

第六届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn

Email: 2013@tzmcm.cn

参赛队号 # 2562

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：**2562**

参赛队员（签名）：

队员 1：林炆平

队员 2：张晓鹏

队员 3：亓秦

参赛队教练员（签名）： 无

参赛队伍组别：本科组

第六届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn

Email: 2013@tzmcm.cn

参赛队号 # 2562

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）：

2562

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）：

第六届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn

Email：2013@tzmcm.cn

参赛队号 # 2562

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 MP3 MDCT 学习分类器 螺旋情感模型 K 近邻法

摘 要：

音乐的自动分类是现代检索技术的一个研究内容,也是音乐可视化研究中亟待解决的问题。根据对网络音乐的调查,目前值得关注的分类方式有两种,一种是当下仍占主流的传统分类方式(如流行,古典等),一种是日益受到关注的情感分类方式(如伤感,欢快等)。针对这两种不同的分类方式,本文将建立两种模型加以分析。

模型一,针对传统的分类方式。了解到现下音乐的格式大部分是 MP3 格式的,所以,模型一只针对 MP3 格式的音乐文件进行分类。针对 MP3 编解码的特点,用基于离散余弦变换(MDCT)系数域 3 种特征参数的特征片段提取方法提取 MP3 音乐特征片段,表示为特征向量,选择用 LCS(学习分类器)进行分类,从而达到分类目的。

模型二,螺旋情感模型,以音乐情感为主导,通过 matlab 提取音频片段的 FFT 值以提取特征量音乐强度,再结合前后搜索法、最大公约数法等提取节奏速度,对此二维特征用层次 K 近邻法构建了歌曲情感识别算法,最后通过 matlab 与 java 程序求取参数并验证。最终验证音乐可较严格地分为激情、热情、抒情、忧伤、快乐、轻盈、柔和与平静 8 类。

第六届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn

Email: 2013@tzmcm.cn

参赛队号 # 2562

参赛队号： 2562

所选题目： B 题

参赛密码 _____
(由组委会填写)

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%)

第六届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn

Email：2013@tzmcm.cn

参赛队号 # 2562

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。请你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。

二、问题背景

随着计算机技术和通信技术的飞速发展, 各式各样的信息急速的增长, 人们也时时刻刻的能接触到大量多媒体形式的内容, 如图像、音频、视频等。但是随着数据量的快速增长, 如何自动对这类内容进行快速管理就成为了一个亟待解决的问题。特别是对身边大量的音乐信息, 人们需要快速高效的方法对它们进行分类和管理, 以便更好的应用在音乐推荐、KTV点唱及在线选歌等诸多领域中。

由于MP3音乐在数字音乐世界中大为流行, 因此在互联网上有越来越多的MP3歌曲可以自由下载, 于是MP3的音乐收藏于整理变成了一个重要的课题。对不明身份的MP3音乐文件进行自动分类与整理时本文关注的问题。本文利用MP3音乐文件在编解码过程中的核心阶段多相位滤波以及MDCT变换体现的特点, 对MP3音乐进行特征表示, 并采用LCS针对音乐特征进行分类。

歌曲情感识别属于音乐情感计算的研究范畴, 音乐情感计算的基本方法是将音乐情感归并为(诸如平静、悲伤、激动和愉悦等)几大类, 然后建立音乐情感模型, 再从音乐中提取情感特征, 最后采用分类算法进行情感识别。常用的音乐情感模型有离散形式的Hevner 情感模型, 该模型是针对西方音乐作品提出的。由于东西方文化背景和思维方式的差异, 它不完全符合中国人的情感表达方式。

三、问题分析

由题意知, 该题目的问题实质是建立一种模型, 实现对音乐的分类。调查得知, 当下的音乐分类标准很多, 一种是传统的音乐分类方法, 如(流行, 古典, 摇滚...)但考虑到当下听众的客观需求以及现实状况, 基于情感标签的分类方法(如伤感, 快乐...)也越来越受到重视。所以本文将从这两种思路出发, 建立两种模型, 分别实现

第六届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn

Email: 2013@tzmcm.cn

参赛队号 # 2562

对音乐的分类。

针对传统音乐分类，本文采取模型一的分类方法。模型一针对 MP3 格式音乐的编解码特点，建立模型进行分类。在 MDCT 域提取 MP3 音乐特征片段，对提取的音乐片段，构造音乐特征向量，再将音乐向量与音乐类别建立联系，可用 LCS（学习分类器）进行分类，从而达到分类目的。

从模型二的角度来看，时下流行的 Hevner 情感模型不合中国人审美，Thayer 情感模型不够精确，因此本文通过在极坐标上对 Thayer 扩展，建立螺旋情感模型。由于来自 Thayer，本模型亦有两个属性，分别为音乐强度和节奏速度。根据最终训练数据的“同类型聚集”特性，可用 K 邻近法分析。

四、模型的假设

1、总体假设：

- 1) 音乐文件在处理分类过程中不会有损坏；
- 2) 所取的训练数据不会有过多的极端的多种类型复杂音乐；
- 3) 所有数据均为原始数据，来源真实可靠。

2、模型一补充假设：

假设所选音乐文件都有高潮部分，相对平缓部分，背景声组成

3、模型二补充假设：

- 1) 大多数音乐的频谱能量差不会有极端的集中出现；
- 2) 假设所取音乐都有主体情感部分；

五、符号说明

模型一符号说明

序号	符号	意义
1	f_a	MP3 文件的取样率
2	f_{\max}	最大频率
3	$RF_{(f)}$	真实频率值
4	TTH	当前 frame 的能量值
5	QER	相邻两个 frame 的关系
6	f_c	定点频率
7	$F\langle f_c \rangle$	频率 f_c 在频率序列上的索引
8	SC	平衡频率点
9	NoW	窗口数
10	NoF	窗口参数个数
11	n	窗口内参数个数

