

基于 k 平面分段回归的音乐情感分类

邵 曦,汪慧敏

(南京邮电大学 通信与信息工程学院,江苏 南京 210003)

摘 要:为了提高基于回归的音乐情感分类准确率,文中运用了 k 平面分段回归的方法,在音乐特征与音乐情感组成的高维空间内,通过多次迭代寻找超平面的方法直接求解非线性回归问题,进而预测二维情感变量值 Valence 与 Arousal,并通过该二维情感变量值进行音乐情感分类。为了验证分类系统的性能,实验中按 MIREX 分类标准建立有 5 类音乐情感的音乐库,对其 300 首音乐样本进行分类,与传统的多元线性回归和支持向量回归相比分类准确率有了一定提高。表明 k 平面分段回归的方法可以有效运用于音乐情感分类。

关键词:音乐情感分类;回归分析; k 平面分段回归;支持向量回归

中图分类号:TP301

文献标识码:A

文章编号:1673-629X(2015)06-0166-05

doi:10.3969/j.issn.1673-629X.2015.06.037

Music Emotion Classification Based on k -plane Piecewise Regression

SHAO Xi, WANG Hui-min

(College of Telecommunications and Information Engineering, Nanjing University of Posts
and Telecommunications, Nanjing 210003, China)

Abstract: In this paper, a piecewise regression approach of k -plane is employed in order to improve the classification accuracy of music emotion based on regression. It solves the nonlinear regression problem directly through several iterations in high dimensional space consisted by music feature and music emotion, predicting the valence and arousal values in the emotion model and classifying the music emotion. To verify the performance of classifier, test the classifier on 300 music sample from a music dataset which is commonly employed in MIREX, and the testing results on classification accuracy of the proposed approach are compared with the results from multiple linear regression and support vector regression. The experimental results show that k -plane piecewise regression approach can achieve the higher accuracy than the other two. That is to say the method of k -plane piecewise regression can be effectively applied to music emotion classification.

Key words: music emotion classification; regression analysis; k -plane piecewise regression; support vector regression

0 引 言

音乐是人们生活不可分割的一部分,而在数字化时代,从海洋般浩瀚的音乐曲库中挑选一首自己想要的歌曲变得越来越困难。因此很多音乐软件都提出了分类功能,其中分类标准有语言、风格、流派、心情等。对于用户来说,风格、流派等分类方法正确率虽然很高,但是人们更愿意在选择音乐的时候听到符合自己心情的歌曲。然而音乐按情感分类,较之风格、流派等分类方法有更多的主观因素,是一件相对困难的事情,目前分类正确率并不是很高。

由于音乐情感的主观性,不同的人对同一首歌有

不同的感受,需要一个统一的区分标准。Thayer^[1]提出的 Arousal-Valence 情感盘一般被用于音乐情感的界定。如图 1 所示, Arousal 表示音乐情感的激烈程度,值越高说明激烈程度越高,人们感受越强烈,反之,值越低说明激烈程度越低,人们感受越微弱; Valence 表示音乐的愉悦程度,值越高说明给人愉悦度越高,人们感受正面情绪越多,值越低说明给人愉悦度越低,人们感受负面情绪越多。

传统的分类方法是采用经典的机器学习方法,如 Wang Muyuan^[2]等采用了 SVM(支持向量机)算法,使用 libsvm 库作为分类的工具; Shi Yuanyuan^[3]等使用

收稿日期:2014-07-01

修回日期:2014-10-10

网络出版时间:2015-05-06

基金项目:国家自然科学基金资助项目(60902065)

作者简介:邵 曦(1976-),男,博士研究生,副教授,研究方向为多媒体信息系统与多媒体通信;汪慧敏(1990-),女,硕士研究生,研究方向为多媒体信息检索。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20150506.1621.009.html>

了双层分类的方法,先用 GMM 弱分类器构造第一层分类,然后再次提取新特征用于第二层 AdaBoost^[4] 分类。机器学习的缺点是无法将音乐情感的主观因素考虑在内,效果不是很好。

