

参赛队号 #2183

## 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：**2183**

参赛队员（签名）：

队员 1：

队员 2：

队员 3：

参赛队教练员（签名）：

参赛队伍组别：本科组

参赛队号 #2183  
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）：

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 #2183

# 2013 年第六届“认证杯”数学中国 数学建模网络挑战赛

题 目 流行音乐时代分类的多模态模型关 键 词 神经网络 声学特征 标签特征 旋律特征 语义特征

## 摘 要：

不同时代的音乐特征具有鲜明的时代特色，由于社会文化的发展，流行音乐日趋多元化，造成类别混乱，难以划分等诸多不足。本文采用循序渐进、综合加权的方法对流行音乐的时代给出多模型优势互补的分类方法。

模型考虑各个时代音乐的声学、主旋律、标签和语义特征。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，最终对音乐进行时代划分。

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不同时代的音乐划分。

第二针对旋律特征，以 MIDI 乐曲为数据源提取出乐曲的旋律，然后对旋律数字化，通过经验公式对音乐进行旋律重合匹配，得出不同时代音乐的旋律，最后利用分类器进行不同时代的音乐划分。

第三针对标签特征，从 last.fm 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，将标签属于每个时代的概率转换为对应的音乐属于每个时代的概率。

第四针对语义特征，基于语义音乐分类模型，使用音乐名称和艺术家姓名这些与音乐有关的语义信息，通过搜索网络资源，计算音乐与不同音乐之间的联系紧密度并以此为依据进行音乐的时代分类，最后利用分类器进行不同时代的音乐划分。

最终将基于四种特点所求得分类结果进行加权结合，得到流行时代音乐分类的最优解。由于没有一种模型能应对所有情况，只有扩大其优点，针对性解决，优势互补才能达到理想的分类效果，本文在模型的优化中将多种模型综合，提出了多模态音乐时代分类方法，使分类更加自然、合理、准确。

参赛队号： 2183

所选题目： B 题

参赛密码 _____ (由组委会填写)
------------------------

参赛队号 #2183

**Abstract:**

The musical characteristics of the different eras with distinctive character, social and cultural development, pop music become more diversified, resulting in Category confusing, difficult to draw many deficiencies. In this paper, step-by-step method of weighted multi-model classification methods complement each other given the era of pop music.

The model takes into account each era music acoustics, the main theme, tags and semantic features. Audition software processing music data in the use of neural networks and other methods for data processing, the final era of music division.

First for the acoustic characteristics of the first use of the software to handle the music, and the use of neural network structure gives the error back propagation training Training algorithm classifier, last classifier training and music division of a different era.

For melody features, MIDI song data sources to extract the melody, the melody and then digitized, By the empirical formula of music melody coincidence match, come to the melody of a different era, the final classifier different era music division.

Third for the label features, from last.fm music sites using the Audioscrobbler extraction Music label tag vector by the inverse document frequency and LSA dimensionality reduction pretreatment, and finally use the vector the LDA model processing dimensionality reduction, the label belongs to every Age probability conversion for the music corresponding to the probability of belonging to every age.

Fourth semantic features, classification model based on semantic music, music title and artist name and music-related semantic information by searching network resources, calculate the tightness of the link between music and music and music as a basis for the classification of the times, the last classifier different era music division.

Will eventually be based on the classification results obtained by the four characteristics of the weighted combination of the optimal solution of the popular music of the times classification. The absence of a model able to cope with all situations, only to expand its advantages, targeted solutions complement each other to achieve the desired classification results in the optimization of the model will be a variety of models that incorporate multi-modal music era classification, more natural, reasonable and accurate classification.

参赛队号 #2183

## 目 录

一	问题重述 .....	1
二	问题分析 .....	1
2.1	问题分析 .....	1
2.2	基于时代的流行音乐分类 .....	1
2.3	模型框图 .....	2
三	模型假设 .....	3
四	变量与符号说明 .....	3
五	模型建立与求解 .....	4
5.1	声学特征 .....	4
5.1.1	数据的获取 .....	4
5.1.2	数据的处理 .....	4
5.1.3	分类器 .....	6
5.1.4	结果和讨论 .....	7
5.2	旋律特征 .....	7
5.2.1	数据获取与分析 .....	7
5.2.2	乐曲的数据挖掘 .....	8
5.2.3	结果与讨论 .....	9
5.3	标签特征 .....	9
5.3.1	数据值获取 .....	9
5.3.2	数据值处理 .....	9
5.3.3	实验结果 .....	10
5.4	语义特征 .....	11
5.4.1	原始数据采集 .....	11
5.4.2	互联网语义关系特征向量的表示 .....	13
5.4.3	数据分析和结论 .....	13
5.5	四种特征的加权结合 .....	13
六	模型评估与改进 .....	14
6.1	模型优点 .....	14
6.2	模型改进 .....	14
七	参考文献 .....	15
八	附录 .....	16
附录 A	神经网络 Matlab 代码 .....	16
附录 B	5 幅波形图 Matlab 代码 .....	16
附录 C	Wave 文件格式 .....	17

参赛队号 #2183

## 一 问题重述

近几十年来，流行音乐的时代经历了相当复杂的演变过程。每个时代都有其代表性的人物和时代，歌曲的成功要素也随时间不断改变。请你建立合理的数学模型，对曾经走红过的流行歌曲，通过其原唱录音或曲谱来推断其所处时代。并请你写一篇短文，向大众简述（在某个时期内）流行歌曲的“成功要素”是如何变化的。

## 二 问题分析

### 2.1 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时代标签。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，时代音乐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。

所建模型需要考虑各个时代音乐的声学、主旋律、标签和语义特征，用软件处理音乐获取数据，再利用神经网络等方法进行数据处理，最终对音乐进行时代划分。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不同时代的音乐划分。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，将标签属于每个时代的概率转换为对应的音乐属于每个时代的概率。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，得出不同时代音乐的旋律。基于语义音乐分类模型，使用音乐名称和艺术家姓名这些与音乐有关的语义信息，通过搜索网络资源，计算音乐与不同音乐之间的联系紧密度并以此为依据进行音乐的时代分类。最后利用分类器进行不同时代的音乐划分。