第六届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn

Email：2013@tzmcm.cn

参赛队号 # 2562

12	MDCT 【i】	MP3 文件进行解码过程中进行到 IMDCT 前第 i 个频带的系数
13	MDCT _{mean}	MDCT 系数的均值
14	MDCT _{deviation}	MDCT 系数的均方差

模型二符号说明：

序号	符号	意义
1	$S(n)$	为第 n 帧的声音强度
2	S_v	歌曲片段的聲音强度
3	$F_k(n)$	输入歌曲片段的第 k 帧的 FFT（快速傅里叶变换）的值
4	L	输入的歌曲片段的帧长度
5	L_p	输入歌曲片段的长度
6	$E(m, n)$	歌曲片段第 n 子频带的第 m 帧的频谱能量
7	$F_k(m, n)$	歌曲片段的第 n 子频带的第 m 帧的第 k 个 FFT 的值
8	$D_{\text{value}}(m, n)$	歌曲片段第 n 子频带中的第 m 帧与第 m-1 帧的频谱能量差值
9	$D_{\text{result-value}}(m, n)$	歌曲片段第 n 子频带的第 m 帧中的击点的个数
10	$H(m, n)$	表示第 n 子频带的第 m 帧是否有击点的标志位
11	$H_{\text{result}}(m)$	表示第 m 帧中有多少个子频带上有击点
12	f	表示节奏击点序列的周期
13	dis	二维空间中两点的欧几里得距离

六、模型的建立

（一）模型一的建立：

1. MP3 音乐片段提取过程

1.1>音乐特征片段提取方法

经过查资料得，MPEG 标准的音频部分制定了 3 层压缩算法，其中 MP3 是 Mpeg 一 1 Layer 3 的简称。MP3 编码过程分为 5 个部分：1) 多相位分析滤波；2) MDCT (改良离散余弦变换)；3) 第 2 层声线模型建立；4) 量化编码；5) 哈夫曼编码和 CRC 编码校验。其核心步骤是多相位滤波和 MDCT 变换。

和编码相反，MP3 解码主要经过以下几个步骤：对于输入的码流，首先解析文件头信息、得到量化因子，然后解霍夫曼编码、反量化，熏组、进行立体声处理、降低真处

第六届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn

Email：2013@tzmcm.cn

参赛队号 # 2562

理，最后经过改良离散余弦反变换(IMDCT)和多相合成滤波器，输出 PCM 码流。

为了减少运算量，提到信息检索效率，故选择在 MDCT 域上进行特征提取。在 MP3 音乐特征提取过程中，需找到在本文件中最具有“特征”的片段，经过了解，选择以下 3 个特征参数：1>TTH 2>QER 3> SC, 对音乐进行特征片段提取。

根据第二层声音心理模型，MDCT 运算时的窗口数目的选择方法如下表 1：

表 1 MDCT 运算的窗口数目

窗口模式	窗 口 数 (NoW)	窗 口 参 数 个数 (NoF)
长窗	1	576
短窗	3	216

假设 f_a 为 MP3 文件的取样率，根据尼奎斯特 (Nyquist) 理论，信号最大频率 $f_{\max} = f_a / 2$ 。由于 MP3 使用线性空间频率，真实的频率值 $RF_{(f)}$ 可表示为非混合式和混合式，即非混合式为

$$RF(n) = (n+1) \times f_{\max} / NoF \quad (1)$$

混合式为

$$RF_{(n)} = \begin{cases} (n+1) \times f_{\max} / NoF, & n < 36 \\ \frac{f_{\max}}{16} + \frac{(f-35) \times f_{\max}}{192}, & n \geq 36 \end{cases} \quad (2)$$

(1) (2) 式中, n 表示窗口内参数个数 ($0 \leq n \leq NoF$)

TTH 定义了当前 frame 的能量值，

其计算式为

$$TTH_j = \sqrt{\sum_w \sum_n (MDCT_j(w, n))^2} \quad (3)$$

这是第一个分段依据参数去判定当前 frame 是否处于平缓期。
QER 定义了相邻两个 frame 间的关系，其计算式为

第六届数学中国数学建模网络挑战赛

地址: 数学中国数学建模网络挑战赛组委会

电话: 0471-4969085

邮编: 010021

网址: www.tzmcm.cn

Email: 2013@tzmcm.cn

参赛队号 # 2562

$$QER(f_c) = \frac{\sqrt{\sum_w \sum_{n=0}^{NoF} (MDC\mathcal{T}(w,n))^2}}{\sqrt{\sum_w \sum_{n=f < f_c}^{NoF} (MDC\mathcal{T}(w,n))^2}} \quad (4)$$

(4) 式中, f_c 为定点频率 ($f_c \leq f_{\max}$); $f < f_c$ 为频率 f_c 在频率序列上的索引, 而且

$$RF_{(f < f_c)} \leq f_c \leq RF_{(f < f_c + 1)}$$

SC 定义了平衡频率点, 也就是频率序列上的某个值, 其计算式为

$$f_{sc} = \frac{\sum_w \sum_n^{NoF} (MDC\mathcal{T}_f(w,n) \times RF_{(n)})}{\sum_w \sum_n^{NoF} MDC\mathcal{T}_f(w,n)} \quad (5)$$

1.2>特征片段提取

假设音乐文件由高潮部分, 相对平缓部分, 背景音乐声 3 部分组成, MP3 音乐特征提取问题实质就是将一个个的 frame 归类为 3 部分中的一个, 然后将比较密集的归为音乐高潮部分的 frame 融合在一起, 组成 MP3 音乐特征片段。

步骤 1>. 初始分类。每个 granule 被分成为高潮部分, 平缓部分或背景声的分类之一。根据 (3) 式计算 TTH 的值。设定阈值 T_{TTH} 来判定当前 granule 是否属于背景声。

$$T_{TTH} = E_{\min} + \lambda_s \times (E_{\mu} - E_{\min}), \quad 0 < \lambda_s \leq 1 \quad (6)$$

最小能量值 (E_{\min}), 最大能量值 (E_{\max}) 和 granule 能量值 (E_{μ}) 在第 1 次对所有 granule 进行能量值计算时进行统计。 λ_s 称为静音参数, 决定了静音阈值。当计算某个 granule 的 TTH 值小于阈值 T_{TTH} 时, 认为此 granule 属于背景声, 以此区分每个 granule 属于背景声或非背景声。

