参赛队号 #1009

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)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： 1009

参赛队员 (签名) ：

队员 1：马 超

队员 2：黄江楠

队员 3：徐瑾辉

参赛队教练员 (签名)：

参赛队伍组别： 本科组

数学中国整理(www.madio.net)

第六届“认证杯”数学中国

数学建模网络挑战赛

编 号 专 用 页

数学中国提供（www.madio.net）

参赛队伍的参赛队号：（请各个参赛队提前填写好）：

1009

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 #1009

2013 年第六届“认证杯”数学中国
数学建模网络挑战赛

题　　目：基于音频指纹与用户指纹的开放式音乐分类模型

关 键 词：Mel 倒谱系数　SOM 神经网络　分形维数　核密度估计

摘　　要：

本文通过音频指纹与用户指纹的思想，对音乐风格的开放式分类问题进行了建模分析。

在模型一中，我们参考借鉴语音识别技术，使用 Mel 倒谱系数对音乐的音频信息进行特征提取。由于这种特征不依赖于信号的性质，对输入信号不做任何的假设和限制，所以非常利于下一步工作的开展。

在模型二中，我们基于 Mel 倒谱系数与 SOM 无监督神经网络，提出了一种开放式音乐风格分类类型。该无监督分类模型考虑了不同音乐类别间有交叉重叠的属性，不仅摆脱了传统音乐风格概念下的桎梏，同时对于新产生的音乐类别也能进行恰当的分类，避免了传统模型的简单、机械地划分。我们对总计 11 个类别的音乐样本进行分类，开放式分类正确率达 93%。

在模型三中，基于文献中音频信号的统计自相似性特点，我们借鉴了分形几何中的概念，将音频信号看成是一个具有分形特征的数据集，提出利用音乐信号的计盒维数作为其音频指纹，充分刻画了音频信号的整体特征，以及部分与整体的统计自相似性特征，具有抗噪的特点；而且，由于可以将音频的整体信息全部用一个介于 1 和 2 之间的实数来表示，所以与传统的 MFCC 特征提取相比，极大地起到了压缩音频信息的效果，使得基于内容的在线音乐推荐算法成为可能。我们通过对 110 个音乐样本的计盒维数的分布进行了分析，证实了将计盒维数作为音乐的音频指纹的有效性。我们就是否应该利用音乐的外部信息进行了探讨。

在模型四中，我们基于计盒维数和概率统计知识，提出了两种反映用户音乐喜好特征的“用户指纹”的定义，并应用 Parzen 核密度估计的方法，针对具体的用户数据，对其用户指纹的密度函数进行了估计。该种用户指纹能够面向个性化推荐系统，使得基于内容和基于协同过滤结合的音乐推荐算法成为了可能。

数学中国竞赛组委会版权所有

参赛队号： 1009

参赛密码 _____
（由组委会填写）

所选题目： B　题

英文摘要（选填）

（此摘要非论文必须部分，选填可加分，加分不超过论文总分的 5%）

The Present Paper proposes that we can analyze and make models for the open types of classifications of the music styles, by means of referencing the concepts of the audio fingerprinting and the user’s fingerprinting.

In the first model, we reference the speech recognition technology, collect and package the characteristic audio information via using MFCC. This characteristic doesn’t rely on the nature of signal and there are no assumption and limitation when we input signal. So it is very convenient for us to continue researching.

In the second model, based on MFCC and non-supervise SOM neural network, we raise a new music style. Because of considering the juxtaposition among the music, this non-supervised classification model gets rid of the shortcomings of the traditional music classifications and for the new music styles, it can mark off correctly at the same time. We classify eleven samples totally, whose styles are different, and the accuracy rate reaches 93% via this way.

In the third model, based on the statistical self-similarity of the sound signal, we reference the concepts in fractal geometry, and consider the sound signal as a dataset with fractal features. We raise that use box-counting dimension of the sound signal to be its audio fingerprinting, which is also antinoise. In this way, it can fully depict not only the overall features of the sound signal, but also the statistical self-similarity between the part and the whole. Meanwhile, we can use a real number between 1 and 2 to represent all the information of the sound signal. So, comparing with the traditional ways of extracting MFCC, the effect is more obvious. We prove the effectiveness of using box-counting dimension of the sound signal to be its audio fingerprinting. And then, we discuss whether to consider external information of the music.

In the fourth model, based on the knowledge of probability statistics and the box-counting dimension, we raise two definitions about the user’s fingerprinting, which reflect the characteristic of the users’ like. Aiming at specific user data, we apply the kernel density estimation to estimate the density function of the user’s fingerprinting. In this way, it makes possible to make a new algorithm of music recommendation.

数学中国提供（www.madio.net）

一、问题重述与问题分析

1.1 问题背景

随着互联网的发展，以及人民大众用音乐来满足自己对业余生活的追求，数字音乐现在呈爆炸式增长的姿态已呈现在我们面前。因此，流行音乐的主要传播媒介已经从传统的电台以及唱片过渡到网络下载以及自动的网络电台。现在听众通过网络电台，如豆瓣，可以收听到自己喜欢类别的音乐。但是，每个人对音乐的喜好可以横跨若干种风格，且这些风格之间的区别是很大的。所以，如何区分音乐风格就显得十分之重要。

1.2 问题重述与问题分析

为了从另外一个角度重述问题，我们首先引入如下的音频指纹概念：

音频指纹是指从一段音频采样中提取的独特的信息，可以用于辨识不同的声音采样，是一项基于内容的识别技术（Content-Base Identification，CBID）[1]。从相同采样不同编码格式的音频文件提取的音频指纹信息是相似的。

这样一来传统的基于风格流派的分类方法，本质上是一种较为粗糙的音频指纹。

借助于音频指纹的概念，用一种自然、合理的分类方法对音乐进行风格区分的问题，就可以重述为：建立合理的数学模型，探寻一种更为方便、有效的提取音频指纹的科学方法，使得在分类下，既不会造成类别之间的关系模糊与混乱，也不会分类过于粗略或精细。一种优良的音频指纹，应具有面向个性化推荐平台的特征，方便给网络电台的推荐功能和其它可能的用途，（包括对流行音乐市场的分析、基于流行音乐的大众审美研究）等方面提供支持。

1.3 问题意义

当今的主流音乐传播途径正逐步转向基于音乐的社区网络，基于流派的音频指纹缺少了和用户主观感受的融合(如,每个人喜好的音乐可能横跨若干种风格；同时，每个人的喜好也会随着时间的变化而改变），不利于用来描述用户的喜好特征（后文称之为用户指纹）。

同时，随着我国经济社会的发展、人民的文化需求日益提升，音乐界也在随之发生着日新月异的变化：新歌手、新风格、新流派不断地产生，不少新作品已经无法用传统的概念加以归类；机械地强行归类将遏制这类歌曲的传播与流行，这对于流行音乐的发展来说，将是致命的。为此，我们希望探寻更加科学、有效的音频指纹提取方法。

二、模型假设

1、假设我们所收集的数据和音乐样本都足够准确；并且文件格式转换与压

缩处理没有从本质上改变音乐样本的特征；

2、假设每个音乐样本所截取的音频片段代表了整首歌的大部分信息；

3、假设音乐信号的数据集具有分形特征；我们在后文将提到，这一假设有着一定的研究支撑；

4、不考虑音乐样本的外部信息（如歌手、歌词、歌名、标签、歌曲排行榜等），只考虑音乐内部信息（如节拍、音色、结构、音高）；关于这一假设的合理性，我们将会在后文进行分析讨论。

三、数据预处理

在我们收集到的总计 11 类、110 首的音乐作品样本中，其格式均为 mp3 格式，采样频率均为 44.1kHz，具有双声道立体效果。我们利用格式转换软件，将这 110 首 mp3 格式的音乐全部转换为 wav 格式。同时，为了降低运算量，我们对转换后的音乐文件进行了降低采样率以及转换为单声道的预处理。对于每首音乐，我们统一截取 20 秒的高潮部分，作为整首音乐的代表。

数学中国提供（www.madio.net）

四、符号说明

符号	解释及含义说明
$Mel(f)$	Mel 标度频率
f	频率，单位为 Hz
$X_a(k)$	音频信号的傅里叶变换
$x(n)$	输入的语音信号
N	傅里叶变换的点数
$f(m)$	中心频率
$H_m(k)$	三角滤波器的频率响应，其中 $\sum_{m=0}^{M-1} H_m(k)=1$
$S(m)$	每个滤波器组输出的对数能量
$C(n)$	经离散余弦变换(DCT)得到的 MFCC 系数
d_t	第 t 个一阶差分

C_t	第t个倒谱系数
Q	倒谱系数的阶数，取值范围为[12, 16]
K	一阶导数的时间差
j^*	胜出神经元, 即具有最小距离的神经元
S_j	选取的输出神经元 j^* 个“邻接神经元”的集合
W_{ij}	输入层的i 神经元和映射层的 j 神经元之间的权值
Δw_{ij}	修正的输出神经元 j^*及其“邻接神经元”的权值
η	一个大于 0 小于 1 的常数，且会随着时间变化逐渐下降到 0
$H^s(E)$	豪斯多夫外测度
$\dim_H E$	豪斯多夫维
$\dim_{box}(S)$	计盒维数
λ	样本间隔
f_s	采样率
$D_B(f)$	简化后的计盒维数
$f(x)$	该用户的基于计盒维数的用户音乐库指纹。
$f_n(x)$	总体未知密度 f 的一个核函数
h_n	窗宽度