所建数学模型需要给出对流行音乐的时代给出一个自然、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。

### 2.2 基于时代的流行音乐分类

中国流行音乐的发展迄今已有近 80 年，现在所谓的“流行音乐”，可以追溯到上世纪 20、30 年代的大上海。当时有声电影的出现取代了无声片，电影插曲的需求量增加，从而刺激了流行歌曲的创作，出现了《秋水伊人》等佳作，开启了现代流行歌曲的先河。

由于不同文化背景和教育方式，人们对流行音乐的理解存在不同的看法。目前业界较为认可的是《中国大百科全书音乐舞蹈卷》中对流行音乐、通俗音乐的解释，“通俗音乐泛指一种通俗易懂、轻松活泼、易于流行、拥有广大听众的音乐，它有别于严肃音乐、古典音乐和传统的民间音乐，亦称流行音乐。”在本文中所指的流行音乐亦指流行歌曲。

中国流行音乐按照按时间划分大致分为五个发展时期。

20 世纪 20-40 年代的流行音乐。此时的流行音乐以半殖民地“十里洋场”的生活为题材，多以描写个人情感问题为主。创作手法主要以西方流行音乐为蓝本，不同时代的

## 参赛队号 #2183

音乐有民族化（民歌和戏曲素材）和中西结合两种趋势。由黎锦辉创作的《毛毛雨》是被音乐界公认的中国第一首流行歌曲。与其同一时期的刘雪庵、陈歌辛、姚敏、周旋等一批优秀的词、曲作家和歌唱演员创作和演绎的流行音乐作品将中国流行音乐推向第一个高峰。与此同时，在内地带有本土时代的音乐创作从 30-40 年代开始自发形成了，带有明显的地域色彩和乡土气息，很多音乐素材都是从当地的民歌，戏剧等演变而来的，这个时期，这个类型的音乐，也是主动的过程、自发的过程，但已经受到了解放区政治宣传需要的影响，比如《南泥湾》、《大生产》《游击队歌》等等作品。

20 世纪 50-70 年代的流行音乐。40 年代末期，共产党掌握了中国的政治主动权，在我国一批有深远政治战略目光的政治家的推动和影响下，文化宣传机构开始有计划，有目的的开始这种带有政治宣传色彩的文艺作品的创作，带有了更加鲜明的政治色彩。到了 66-76 年的十年间，中国的流行音乐已经没有真正意义上的流行因素，一切都是为了政治服务。

20 世纪 80 年代的流行音乐，80 年代初期，在“港台风”的带领下，中国大陆真正的流行歌曲迅速发展起来。起初，大陆歌手以模仿港台华语流行歌曲为主。80 年代中期，以“西北风”为基调的流行歌曲出现。如《信天游》、《黄土高坡》等。

20 世纪 90 年代出现流行音乐，则是多种时代并存，如《爱情鸟》《涛声依旧》等。在政治上的监管放松之后，1994 年，在港台流行歌曲依旧如火如荼时，大陆原创流行音乐也开始了一次井喷式的大爆发，宣告了内地流行音乐的复苏。这其中，以崔健、窦唯、何勇、张楚、黑豹乐队、唐朝乐队为代表的摇滚乐，以老狼、高晓松为代表的校园民谣，以陈琳为代表的都市情歌，以杨钰莹、毛宁为代表的青春偶像歌星，迅速成长为大陆流行乐坛的中坚力量，至今他们的影响力在 80 代人的心目中仍然难以磨灭。这个时期的中国流行音乐，虽然现在听起来显得老土，时代也很不统一，但是有个鲜明的特点，就是都具有明显的独立性，它不是源自任何时代的，听不到当时对于香港和台湾，日本的模仿，那是我们自己的声音。

21 世纪出现的流行音乐，中国流行音乐慢慢融入世界流行音乐繁荣主流大潮，流行音乐坛开始有一种‘中国风’的趋势，像京剧、昆剧等等都在歌曲中出现，其中有王力宏的《在梅边》、陈升的《北京一夜》等，并于国际接轨，发展空间变得更加广阔。