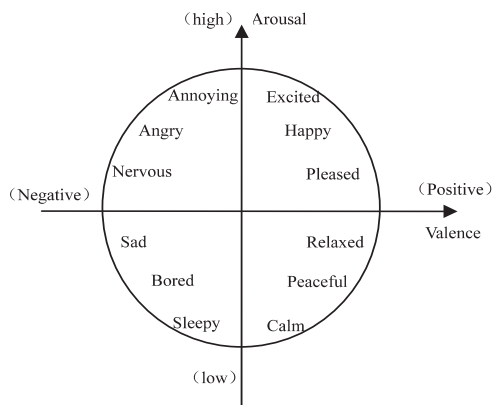


图1 Thayer's Arousal-Valence 情感

Yang Yi-Hsuan 等将音乐情感的主观性考虑进分类方法中,在文献[5]中采用了模糊最近邻均值法(FNM)与模糊 K 近邻法(FKNN)两种分类器,计算一个音乐片段属于某种情感类别的程度,并以此来追踪音乐情感。但是此方法并没有被广泛使用。

Yang Yi-Hsuan 等在文献[6]中提出将音乐情感看成 Valence-Arousal 情感平面内的二元连续变量,由此将音乐情感分类转化成在情感平面上的回归问题,分别构造 V 和 A 回归器来预测 V 值和 A 值。常见的回归算法有多元线性回归(MLR)^[7]、AdaBoost 回归^[8] 与支持向量回归^[9-10] (SVR) 等,在 Yang Yi-Hsuan 的实验中 SVR 得到的回归准确率最高, V 值和 A 值准确率分别达到了 68% 和 84%。

文中采用的就是基于回归的音乐情感分类方法。音乐情感回归由于音乐特征的多维性,一般将其看成非线性回归问题。SVR 模型引入了核函数,将低维空间的非线性问题转化为高维空间的线性问题,避免了在高维空间运算时存在的“维数灾难”,在非线性回归问题方面都有很好的应用。然而核函数以及核参数的选取,目前为止还没有一个统一的指导标准,多数情况下都是通过多次试验来确定最佳参数,由此增加了实验的不确定性。为此文中采用了直接求解非线性回归问题的方法— k 平面分段回归,这也是首次将 k 平面分段回归运用于音乐情感分类方面。

1 基于回归的音乐情感分类系统

基于回归的音乐情感分类方法是将音乐情感看成 Valence-Arousal 情感盘上的二维情感连续变量,将情感分类问题转化成回归分析并预测 V 、 A 值的问题。

假设 N 个输入变量 (x_i, y_i) , $1 \leq i \leq N$, x_i 是第 i 个

样本的特征向量, $y_i \in \mathfrak{R}$ 是待预测值,回归方法的目的是寻找一个回归器 $R(\cdot)$ 使得预测误差最小:

$$R \mid \min \sum_{i=1}^N |y_i - R(x_i)| \quad (1)$$

为了将音乐情感分类问题转化成二维情感连续变量的回归问题,需要考虑以下几点:

(1) \mathfrak{R} 的范围:将 Valence、Arousal 看成二维连续情感变量, V 、 A 值域是 $[-1, 1]$ 。

(2) 标准:志愿者在经过情感分类的训练后,为每个输入样本标注 V 、 A 值,最后取每个志愿者标注的平均值,以此作为训练样本的标准。

(3) 回归器的数量:没有考虑 V 、 A 值的关系,默认它们是相互独立的,因此使用两个回归器分别进行回归分析。

(4) 特征提取^[11-12]:提取与音乐情感相关的音乐特征,具体的提取特征在实验部分给出。

在训练部分生成两个回归器以后,对输入的测试样本分别进行 V 、 A 值的预测,以 (V, A) 对应的点在情感平面的位置确定所属的情感类别。基于回归的音乐情感分类系统具体流程如图2所示。

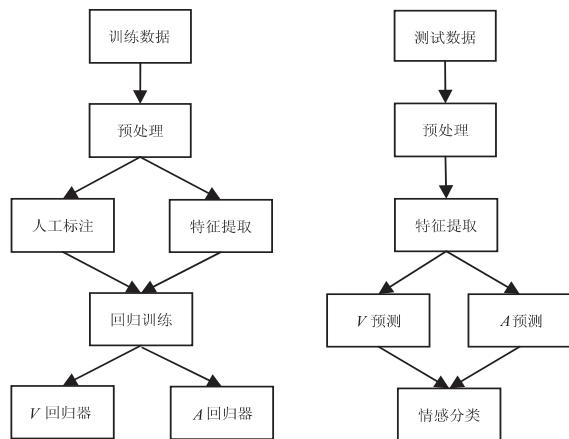


图2 基于回归的音乐情感分类系统

2 k 平面分段回归

分段回归的思想是按数据的变化规律将其分为几个区间,在每个区间上分别进行回归分析。在分段回归算法中,最重要的是区间的选取,传统的求解方法是先找到拐点来确定分成几段,在自变量为两个或多个的情况下,还要根据自变量的影响按显著程度进行排序,然后选择最显著因子进行分段。 k 平面分段回归的思想来源于 k 平面聚类算法,它的好处是不用进行拐点的选取,对样本点进行聚类分析直接求解非线性回归问题。