步骤 2>. 对已经确认为非背景的音乐片段进行 QER 计算。将 QER 的计算结果从高到低排列, 由于 QER 表示的是相邻 frame 之间的关系, QER 系数越大代表变化越大; 将 QER 大的 frame 作为划分高潮和平缓的边界, 随后以当前 frame 为中心, 计算改 frame 相邻 frame 的能量值, 较大能量值的 frame 代表高潮部分, 较小的代表平缓部分; 最后根据当前的 QER 较大的值得位置寻找狭义的较大的 QER 锁子啊位置, 2 个 QER 之间的片段可以作为特征片段。

步骤 3>. 对密集的属于高潮部分的连续 frame 进行作何, 形成 MP3 特征片段。

以下实际进行以上实验步骤。

一个完整的 frame 进行 2 次相位滤波, 得到 36×32 个 (36 是按照时域排列的, 32

第六届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn

Email: 2013@tzmcm.cn

参赛队号 # 2562

是按照频率排列的)滤波后的样本。这里可以将 MDCT 域参数当做特征向量, 一个 frame 在进行 MDCT 运算后共有 576 个系数, 表示这个 frame 需要有一个 576 维的向量。

定义 MDCT **【i】** 代表着 MP3 文件进行解码过程中进行到 IMDCT 前第 i 个频带的系数:

$$MC_i = (MDCT[i])^2, i = 1, 2, \dots, 576 \quad (7)$$

同理, 假设连续的 n 个 frame 进行特征表示, 那么 PMCV (phoneme MDCT coefficient vector) 表示一个序列上连续 n 个 frame 的特征向量:

$$PMCV = (\sum_{i=1}^n MC_1, \sum_{i=1}^n MC_2, \dots, \sum_{i=1}^n MC_{576}) \quad (8)$$

定义 PMCV 可以方便地计算 2 个特征向量之间的距离, 其计算式为

$$dist(X, Y) = \sqrt{(x_1 - y_1)^2 + \dots + (x_m - y_m)^2} \quad (9)$$

(9) 式中, m=576.

对摇滚, 流行 2 种风格, 各 3 首歌进行特征提取, 得到如下结果。

表 1 采用摇滚乐风格歌曲的比较 (随机片段)

歌曲	随机片段 1-7							均值	方差
曲 1	6737	8645	9630	1200 4	7321	1356 8	8230	8900	2301 .2
曲 2	8150	8620	1145 3	1609 3	9807	1290 8	6787	9478	3023 .1
曲 3	1460 0	7200	8301	9786	9757	9870	1512 2	1126 3	2784 .3

表 2 采用摇滚乐风格歌曲的比较

歌曲	随机片段 1-7							均值	方差
曲 1	6670	2356	5676	4565	2323	1203	3345	3501	1684 .1
曲 2	3907	5832	4323	3489	5807	2908	2655	4177	1082 .5
曲 3	5303	4200	2178	3332	9997	9870	8977	5737	2866 .3

第六届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn

Email: 2013@tzmcm.cn

参赛队号 # 2562

表 3 采用流行风格歌曲的比较（随机片段）

歌曲	随机片段 1-7							均值	方差
曲 1	9687	6857	5878	1320 5	9652	8723	7893	9910	2859 .1
曲 2	8985	1232 0	1354 8	8924	7873	8345	1978 7	1444 1	6683 .0
曲 3	5600	6598	8301	9756	9997	2345 3	1232 2	1315 0	6613 .4

表 4 采用流行风格歌曲的比较（选择片段）

歌曲	随机片段 1-7							均值	方差
曲 1	5625	6548	1253	1565	1523	1603	2568	3121	1880 .1
曲 2	7452	6895	2345	6763	2769	4376	4121	5260	1867 .0
曲 3	5303	4200	2732	3476	5643	1889	3465	4136	1262 .5

2. 基于学习分类器（LCS）的分类方法

标准学习分类器的框架爱如图 1 所示，在一次学习过程中，首先由探测器(Detector)检测当前环境（Environment）得到环境状态（State），并将之按一定格式编码放入消息表（Message List）中，而后通过消息（Message）的匹配从规则库中选择出匹配集；

第六届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn

Email：2013@tzmcm.cn

参赛队号 # 2562

如果没有则通过遗传算法生成，并将产生的新分类器置入规则库中（若规则库满则根据一定的算法做删除操作），这些分类器产生新的消息，可能触发分类器也可能引发动作（Action）；如果是前者则不断重复上述过程知道产生动作，动作出现后通过作用器（Effector）作用于环境。如果从环境中得到奖赏（Reward），则根据定义的算法进行奖赏分配。如果分类器符合一定条件，则通过遗传算法生成新分类器。本轮学习完毕。