## 2.3 模型框图

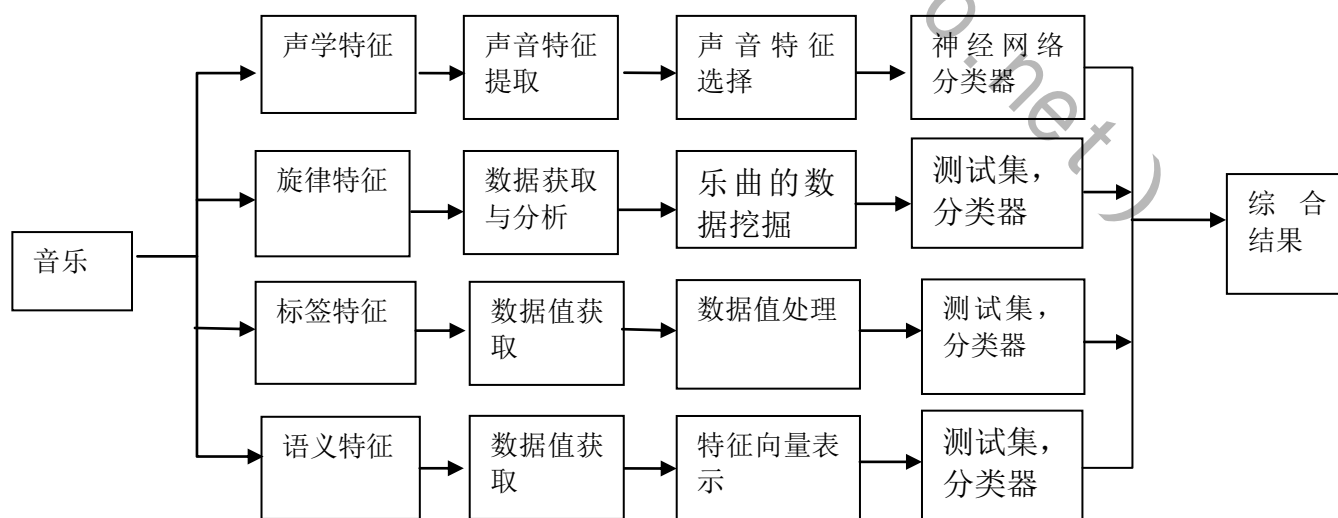


图 1 模型总体框图

参赛队号 #2183

### 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. 假设我们在音乐库中得到的音乐具有代表性；
3. 假设音波其中的音乐信息，没有进行一定的损坏；
4. 假设音符的长短对旋律的数值分析影响不大。
5. 假设每个音乐收听者对歌曲的选择只受歌曲类型与主观因素的影响。
6. 假设不同时代的音乐类型只受音乐本身的物理属性以及乐理属性的影响。
7. 假设不同时代的音乐分类只考虑本文中所考虑涉及的几个描述项。
8. 假设只考虑文章中所涉及的若干流行不同时代的音乐分类类型。
9. 假设对不同时代的音乐的分类不受乐器音质的影响。

### 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 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
$\bar{Y}_{tj}$	第 j 层节点的输出的期望
$Y_{ti}$	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
$f_i$	时代的趋向值
m	测试集的曲目数量
n	测试集中被正确识别的曲目数量
Ac	识别准确率
P	两个词语同时出现在一个网页中的概率
M(a, b)	词语 a 和词语 b 同时出现的网页的个数
C(a)	只有 a 出现的网页的个数
T	流行音乐的互联网语义关系特征向量
$\omega_{ts}$	标签在矩阵中的权重
$f_{ts}$	歌曲 s 中标签 t 出现的次数



参赛队号 #2183

$d_{fs}$	标签 $t$ 的歌曲数目
$N$	歌曲总数
$U_k$	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
$pr(c_k   t_j)$	每一个标签属于某个年代的概率

表 1 变量与符号说明

## 五 模型建立与求解

### 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不同时代的音乐。但相同的歌曲以不同速度或不同人的演唱会影响其时代分类，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： $LH_r = H - H_{\min}$ ， $L_r = \frac{L}{L_{\min}}$  其中  $H_{\min}$  为最低音， $L_{\min}$  为最短音节。

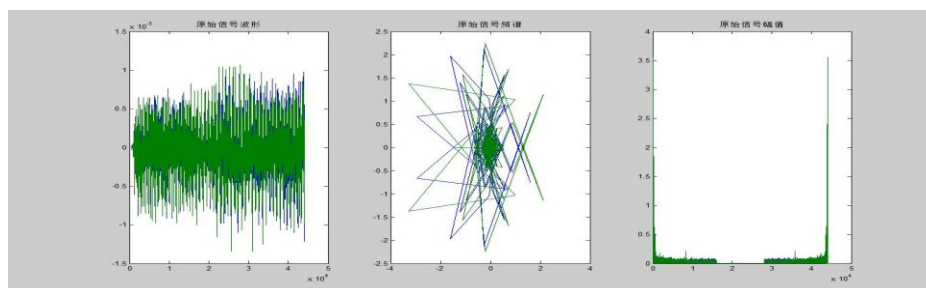
#### 5.1.1 数据的获取

本模型选取整首音乐作为处理对象，数据集共包括5个时代阶段流行音乐分类，即20-40年代的流行音乐，50-70年代的流行音乐，80年代的流行音乐，90年代的流行音乐，21世纪的流行音乐。

其中每个时代的流行音乐由100首音乐组成。这些音乐是根据网站mp3.com列出的每个时代的经典专辑和经典曲目列表在Last.fm上免费下载获得。然后利用软件Audition处理音乐获取音高的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。

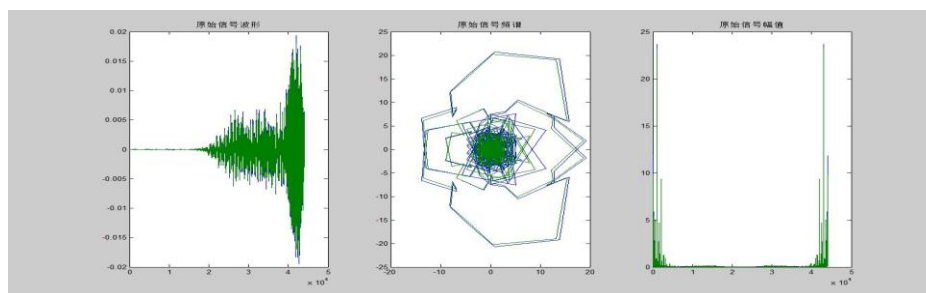
#### 5.1.2 数据的处理

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，可以获得每一种不同时代的音乐的音节数据，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：

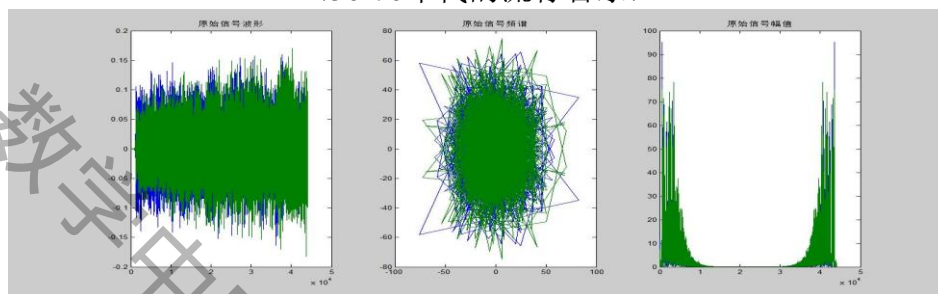


(20-40年代的流行音乐)