数学中国提供（www.madio.net）

五、模型的建立与求解

5.1 基于听觉模型的 MFCC 参数[2]

Mel 倒谱系数（MFCC）由于其利用了人耳听觉的特性，在语音识别技术中被广泛应用于提取语音特征[3]。因此，我们应用 MFCC 来提取音乐的特征。

5.1.1 Mel 倒谱参数原理

根据人的听觉机理的研究发现，人耳对不同频率的声波有不同的听觉灵敏度。从 200Hz 到 5kHz 之间的语音信号对语音的清晰度影响最大。低音掩蔽高音容易，反之则困难[4]。据此，人们从低频到高频这一段频带内按临界带宽的大小由密到稀，安排一组带通滤波器，对输入信号进行滤波。将每个带通滤波器输出的信号能量作为信号的基本特征，对此特征经过进一步处理后就可作为语音的输入特征。由于这种特征不依赖于信号的性质，对输入信号不做任何的假设和限制，又利用了听觉模型的研究成果，因而被广泛应用于语音识别中。

MFCC 是在 Mel 标度频率域提取出来的倒谱参数，Mel 标度描述了人耳频率的非线性特性，它与频率的关系可用下式近似表示：

$$Mel(f) = 2595 * \lg(1 + f / 700)$$

式中 f 为频率，单位为 Hz。下图显示了 Mel 频率与线性频率的关系：

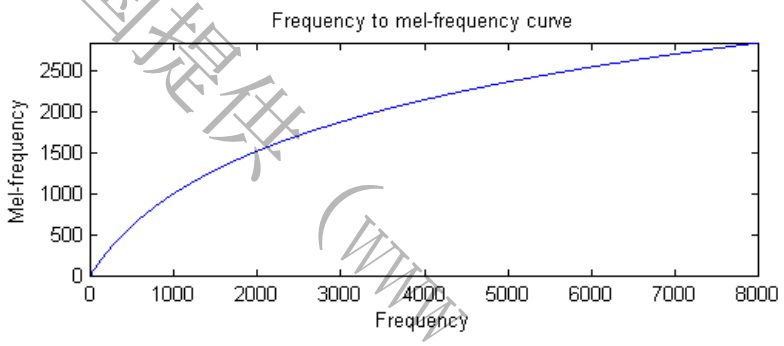

Figure 1 Mel 频率与线性频率的关系

求 Mel 倒谱系数的方法是将时域信号做时频变换后，对其对数能量谱用依照 Mel 刻度分布的三角滤波器组做卷积，再对滤波器组的输出向量做离散余弦变换(DCT)，这样得到的前 N 维向量称为 MFCC。

Mel 倒谱系数的提取过程如下图所示：

数学中国提供（www.madio.net）

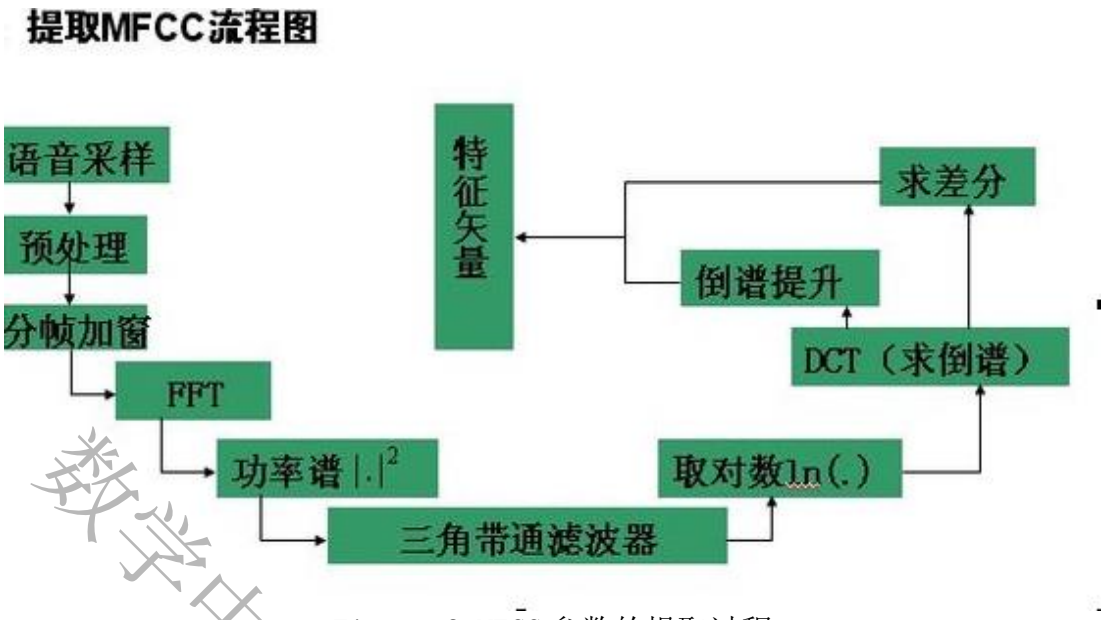

Figure 2 MFCC 参数的提取过程

5.1.2 MFCC 系数计算过程[5]

计算过程如下：

(1)对输入的音频信号进行分帧、加窗，然后作离散傅里叶变换，获得频谱分布信息。设音频信号的离散傅里叶变换（DFT）为：

$$X_a(k)=\sum_{n=0}^{N-1}x(n)e^{-j2\pi k/N},0\le k\le N$$

式中 $x(n)$ 为输入的语音信号， N 表示傅里叶变换的点数。

(2)再求频谱幅度的平方，得到能量谱。

(3)将能量谱通过一组 Mel 尺度的三角形滤波器组。

定义一个有 M 个滤波器的滤波器组(滤波器的个数和临界带的个数相近)，采用的滤波器为三角滤波器，中心频率为 $f(m),m=1,2,\cdots M$ ，本文取 $M=24$ 。各 $f(m)$ 之间的间隔随着 m 值的减小而缩小，随着 m 值的增大而增宽，如下图所示：

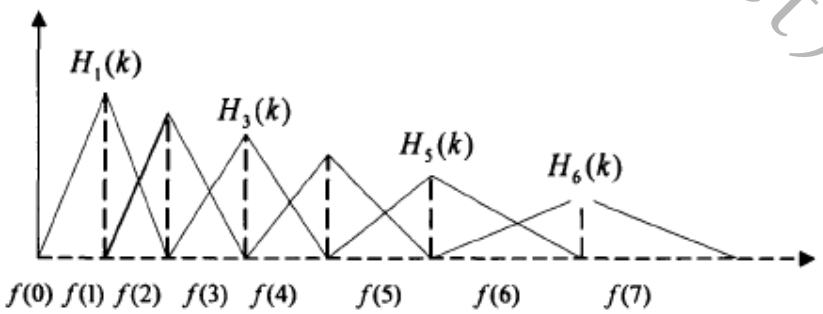

Figure 3 Mel 频率滤波器组

三角滤波器的频率响应定义为:

$$H_m(k)=\begin{cases}0, k<f(m-1)\\ \dfrac{2(k-f(m-1))}{(f(m+1)-f(m-1))(f(m)-f(m-1))}, f(m-1)\le k\le f(m)\\ \dfrac{2(f(m+1)-k)}{(f(m+1)-f(m-1))(f(m+1)-f(m))}, f(m)\le k\le f(m+1)\\ 0, k\ge f(m+1)\end{cases}$$

其中 $\sum_{m=0}^{M-1}H_m(k)=1$。

(4)计算每个滤波器组输出的对数能量为:

$$S(m)=\ln\left(\sum_{k=0}^{N-1}|X_a(k)|^2 H_m(k)\right), 0\le m<M$$

(5)经离散余弦变换(DCT)得到 MFCC 系数：

$$C(n)=\sum_{m=0}^{N-1}S(m)\cos\left(\frac{\pi n(m-0.5)}{M}\right), 0\le n<M$$

MFCC 系数阶数通常取 12-16，本文选取 12 阶倒谱系数。

标准的倒谱参数 MFCC 只反映了音频参数的静态特性，音频的动态特性可以用这些静态特征的差分谱来描述。实验证明:把动、静态特征结合起来能有效提高系统的识别性能。差分参数的计算可以采用下面的公式：

$$d_t=\begin{cases}C_{t+1}-C_t, t<K\\ \dfrac{\sum_{k=1}^{K}k(C_{t+k}-C_{t-k})}{\sqrt{2\sum_{k=1}^{K}k^2}}, \text{其它}\\ C_t-C_{t-1}, t\ge Q-K\end{cases}$$

式中，d_t 表示第 t 个一阶差分；C_t 表示第 t 个倒谱系数；Q 表示倒谱系数的阶数；K 表示一阶导数的时间差。将 MFCC 系数与差分 MFCC 系数二者合并，就得到了音频信号的特征矢量。

5.2 基于 CFCC 系数和 SOM 无监督神经网络的开放式音乐分类法

5.2.1 建模思路：

在利用 MFCC 系数对音乐音频提取了特征之后，接下来我们应用人工智能的学习算法来进行音乐类别的分类。正如前文中所提到的那样，在面向网络电台的音乐自动推荐功能时，传统的分类存在许多问题，一个用户所喜好的音乐类型也往往横跨多个风格，难以基于传统标签进行自动推荐。此外，随着音乐界的迅猛发展，新的音乐风格、音乐类型将越来越多，传统的有监督学习算法在面对全新的音乐风格的样本时，只能做简单而机械的划分，这显然是不符合现实需求的。