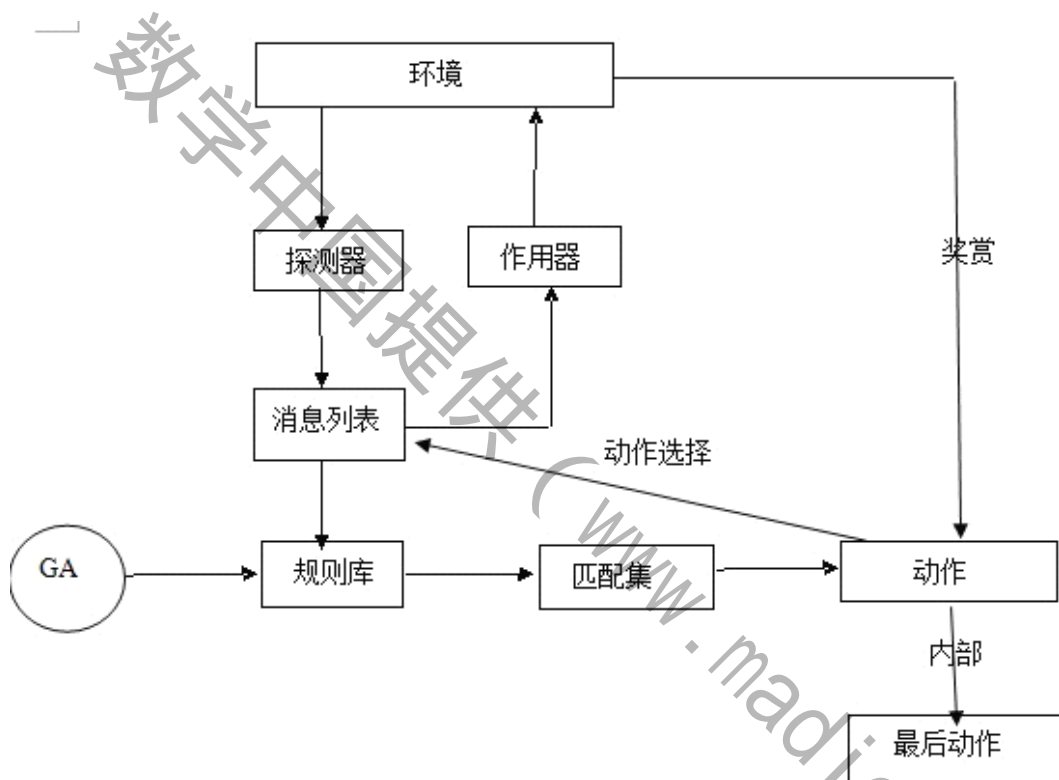


图1 标准学习期框架图

对上述的音乐特征提取，可以采用学习分类器对已经提取的音乐特征向量进行分类。其基本步骤是：

- 1) 把提取的特征组织成分类学习系统所需要的格式，（形如：属性 1，属性 2，…，属性 n ，类别）。
- 2) 以训练样本的属性作为输入，与规则库中的规则匹配。规则的形式如： $\{[c_1, s_1], [c_2, s_2], \dots, [c_n, s_n], \text{Action}\}$ ，这里 c 表示中心， s 表示偏离中心的振幅。如存在相匹配的规则（否则以此样本为参照，通过 GA 产生相应的规则，并将之加入规则库中，如规则库满则此处可能存在删除操作，见步骤 4），则根据一定的标准选择出动作集。
- 3) 将步骤 2) 所产生的动作作用于环境（这里即与真实的类别进行比较），并将环境的反馈作用于规则库，调整相应规则的参数或权重。如果调整后的规则的权重低于一定的阈值，则将其删除。
- 4) 在对训练样本重复步骤 2)、3) 完成训练之后，以未知样本作为输入，从而标记出它的

第六届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn

Email: 2013@tzmcm.cn

参赛队号 # 2562

类别(步骤与 2)、3) 相似, 但没有环境反馈规则更新与 GA)。

对于一首 MP3 音乐歌曲, 特征片段有很多处, 进行特征向量表示时, 我们将特征片段的特征向量(包含若干个 frame)排列成序列, 歌曲由特征向量序列表示。特征向量序列的表示由一个一个 frame 排列而成, 而这些连续的 frame 来自于歌曲的特征明显的地方。

依据 MP3 音乐编解码特性, 对音乐风格分类, 将音乐类库中的音乐文件划分成若干风格面对不同风格的音乐文件进行归类。在针对音乐风格进行分类时, 取频率范围 100Hz—5kHz 的 18 个频率段。定义 1 个 frame 组包含 100 个连续的 frame, 来简化特征向量的表示。那么, 1 组 frame 进行 MDCT 变换后, 敏感声音频率的 MDCT 系数的均值为

$$MDCT_{mean} = \frac{\sum_{i=1}^{100} (\sum_{j=100}^{5000} (MDCT[i, j]))}{18 \times 18 \times 100} \quad (11)$$

(11) 式中, $j=100, 125, 160, \dots, 5000$.

均方差为

$$MDCT_{deviation} = \sqrt{\frac{\sum_{i=1}^{18 \times 18 \times 100} (MDCT_i - MDCT_{mean})^2}{18 \times 18 \times 100}} \quad (12)$$

那么, 由于每个特征向量片段含有 n 个 frame 组, 1 个特征片段音乐特征向量为

$$MDCT_{segment} = \{MDCT_{mean}[1], MDCT_{deviation}[1], \dots, MDCT_{mean}[n], MDCT_{deviation}[n]\} \quad (13)$$

由于一首歌曲有 m 个特征片段, 每个特征片段含有 N 个 frame 组, 那么一首歌曲的音乐特征向量重新表示为

$$\{MDCT_{segment}[1], \dots, MDCT_{segment}[m], Classification\} \quad (14)$$

(14) 式中, Classification 表示音乐类别。

(二) 模型二的建立:

建立以歌曲声音强度和节奏速度为影响因素的螺旋情感模型, 如图 2 所示。由歌曲声音强度为极坐标构建一个螺旋情感空间, 极坐标表示声音强度的高低, 而围绕螺旋线顺时针方向表示节奏速度的快慢。根据声音强度的大小和节奏速度的快慢, 将歌曲情感分成 8 大类, 即鸡表 1 所示的激情、热情、抒情、忧伤、快乐、轻盈、柔和与平静。