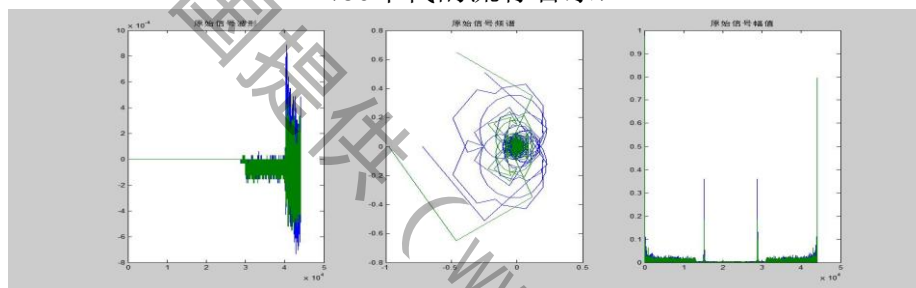
参赛队号 #2183



(50-70年代的流行音乐)



(80年代的流行音乐)



(90年代的流行音乐)

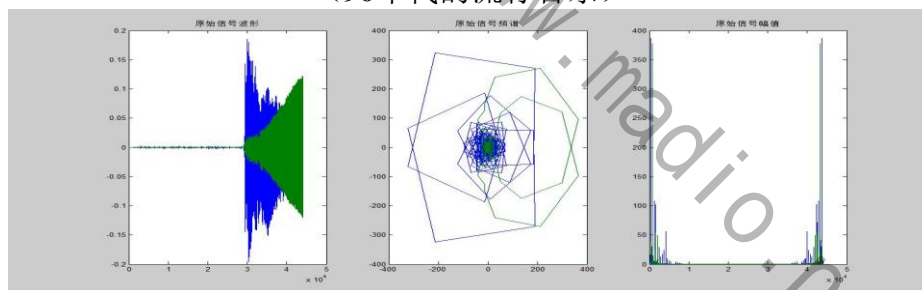


图2 (21世纪的流行音乐)

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的输入值，以获取不同时代音乐的特征值。

对于一种时代音乐的二维特征值有：
$$f(x, z) = \begin{cases} 1, (x, y) \in I \\ 0, (x, y) \notin I \end{cases}$$

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，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

参赛队号 #2183

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：

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

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  $\theta \in [0, \theta]$ 。

$$\text{令 } \begin{cases} a = a_i, i = 1, 2, \dots, M \\ \theta = \theta_j, \frac{(j-1)\pi}{N}, j = 1, 2, \dots, N \end{cases}, \text{ M 为径向离散量, N 为旋转角。可以得到音乐的二维特征}$$

值离散的  $G_0(r, \theta)$ ，音乐二维特征向量为：

$$F_c = [G(a_1, \theta_1), G(a_2, \theta_2), \dots, G(a_M, \theta_N)]^T$$

基于上式建立一个特征向量，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

我们采用一种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分类器。新设计的非毗邻层连接神经网络的拓扑结构，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，可见图 3：

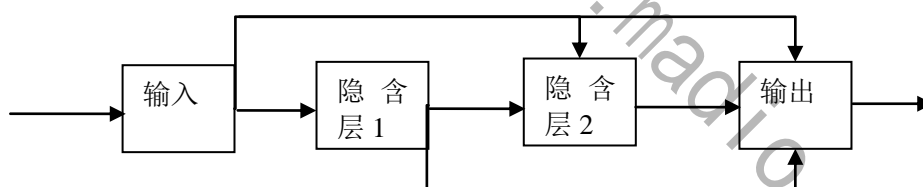


图3 生物回路模型

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：

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

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

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

误差反转计算和连接权重计算如下： $E_p = \frac{1}{2} \sum_{j=1}^{N_p} (\bar{Y}_{ij} - Y_{ii})^2$ ，其中  $\bar{Y}_{ij}$  为第 j 层节点的输出的期望， $Y_{ii}$  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：

参赛队号 #2183

$netY_{Kj} = \sum_{L=1}^{K-1} \sum_{i=1}^{N_L} W_{KLij} Y_{Li}$ ，上式为K层节点的输入和函数，对于K层i节点，误差为：

$$e_{Kj} = -\frac{\partial E_p}{\partial Y_{Kj}} \bullet \frac{\partial Y_{Kj}}{\partial netY_{Kj}} = -\frac{\partial E_p}{\partial Y_{Kj}} \bullet f'(netY_{Kj})$$

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

如果K=P, 对于输出层节点得：  $e_{Pj} = (\overline{Y_{Pj}} - Y_{Pj})(netY_{Pj})$

如果K<P, 对于隐含层及节点得：  $e_{Kj} = f'(netY_{Kj}) \sum_{L=K+1}^P \sum_{i=1}^{N_L} e_{Li} W_{LKij}$

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20-40年代，50-70年代，80年代，90年代和21世纪共5种时代音乐分类，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，分类器以每种时代各50首乐曲训练，另外50首乐曲样本进行测试。实验结果见下表二：

作品年代	20-40年代	50-70年代	80年代	90年代	21世纪
正确识别率	89.92%	85.42%	84.22%	73.67%	71.24%

表2 基于声学特点时代划分的实验结果

## 5.2 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，可能有的曲子中有很多不同的乐段在曲子中都有其对应的达到一定量的相同的乐段，所以我们需要考虑一定的阈值，判别哪段小型旋律是具有区别于其它时代时代的特点。

### 5.2.1 数据获取与分析

利用Audition软件将收集的五类歌曲转化为wave格式，采样率11.025kHz, 16 bit/sample。在特征提取之前，要对每一段音乐行预处理。

首先对信号进行预加重(参数为0.96)，以提升高频部分的能量，减小其能量的损耗；然后对每一首音乐进行分帧，帧长256点(约23 ms)，相邻帧之间有128点(50%)的重叠，每一帧都用汉明窗进行处理；这样得到了第一步的乐曲，再将其转换为MIDI格式的音乐，因为MIDI文件提供的信息是相当丰富的，包括了演奏的各种乐器及其所在的通道号。MIDI通过事件的方式记录音符的发音时间、音高、音强等。通过这些事件提供的信息，就能比较准确地提取乐曲的旋律。

稍事对乐曲的纠正，以使其内部乐器、音色、音域基本保持不变同时还能保留相对

## 参赛队号 #2183

丰富的音乐信息。曲子的旋律主要是由各个不同类型的音符以一定的顺序由高到低，由左到右依次在五线谱中排列，所以这里面总共有两个指标：音符本身的音节长短和许多音符共同组成的音序的框架。