为此，我们在这里将采用无监督算法。与一般的分类算法相比，无监督算法旨在构造一个开放式的分类体系，其优势有以下两点[6]：

（1） 当出现新的音乐类型时，依据无监督分类方法，不做简单、机械的划分；
（2） 考虑了不同音乐类别间有交叉重叠的属性，从而可以为进一步的有监督分类提供分类依据。

5.2.2 SOM 无监督神经网络[7]

SOM 神经网络即自组织特征映射网络，该网络是由一个全连接的神经元陈列组成的无教师、自组织、自学习网络。SOM 神经网络既可以学习训练数据输入向量的分布特征，也可以学习训练数据输入向量的拓扑结构。其中，具有最小距离的神经元，称为胜出神经元，记为 j^*。在权值更新过程中，不仅胜出神经元的权值向量得到更新，而且其近邻神经元的权值向量也按照某个“近邻函数”进行更新。因此，在竞争层的神经元位置演变的过程中，每个区域代表一类输入向量。通过训练，可以使得每个权值向量都位于输入向量聚类的中心，一旦 SOM 完成训练，就可以将之用于对训练数据或其他数据进行聚类。其学习算法步骤归纳如下：

（1）网络初始化

用随机数设定输入层和映射层之间的权值的初始值。对 j 个输入神经元到输出神经元的连接权值赋予较小的权值。选取输出神经元 j^* 个“邻接神经元”的集合 S_j。区域 $S_j(t)$ 随着时间的增长而不断缩小。

（2）输入向量的输入

把输入向量 $X=(x_1,x_2,x_3\cdots x_m)^T$ 输入给输入层。

（3）计算欧式距离

在映射层，计算各神经元的权值向量和输入向量的欧式距离。映射层的第 j 个神经元和输入向量的距离，公式如下：

$$d_j=\|X-W_j\|=\sqrt{\sum_{i=1}^{m}(x_i(t)-w_{ij}(t))^2}$$

其中，w_{ij} 为输入层的 i 神经元和映射层的 j 神经元之间的权值。

（4）权值的学习

按下式修正输出神经元 j^* 及其“邻接神经元”的权值：

$$\Delta w_{ij}=w_{ij}(t+1)-w_{ij}(t)=\eta(t)(x_i(t)-w_{ij}(t))$$

式中，η 为一个大于 0 小于 1 的常数，随着时间变化逐渐下降到 0，

数学中国提供（www.madio.net）

$\eta(t)=1/t$ 或 $\eta(t)=0.2(1-t/10000)$

（5）计算输出 O_k

$$O_k = f(\min_j \| X - W_j \|)$$

式中， $f(x)$ 一般为 0-1 函数或者其它非线形函数[8]。根据前文，我们搜集了 11 类音乐中的经典歌曲，分别为：流行，摇滚，说唱，爵士，布鲁斯，民谣，乡村，新世纪，钢琴，古典，电子，分别记为 A，B，C，…，K 类，每种均有 10 首歌曲。首先我们把流行，摇滚，说唱，爵士，布鲁斯，古典，乡村，钢琴中每种随机抽取 5 首歌曲作为训练集，分别记为 A1，A2，…，A5；B1，B2，…，B5；以此类推。每种剩下的 5 首作为测试集，分别记为 A6，A7，…，A10；B6，B7，…，B10；以此类推。用于测定训练好的 SOM 神经网络分类的能力。其次，我们把剩下的三种，新世纪，民谣，电子作为新出现的种类，来测试 SOM 对于新生类别的分类能力，用以确定歌曲不同风格之间的相似度。

我们现在利用格式转换软件，将这 11 组音频数据，即 110 首 mp3 格式音乐转换为 wav 格式。同时，为了降低运算量，我们对转换后的音乐文件进行降低采样率以及转换为单声道的预处理。对于每首音乐，我们统一截取 20 秒高潮部分，通过 MATLAB 软件计算其 MFCC 参数，并将 MFCC 参数作为输入值，运用前文提到的 SOM 神经网络进行训练（MATLAB 编程情见附录），训练完成得到音乐样本的无监督分类。在经过迭代后，SOM 神经网络得到不同类别的音乐样本。

