

# 基于音乐内容和歌词的音乐情感分类研究

邵 曦,陶凯云

(南京邮电大学 通信与信息工程学院,江苏 南京 210003)

**摘要:**针对音乐情感分类问题,为了弥补仅仅利用音乐内容进行音乐情感分类的单一模态分类方法的不足,文中提出了结合音乐内容和歌词的多模态音乐情感分类的方法。主要探讨了如何利用歌词对音乐进行情感分类以及结合歌词和音乐内容以达到提高分类准确率的效果。对歌词进行特征选择时,分别利用 CHI 特征选择算法和潜在语义分析(LSA)对歌词进行降维处理,有效去除了噪声,提高了分类效率。针对多模态融合问题,在传统的 LFSM 融合方法的基础上,提出了改进的 LFSM 融合方法,并通过实验验证了该方法的可行性;同时将该方法与其他传统的融合方法的分类效果进行了比较。结果表明,改进的 LFSM 融合方法的分类准确率最高,达到了 79.51%,验证了该方法的有效性。

**关键词:**音乐情感分类;CHI 特征选择;潜在语义分析;多模态融合

中图分类号:TP39

文献标识码:A

文章编号:1673-629X(2015)08-0184-04

doi:10.3969/j.issn.1673-629X.2015.08.039

## Research on Music Emotion Classification Based on Music Content and Lyrics

SHAO Xi,TAO Kai-yun

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

**Abstract:**According to the music emotion classification,an approach of multi-modal music emotion category combining music content and lyrics is proposed to compensate for lack of the single modal music emotion classification method that only uses music content for classification. Mainly discuss how to use lyrics for music emotion classification and combine music lyrics and content to improve the classification accuracy. Using feature selection algorithm based on CHI and quadratic dimension reduction method based on Latent Semantic Analysis (LSA) effectively improves the efficiency of text classification. For multi-modal fusion problem,propose an improved LFSM fusion method based on the traditional LFSM fusion method,and verify its feasibility through some experiments and compare the improved LFSM fusion method with the others. The results show that the accuracy of the improved method is highest,reaching 79.51%,that verify the effectiveness of the method.

**Key words:**music emotion classification;CHI feature selection;LSA;multi-modal fusion

## 0 引言

伴随着互联网技术的快速发展和普及,数字音乐呈现出爆炸式的增长,使得用来处理音乐数据库的音乐信息检索(MIR)系统受到了越来越多的关注。越来越多的人希望通过与音乐内容相关的信息来检索音乐,例如基于流派、情感等高级语义的检索<sup>[1]</sup>。

音乐是情感的载体,情感是音乐的内涵和本质特征<sup>[2]</sup>,对音乐情感的自动识别是近年来 MIR 系统研究的热点问题。音乐情感分类的一个典型方法就是提取

音乐的底层声学特征,然后应用机器学习来识别嵌入在音乐信号中的情感。然而,这种单一模态的分类方法的准确度往往不能令人满意,同时也满足不了实际 MIR 系统发展的需要<sup>[3]</sup>。通过进一步研究发现,除了音乐本身,歌词作为音乐的补充,同样包含了丰富的情感信息,对音乐的情感分类具有积极的影响。文献[4-5]也表明,融合音频和歌词的情感分类效果在一定程度上要优于基于单一音频特征或歌词特征的情感分类效果。所以,文中研究了多模态的音乐情感分类,在

收稿日期:2014-09-28

修回日期:2014-12-30

网络出版时间:2015-07-21

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;陶凯云(1990-),女,硕士研究生,研究方向为现代语音处理与通信技术。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20150721.1453.062.html>