对于音符本身而言，音节长短不一，有四分音符，八分音符，十六分音符，以及符点二分音符，全音符等等，它们一共占几拍，就按照其所占节拍来计量，

例如：

四分音符-----1  
八分音符-----1/2  
十六分音符-----1/4  
符点二分音符----5/2  
全音符-----4

由于还有升调和降调，所以升降调也算为半个音。例如，在钢琴中，相邻白键为一个全音，一个黑键为一个半音。

对于许多音符共同组成的音序，在这里，我们不能把它看做是单个的因素，由于有了共同的集合才有旋律，因而需要将其结合在一起，本文将相邻的音符联系在一起，以前者占后者的分数为值，若前者为双音，则同理可得，前者两个音符相加，与后者做一差值。

将上述两者结合形成一个函数 $f(x_i, y_i)$ ,  $i=0, 1, 2, \dots$  下图5为其中一首歌曲的五线谱片段：

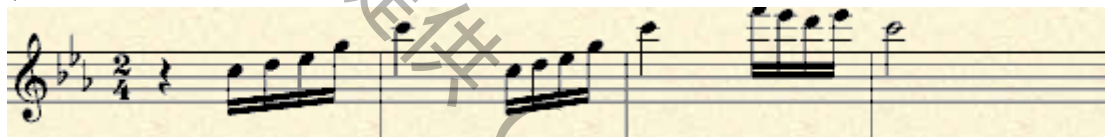


图4 旋律片段

上述的旋律的表达可变为：

$(1/4, 1), (1/4, 1), (1/4, 2), (1/4, 3), (1, 7), (1/4, 1), (1/4, 1), (1/4, 2),$   
 $(1/4, 3), (1, 3), (1/4, -1), (1/4, -1), (1/4, 1), (1/4, 2), (2, 0)$

由于人们平时听到音乐之后，对音乐中音符的音长不如对每一个音符的音高那么敏感，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：

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

### 5.2.2 乐曲的数据挖掘

划分乐曲的时代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： $K_i (0 < i \leq n, n = |H|)$  计算其时代的趋向值 $f_i$ 。设 $r_{Ai}, r_{Bi}$  分别是模式 $K_i$  分别在 $a, b$ 两类时代讨论中出现的频率。下面是一个经验公式，它可以保证在两种时代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，不同时代乐曲对各自时代所拥有相同乐段的出现频率是不一样的，这是由特定时代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时代没有什么作用的相同乐段，使得 $H$ 精简化。

参赛队号 #2183

将H中各模式的时代趋向值排列组成向量 $\alpha: \alpha = (f_1, \dots, f_m)'$  ( $0 < i \leq m, m = |H|$ ) 扫描待测时代的乐曲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的时代分值:

$$F = (\alpha, \beta), (\alpha, \beta) = \sum_{i=1}^m f_i t_i, \text{ 当 } F > 0 \text{ 时表示两乐曲属于一时代, 反之则不属于一类。}$$

### 5.2.3 结果与讨论

识别的正确率的计算公式是:  $Ac = \frac{n}{m}$

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。

从旋律片断、节奏指导乐曲时代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, 对旋律片断、节奏进行加权处理。最终计算出每一种时代时代的正确识别率都会得到8个值, 我们最后将采用8个值的平均值, 将得到总共五个正确识别率。其结果如下表所示:

作品年代	20-40年代	50-70年代	80年代	90年代	21世纪
正确识别率	85.28%	86.21%	79.29%	74.38%	70.16%

表3 基于旋律特点时代划分的实验结果

## 5.3 标签特征

Geleijnse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, 每一个标签里面都有一个表示该首音乐的创作年代的信息。

### 5.3.1 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。

```
<toptags artist=" B.B.King " track="Three O' clock Blues">
  <tag>
    <name>blues</name>
    <count>100</count>
    <url>www.last.fm/tag/blues </url>
  </tag>
  <tag>
    <name>blues rock</name>
    <count>20</count>
    <url>www.last.fm/tag/blues rouck</url>
  </tag>
  ...
</toptags>
```

图5 Track.GetTopTags

### 5.3.2 数据值处理

先利用反文档频率的方法和LSA 方法对取得的音乐标签进行预处理降维, 定义集合

## 参赛队号 #2183

$\chi = \{x_1, x_2, \dots, x_n\}$  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  $C = \{c_1, c_2, \dots, c_k\}$  为音乐的年代, 使得对  $\forall x_i \in \chi$  都有  $C(x_i) \in c$ 。基于标签的音乐年代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:

$c_k = \arg \max_{c_k} \sum_{j=1}^t [Freq(t_j, x_i) \times pr(c_k | t_j)]$ ,  $Freq(t_j, x_i)$  其中, 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  $i$  个标签词在第  $j$  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:

$$\omega_{ts} = \begin{cases} (1 + \log_2 f_{ts}) \log_2 \frac{N}{d_{ts}}, & f_{ts} > 0; \\ 0, & f_{ts} \leq 0; \end{cases} \quad \text{其中 } \omega_{ts} \text{ 为该标签在矩阵中的权重, } f_{ts} \text{ 为歌曲 } s \text{ 中}$$

标签  $t$  出现的次数,  $d_{ts}$  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9028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。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:

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

$$\text{对上式计算边缘概率得到: } p(D | \alpha, \beta) = \prod_{d=1}^M \int p(\theta_d | \alpha) \left( \prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn} | \theta_d) p(w_{dn} | z_{dn}, \beta) \right) d\theta_d$$

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年代的概率, 用  $pr(c_k | t_j)$  来表示。

### 5.3.3 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



参赛队号 #2183

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:

作品年代	20-40年代	50-70年代	80年代	90年代	21世纪
正确识别率	90.16%	88.54%	85.31%	82.13%	83.47%

表4 基于标签特点时代划分的实验结果

## 5.4 语义特征

我们提出的基于互联网语义关系来对流行不同时代的音乐进行分类是借鉴了 C. McKay 等人在文章中提出的音乐的文化特征和甄超等人提出的“语义特征”这一系列概念。这类关系包含了相当松散的语义关联性, 但是它对于本次模型的建立仍然是有用的, 我们可以使用语义网络来呈现逻辑上的描述。当我们在网络上抓取数据的时候, 如果一段流行音乐总与某个特定的时代出现在一起, 那么我们就认为此流行音乐与这个时代有着非常紧密的联系, 从而我们就可以推断该流行音乐就是属于这个时代的。