第六届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn

Email: 2013@tzmcm.cn

参赛队号 # 2562

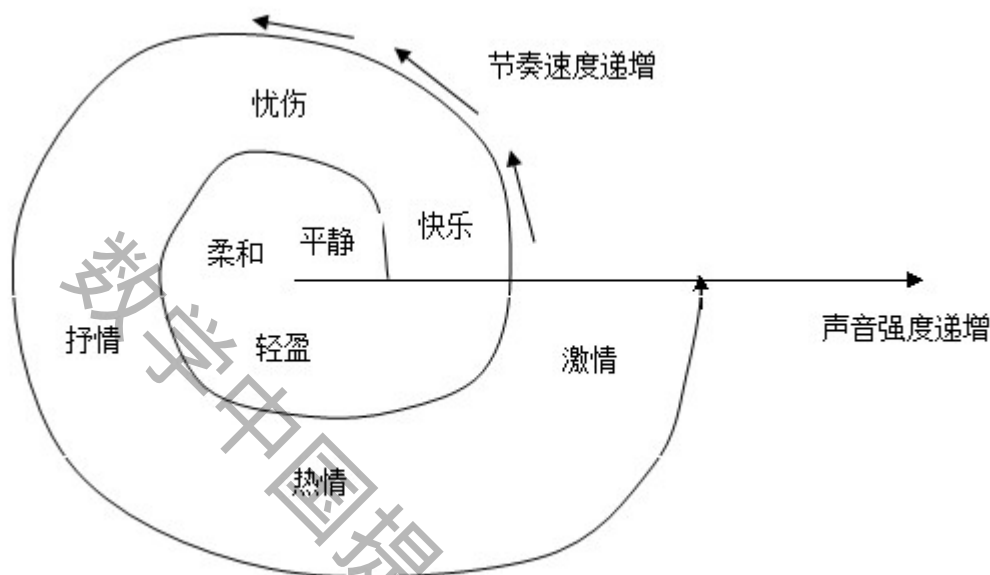


图 2

本文着重采用螺旋情感模型和层次 K 近邻分类器法构建了歌曲情感识别算法。首先提取样本集的声音强度和节奏速度，然后根据螺旋情感模型与相应算法得到特征库，通过测试样本的声音强度和节奏速度与特征库中的特征进行层次 K 近邻分类，得出的分类结果就是歌曲的情感类型。

1、特征提取：

1) 声音强度的提取：

通过 matlab 提取音频片段的第 k 帧的 FFT 的值 $F_k(n)$ ，并提取该片段相应的长度 L_p 和帧长度 L ，

则通过以下公式可得第 n 帧的声音强度：

$$S(n) = \frac{1}{L} \sum_{k=1}^L |\lg(F_k(n))|$$

对此求和可得整个片段的声音强度：

$$S_w = \frac{1}{L} \sum_{k=1}^{L_p} S(n)$$

2) 歌曲片段节奏的提取：

节奏大体可概括为我们所说的节拍强弱感，歌曲的能量在节奏击点处必然会有一定的激变，这种激变发生的周期在一定程度上可以用来表示歌曲的节奏速度（我们可以只考虑达到一定程度的激变，这样既可简化问题又不会严重影响结果的正确性）。因此，我们可以采用一种多频带节奏提取算法来计算歌曲的节奏速度。

第六届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn

Email：2013@tzmcm.cn

参赛队号 #2562

我们的方法分以下几个步骤：

A) 从 $1/2$ 采样频率开始，将歌曲片段分成 8 个子频带 ($[f_s/2-f_s/4]$ 、 $[f_s/4-f_s/8]$ 、 \dots 、 $[f_s/2^8-0]$)；

B) 通过 matlab 计算每个子频带每一帧的 FFT 值（这些 FFT 值可作为当前子频带的帧能量序列）；

C) 对每个子频带的歌曲片段进行预处理，并提取频谱能量的激变点处的能量差，该步骤如下：

a) 求得每一帧的频谱能量：

$$E(m, n) = \sum_{k=1}^N |\lg F_k(m, n)|$$

这里的 $F_k(m, n)$ 在 B) 中求得，这里的 N 为当前帧所有的 FFT 数；

b) 求得相邻两帧的能量差值，即所求的激变点的能量差：

$$D_{\text{value}}(m, n) = E(m, n) - E(m-1, n)$$

图 3 显示了测试歌曲《中国话》(以下称为案例)中第 1 子频的 $D_{\text{value}}(m, n)$ 值的集合的图像。

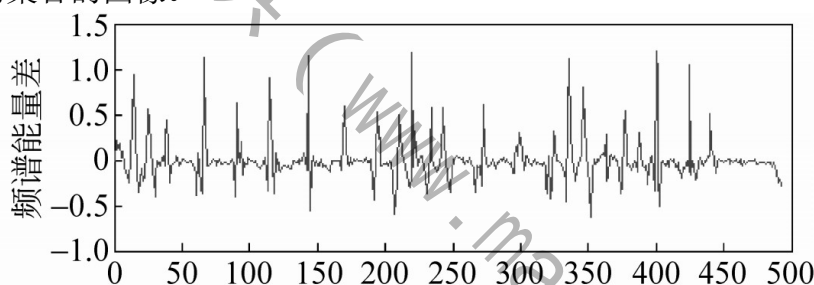


图 3

D) 对上一步得到的激变点能量差进行后处理，对冗余的数据进行剔除以简化计算，并从数据层减少可能造成严重误差的危险因子，只保留峰值序列，得到子带击点序列。

步骤如下：

a) 只留取达到一定程度的正的峰值序列，对负值的、低于平均峰值 $1/2$ 的峰值置为 0；

案例结果如图 4 所示：

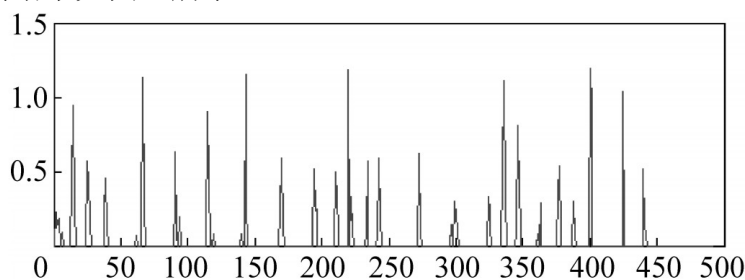


图 4

b) 采用参数为 4 的前后搜索法保留相邻 4 帧的最大峰值作为击点，其余峰值置为 0，

第六届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn

Email：2013@tzmcm.cn

参赛队号 # 2562

公式如下：

$$D_{front-value}(m, n) = \begin{cases} \max(D_{value}(k, n)) & m-2 < k < m+2 \\ 0 & other \end{cases}$$

$$D_{back-value}(m, n) = \begin{cases} \max(D_{value}(k, n)) & N-m-2 < k < N-m+2 \\ 0 & other \end{cases}$$

$$D_{result-value}(m, n) = \sqrt{D_{front-value}(m, n) * D_{back-value}(m, n)}$$

E) 经过上述操作后，得到 8 个子频带的击点序列，对这些击点序列按照一定规则整合成一个综合的击点序列，此击点序列就是节奏击点序列。具体处理如下：

a) 依次判断每个子频带的每一帧上是否有有击点的标志位（1 表示有，0 表示无）：

$$H(m, n) = \begin{cases} 1 & if D_{result-value}(m, n) > 0 \\ 0 & other \end{cases}$$