下图，为训练结束时的 SOM 网络拓扑结构：

数学中国论坛（www.madio.net）

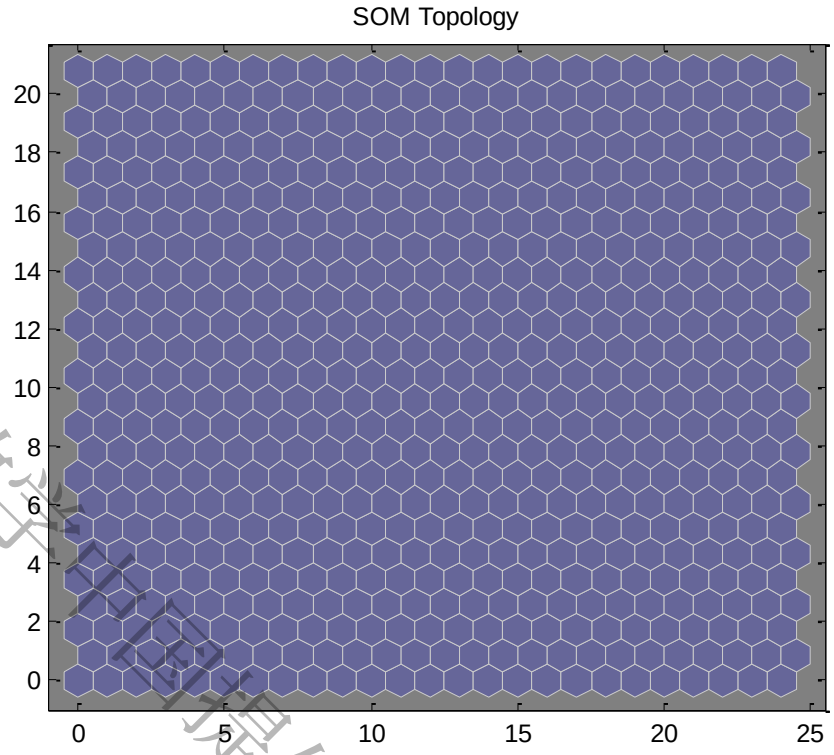

Figure 4 训练结束时的 SOM 网络拓扑结构

下图为临近神经元之间的距离情况：

数学中国提供（www.madio.net）

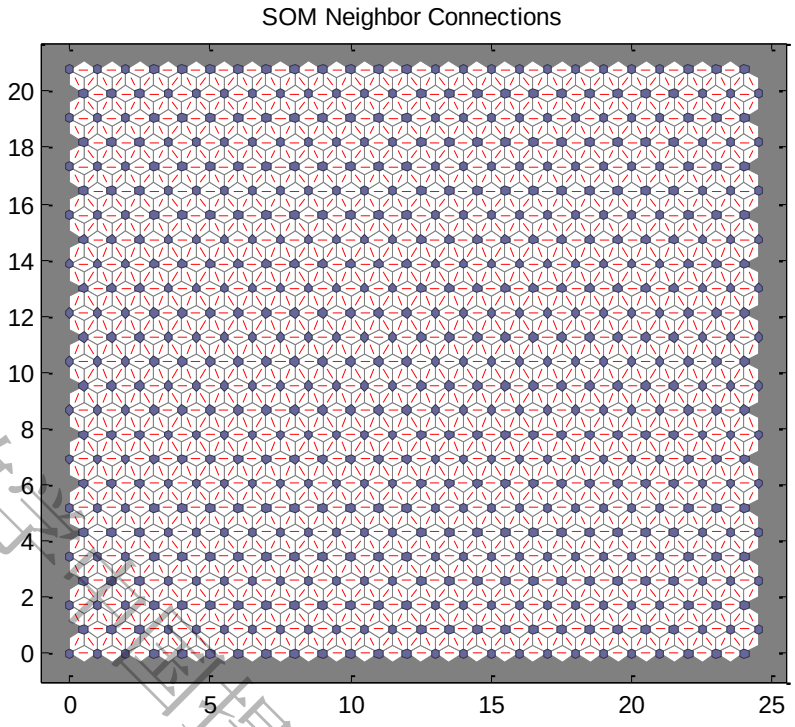

Figure 5 临近神经元之间的距离情况

下图中，蓝色表示神经元，红色的线表示神经元直接的连接，每个菱形中的颜色表示神经元之间距离的远近，从黄色到黑色，颜色越深说明神经元之间的距离越远。

数学中国提供（www.madio.net）

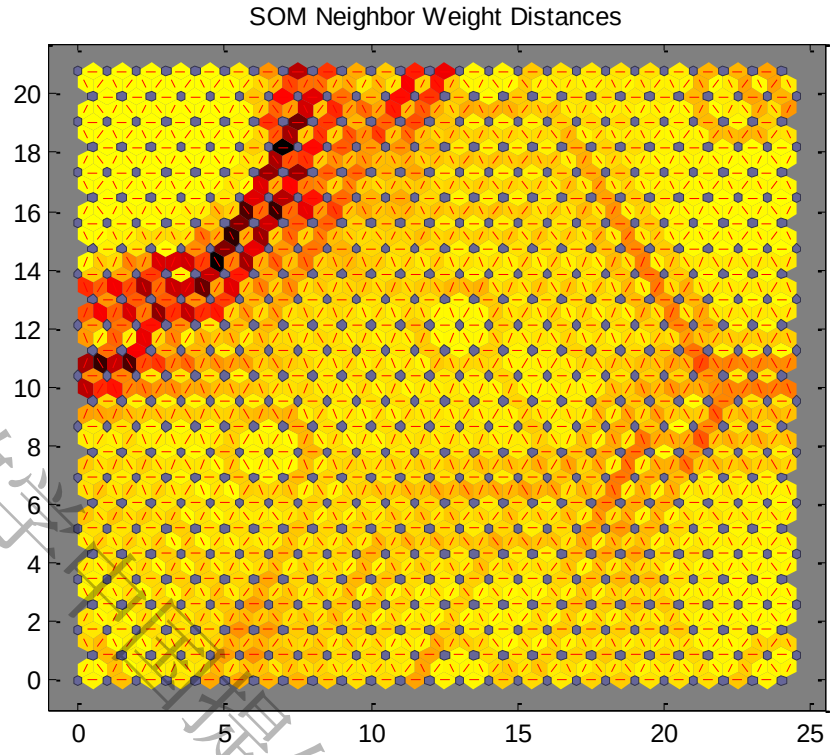

Figure 6 神经元间的距离与关系

下图中，蓝色表示竞争胜利的神经元

数学中国提供（www.madio.net）

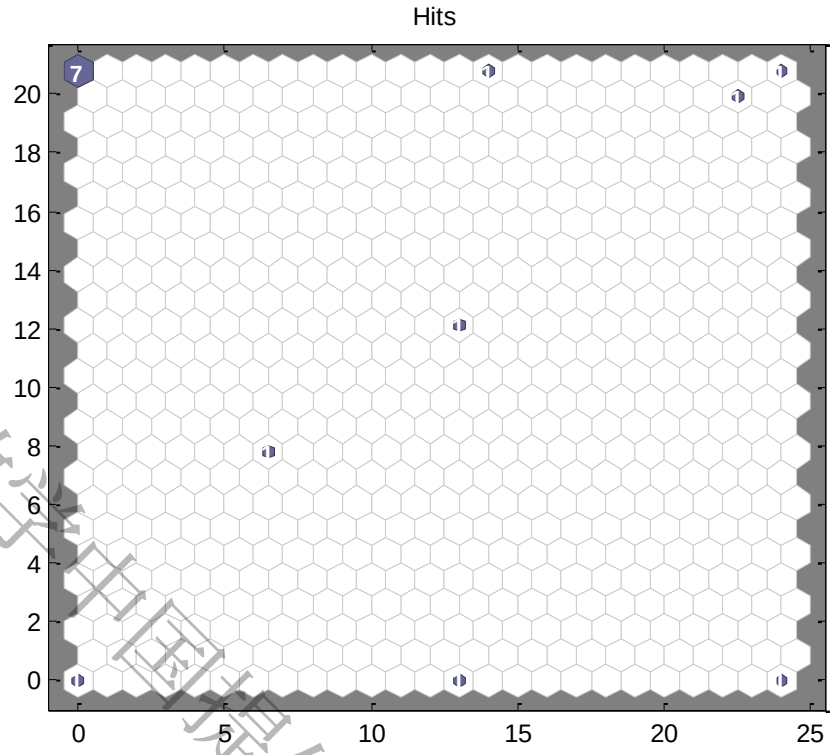

Figure 7 竞争胜利的神经元的图示

5.2.3 结果及其分析

结果 1：

下表是我们对 40 个音乐样品进行 SOM 训练输出的分类结果：

类别	音乐样品
1	A1, A3, A4, A5, B1, B2, B3, B4, B5, C5
2	C1, C2, C3, D1, D2, D3, D4, D5, E1, E2, E3, E4, E5
3	I1, I2, I3, I4, I5, J1, J2, J3, J4, J5
4	A2, C4, G1, G2, G3, G4, G5

结果 1 的分析：

从结果上看，SOM 所分的第一大类主要是流行音乐与摇滚乐；第二大类主要是说唱，爵士和布鲁斯；第三大类是古典和钢琴；第四大类是说唱和乡村。从结果上看，大致符合我们之前所介绍的音乐类别的知识（如不少说唱、爵士、布

数学中国提供（www.madio.net）

鲁斯样本被分到一类；事实上，这三类流派均有着相同的根基）。若强行根据样本本身的“风格标签”进行分析，我们发现存在一定的名义上的“错分”现象。如说唱类样本 C5 被“错”分到了第一类（流行与摇滚）。但事实上，经过我们的试听发现，说唱样本 C5 的整体风格确实偏向于流行，与其他说唱样本整体差异较大。

结果 2：

下表是我们应用训练完成的 SOM 分类器，对 40 个已知类型的测试集音乐样本进行分类的结果：

类别	音乐样品
1	A7, A8, A9, A10, B6, B8, B9, B10, G10
2	C7，C8，C10，D6, D7, D8, D9, D10, E6, E7, E9, E10
3	E8， I6, I7, I8, I9, I10， J6, J7, J8, J9, J10
4	A6, B7, C6, C9, G6, G7, G8, G9

结果 2 分析：

我们运用 SPSS 软件对 SOM 神经网络分类结果以及样本自带的风格标签进行相关性分析。结果显示，和 SOM 神经网络分类结果与样本风格标签有 75.96％相关。这说明，SOM 神经网络分类结果与传统风格标签一脉相承，同时又能打破传统风格标签的桎梏，依据音乐本身的音频特征进行恰当的分类。

结果 3：

对于三类 SOM 神经网络未知类型的音乐样本（电子，新世纪，民谣），SOM 的分类如下：

类别	音乐样品
1	F2, F3, F4, F10, K3, K5, K7, K9, K10
2	H8, H9, K4, K8
3	H1, H2, H3, H4, H5, H7, H10, K1, K6
4	F1, F5, F6, F7, F8, F9, H6, K2

数学中国提供（www.madio.net）

结果 3 分析：

由计算机结果可知，得益于算法的无监督性，在面对三种从未见过的风格标签类型时，SOM 也能进行恰当的分类。如新世纪音乐主要被分在第三类（即训练集古典与钢琴类样本主要的分类）；民谣主要被分在第一类（即训练集流行、摇滚的主要分类）和第四类（即训练集中乡村、说唱的主要分类）；电子主要被分在第一类（即训练集中的流行、摇滚的主要分类）。我们根据分类结果进行试听，发现这样的分类确实在一定程度上把握了新类型的一定特性。

当然，若按照传统风格标签生硬地进行判断，那么 SOM 均存在一定的名义上的“错分”。我们说明如下：

(1)对于新世纪音乐，由于其风格的多样性，也有部分被分为第二类（H8，H9）和第四类（H6）。通过我们的试听，发现除了 H6 为明显错分外，其余分类均能体现新样本的整体特征。

(2)对于电子乐，有部分样本被“错分”在第三类（K1,K6）。经过我们试听发现，这两首电子乐样本确与其它电子样本差异较大；其他电子样本主要是偏流行或偏说唱的风格（如 K2）；但这两首电子音乐样本为纯音乐，表现手法上与新世纪音乐更为接近。

SOM 网络有着比较好的分类能力，在对流行，摇滚，说唱，爵士，布鲁斯，古典，乡村，钢琴的测试中，经人工听辨，把握风格的准确率达到了 92.727%，而对于训练集没有的类别，新世纪，民谣，电子，SOM 神经网络能够把他们归类于相似的类别。由此可看出，SOM 神经网络是比较适合歌曲的开放式分类的。

5.3 基于分形维数的音频指纹

5.3.1 音频指纹

音频指纹是指从一段音频采样中提取的独特的信息，可以用于辨识不同的声音采样，是一项基于内容的识别技术（Content-Base Identification，CBID）[9]。从相同采样不同编码格式的音频文件提取的音频指纹信息是相似的。

5.3.2 传统风格分类的局限性

在流行音乐中，传统的风格（流派）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。这种传统的风格概念即可以看作是是一种简单而粗糙的音频指纹。然而，正如我们在问题重述中所提到的一样，这种基于流派的音频指纹存在着许多不足之处：分类繁杂、区分度不足、无法从整体上代表音乐作品的特征、不易于与网络电台的推荐功能相结合。

当今的主流音乐传播途径正逐步转向基于音乐的社区网络，基于流派的音频指纹缺少了和用户主观感受的融合(如，每个人喜好的音乐可能横跨若干种风格；同时，每个人的喜好也会随着时间的变化而改变），不利于用来描述用户的喜好特征（后文称之为用户指纹）。同时，随着我国经济社会的发展、人民的文化需

数学中国YY网校频道：159214 数学中国www.madio.net 数学中国公众微信平台：shuxuezhongguo

求日益提升，音乐界也在随之发生着日新月异的变化：新歌手、新风格、新流派不断地产生，不少新作品已经无法用传统的概念加以归类；机械地强行归类将遏制这类歌曲的传播与流行，这对于流行音乐的发展来说，将是致命的。为此，我们希望探寻更加科学、有效的音频指纹提取方法。

5.3.3 分数维：分形几何方法的应用

正如前文所述，Mel 对数倒谱系数 MFCC 特征由于符合人类的听觉特征，是目前在语音领域中应用最广泛的特征抽取方法。但是这类音频指纹属于基于局部提取方法（如：需要将音频分帧），忽略了音频的整体特性（如：帧与帧之间的时间序列关系），在有外界噪音以及干扰的环境下，检索性能急剧下降。声学及空气动力学理论证明了语音信号是一个复杂的非线性过程[10],其中存在着产生混沌的机制,它在一定尺度下局部与整体之间具有统计自相似性,语音信号所具有的分形特征是将分形理论引入语音信号分形分析的基础。同时，后文通过研究发现，不同风格音乐之间的分形特征的确有着明显的差异。因此，分形特征非常有利于音频指纹的提取。