### 5.4.1 原始数据采集

#### a) 关系紧密度

在我们的实验中对于每一首歌曲我们都要通过搜索网络资源计算歌曲名称与各个时代的联系紧密度以及音乐家与各个时代的联系紧密度。在这里, 联系紧密度是我们通过 Google 搜索引擎抓取互联网上相关资源, 然后计算两个词语同时出现在一个网页中的概率而得到, 计算的公式是:  $P = M(a, b) / C(a)$

其中,  $M(a, b)$  代表  $a$  和  $b$  同时出现的网页的个数,  $C(a)$  代表只有  $a$  出现的网页的个数。Google 搜索表达式分别是 “ $a$ ” AND “ $b$ ” 和 “ $a$ ”, 并且采用 Google 搜索工具的精确搜索以提高准确性。

#### b) 训练集数据采集

下面举例说明, 我们从流行音乐训练集当中随机抽取的五个不同时代的流行音乐, 其中音乐与音乐家的对应关系如下表所示。

音乐名称	音乐家
夜上海	周璇
我的祖国	郭兰英
冬天里的一把火	费翔
涛声依旧	毛宁
清明雨上	许嵩

表5 音乐名称与对应的音乐家

根据上述公式定义  $a$ =音乐名称,  $b$ =时代音乐, 分别求出  $M(a, b)$  和  $C(a)$ 。如表5.2所示。其中

20-40年代= (1920-1949);  
 50-70年代= (1950-1979);  
 80年代 = (1980-1989);  
 90年代 = (1990-1999);  
 21世纪 = (2000至今);



参赛队号 #2183

	夜上海	我的祖国	冬天里的一把火	涛声依旧	清明雨上
20-40 年代	1009080	85672400	971514000	7249821700	732648700
50-70 年代	372500	123781100	845456200	9386892000	9248987400
80 年代	361200	73589800	1932745100	7823398000	6763482000
90 年代	345000	62378900	525567700	13424674500	9376498000
21 世纪	212100	99454800	835566700	7632411000	14143871200
单独	1609800	434877000	5186238600	30162387000	39265487300

表6 不同音乐名称与不同年代同时出现和单独出现的网页条数

根据公式定义a=音乐名称，b=音乐年代，以及表五的数据，求出 $P=M(a, b) / C(a)$ 。  
如下表所示：

	夜上海	我的祖国	冬天里的一把火	涛声依旧	清明雨上
20-40 年代	0.62684	0.1970038	0.187325358	0.24035968	0.018658846
50-70 年代	0.2314	0.2846347	0.163019148	0.311211841	0.235550048
80 年代	0.22438	0.1692198	0.372667987	0.259375957	0.172250046
90 年代	0.21431	0.1434403	0.101338897	0.445079977	0.238797444
21 世纪	0.13176	0.2286964	0.161112275	0.253043998	0.360211274

表7 参数a和b分别为音乐名和时代时的概率P

根据公式定义a=音乐家，b=音乐年代，分别求出 $M(a, b)$ ， $C(a)$ ，如下表所示：

	周璇	郭兰英	费翔	毛宁	许嵩
20-40 年代	9023000	322400	1514000	9821700	6487000
50-70 年代	7123500	126231100	5456200	8892000	9874000
80 年代	6423000	78224500	5932745100	3980000	34820000
90 年代	2132000	23378900	325567700	23424674500	98230000
21 世纪	2213200	29454800	435566700	9613242400	11143871200
单独	25914700	249611700	6640849700	33060510600	11293282200

表8 不同音乐家与不同年代同时出现和单独出现的网页条数

根据公式定义a=音乐家，b=音乐年代，以及表七的数据，求出 $P=M(a, b) / C(a)$ 。  
如下表所示：

	周璇	郭兰英	费翔	毛宁	许嵩
20-40 年代	0.348181	0.0012916	0.000227983	0.000297083	0.000574412
50-70 年代	0.274883	0.5057099	0.000821612	0.000268961	0.000874325
80 年代	0.247852	0.3133847	0.893371386	0.000120385	0.003083249
90 年代	0.08227	0.0936611	0.049025007	0.708539405	0.008698091
21 世纪	0.085403	0.1180025	0.065589001	0.290777191	0.986769922

表9 参数a和b分别为音乐家和时代时的概率P

参赛队号 #2183

### 5.4.2 互联网语义关系特征向量的表示

当计算出流行音乐名称、音乐家和流行时代的联系紧密度之后，我们就可以定义流行音乐的语义特征了。例如，对于某段音乐的名称与流行时代的联系紧密度用向量 $X$ 表示，其音乐家与流行时代的联系紧密度用向量 $Y$ 表示。那么流行音乐的互联网语义关系特征向量将根据下面的公式计算得到： $T = m \times X + n \times Y$ （其中 $m+n=1$ ）

### 5.4.3 数据分析和结论

经过几十次的测试，我们得出的实验数据表明，当 $m=0.39$ ， $n=0.61$ 时对流行音乐的时代分类的准确性最高，可以达到83.2%。说明音乐家和流行时代的紧密度 $Y$ 的权重更大，互联网语义关系特征越明显，而流行音乐名称与流行时代之间的互联网语义关系特征稍弱。

作品年代	20-40年代	50-70年代	80年代	90年代	21世纪
正确识别率	82.45%	86.37%	78.66%	80.21%	81.31%

表10 基于标签特点时代划分的实验结果

基于互联网语义关系来对流行音乐进行时代分类是非常好的一个研究方向，实验结果表现较好，但仍需要用大量数据进行佐证。本模型可研究的潜力非常之大，它不仅与语义网络紧密相连，同时又和数据挖掘技术密切相关，这无疑为我们在以后的研究中提供了一个新的方向。