再求取对于每一帧，有多少个子频带上有击点：

$$H_{result}(m) = \sum_{k=1}^8 H(m, k)$$

b) 设置阈值 θ ，若 $H_{result}(m) < \theta$ ，则将 $H_{result}(m)$ 置为 0，视第 m 帧无击点；若 $H_{result}(m) > \theta$ ，则保留 $H_{result}(m)$ 值。经过此判断可得到 N 个 $H_{result}(m)$ 值（ N 为帧数），在此我们将其记为：

$P_{res}(1)$ 、 $P_{res}(2)$ 、 $P_{res}(3)$ …… $P_{res}(N)$ 。（在此阈值 θ 取 4）

对案例进行上述操作可得到图 5：

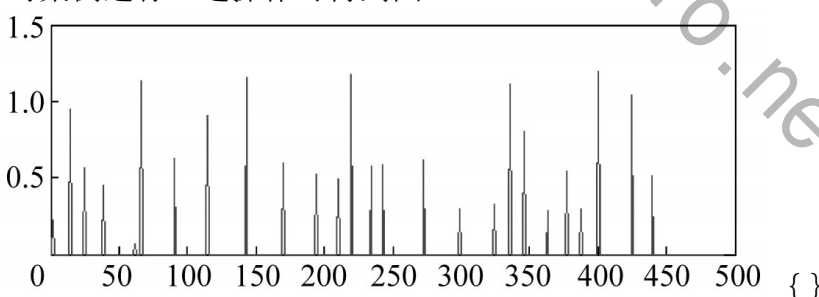


图 5

F) 采用最大公约数法计算节奏击点序列的周期，该周期可与歌曲的节奏速度直接关联公式如下：

$$f = \arg \min_{P_{res}(k)} \sum_{m=1}^N \left| \frac{P_{res}(m)}{P_{res}(k)} - \text{round}\left(\frac{P_{res}(m)}{P_{res}(k)}\right) \right|$$

f 即为所求周期，可用它来代替节奏速度进行接下来的数据分析

第六届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn

Email: 2013@tzmcm.cn

参赛队号 # 2562

2、特征分析：

在后面寻找训练数据集的过程中我们会发现，我们所维护的数据有“同类型多聚集，异类型少交叉”的特点，我们了解到这种特性的数据较适合用以空间距离为分类准则的分类方法进行分类。因此，本文将采用较经典的 K 近邻算法对数据进行分析与测试。

既然本文之前提取了歌曲的两个属性（声音强度和节奏速度），那么本文必然根据这两个属性执行 K 近邻算法进行情感分类。

1) K 近邻算法的步骤如下：

(1) 寻找适当的训练数据集

本文在此使用的训练数据来自于 312 首歌曲中的 510 个经典歌曲片段。已确定这 312 首歌曲的情感单一，一首歌基本上只表达一种感情类型，且情感主体部分比较明确，上述的 510 个片段也都是截取情感主体部分的 30s 来作为训练样本的。我们对这 510 个片段基于螺旋情感模型进行情感标注，标注后的实验数据集分布如表 3 所示：

类别	片段数	总数
平静	55	510
忧伤	56	
柔和	69	
抒情	64	
轻盈	67	
热情	65	
快乐	57	
激情	58	

表 3

对 510 个片段分别进行数据提取并记录，可得实验数据集聚类情况如图 6 所示。在该图中我们不难发现，相同情感类型的歌曲片段聚集在一起，且相邻情感类型之间存在较少交叉，因此我们使用 K 近邻算法是合适的。

第六届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn

Email：2013@tzmcm.cn

参赛队号 # 2562

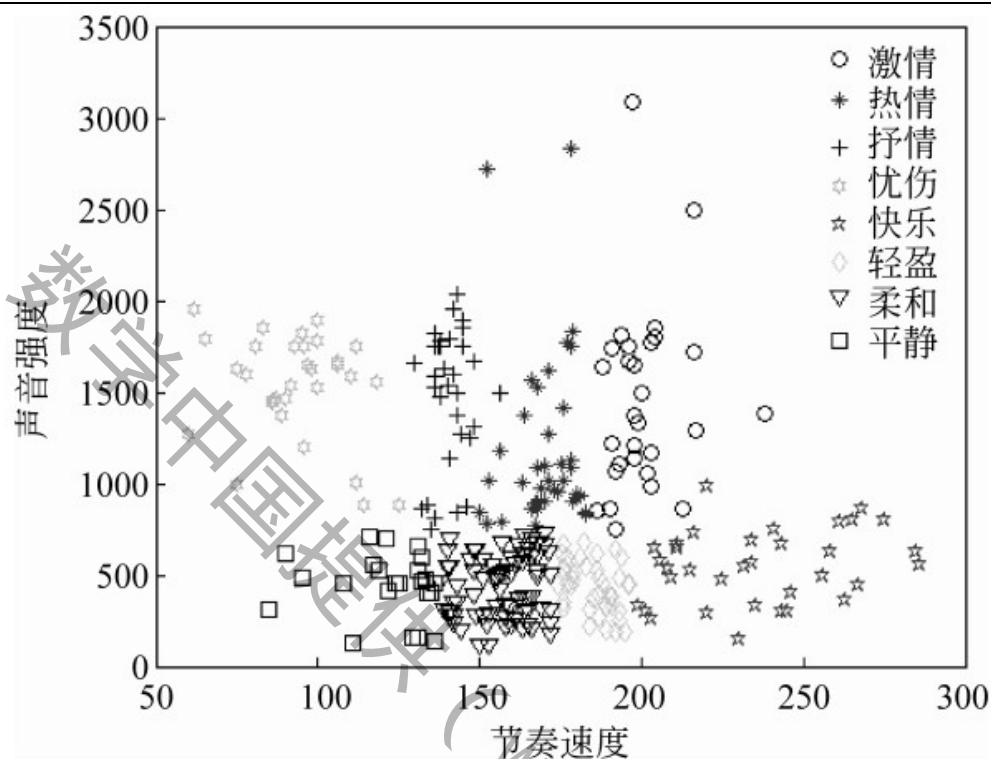


图 6

(2) 确定距离函数

距离函数决定了哪些样本是待分类样本的 K 个最近邻居。在本案例中，样本是空间中点，且为二维空间，所以我们可以使用二维欧几里德距离。

$$\text{即: } dis = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

具体实现时可所有距离进行排序，取前 K 个进行分析。具体算法将在“结果与分析”中出现。

(3) 决定 K 的取值

邻居的个数对分类的结果有一定的影响，一般先确定一个初始值，再进行调整，直到找到合适的值为止。我们对数据进行测试与分析时发现，对本案例 $K=5$ 时为最优。具体过程与结果将在“结果与分析”中出现。