2.1 k 平面聚类算法

聚类算法是非监督模式分类的一种,它的主要目的是将数据集分成若干类,使得属于同一个类的对象

极大程度的相似,不同类的对象极大程度的相异。其中最常用的就是 k -均值类算法,它的聚类方法是椭圆形聚类,以随机选择的 k 个对象为初始聚类中心,求取其他对象与初始聚类中心的距离来划分类别。

Bradley 和 Mangasarian 在 k -均值类算法的基础上推广得到 k 平面聚类算法,以超平面代替 k -均值类算法中类的中心点,在不断迭代计算中求取样本点到据其最近的超平面的最小平方距离之和,最终选定一组超平面^[13]。

假设 d 维空间中的样本集为 $X = \{x_1, x_2, \cdots, x_n\}, x_j \in R^d, j = 1, 2, \cdots, n, k$ 平面聚类算法的目的是要寻找超平面组 $P = \{P_1, P_2, \cdots, P_k\}$,使 n 个样本点到据其最近的超平面的平方距离之和最小。算法的数学模型表述如下:

$$\min \sum_{i=1}^k \sum_{x_j \in C_i} |x\omega x_i^T x_j - \gamma_i|^2 \tag{2}$$

s. t. $\omega_i^T \omega_i = 1, i = 1, 2, \cdots, k$

其中, C_i 表示第 i 个聚类; γ_i 和 ω_i 分别表示超平面 P_i 的截距和单位化法向量; $|\omega_i^T x_j - \gamma_i|$ 是样本点 x_j 到超平面 P_i 的距离^[14]。

2.2 k 平面分段回归

k 平面分段回归设计思想来源于 k 平面聚类算法,通过反复迭代求取距离样本点最近的超平面的回归方程。假设用 $n-1$ 个自变量 $(x_1, x_2, \cdots, x_{n-1})$ 来做因变量 y 的回归分析,则 n 维空间中样本点集为 $A = \{a_1, a_2, \cdots, a_m\}$,其中 $a_j \in R^n, j = 1, 2, \cdots, m$,数学模型如下:

$$\text{Min} ||A\omega - e\gamma||^2 \tag{3}$$

s. t. $\omega^T \omega = 1$

其中, $(\omega, \gamma) \in R^n; e$ 是全 1 的 n 维向量; ω 表示超平面的单位化法向量; γ 表示超平面的截距; $||A\omega - e\gamma||$ 表示样本点到超平面的距离。

构造 Lagrange 函数:

$$L(\omega, \gamma, \lambda) = ||A\omega - e\gamma||^2 - \lambda(\omega^T \omega - 1) \tag{4}$$

分别对 ω, γ, λ 求导并令其导数为零。

$$\frac{\partial L}{\partial \omega} = 2A^T(A\omega - e\gamma) - 2\lambda\omega = 0 \tag{5}$$

$$\frac{\partial L}{\partial \gamma} = -2e^T(A\omega - e\gamma) = 0 \tag{6}$$

$$\frac{\partial L}{\partial \lambda} = \omega^T \omega - 1 = 0 \tag{7}$$

由式(5)得

$$\lambda = \omega^T A^T(I - \frac{ee^T}{m})A\omega \tag{8}$$

由式(6)得

$$\gamma = \frac{e^T A\omega}{m} \tag{9}$$

将式(8)、(9)代入式(5)得

$$A^T(I - \frac{ee^T}{m})A\omega - \omega^T A^T(I - \frac{ee^T}{m})A\omega \cdot \omega = 0 \tag{10}$$

令 $B = A^T(I - \frac{ee^T}{m})A$,则式(10) 表示为

$$B\omega - \omega^T B\omega \cdot \omega = 0 \tag{11}$$

令 $v = \omega^T B\omega$,则式(11) 变为

$$B\omega = v\omega \tag{12}$$

由式(12)可以得出, ω 为 B 的最小特征值相对应的特征向量。

至此,式(3)的数学问题描述如下:

$$\begin{aligned} \text{Min} \sum_{j=1}^k \sum_{i=1}^m (\omega_j^T a_i + r_j)^2 \\ \text{s. t. } ||\omega||^2 = 1 \end{aligned} \tag{13}$$