使用音乐内容的同时使用歌词对音乐进行情感分类,并通过某种融合方法将两者结合起来以提高分类的准

确率。

多模态音乐情感分类的框架如图1所示。

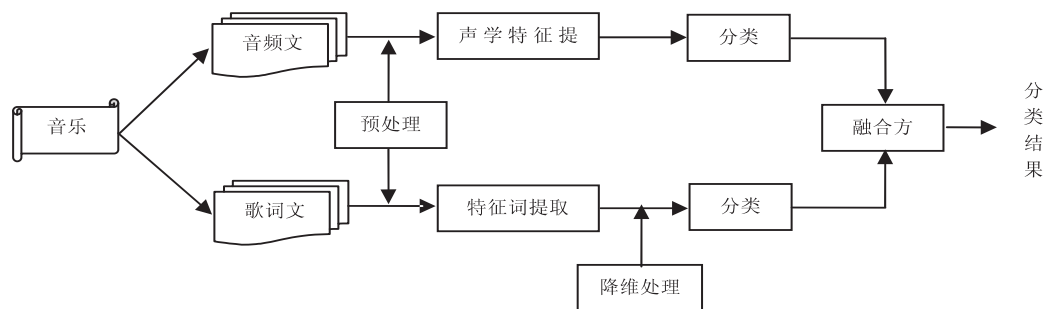


图1 多模态音乐情感分类框架图

## 1 音乐情感模型

想要对音乐进行情感分类,首先需要了解音乐的情感模型。目前比较有代表性的有 Thayer 情感模型和 Hevner 情感模型。

Thayer 情感模型是一个如图2所示的二维情感模型,它基于能量(energy)和压力(stress)两个维度<sup>[6]</sup>。按照能量从平静到充满活力、压力从快乐到焦虑,可将音乐分为焦虑、生机勃勃、沮丧、令人满足4类。

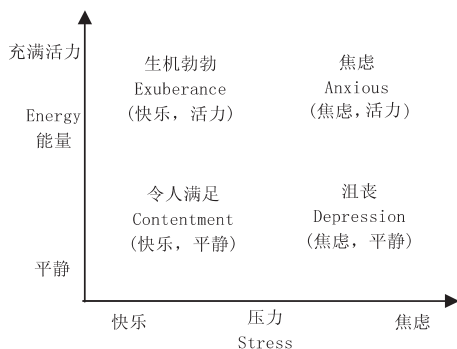


图2 Thayer 二维情感模型

Hevner 情感模型是离散的情感模型,它将情感分为“神圣”、“悲伤”、“向往”、“抒情”、“轻盈”、“欢快”、“热情”、“生机”8类,并且这8个类别根据其相互关系构成了一个环形,故称为 Hevner 情感环模型<sup>[7]</sup>。

由于能量和压力这两个因素,可以较好地与声学特征对应<sup>[1]</sup>,所以文中采用 Thayer 二维情感模型进行情感分类研究。

## 2 基于音乐内容的情感分类

基于音乐内容的情感分类过程主要分为三个阶段:

(1) 预处理过程:将音频文件转化为统一的格式(wav 格式,采样频率 16 kHz,单声道,时长 30 s);

(2) 声学特征提取过程:在这个过程中主要提取一些描述音乐音色、节奏和音高的底层声学特征,主要

包括 20 维的 Mel - Frequency Cepstral Coefficients (MFCC), 21 维的 Perceptual Linear Predictive (PLP) 中频谱相关系数和 9 维的 PLP 中倒谱相关系数。对于每一维特征都要计算其均值和方差,这样每一个音乐片段都可以由一个 100 维的特征向量表示。

(3) 分类过程:使用分类器对特征向量进行处理,从而对音乐进行情感分类。

## 3 基于音乐歌词的情感分类

歌词是包含着丰富情感的文本,根据自然语言处理技术,首先要对歌词进行预处理。预处理过程包括去除停用词、将单词转化为词根等。然后进行文本建模、特征提取、特征选择等等。

### 3.1 歌词的 VSM 表示

为了方便计算机处理和理解歌词文本,需要对歌词进行数字化表示。向量空间模型(VSM)是由 G Salton<sup>[8]</sup>等提出的一种文本表示方法,该模型的核心思想是将每一篇文档映射为向量空间中的一个点。该方法将文档表示成高维空间中的向量,每篇文档对应一个向量,该向量中的每一维对应文档的每一个特征项。假设有一个文本集,共包含  $n$  篇文档,用到了  $m$  个词,构造“词-文档”矩阵  $X_{m,n} = [x_{ij}] = (d_1, d_2, \dots, d_n) = (t_1, t_2, \dots, t_m)^T$ 。其中,  $x_{ij}$  表示特征词  $t_i$  在文档  $d_j$  中的权重,  $t_i$  和  $d_j$  分别代表特征词和文档的列向量。特征权重用于衡量特征词  $t_i$  在文本分类中区分能力的强弱或者对分类的重要程度。文中采用词频-逆文档频率(TF-IDF)来计算特征词的权重。公式如下:

$$\text{TFIDF}_{ij} = \frac{N_{ij}}{N_{*j}} \times \log \frac{D}{D_i}$$

其中,  $\text{TFIDF}_{ij}$  表示特征词  $t_i$  在文档  $d_j$  中所占的权重;  $N_{ij}$  表示特征词  $t_i$  出现在文档  $d_j$  中的次数;  $N_{*j}$  表示文档  $d_j$  中所有词的个数;  $D$  表示文档总数;  $D_i$  表示文本集中包含特征词  $t_i$  的文档数。