(4) 综合 K 个邻居的类别

我们再次选取简单有效的多数法进行选择，即从邻居中选择一个出现频率最高的类别作为最后的结果，如果频率最高的类别不止一个，就选择最近邻居的类别。

2) 求解模型如下：

输入：训练数据集 $D = \{(X_i, Y_i), 1 \leq i \leq N\}$ ，其中 X_i 是第 i 个样本的条件属性， Y_i 是类别，新样本 X ，距离函数 dis 。

伪代码：（附录附有用 java 实现的详细代码）

输出： X 的类别 Y 。

for $i=1$ to N do

 计算 X 和 X_i 之间的距离 $dis(X, X_i)$;

第六届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn

Email: 2013@tzmcm.cn

参赛队号 # 2562

end for

对距离排序，得到 $d(X, X_{i1}) \leq d(X, X_{i2}) \leq \dots \leq d(X, X_{iN})$ ；选择前 K 个样本： $S = \{(X_{i1}, Y_{i1}) \dots (X_{iK}, Y_{iK})\}$ ；统计 S 中每个类别出现的次数，确定 X 的类别 Y 。

六 结果与分析

1、模型一结果与分析：

我们在歌曲中提取 8 个特征片段，同时每个特征片段中提取连续的 4 个 frame 组，根据(11)，(12)，(14)式计算每个音乐文件的 65 维特征向量。实验针对音乐风格进行分类，我们将音乐风格分为 4 种，流行（Pop）、古典（Classical）、新潮（NewAge）、摇滚（Rock），样本共 200 首，每个风格包含 50 首。未指定风格的 MP3 音乐文件 30 首，我们对不明风格的 MP3 音乐文件进行风格分类。

首先对已经表示为特征向量的 MP3 音乐样本文件进行 10 折交叉验证，观察精度和均方差来观察分类器在 MP3 音乐特征样本上的性能表现；然后对未指定风格的 30 首音乐文件进行分类，观察分类器的预测正确率。实验结果如下表所示：

表 5 音乐风格分类实验结果

	10 折交叉验证精度 / %	均方差	对新样本预测的正确率
流行	76	0.185	76.2
古典	70	0.163	70.3
新潮	74	0.200	68.3
摇滚	73	0.193	73.1

2、模型二结果与分析：

1) 对 K 值的确定：

用 Matlab 对歌曲情感识别算法进行训练和测试，实验结果采用交叉验证的方法，数据集的 70% 用作训练数据，30% 用作测试数据。测试时分别对 K 取值 3—9，进行共 7 次测试，记录并比较识别准确率。实验结果如图 7 所示，可看到 $K=5$ 时准确率最高。

第六届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn

Email：2013@tzmcm.cn

参赛队号 #2562

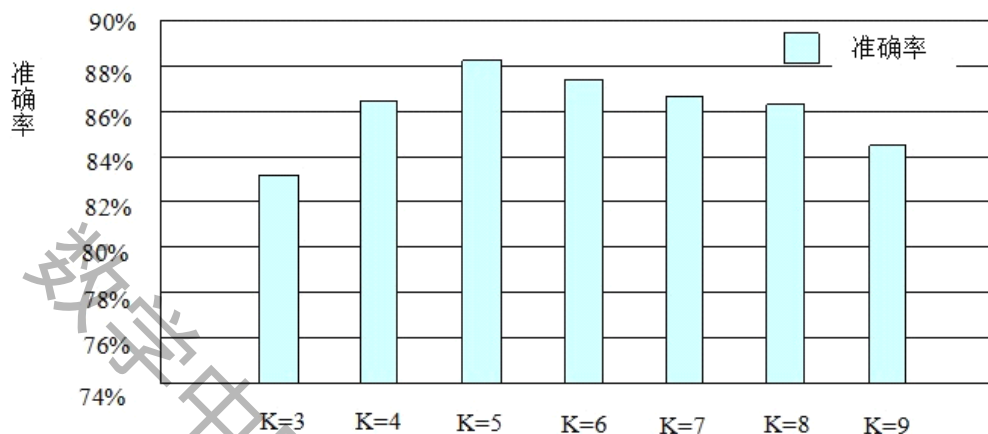


图 7

表 4 给出了基于螺旋模型的歌曲情感识别算法的 K=5 时的实验结果：

类别	平静	忧伤	柔和	抒情	轻盈	热情	快乐	激情
平静	92.9%	0%	7.1%	0%	0%	0%	0%	0%
忧伤	7.7%	76.9%	0%	15.4%	0%	0%	0%	0%
柔和	6.7%	0%	80%	3.3%	6.7%	3.3%	0%	0%
抒情	3.4%	6.9%	3.4%	79.4%	0%	0%	0%	6.9%
轻盈	3.3%	0%	6.7%	3.3%	80.1%	3.3%	0%	3.3%
热情	0%	0%	3.4%	3.4%	0%	89.8%	0%	3.4%
快乐	0%	0%	0%	0%	0%	7.1%	89.3%	3.6%
激情	0%	0%	3.3%	0%	0%	3.3%	3.3%	90.1%
平均准确率	88.2%							

表 4

2) 测试数据的方法与部分测试数据：

使用附录【1】的代码，输入测试数据的声音强度 strength 与节奏速度 speed 运行程序即可得到相应的音乐类别代号（代号为 1 到 8，分别代表：激情、热情、抒情、忧伤、快乐、轻盈、柔和与平静）。