其中, $\omega_j \in R^n, r_j \in R^n$ 分别表示超平面 P_j 的单位法向量和截距; $(\omega_j^T a_i + r_j)$ 表示样本点 a_i 到超平面 P_j 的距离。

k 平面分段回归所要求的超平面 P_j 对应的回归方程为:

$$\omega_j \bullet \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ y \end{bmatrix} = r_j \tag{14}$$

3 实验结果与分析

3.1 建立音乐库

本实验中所使用的音乐库,是依据 The Music Information Retrieval Evaluation eXchange(MIRE-X)^[15]提出的方法创建的。音乐库按表 1 所示将音乐情感分成以下五类,具体建立步骤如下:

表 1 MIREX 的情感分类标准

| Cluster ₁ | Cluster ₂ | Cluster ₃ | Cluster ₄ | Cluster ₅ |
|----------------------|----------------------|----------------------|----------------------|----------------------|
| Rowdy | Amiable | Literate | Witty | Volatile |
| Rousing | Sweet | Wistful | Humorous | Fiery |
| Confident | Fun | Bittersweet | Whimsical | Visceral |
| Boisterous | Rollicking | Autumnal | Wry | Aggressive |
| Passionate | Cheerful | Brooding | Campy | Tense |
| | | Poignant | Quirky | Intense |
| | | | Silly | |

第 1 步:寻找 10 名志愿者,让 10 名志愿者依据给出的五类音乐情感,分别下载属于每类音乐情感的歌曲各 30 首并截取其中最能体现情感的 30 s 片段,由此得到 1 500 首待甄别音乐片段。志愿者下载歌曲前,会让他们听取每类音乐情感的例曲,然后让其下载类

似情感歌曲,选取歌曲过程中尽量不考虑歌词的影响。

第2步:为每类每首歌曲按1~200进行编号,并将10名志愿者分为A、B组,让A、B组成员依次听取对方成员的音乐曲库,如果同意别人的分类,则在此类歌曲序号后面打√,不同意则打×。如果一首歌曲有三人以上打√则保留,否则删除。经过此步骤选取出五类歌曲各200首(为方便起见,如果超出200个的类别则删除超出部分,以下同)。

第3步:让A、B组志愿者分别为步骤2中产生的对方曲库进行V、A值的标注,每人标注500首,为方便起见,V、A值范围取为{-1,-0.8,-0.6,⋯,0,⋯,0.8,1},标注后取平均值。

第4步:对标注完的每类200首歌曲进行甄选,确定了每类情感对应V、A值的范围,并将超出该范围的歌曲删除,此时音乐曲库每类剩下160首歌曲。在人工标注V、A值后,界定五类音乐情感的(V,A)范围如表2、图3所示。

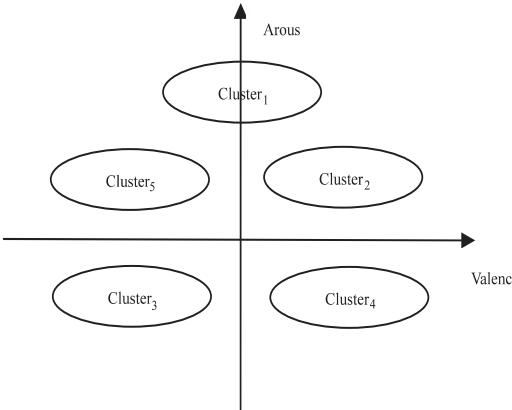


图3 五类音乐情感在Thayer's Valence-Arousal情感盘上的对应

表2 五类音乐情感对应的V、A值

| | Cluster ₁ | Cluster ₂ | Cluster ₃ | Cluster ₄ | Cluster ₅ |
|---|----------------------|----------------------|----------------------|----------------------|----------------------|
| V | [-0.6,0.6] | [0.2,1] | [-1,-0.2] | [0.2,1] | [-1,-0.2] |
| A | [0.4,1] | [0,0.6] | [-0.6,-0] | [-0.6,0] | [0,0.6] |

第5步:将每类歌曲进行1~160的重新编号,用音乐软件统一音乐样本的格式:采样率为1.6 kHz,单声道,wav格式。在每类歌曲中随机抽取100首作为训练曲库,剩下的作为测试曲库。至此,训练曲库共有音乐样本500首,测试曲库共有音乐样本300首。