分形的概念来自于自然，并由曼德勃罗于 1973 年首次提出。所谓分形，是指具有以下特质的几何结构[11]：

1、在任意小的尺度上都能有精细的结构；

2、太不规则，以至无论是其整体或局部都难以用传统欧氏几何的语言来描述；

3、具有（至少是近似的或统计的）自相似形式；

4、一般地，其“分形维数”（通常为豪斯多夫维数）会大于拓扑维数；

5、在多数情况下有着简单的递归定义。

下图是分形的一个经典例子：曼德勃罗集合。

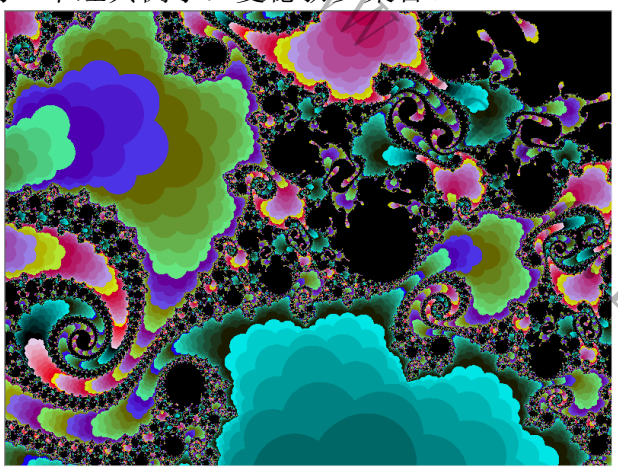

Figure 8 曼德勃罗集合

分数维，是分形理论当中的主要参数，它定量描述了分形集的复杂程度。所谓分数维，是一个描述一个分形对空间填充程度统计量。在信号的领域中，分数维能作为信号的信息度量。分数维没有统一的定义。主要的分数维定义方法有豪斯多夫维数、计盒维数和分配维数等[12]。

分数维能反映大量的有关信号“几何”特征的信息。研究表明不同类型的信号或噪声往往具有不同的分维，在音乐处理中也逐渐得到应用[13]。

豪斯多夫分数维是最常用的分数维之一，通过豪斯多夫维可以给一个几何对象、一维音频信号乃至是任意复杂的点集合比如分形（Fractal）赋予一个维度。其严格数学定义如下[14]：

(1)豪斯多夫外测度： 令(X,d)为一个度量空间，E 为 X 的一个子集，定义

$$H_\delta^s(E)=\inf\left\{\sum_{i=1}^{\infty} diam(A_i)^s\right\}$$

并且 E 能被集族 $(A_j)_k$ 所覆盖。 则 E 的豪斯多夫外测度被定义为：

$$H^s(E)=\lim_{\delta\to 0} H_\delta^s(E)$$

(2)豪斯多夫维：豪斯多夫维被定义为豪斯多夫外测度从零变为非零值跳跃点对应的 s 值。严格的定义为：

$$\dim_H E=\inf\left\{s:H^s(E)=0\right\}=\sup\left\{s:H^s(E)=\infty\right\}$$

然而，豪斯多夫分数维计算起来是非常困难的；我们一般通过计盒维数来近似计算。计盒维数由于计算相对容易，有许多成熟算法[15]。

计盒维数的原理是，将分形放在一个均匀分割的网格上，计算覆盖这个分形所需的最小格子数。通过对网格的逐步精化，查看所需覆盖数目的变化，从而计算出计盒维数。

假设当格子的边长是 ε 时，总共把空间分成 N 个格子，那么计盒维数就是[16]：

$$\dim_{box}(S):=\lim_{\varepsilon\to 0}\frac{\log N(\varepsilon)}{\log(1/\varepsilon)}$$

在参考文献中[17]，针对信号序列，对计盒维数的定义进行了简化：

设信号的采样序列为 $f(t_1), f(t_2), f(t_3),\ldots, f(t_N), f(t_{N-1})$，$N$ 为偶数。令

$$d(\Delta)=\sum_{i=1}^{N}\left|f(t_i)-f(t_{i1})\right|,$$

$$d(2\Delta)=\sum_{i=1}^{N/2}\max\{f(t_{2i-1}), f(t_{2i}), f(t_{2i+1})\}-\min\{f(t_{2i-1}), f(t_{2i}), f(t_{2i+1})\}$$

以及 $N(\lambda)=d(\lambda)/\lambda, N(2\lambda)/2\lambda,$

其中样本间隔 $\lambda=1/f_s, f_s$ 为采样率，那么

$$D_{\mathbf{B}}(f)=\frac{\lg N(\lambda)/N(2\lambda)}{\lg\frac{1/\lambda}{1/2\lambda}}=\frac{\lg N(\lambda)-\lg N(2\lambda)}{\lg 2}$$

数学中国YY网校频道（www.madio.net）

在满足双向 Holder 条件时，用简化算法得到的盒维数满足 $D_B(f)\in[1,2)$，且盒维数对噪声不敏感[18]。

于是，我们采取音频信号的分数维作为一种新的音频指纹。据前文所述，使用计盒维数作为音频指纹，充分刻画了音频信号的整体特征、以及部分与整体的统计自相似性特征，具有抗噪的特点；而且，由于可以将音频的整体信息全部用一个介于 1 和 2 之间的实数来表示，与传统的 MFCC 特征提取相比，这极大地起到了压缩音频信息的效果。

5.3.4 模型应用

为了检验以计盒维数作为音频指纹的合理性与有效性，我们从数十种音乐风格中选取（流行）一定数量的音乐作品进行计盒维数的计算与统计。以虾米网（www.xiami.com）为例，具体的类别如下图所示：

Figure 9 音乐样品的类别图示

由于时间关系，我们在仅在具有代表性的歌曲中随机下载共计 11 种风格、共 110 首 mp3 格式的音乐进行分析，涉及到的歌手（或作曲家）有*周杰伦*、*陈奕迅*、*羽泉*、*邓丽君*、*凤凰传奇*、*Avril Lavige*、*Brad Paisley*、*Osis*、*拔头士*、*迈克尔杰克逊*、*Elivs Presley*、*Eagles*、*Linkinpark*、*W.C.Handy*、*Norah Jones*、*Yanni*、*Sarah Brightman*、*Kitaro*、*Secret Garden*、*莫扎特*、*比才*、*德彪西*、*巴赫*等。

这 110 首 mp3 格式的作品采样频率均为 44.1kHz，具有双声道立体效果。我们利用格式转换软件，将这 110 首 mp3 格式音乐转换为 wav 格式。同时，为了

数学中国提供（www.madio.net）

降低运算量，我们队转换后的音乐文件进行降低采样率以及转换为单声道的预处理。对于每首音乐，我们统一截取 20 秒高潮部分，通过 MATLAB 软件计算其计盒维数。在总计共 110 个样本中，我们按类别，将每个音乐样本的计盒维数绘制出来，如下图所示：

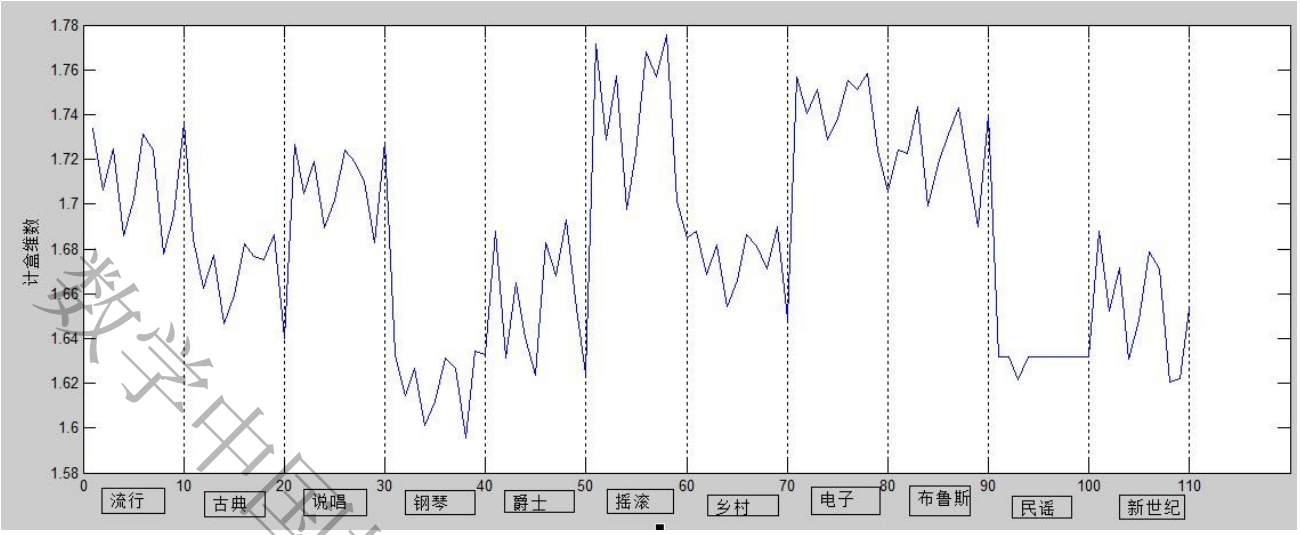

Figure 10 音乐样本的计盒维数图示

根据上图，可以看出，不同传统音乐风格的计盒维数均存在较明显差异，如钢琴类的 10 个样本的计盒维数都处于较低水平；而摇滚类的 10 个样本的计盒维数则非常高。同时，我们还可以发现，同一种音乐风格内部差异较小。