## 5.5 四种特征的加权结合

以 $p_A(c_i | x_{ij})$ 表示使用特征 $i$ 中求得的歌曲年代 $j$ 数值 $x_{ij}$ 占该特征所有年代总数值 $c_i$ 的概率。则最终利用加权结合的方法求得的歌曲 $x_{ij}$ 属于年代 $c_i$ 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：

作品年代 分类特征	20-40年代	50-70年代	80年代	90年代	21世纪
声学特征	89.92%	85.42%	84.22%	73.67%	71.24%
旋律特征	85.28%	86.21%	79.29%	74.38%	70.16%
标签特征	90.16%	88.54%	85.31%	82.13%	83.47%
语义特征	82.45%	86.37%	78.66%	80.21%	81.31%

表 11 四种分类结果统计

将四种分类的结果进行权重的计算，其结果如下表所示：

权重	20-40 年代	50-70 年代	80 年代	90 年代	21 世纪
声学特征	0.2223	0.2112	0.2082	0.1821	0.1761
旋律特征	0.2157	0.2181	0.2006	0.1882	0.1775
标签特征	0.2099	0.2061	0.1986	0.1912	0.1943
语义特征	0.2016	0.2112	0.1923	0.1961	0.1988
总体特征	0.2123	0.2115	0.1999	0.1894	0.1869

表 12 分类权重

参赛队号 #2183

利用 Excel 软件最终得到了分组条形图如下：

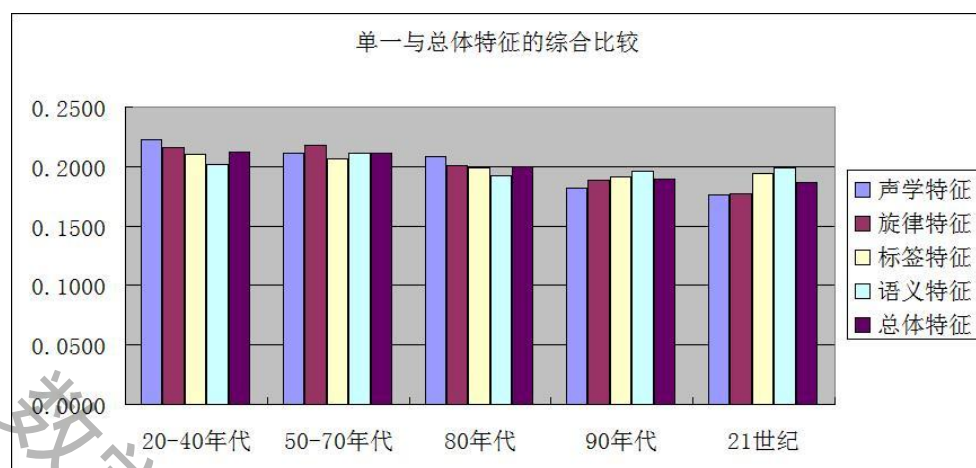


图6 分组条形图

从图中可以看出，综合分类准确率要高于使用单一方法求得的准确率。

## 六 模型评估与改进

### 6.1 模型优点

1. 为了对五种时代音乐进行分析和获得时代音乐的划分标准。首先考虑各个音乐年代的声学、主旋律、标签和语义特征。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，将标签属于每个年代的概率转换为对应的音乐属于每个年代的概率。最终将基于四种特点所求得的结果进行分类的准确率较高，结果令人满意。

2. 本文借鉴语音信号处理技术和语音识别技术结合了神经网络算法提出了一种快速的音乐分类方法，并借助 Mat lab 平台进行了实验。利用统计规律提取音频特征向量，并可以对未知分类的音频利用神经网络算法进行快速的分类。

### 6.2 模型改进

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年代分类，应该利用更多的社会化资源对音乐年代分类进行划分研究。

2. 很多地方用到一些参数，这些参数的选择往往是经验性的，必然会有局限性，如果分类统计的规模和应用环境等发生变化，参数的选择是否也应该变化呢？能否从理论上找到更好的方法、或实现一种自动选择的算法呢？这有待我们去进一步研究。

3. 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。但具有相似时代的旋律，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。

参赛队号 #2183

## 七 参考文献

【01】姚斯强, 胡剑凌, 一种基于线性判别分析和支持向量机的音乐分类方法, 《电声技术》, 卷期号: 2006 年第12 期

【02】文杰, 基于SVM—HMM混合模型的音乐分类研究, 中山大学硕士学位论文 卷期号: 2005年5月

【03】牛滨, 孔令志, 罗森林, 潘丽敏, 郭亮, 基于MFCC和GMM的个性音乐推荐模型, 北京理工大学学报, 卷期号: 第29卷第4期  
2009年4月

【04】卢坚, 陈毅松, 孙正兴, 基于隐马尔可夫模型的音频自动分类[J]. 软件学报, 卷期号: 2002, 8( 13): 1 593 1 597.

【05】张燕, 唐振民, 李燕萍, 邹益, 基于MFCC和HMM的音乐分类方法研究, 南京师范大学学报(工程技术版)卷期号: 第8卷第4期2008年12月

【06】项慨, 基于频谱对比度特征的不同时代的音乐分类, Microcomputer Applications Vol. 21, No. 3, 2005

【07】马希荣, 梁景莲, 基于情感音乐模板的音乐检索系统研究, 计算机科学, 卷期号: 第36 卷 第1 期2009 年1 月

【08】秦丹, 马光志, 基于挖掘技术的不同时代的音乐识别系统, 计算机工程与设计 卷期号: 第26卷第11期计2005 年 11月

【09】金毅, 黄敏, 基于旋律的音乐检索, 情报学报卷期号: 2003 (03)

【10】刘丹. 张乃尧. 朱汉城音乐特征识别的研究综述[期刊论文]-计算机工程与应用 2002 (24)

【11】Kuo Fang Fei. Shan Man kwan A personalized music filtering system based on melody styleclassification 2002

【12】洪军, 倪宏, 邓浩江, 严勤, 一种基于Web 挖掘的时代音乐分类方法, 《微计算机信息》(管控一体化)卷期号: 2009 年第 25 卷第 9-3 期 文章编号: 1008-0570(2009)09-3-0168-02