3.2 特征提取

文中对音乐片段进行特征提取时,选择的是在音乐分类中广泛使用的Mel Frequency Cepstrum Coefficients(MFCC),基于传统PLP提出的RelAtive SpecTrA-Perceptual Linear Predictive(RASTA-PLP)两个特征。MFCC系数20维,PLP频谱相关系数21维,PLP倒谱相关系数9维,共50维,然后对这50维音乐特征分别

求均值和方差,最终每个音乐片段的特征向量维度是100维。

3.3 训练回归器

训练部分的目的是得到V值和A值的回归器,文中训练回归器的主要算法是k平面分段回归算法,另外还使用线性回归算法、SVR算法与之进行比较。文中算法都是用Matlab实现完成。

k平面分段回归算法在迭代过程中计算出每个样本到所有超平面的距离,并将样本点归入距离最近的超平面,然后计算新的一组法向量与截距,反复迭代直到找到满足所有条件的法向量与截距为止。具体算法步骤如下:

第1步:随机获取一组法向量和截距(ω,r),其中ω为单位向量。

第2步:将(ω,r)带入(ω_j^Ta_i+r_j),求出m个样本点到所有超平面的距离,并求出所有距离中样本点到哪个超平面距离最小,并将样本点归入据其最近的超平面。

第3步:由 $r = \frac{e^T A \omega}{m}$ 、ω为B的最小特征值对应的特征向量迭代求出下一组法向量和截距(ω,r)。

第4步:重复第1步与第2步,直到每个样本点归入的超平面不变为止。

第5步:求出最终的法向量和截距(ω,r)组,并以此求出超平面组。

以V值回归器为例,由于音乐片段的特征向量维度是100维,设为(x₁,x₂,⋯,x₁₀₀),加上V值的维度,k平面分段回归要求取的超平面维度为101维,代入式(11),假设样本点可以分为k个超平面(2<k<10),m表示音乐片段的个数500,a_i为101维的样本点,ω_j、r_j为待求的单位法向量、截距组。根据k平面分段回归的迭代算法最终分别求出V值、A值的超平面组。

线性回归的方法是从500首音乐片段中随机抽取100首,用Matlab自带的regress函数进行回归,求出V值和A值分别对应的回归系数。

SVR方法选用核函数为高斯径向函数(RBF),通过多次试验比较,选取V值回归器核参数γ=1.5,A值回归器核参数γ=0.8时,回归效果最好。

3.4 测试结果

文中分别用平均绝对误差(MAE)、回归值准确率(Ac)以及情感分类准确率(Ac of classification)作为基于三种不同回归方法的音乐情感分类系统的评价准则。

平均绝对误差通常用来评判回归模型的预测精度,回归值准确率由表2将V、A分别划分成的三类来计算求得,情感分类准确率用来判定整个分类系统。

实验评测结果如表 3 所示。

表 3 三种回归算法的有效率

| | MAE of Valence | MAE of Arousal | Ac of Valence/ % | Ac of Arousal/ % | Ac of Classifi- cation/% |
|---------------------|-------------------|-------------------|------------------------|------------------------|--------------------------------|
| 线性回归 | 0.282 6 | 0.277 6 | 61.4 | 70.4 | 56.11 |
| 支持向 量回归 | 0.233 3 | 0.217 3 | 70.6 | 83.3 | 63.4 |
| <i>k</i> 平面 分段回归 | 0.198 7 | 0.180 3 | 72.4 | 86.7 | 67.66 |

由此可以看到：

(1)在三种回归算法中,*k* 平面分段回归算法的准确率最高,*V*、*A* 值的回归准确率分别达到了 72.4% 和 86.7%,比支持向量回归的准确率高,并且 *k* 平面分段回归可以直接求解非线性问题,不需要像 SVR 方法一样进行核函数、核参数的选择。

(2)基于 *k* 平面分段回归的情感分类准确率也最高,达到了 67.66%,在分成五类的音乐情感分类系统中已经有所提高。

(3)可以看到 Arousal 值比 Valence 值更容易预测,准确率都略高,这个与现有的音乐情感分类的结果是一致的,如何提高 Valence 值的预测准确率还是一个值得研究的问题。

