站Last.fm中的个性化标签对音乐进行流派分类,应该利用更多的社会化资源对音乐流派分类进行划分研究。

2. 旋律特征在一定程度上实现了旋律的近似匹配，然而这种近似只是对节奏而言的，音符信息仍没有精确匹配。但具有相似风格的旋律，还有很大的距离。这也是本模型需要改进的方向。数据挖掘所产生的有效频繁模式数量较少，影响了识别正确率的提高。

## 七、参考文献

- [1] 甄超，郑涛，许洁萍，基于音乐语义信息的音乐流派分类研究，第五届全国信息检索学术会议论文集 40-123，2009-11-14
- [2] 韩敏，唐常杰，段磊，李川，巩杰，基于 TF-IDF 相似度的标签聚类方法，计算机科学与探索，240-246, 2010 年 3 月 6 日
- [3] 刘璇，许洁萍，陈捷，以 web 标签为基础的相识歌曲研究 《山东大学学报（理学版）》，第 47 卷 第五期：53-58，2012 年 5 月
- [4] 甄超，宋爽，许洁平，多模态音乐流派分类研究，计算机科学与探索 第 01 期：50-59，2011 年 5 月 1 日
- [5] 秦丹，马志光. 基于挖掘技术的音乐风格识别系统，计算机工程与设计，第 26 卷，第 11 期：3094-3096，2005. 11. 26
- [6] Setiono R, liu H. Neural-network feature selector[J]. IEEE Transactions on Neural Neural Networks, 1997, 8(3):654-662.
- [7] Kohavi R, Frasca B. Useful feature subsets and rough set reducts[C]//The Third International Workshop on Rough Sets and Computong, 1994:310-317.
- [8] GELEIJNSE G, SCHEDL M, KNEES P. The quest for ground truth in musical Artist tagging in the social web era[C]// proceedings of the 8th Internatal Conference on Music Information Retrieval (ISM IR' 07) .Vienna, Austria:

[s. n. ], 2007:525-530

- [9] 韩敏, 唐常杰, 段磊, 李川, 巩杰, 基于 TF-IDF 相似度的标签聚类方法, 计算机科学与探索, 240-246, 2010 年 3 月 6 日

## 八、附录

### 附录 1: 神经网络代码

```
%利用神经网络进行函数逼近
clear all
x=0:0.1*pi:4*pi;
y=sin(x);
%设定迭代次数
net.trainparam.epochs=10000;
%网络初始化
net=newff([0,4*pi],[8,8,8,8,1],{'tansig','logsig','logsig','tansig','tansig'
'});
%训练网络
[net,tr,y1,e]=train(net,x,y);
X=0:0.01*pi:4*pi;
%网络泛化
y2=sim(net,X);
subplot(2,1,2);
plot(X,y2);
title('网络产生')
grid on
subplot(2,1,1);
plot(x,y,'o');
title('原始数据')
grid on
%利用神经网络进行分类
clear all
```

## 参赛队号#1129

```

x=[1.24 1.36 1.38 1.38 1.38 1.40 1.48 1.54 1.56 1.14 1.18 1.20 1.26 1.28 1.30;...
    1.72 1.74 1.64 1.82 1.90 1.70 1.82 1.82 2.08 1.78 1.96 1.86 2.0 2.0 1.96];
y=[1 1 1 1 1 1 1 1 1 0 0 0 0 0 0;...
    0 0 0 0 0 0 0 0 0 1 1 1 1 1 1];
xmin1=min(x(1,:));
xmax1=max(x(1,:));
xmin2=min(x(2,:));
xmax2=max(x(2,:));
%设定迭代次数
net.trainparam.epochs=10000;
%网络初始化
net=newff([xmin1,xmax1;xmin2,xmax2],[8,8,8,2],{'tansig','logsig','logsig',
'tansig','tansig'});
%训练网络
[net,tr,y1,e]=train(net,x,y);

X=[1.24 1.28 1.40;...
    1.80 1.84 2.04];
%网络泛化
y2=sim(net,X)

plot(x(1,1:9),x(2,1:9),'o',x(1,10:15),x(2,10:15),'*',X(1,:),X(2,:), 'v')
grid on

```

## 附录 2：6 幅波形图的代码

```

[x,fs,bits]=wavread('11.wav',1*44100);
sound(x,fs,bits);
X=fft(x,1*44100);
magX=abs(X);
angX=angle(X);
subplot(131);plot(x);title('原始信号波形');
subplot(132);plot(X);title('原始信号频谱');
subplot(133);plot(magX);title('原始信号幅值');
figure
[x,fs,bits]=wavread('12.wav',1*44100);
sound(x,fs,bits);
X=fft(x,1*44100);
magX=abs(X);
angX=angle(X);
subplot(131);plot(x);title('原始信号波形');
subplot(132);plot(X);title('原始信号频谱');
subplot(133);plot(magX);title('原始信号幅值');

```

## 参赛队号#1129

```
figure
[x,fs,bits]=wavread('13.wav',1*44100);
sound(x,fs,bits);
X=fft(x,1*44100);
magX=abs(X);
angX=angle(X);
subplot(131);plot(x);title('原始信号波形');
subplot(132);plot(X); title('原始信号频谱');
subplot(133);plot(magX);title('原始信号幅值');
figure
[x,fs,bits]=wavread('14.wav',1*44100);
sound(x,fs,bits);
X=fft(x,1*44100);
magX=abs(X);
angX=angle(X);
subplot(131);plot(x);title('原始信号波形');
subplot(132);plot(X); title('原始信号频谱');
subplot(133);plot(magX);title('原始信号幅值');
figure
[x,fs,bits]=wavread('15.wav',1*44100);
sound(x,fs,bits);
X=fft(x,1*44100);
magX=abs(X);
angX=angle(X);
subplot(131);plot(x);title('原始信号波形');
subplot(132);plot(X); title('原始信号频谱');
subplot(133);plot(magX);title('原始信号幅值');
figure
[x,fs,bits]=wavread('16.wav',1*44100);
sound(x,fs,bits);
X=fft(x,1*44100);
magX=abs(X);
angX=angle(X);
subplot(131);plot(x);title('原始信号波形');
subplot(132);plot(X); title('原始信号频谱');
subplot(133);plot(magX);title('原始信号幅值');
>>
```



数学中国提供 (www.madio.net)