例如（部分测试数据）：

输入的声音强度为 562，节奏速度为 148，可得音乐类型为 7，即柔和型；
 输入的声音强度为 1042，节奏速度为 139，可得音乐类型为 3，即抒情型；
 输入的声音强度为 468，节奏速度为 237，可得音乐类型为 5，即快乐型；
 输入的声音强度为 321，节奏速度为 263，可得音乐类型为 5，即快乐型；
 输入的声音强度为 1562，节奏速度为 200，可得音乐类型为 1，即激情型；
 输入的声音强度为 383，节奏速度为 153，可得音乐类型为 7，即柔和型；
 输入的声音强度为 446，节奏速度为 188，可得音乐类型为 6，即轻盈型；
 输入的声音强度为 196，节奏速度为 170，可得音乐类型为 7，即柔和型；
 输入的声音强度为 1466，节奏速度为 168，可得音乐类型为 2，即热情型；

第六届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn

Email: 2013@tzmcm.cn

参赛队号 # 2562

输入的声音强度为 1699，节奏速度为 213，可得音乐类型为 1，即激情型；
输入的声音强度为 912，节奏速度为 256，可得音乐类型为 5，即快乐型；
输入的声音强度为 1255，节奏速度为 205，可得音乐类型为 1，即激情型；
输入的声音强度为 1430，节奏速度为 68，可得音乐类型为 4，即忧伤型；
输入的声音强度为 1335，节奏速度为 46，可得音乐类型为 4，即忧伤型；
输入的声音强度为 368，节奏速度为 102，可得音乐类型为 8，即平静型；
输入的声音强度为 1035，节奏速度为 157，可得音乐类型为 2，即热情型；
输入的声音强度为 1463，节奏速度为 243，可得音乐类型为 1，即激情型；
输入的声音强度为 689，节奏速度为 231，可得音乐类型为 5，即快乐型；
输入的声音强度为 467，节奏速度为 139，可得音乐类型为 7，即柔和型；
输入的声音强度为 1560，节奏速度为 72，可得音乐类型为 4，即忧伤型；
输入的声音强度为 360，节奏速度为 178，可得音乐类型为 6，即轻盈型；
输入的声音强度为 458，节奏速度为 86，可得音乐类型为 8，即平静型；

七、模型的优缺点分析

1、模型一优缺点分析：

优点：MP3 音乐分类方法借鉴了语音识别技术和音频分类技术，同时针对音乐分类系统采用的特有媒介进行音乐的特征分析，利用常用的机器学习分类方法进行分类。本文采用的特征片段提取方法可以在较短时间里找到最具代表性的片段，从而可以更加准确快速的进行分类。

缺点：采用的分类方法数量偏少，而且音乐的特征数量也偏少，特征的有效性有待提高。

2、模型二优缺点分析：

优点：螺旋情感模型符合中国人的认知，且情感分类较宽泛，即每一种类型所能表达的情感较宽泛，并且各种情感类型之间不存在重叠，较易区分歌曲的各种情感类型；同时基于螺旋模型的各种情感类别的识别率都较高，且不会出现两种情感类别之间误差较大的情况。

缺点：特征分析的 K 邻近算法的效率有待提高，即使 $K=5$ 也需要对所有训练数据进行排序比较，如此应对千万级数据时可能会力不从心。

八、模型的推广

商业方面，以上两个模型均可用于网络音乐分类，根据音乐本身的性质重新严格化每一首歌的类别。

学术方面，模型二还可推广到诸多模糊属性的处理与分类，如对视频等的分类。

第六届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn

Email: 2013@tzmcm.cn

参赛队号 # 2562

参考文献：

- [1] 孙晓煜等 自动化音乐情感分类问题的研究
<http://www.docin.com/p-59770473.html> 2013 年 4 月 12 日
- [2] 张艳鹏、孙博文等 基于分形维数的音乐自动分类方法
<http://wenku.baidu.com/view/ce0f4c2c3169a4517723a35d.html> 2013年4月13日
- [3] 王志强、熊子东等 基于螺旋模型的歌曲情感识别算法研究
<http://www.tup.com.cn/Resource/tsyz/043737-01.pdf> 2013年4月13日
- [4] Taesam Lee、Taha B.M.J. Ouarda等 Identification of model order and numQER of neighbors for k-nearest neighbor resampling
<http://wenku.baidu.com/view/6ddb2c838762caaedd33d45c.html> 2013年4月13日
- [5] 张家发等 基于 MDCT 域特征的 MP3 音乐分类
<http://wenku.baidu.com/view/2b30d4ee172ded630b1cb6f8.html> 2013 年 4 月 13 日
- [6] 胡景凯等 基于学习分类器(LCS)的 MP3 音乐分类方法
<http://www.doc88.com/p-90693167177.html> 2013 年 4 月 13 日

第六届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn

Email：2013@tzmcm.cn

参赛队号 # 2562

附录【1】

```
import java.util.Arrays;
public class K {
    Music curdata;//当前需要测试的数据
    Music[] trainDatas;//存储了所有训练数据
    class Music implements Comparable<Music>{
        int sort;//音乐类型的代号，1到8依次表示：激情、热情、抒情、忧伤、快乐、轻盈、柔和与平静
        double strength;//表示声音强度
        double speed;//表示节奏速度
        double dis;//表示与当前测试数据的距离
        @Override
        public int compareTo(Music o) {
            // TODO Auto-generated method stub
            if(dis<o.dis) return -1;
            else return 1;
        }
    }
    double getDistance(Music x,Music y){
        double temp=Math.pow(x.speed-y.speed,
2)+Math.pow(x.strength-y.strength, 2);
        return Math.sqrt(temp);
    }
    int make(){ //其返回值即为音乐类型的代号
        for(int i=0;i<trainDatas.length;i++){
            trainDatas[i].dis=getDistance(trainDatas[i], curdata);
        }
        Arrays.sort(trainDatas);
        int[] style=new int[8];
        for(int i=0;i<5;i++){
            style[trainDatas[i].sort-1]++;
        }
        int result=0;
        for(int i=0;i<7;i++){
            if(style[i]<style[i+1])
                result=i+1;
        }
    }
}
```

第六届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会

电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn

Email: 2013@tzmcm.cn

参赛队号 # 2562

```
}  
    return (result+1);  
}  
  
}
```

数学中国提供 (www.madio.net)