图 4 是 *V*、*A* 人工标注值与预测值的分布情况(*k* 平面分段回归算法)。

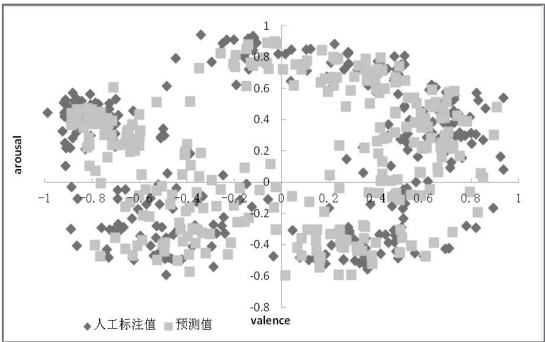


图 4 *k* 平面分段回归的人工标注值和预测值对比

4 结束语

文中将音乐情感看成 Valence–Arousal 平面上的连续二维情感变量,然后用回归的方法预测音乐的 *V*、*A* 值,以此达到音乐情感分类的目的。有了 *V* 值和 *A* 值的预测方法,就可以将任何一首音乐转变成情感平面上的一个点,除了便于情感分类,同时也易于音乐的检索和管理。

文中首次将 *k* 平面分段回归应用于高维空间, *V*、*A* 值的回归准确率分别达到了 72.4% 和 86.7%,与多

元线性回归(MLR)、支持向量回归(SVR)相比准确率更高。一是说明 *k* 平面分段回归在高维空间解决非线性问题也有良好表现,二是说明在基于回归的音乐情感分类系统中, *k* 平面分段回归也是一个很好的选择。

参考文献：

[1] Thayer R E. The biopsychology of mood and arousal[M]. Oxford;Oxford Univ Press,1989.

[2] Wang Muyuan,Huang Naiyao,Zhu Hangcheng. Emotion recognition of western tonal music using support vector machine[J]. Chinese Journal of Electronics,2006,15(1):74–78.

[3] Shi Yuanyuan,Zhu Xuan,Kim Hyoung–Gook,et al. A tempo feature via modulation spectrum analysis and its application to music emotion classification[C]//Proc of IEEE international conference on multimedia and expo. Toronto: IEEE, 2006: 1085–1088.

[4] 王 磊,杜利民,王劲林. 基于 AdaBoost 的音乐情绪分类[J]. 电子与信息学报,2007,29(9):2067–2072.

[5] Yang Yi–Hsuan,Liu Chia–Chu,Chen H. Music emotion classification;a fuzzy approach[C]//Proc of 14th ACM international conference on multimedia. Santa Barbara;ACM,2006: 81–84.

[6] Yang Yi–Hsuan,Lin Yu–Ching,Su Yafan,et al. Music emotion classification;a regression approach[C]//Proc of IEEE international conference on multimedia and expo. Beijing: IEEE,2007:208–211.

[7] Zhang Jiange. Sensor situation based on the multiple linear regression forecast[C]//Proc of IEEE international conference on computer science and automation engineering. [s. l.]: IEEE,2011:47–50.

[8] 李雅芹,杨慧中. 基于改进的 Adaboost. RT 模糊支持向量回归机集成算法[C]//中国智能自动化会议论文集. 出版地不详;出版者不详,2009:1123–1129.

[9] 苏高利,邓芳萍. 关于支持向量回归机的模型选择[J]. 科技通报,2006,22(2):154–158.

[10] 王 强,陈英武,邢立宁. 支持向量回归参数的混合选择[J]. 计算机工程,2007,33(15):40–42.

[11] 方 杰. 语音特征值提取方法的研究[D]. 无锡:江南大学,2005.

[12] 徐 欣,周 运,邵 曦. 基于音乐情感特征提取的音乐检索分析[J]. 信息通信,2011(5):9–12.

[13] 王 颖,陈松灿,张道强,等. 模糊 *k*–平面聚类算法[J]. 模式识别与人工智能,2007,20(5):704–710.

[14] Mangasarian O L. Arbitrary–norm separating plane[J]. Operations Research Letters,1999,24(1/2):15–23.

[15] Hu Xiao,Downie J S,Laurier C,et al. The 2007 MIREX audio mood classification task;lessons learned[C]//Proc of 9th international conference on music information retrieval. Philadelphia,PA,USA:[s. n.],2008.

基于k平面分段回归的音乐情感分类

作者：[邵曦](#)，[汪慧敏](#)，[SHAO Xi](#)，[WANG Hui-min](#)
作者单位：[南京邮电大学 通信与信息工程学院, 江苏 南京, 210003](#)
刊名：[计算机技术与发展](#)[ISTIC](#)
英文刊名：[Computer Technology and Development](#)
年，卷(期)：2015(6)

引用本文格式：[邵曦](#). [汪慧敏](#). [SHAO Xi](#). [WANG Hui-min](#) [基于k平面分段回归的音乐情感分类](#)[期刊论文]-[计算机技术与发展](#) 2015(6)