【13】刘怡, 蔚磊, 刘子利, 王鹏, 中国民歌地域时代分类中的特征选择, 电子学报, 卷期号: 第12A期2008年12月

【14】HU Jingkai, WU Lei, GAO Yang. MP3 music classification method based on LCS[J], Journal of Chongqing University of Posts and Telecommunications, 2007, Vol. 19 No. 4

【15】周开利, 康耀红. 神经网络模型及其Matlab仿真程序设计[M], 清华大学出版社, 2005

【16】Martin T. Hagan, Howard B. Demuth, Mark H. Beale, Neural Network Design[M], PWS Publishing Company, 1995

【17】陈峰, Blackfin系列DSP原理与系统设计[M], 电子工业出版社, 2004

【18】赵力. 语音信号处理[M], 机械工业出版社, 2005

【19】韩纪庆, 张磊, 郑轶然. 语音信号处理[M], 清华大学出版社, 2004

【20】张一彬, 周杰, 边肇祺基于样本的流行歌曲关键段分割方法, 电子学报 2006 (02)

【21】Liu Dan;Lu Lie;Zhang Hong-Jiang Automatic mood detection acoustic music data 2003

【22】Lu Lie;Liu Dan;Zhong Hong-Jiang Automatic mood detection and tracking of music audio signals2006(01)

参赛队号 #2183

## 八 附录

### 附录A 神经网络 Matlab 代码

其中  $P$  为标准音乐分类下的参数统计， $x$  为测试数据。

```
[Pn,minp,maxp,Tn,mint,maxt]=premnmx(P,T)
net=newff(minmax(Pn),[10,1],{'tansig','purelin'},'traingdm');
% 当前输入层权值和阈值
inputWeights=net.IW{1,1};
inputbias=net.b{1};
% 当前网络层权值和阈值
layerWeights=net.LW{2,1};
layerbias=net.b{2};
% 设置训练参数
net.trainParam.show = 50;
net.trainParam.lr = 0.05;
net.trainParam.mc = 0.9;
net.trainParam.epochs = 10000;
net.trainParam.goal = 1e-3;
% 调用 TRAINGDM 算法训练 BP 网络
[net,tr]=train(net,Pn,Tn);
% 对 BP 网络进行仿真
A = sim(net,Pn);
% 计算仿真误差
E = Tn - A;
MSE=mse(E);
x=[. . .];%测试
[Xn]=tramnmx(x,minp,maxp)
sim(net,Xn)
```

### 附录B 5幅波形图Matlab代码

```
[x,fs,bits]=wavread('ysh.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('wdzg.wav',1*44100);
sound(x,fs,bits);
X=fft(x,1*44100);
magX=abs(X);
angX=angle(X);
```

## 参赛队号 #2183

```
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('dtldybh.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('tsyj.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('qmys.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');
```

## 附录C Wave文件格式

WAVE文件作为多媒体中使用的声波文件格式之一，它是以RIFF格式为标准的。RIFF是英文Resource Interchange File Format的缩写，每个WAVE文件的头四个字节便是“RIFF”。WAVE文件由文件头和数据体两大部分组成。其中文件头又分为RIFF / WAV文件标识段和声音数据格式说明段两部分。WAVE文件各部分内容及格式见附表。

常见的声音文件主要有两种，分别对应于单声道(11.025KHz采样率、8Bit的采样值)和双声道(44.1KHz采样率、16Bit的采样值)。采样率是指：声音信号在“模—数”转换过程中单位时间内采样的次数。采样值是指每一次采样周期内声音模拟信号的积分值。

对于单声道声音文件，采样数据为八位的短整数(short int 00H, FFH)；而对于双声道立体声声音文件，每次采样数据为一个16位的整数(int)，高八位和低八位分别代表左右两个声道。

WAVE文件数据块包含以脉冲编码调制(PCM)格式表示的样本。WAVE文件是由样本组织而成的。在单声道WAVE文件中，声道0代表左声道，声道1代表右声道。在多声道WAVE文件中，样本是交替出现的。

WAVE 文件格式说明表：

	偏移地址	字节	数据类型	内容
--	------	----	------	----

## 参赛队号 #2183

文 件 头	00H	4	char	“RIFF” 标志
	04H	4	Long int	文件长度
	08H	4	char	“WAVE” 标志
	0CH	4	char	“fmt” 标志
	10H	4		过渡字节(不定)
	14H	2	int	格式类别(10H为P嘯形式的声音数据)
	16H	2	int	通道数, 单声道为1, 双声道为2
	18H	2	int	采样率(每秒样本数), 表示每个通道的播放速度,
	1CH	4	Long int	波形音频数据传送速率, 其值为通道数*每秒数据位数*每样本的数据位数 / 8。播放软件利用此值可以估计缓冲区的大小。
	20H	2	int	数据块的调整数(按字节算的), 其值为通道数*每样本的数据位值 / 8。播放软件需要一次处理多个该值大小的字节数据, 以便将其值用于缓冲区的调整。
	22H	2		每样本的数据位数, 表示每个声道中各个样本的数据位数。如果有多个声道, 对每个声道而言, 样本大小都一样。
	24H	4	char	数据标记符 “data”
	28H	4	Long int	语音数据的长度

PCM 数据的存放方式:

	样本 1		样本 2	
8 位单声道	0 声道		0 声道	
8 位立体声	0 声道(左)	1 声道(右)	8 位立体声	0 声道(左)
16 位单声道	0 声道低字节	0 声道高字节	16 位单声道	0 声道低字节
16 位立体声	0 声道(左)低字节	0 声道(左)高字节	16 位立体声	0 声道(左)低字节

WAVE文件的每个样本值包含在一个整数*i*中,*i*的长度为容纳指定样本长度所需的最小字节数。首先存储低有效字节, 表示样本幅度的位放在*i*的高有效位上, 剩下的位置为0, 这样8位和16位的PCM波形样本的数据格式如下所示:

样本大小	数据格式	最大值	最小值
8位PCM	unsigned int	225	0
16位PCM	int	32767	-32767