这说明，用计盒维数作为音频指纹是合适的。然而，注意到，部分样本的计盒维数与其所属类别的均值存在较大差异，如隶属于新世纪类别的作曲家 *Joe Hisaishi* 的作品 *Summer* 的计盒维数与钢琴类的维数均值更加接近。就听者的感官体验上，这是显然的，因为 *Summer* 这首作品是以钢琴作为主奏乐器。这说明，用计盒维数来刻画音乐的音频指纹，似乎更能打破不同风格标签间的界限，从整体上把握一首乐曲的内在本质。事实上，在下文中，我们将利用这一点，刻画反映用户的音乐兴趣特征的“用户指纹”，并能更好地横跨不同的音乐类别为用户推荐音乐。

5.3.5 是否应该利用歌曲的外部信息？

无论是Mel倒谱系数，还是前文中我们所提出的计盒维数，这些音频指纹均是基于音乐作品自身的内部信息——即诸如节拍、音色、结构、音高等音频信息的集合。虽然内部信息是一首音乐作品中最本质，最核心的内容，但这似乎忽略了歌曲的外部信息——即诸如歌手、歌曲名称、歌词、标签、歌曲排行榜等听众可以提前获取的先验信息。我们需要知道，这样的处理方式在网络推荐的应用当中，是否会存有隐患？

个性化推荐系统之所以被提出，主要是为了解决信息过载问题导致的用户的流失[19]，并帮助用户从信息的海洋中获取有用的信息。在传统的音乐检索与推荐技术中，多是基于音乐作品的外部信息（如歌手、歌曲名称、歌词、标签、歌曲排行榜信息）。然而，由于这些信息本身多易于检索、易于传播、易于自主获

得（如，歌迷常会定期关注某歌手的动态；位置显眼的歌曲排行榜常会吸引用户的自主关注；等等），因此外部信息并非个性化推荐系统所要考虑的主要问题。事实上，基于外部信息的推荐方法实际上并没有包含太多可供用户进一步挖掘的信息，不能有效地利用用户可能感兴趣的潜在信息。在实际应用中，由于音乐文件的内部信息（如MFCC系数）存在着维度较高、检索速度慢的问题，因此无法应用基于内容的过滤推荐算法。基于协同过滤的推荐算法常弥补了外部信息推荐的不足，但依然存在稀疏性问题和最初评价问题[20]。为此，在这里我们指出，就面向个性化推荐系统的角度而言，使用基于音乐内部信息的音频指纹，并没有太大的隐患；事实上，通过下文的进一步讨论我们还将发现，基于分数维这一内部信息的音频指纹在应用于个性化推荐当中将会是非常有效的。

5.4 基于分形维数和核密度估计的用户指纹建模

5.4.1 用户指纹建模思路

正如前文所指出的那样，当今的主流音乐传播途径正逐步转向基于音乐的社区网络，而提取音频指纹的最大目的之一就是为了便于为网络电台的推荐功能和流行音乐市场分析提供一定的支持。在这里，我们引入一个新的概念——用户指纹。与音频指纹的概念类似，我们定义：

定义 用户指纹是一类基于音频指纹、能够反映用户的音乐喜好特征的参数；该参数通常具有一些较优良的性质，如：

- 能够充分反映用户的喜好特征；
- 在用户指纹的指导下，能较容易地匹配相应的音频指纹，并选取相应的音乐推荐给用户；
- 用户指纹相对容易计算和储存；
- 用户的喜好特征通常会随着时间的推移而改变，因此用户指纹也应能够定期更新，并对用户指纹进行遗忘处理[21]；等等。

在前文的建模中，我们曾提出使用计盒维数作为音乐作品的音频指纹，该特征能打破传统风格分类的桎梏、充分刻画音乐作品的整体特征（以及总体与局部的自相似性），并在实证中初步验证了其有效性。同时，由于我们仅使用计盒维数就充分区分了音乐之间内在的内部信息特征，因此基于音乐内容过滤的推荐（或与协同过滤推荐结合）变得可以实现。综上所述，计盒维数非常适用于面向推荐系统的用户指纹建模。接下来，**由于时间关系**，我们仅就基于计盒维数的用户指纹进行建模。

5.4.2 基于计盒维数的用户音乐库指纹与用户点播指纹

我们根据前一节的讨论，提出以下基于计盒维数的**用户音乐库指纹**：

定义 1：设用户的在线音乐库有 n 首音乐，其计盒维数分别为 $x_1, x_2 \cdots\cdots x_n$。我们将 $x_1, x_2 \cdots\cdots x_n$ 看作是从一维总体 X 中抽取出来的独立同分布样本，设 X 的密度函数

为 $f(x), x \in [1,2)$ ，则我们称函数 $f(x)$ 为该用户的**基于计盒维数的用户音乐库指纹。**

考虑到用户的点播操作样本数将远超过其音乐库中音乐的数目，为了更好地抽取用户特征，我们类似地定义**用户点播指纹**：

定义 2：设用户的在线有 n 次播放操作，其对应的音乐作品的计盒维数分别为 $x_1, x_2 \ldots\ldots x_n$ 。我们将 $x_1, x_2 \ldots\ldots x_n$ 看做是从一维总体 X 中抽取出来的独立同分布样本，设 X 的密度函数为 $f(x), x \in [1,2)$ ，则我们称函数 $f(x)$ 为该用户的**基于计盒维数的用户点播指纹。**

这样，基于计盒维数的用户指纹的计算，就转化为了对未知密度函数 $f(x)$ 的估计问题。这是一个非参数的估计问题，我们在下一节将着重讨论 $f(x)$ 的估计问题。

5.4.3 基于 Parzen 核密度估计的用户指纹计算

由于总体 X 的分布 $f(x)$ 未知，所以我们使用非参数估计的方法来确定盒维度这个随机变量的总体分布 $f(x)$ 。这里我们采用 Parzen 核密度估计的非参数估计方法。

核密度估计（kernel density estimation），亦称 Parzen 窗法，由 *Rosenblatt* 和 *Parzen* 所提出，属于非参数检验的方法之一。其目的是，给定数据 $x_1, x_2 \ldots\ldots x_n$ ，估计出该总体的概率密度函数。我们先给出核密度估计的定义[22]：

定义 设 $x_1, x_2 \ldots\ldots x_n$ 是从一维总体 X 中抽取出来的独立同分布样本，X 是具有未知的密度函数 f 的一维实函数，则 f 的核密度估计为：

$$f_n(x) = \frac{1}{nh_n} \sum_{j=1}^{n} K\left(\frac{x - x_j}{h_n}\right)$$

其中 K 为□ 上给定的核函数，h_n 为窗宽度， n 为样本数目， $f_n(x)$ 为总体未知密度 f 的一个核函数。为了保证 $f_n(x)$ 作为密度函数估计的合理性，要求核函数满足

数学中国提供（www.madio.net）

$$K(x)\geq 0,\int_{-\infty}^{+\infty}K(x)=1$$

核函数 $K(\frac{x-x_j}{h_n})$ 的形状和值域控制着用来估计 $f_n(x)$ 在点 x 的值时所用数据点的个数和利用的程度。常用的核函数如下：

均匀核	$\frac{1}{2}I(\lvert u\rvert\leq 1)$
三角核	$(1-\lvert u\rvert)I(\lvert u\rvert\leq 1)$
Epanechikov	$\frac{3}{4}(1-u^2)I(\lvert u\rvert\leq 1)$
四次方核	$\frac{15}{16}(1-u^2)^2I(\lvert u\rvert\leq 1)$
三权核	$\frac{35}{32}(1-u^2)^3I(\lvert u\rvert\leq 1)$
高斯核	$\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}$
余弦核	$\frac{\pi}{4}\cos(\frac{\pi}{2}u)I(\lvert u\rvert\leq 1)$
指数核	$e^{\lvert u\rvert}$

我们选取高斯函数作为核函数，即 $K(u)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}$。由其表达式易知，如果 x_i 离 x 越近，则 $\frac{x-x_i}{h_n}$ 越接近于零，这时正态密度的值 $K(\frac{x-x_i}{h_n})=\varphi(\frac{x-x_i}{h_n})$ 越大。因为正态密度的值域为整个实轴，所以所有的数据都来估计 $f(x)$ 的值，只不过离 x 点越近的点对估计的影响越大。当 h 值小时只有接近 x 的点作用大，h 值越大，则远一些的点的作用也增加。

这里我们做以下说明：

说明（1）对每个观察 x_i 限制在高为 $\frac{1}{nh_n}$，宽为 h_n 的窗内，而估计值为 n 个这种窗之和。因而 h_n 正是这 n 个窗的公共窗宽参数。

说明（2）在给定样本之后，一个核估计性能的好坏，取决于核及窗宽的选取是

数学中国提供（www.madio.net）

否适当。下图展示了，在不同的h_1值和n值下的对正态分布的核密度估计效果，可以看出，当的$h_1=4$和$n=256$时，估计效果较好。

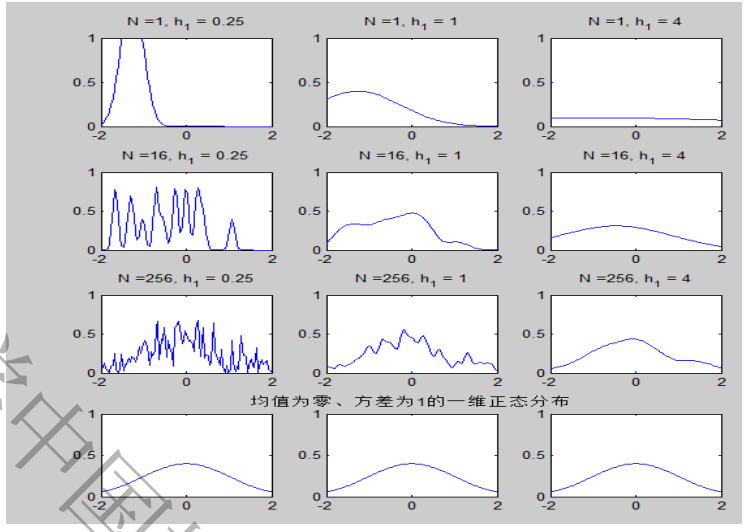

Figure 11 在不同的参数值下的对正态分布的核密度估计效果

而对于双峰均匀分布的估计如下图所示：可见，在$h_1=1$和$n=256$时，估计效果较好。

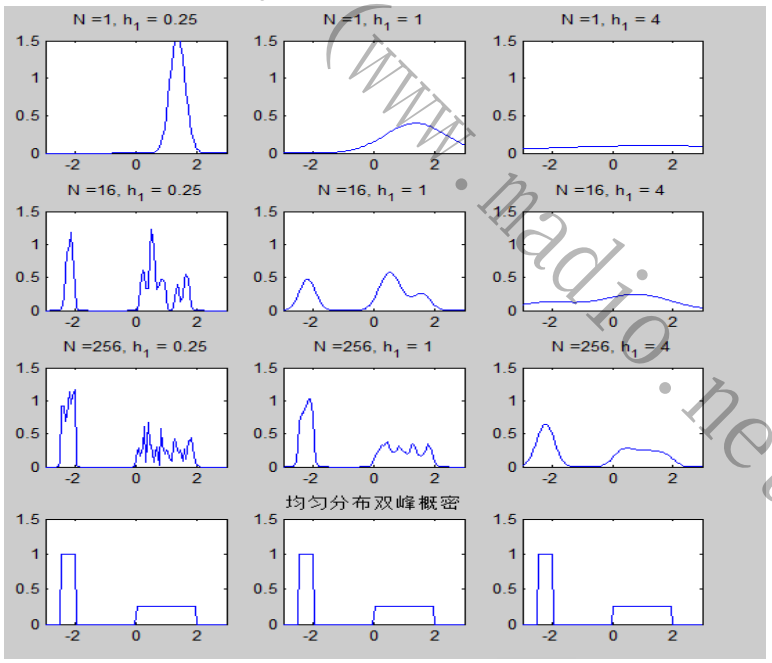

Figure 12 双峰均匀分布的估计

从直观上看，核估计在观察点x_i有一“碰撞”，估计量是这些“碰撞”的宽度，当h_n选得过大，由于x_i经过平移压缩之后使分布的主要部分的某些特征（如

多峰性）被掩盖起来了，估计量有较大偏差；如h_n太小，整个估计特别是尾部出现较大的干扰，从而有增大方差的趋势。

说明（3）下面介绍一种求最佳窗宽的方法

令

$$MISE(f_n)=E\{\int[f_n(x)-f(x)]^2dx\}$$

其中$f(x)$为总体的真实分布密度，$MISE$是关于窗宽h的函数，求它的最小值点，可以得出最佳窗宽的估计值。

5.4.4 模型的应用：用户指纹实例

作为基于计盒维数的用户指纹的应用，我们考虑几个来自于实际在线音乐社交网站的例子。由于时间关系以及权限原因，我们无法获取一些在线音乐网站的用户点播行为数据。因此，我们以一位来自虾米音乐网的用户为例，得到了该用户近一周来的音乐库情况：

Figure 13 虾米音乐网某用户近一周来的音乐库情况

根据这位用户的播放记录，我们应用 Parzen 核密度估计方法进行用户指纹估计，结果如下图所示：

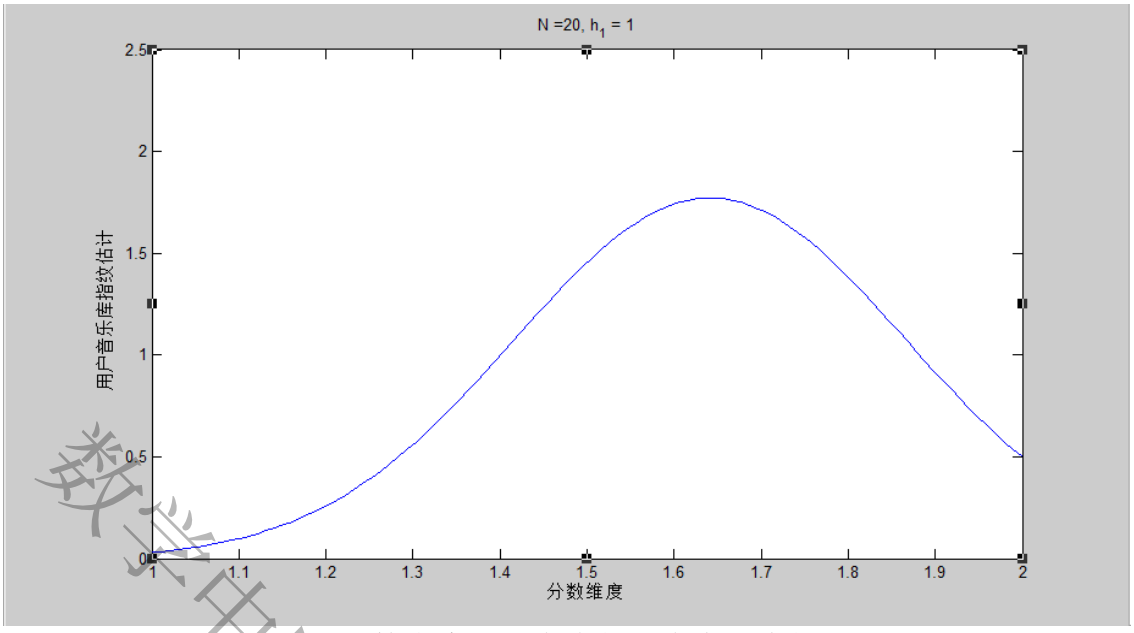

Figure 14 核密度估计产生的用户音乐库指纹估计

进一步计算得，该用户的音乐库指纹估计的峰度为 1.5949，偏度为-0.0719，为峰度不足左偏态分布。

数学中国提供（www.madio.net）

六、模型评价

6.1 模型优点：

1. 使用音频指纹与用户指纹的思想，专注于寻求一种面向个性化推荐系统、开放式的音乐分类方法，将音乐分类模型与网络电台的实际应用紧密结合起来，具有继续深化理论研究、研发应用模型的潜力；
2. 提出了基于 Mel 倒谱系数（MFCC）与 SOM 无监督神经网络的开放式分类方法，得益于无监督算法的特性，考虑了不同音乐类别间有交叉重叠的属性，不仅摆脱了传统音乐风格概念下的桎梏，同时对于新产生的音乐类别也能进行恰当的分类，避免了传统模型简单，机械的划分；
3. 利用音乐信号的计盒维数作为音乐的音频指纹，充分刻画了音频信号的整体特征，以及部分与整体的统计自相似性特征，具有抗噪的特点；而且，由于可以将音频的整体信息全部用一个介于 1 和 2 之间的实数来表示，所以与传统的 MFCC 特征提取相比，极大地起到了压缩音频信息的效果，使得基于内容的在线音乐推荐算法成为可能。
4. 基于计盒维数的优良特性，提出了“用户指纹”的概念，将用户音乐喜好特征的预测和挖掘转化为对密度函数的非参数估计问题。同时，提出并使用 Parzen 核密度估计法，对用户指纹进行估计，效果优良。

6.2 模型有待改进和研究的地方：

1. 是否应该考虑音乐的外部信息（如歌手、歌词、歌名、标签、歌曲排行榜等），仍是一个有待深入讨论和建模分析的地方；

2. 在模型中，多截取音乐样本的高潮部分，这常常不能完全代表整首音乐的特征；
3. 计盒维数给出了豪斯多夫维数上界的一个估计；其准确性需要进一步的讨论；
4. 如何将基于计盒维数的音频指纹与用户指纹运用于基于内容的推荐算法（或和协同过滤推荐算法的结合），是下一步工作的一个重点；
5. 获取更多的在线音乐社交网站的用户数据，进行更多、更完善的数值试验，以说明模型和算法的有效性，是下一步工作的一个重点；
6. 本文提出的基于计盒维数的音频指纹与用户指纹可以被进一步利用，以进行音乐市场的分析以及大众音乐审美的研究，甚至是流行音乐发展史中计盒维数的变化的研究，这也是下一步工作的一个重点。

七、参考文献

[1] Cano P，Batlle E，Kalker T，et al. *A review of audio fingerprinting[J]*.Journal of VLSI Signal Processing，2005，41：P271-284.
[2]佚名，MFCC---语音 MFCC 特征的原理介绍，百度文库，http://wenku.baidu.com/view/7ef5ebfb04a1b0717fd5dd1a.html，2013 年 4 月 13 日
[3] 吴斌，沈廷根，宋雪桦，等. 基于噪声环境下的 MFCC 特征提取[J]. 微计算机信息，2008，1：P095.
[4] 边肇祺，张学工等.模式识别 第三版 清华大学出版社 2010 年 8 月
[5] 佚名，MFCC---语音 MFCC 特征的原理介绍，百度文库，http://wenku.baidu.com/view/7ef5ebfb04a1b0717fd5dd1a.html，2013 年 4 月 13 日
[6] 关欣，音乐信号自动分类相关算法研究，天津大学博士学位论文，2008 年 6 月
[7] 史峰等，MATLAB 神经网络 30 个案例分析，北京：北京航空航天大学出版社，2010
[8] 史峰等，MATLAB 神经网络 30 个案例分析，北京：北京航空航天大学出版社，2010
[9] Cano P，Batlle E，Kalker T，et al. *A review of audio fingerprinting[J]*.Journal of VLSI Signal Processing，2005，41：P271-284.
[10] 柯世杰，岳振军，分形理论在语音信号处理中的应用，http://wenku.baidu.com/view/0d4c90333968011ca300917a.html，2013 年 4 月 12 日
[11] Falconer, Kenneth. Fractal Geometry: *Mathematical Foundations and Applications*. John Wiley & Sons, Ltd. 2003: xxv. ISBN 0-470-84862-6.
[12] 佚名，分形维数，维基百科，http://zh.wikipedia.org/wiki/%E5%88%86%E5%BD%A2%E7%BB%B4%E6%95%B0 ，2013 年 4 月 13 日
[13] Bigerelle M，Iost A. *Fractal dimension and classification of*

music[J].Chaos，Solitons and Fractals，2000，11（14）：P2179-2192.
[14] 佚名，豪斯多夫维数，维基百科，http://zh.wikipedia.org/wiki/%E8%B1%AA%E6%96%AF%E5%A4%9A%E5%A4%AB%E7%BB%B4%E6%95%B0，2013 年 4 月 13 日
[15] Boshoff H F V. *A Fast Box Counting Algorithm for Determining the Fractal Dimension of Sampled Continuous Functions[C]//Proceedings of the 1992 South African Symposium on Communication sand Signal Processing*. [S. l.]: ACM Press, 1992.
[16] 佚名，计盒维数，维基百科，http://zh.wikipedia.org/wiki/%E8%AE%A1%E7%9B%92%E7%BB%B4%E6%95%B0 ，2013 年 4 月 13 日
[17] 吕铁军，郭双冰，肖先赐. 调制信号的分形特征研究[J].中国科学：E 辑，2001，31(6)：508-513.Lü Tiejun, Guo Shuangbing, Xiao Xianci. *Research on fractal features of the modulated signal [J]*. Science in China: Series E, 2001, 31(6): P508-513. (in Chinese)
[18] 佚名，分形盒维数，百度文库，http://wenku.baidu.com/view/e59b280b6c85ec3a87c2c5b4.html，2013 年 4 月 13 日
[19] 佚名，信息过载，百度百科，http://baike.baidu.com/view/1150171.htm，2013 年 4 月 12 日
[20] 佚名，协同过滤，维基百科，http://zh.wikipedia.org/wiki/%E5%8D%94%E5%90%8C%E9%81%8E%E6%BF%BE#.E7.BC.BA.E9.BB.9E，2013年4月13日
[21] 于洪，李转运，基于遗忘曲线的协同过滤算法，南京大学学报自然科学版，2010，2010 年第五期：P520-527
[22] 王星 非参数统计 ，北京：清华大学出版社 2009

附录

1、MFCC 系数提取程序

```
function r =mfcc(s, fs)
nbFrame = floor((l - n) / m) + 1;

for i = 1:n
        for  j = 1:nbFrame
                M(i,  j) = s(((j - 1) * m) + i);
        end
end

h = hamming(n);

M2 =  diag(h) * M;
```

数学中国提供（www.madio.net）

```
for i = 1:nbFrame
        frame(:,i)  = fft(M2(:, i));
end

t = n / 2;
tmax = 1 / fs;

m = melfb(24, n, fs);
n2 = 1 + floor(n / 2);
z = m * abs(frame(1:n2, :)).^2;

r = dct(log(z));
r=r(1:15,:);

[x1 x2]=vad(x);
r=r(x1-2:x2-2,:);

d = zeros(size(r));
for i=2:size(r,1)-2
  d(i,:) = r(i+1,:)-r(i-1,:);
end
d = d / 2;

r = [r d];
r = r(3:size(r,1)-2,:);
```

2、计盒维数计算函数程序：

```
function D=FractalDim(y,cellmax)


if cellmax<length(y)
error('cellmax must be larger than input signal!')
end
L=length(y);  y_min=min(y);
y_shift=y-y_min;
x_ord=[0:L-1]./(L-1);
xx_ord=[0:cellmax]./(cellmax);
y_interp=interp1(x_ord,y_shift,xx_ord);
ys_max=max(y_interp);
factory=cellmax/ys_max;
yy=abs(y_interp*factory);

t=log2(cellmax)+1;
```

数学中国提供（www.madio.net）

```
Ne=0;
cellsize=2^(e-1);
NumSeg(e)=cellmax/cellsize;

for j=1:NumSeg(e)
begin=cellsize*(j-1)+1;
tail=cellsize*j+1;
seg=[begin:tail];
yy_max=max(yy(seg));
yy_min=min(yy(seg));
up=ceil(yy_max/cellsize);
down=floor(yy_min/cellsize);
Ns=up-down;
Ne=Ne+Ns;

end

N(e)=Ne;
end

r=-diff(log2(N));
id=find(r<=2&r>=1);
Ne=N(id);
e=NumSeg(id);

P=polyfit(log2(e),log2(Ne),1);
D=P(1);
```

3、核密度估计函数：

```
function p = Parzen(x, X, h1, N)
        hN  = h1 / sqrt(N);
        sum  = zeros(1, 100);
        for  i = 1:N
                sum  = sum + normpdf((x - X(i))/hN, 0, 1);
        end
        p  = sum/(N * hN);
end
```

4、SOM 神经网络分类

```
%%  清空环境变量
clc
clear
```

数学中国提供（www.madio.net）

```
%%  录入输入数据
%  载入数据
num=110;

disp('正在计算识别模型...')
for  i=1:num
      fname  =  sprintf('%ss%d.wav',  'data\train\',  i);
      [s,  fs]  =  wavread(fname);
        v  =mfcc(s,  fs);
        [n,m]=size(v);
        a=reshape(v,1,n*m);
        b=a(1:15);
ref(i).p  =  b;

end
p=ref(1).p;
for  i=1:num-1
       p=[p;  ref(i+1).p];
end


%转置后符合神经网络的输入格式
p=p';

%%  网络建立和训练
%  newsom 建立 SOM 网络。minmax（P）取输入的最大最小值。竞争层为 6*6=36
个神经元
net=newsom(minmax(p),[25  25]);
plotsom(net.layers{1}.positions)
%  5 次训练的步数
a=[10  30  50  100  200  500  1000];
%  随机初始化一个 1*10 向量。
yc=rands(7,15);
%%  进行训练
%  训练次数为 10 次
net.trainparam.epochs=a(1);
%  训练网络和查看分类
net=train(net,p);
y=sim(net,p);
yc(1,:)=vec2ind(y);
plotsom(net.IW{1,1},net.layers{1}.distances)

%  训练次数为 30 次
```

数学中国提供（www.madio.net）

```
net.trainparam.epochs=a(2);
%  训练网络和查看分类
net=train(net,p);
y=sim(net,p);
yc(2,:)=vec2ind(y);
plotsom(net.IW{1,1},net.layers{1}.distances)

%  训练次数为 50 次
net.trainparam.epochs=a(3);
%  训练网络和查看分类
net=train(net,p);
y=sim(net,p);
yc(3,:)=vec2ind(y);
plotsom(net.IW{1,1},net.layers{1}.distances)


%  训练次数为 100 次
net.trainparam.epochs=a(4);
%  训练网络和查看分类
net=train(net,p);
y=sim(net,p);
yc(4,:)=vec2ind(y);
plotsom(net.IW{1,1},net.layers{1}.distances)


%  训练次数为 200 次
net.trainparam.epochs=a(5);
%  训练网络和查看分类
net=train(net,p);
y=sim(net,p);
yc(5,:)=vec2ind(y);
plotsom(net.IW{1,1},net.layers{1}.distances)

%  训练次数为 500 次
net.trainparam.epochs=a(6);
%  训练网络和查看分类
net=train(net,p);
y=sim(net,p);
yc(6,:)=vec2ind(y);
plotsom(net.IW{1,1},net.layers{1}.distances)

%  训练次数为 1000 次
net.trainparam.epochs=a(7);
```

数学中国提供（www.madio.net）

```
%  训练网络和查看分类
net=train(net,p);
y=sim(net,p);
yc(7,:)=vec2ind(y);
plotsom(net.IW{1,1},net.layers{1}.distances)
yc

%%  网络作分类的预测
%  测试样本输入
disp('正在计算测试识别的结果...')
for  i=1:num
      fname  =  sprintf('%ss%d.wav',  'data\test\',  i);
      [s,  fs]  =  wavread(fname);
        v  =mfcc(s,  fs);
        [n,m]=size(v);
        a_test=reshape(v,1,n*m);
        b_test=a_test(1:15);
        test(i).p_test  =  b_test;
end
p_test=test(1).p_test;
for  i=1:num-1
        p_test=[p_test  ;test(i+1).p_test];
end
p_test=p_test';
  %  sim(  )来做网络仿真
r=sim(net,p_test);
%  变换函数  将单值向量转变成下标向量。
rr=vec2ind(r)

%%  网络神经元分布情况
%  查看网络拓扑学结构
plotsomtop(net)
%  查看临近神经元直接的距离情况
plotsomnd(net)
%  查看每个神经元的分类情况
plotsomhits(net,p)
```

数学中国提供（www.madio.net）