### 3.2 CHI 特征选择方法

特征选择是为了解决文本分类中存在的两个主要

问题:特征空间的高维性和文本向量空间特征的稀疏性;而 CHI 特征选择方法是文本分类中比较常用的特征选择方法。它度量特征词与类别之间的相关程度,并假设特征词与类别之间的分布满足一阶的 $\chi^2$ 分布<sup>[9]</sup>。特征词对于某类别的 CHI 统计值越大,说明它与该类别之间的相关性越强。特征词  $t$  对类别  $c_i$  的 CHI 统计值的计算方法定义为:

$$\chi^2(t, c_i) = \frac{N \times (AD - BC)^2}{(A + B) \times (C + D) \times (A + C) \times (B + D)}$$

其中,  $N$  表示语料库中所有文本总数;  $A$  表示属于类  $c_i$  且包含特征词  $t$  的文档总数;  $B$  表示不属于类  $c_i$  但包含特征词  $t$  的文档总数;  $C$  表示属于类  $c_i$  但不包含特征词  $t$  的文档总数;  $D$  表示不属于类  $c_i$  且不包含特征词  $t$  的文档总数。

计算完每个特征词的 CHI 统计值之后,将所有特征词按照 CHI 值从大到小排序,选取前  $k$  个词作为特征子集,“词-文档”矩阵  $X$  维数变为  $k \times n$  维。

3.3 潜在语义分析

传统的 VSM 假设词之间是相互独立的,它认为两个文本的相似度仅取决于它们拥有的相同词的多少,而忽略了上下文语境对词义的影响,从而产生所谓的同义和多义的问题<sup>[10]</sup>,进而影响分类精度。另外,经过上述特征选择之后,虽然文本向量空间的维数得到了一定的减少,但依然很高<sup>[11]</sup>,这就需要进行第二次降维处理,进一步减少噪声,提高分类精度。

为了解决上述问题,文中采用潜在语义分析(LSA)来进行二次降维。LSA 通过奇异值分解(SVD)将文档在高维 VSM 中的表示,映射到低维的“概念”空间,即潜在语义空间,使得原本稀疏的数据不再稀疏,并呈现出一些潜在的语义结构,同时有效地缩小了问题的规模<sup>[12]</sup>。

LSA 的具体过程如下所述:

(1)对  $X$  做 SVD 分解,  $X = U\Sigma V^T$ , 其中  $U, V$  是正交矩阵,  $\Sigma$  是由  $X$  的奇异值组成的对角阵:

$$\Sigma = \text{diag}(\delta_1, \delta_2, \dots, \delta_r) \quad \delta_1 \geq \delta_2 \geq \dots \geq \delta_r > 0$$

(2)取  $\Sigma$  中前  $p$  个最大的奇异值构成  $p \times p$  的  $\Sigma_p$ , 取  $U$  和  $V$  中前  $p$  列构成  $k \times p$  的  $U_p$  和  $n \times p$  的  $V_p$ , 构建  $X$  的近似矩阵,即  $X \approx X_p = U_p \Sigma_p V_p^T$ ;

(3)对于待分类文本,在经过预处理生成初始文本向量  $d$  之后,同样可以将  $d$  投影到潜在语义空间。具体计算公式为:

$$d^* = dU_p \Sigma_p^{-1}$$

4 多模态融合

结合音乐内容和歌词的多模态的音乐情感分类方

法主要是通过结合音乐内容的情感分类结果和歌词的情感分类结果,再重新确定音乐的情感类别。主要有以下几种融合方法:

(1)线性结合晚融合法(Late Fusion by Linear Combination, LFLC)。

LFLC 方法<sup>[13]</sup>是分别对音乐内容和歌词进行分类,预测出每一类的概率,然后对概率进行线性叠加,最后得出音乐的情感类别。参数  $\alpha \in [0, 1]$  表示两种模态各占的权重( $\alpha > 0.5$  表示歌词占的比重大于音乐内容)。例如,一首歌曲基于音乐内容和歌词的情感预测值分别为  $\{0, 0.1, 0.5, 0.4\}$ ,  $\{0, 0.1, 0.7, 0.2\}$ , 当  $\alpha = 0.5$  时线性组合结果为  $\{0, 0.1, 0.6, 0.3\}$ , 则最终被分为第 3 类。

(2)子任务结合晚融合法(Late Fusion by Subtask Merging, LFSM)。

LFSM 方法<sup>[13]</sup>是基于二维情感模型的融合方法,它认为音乐内容在能量上有较好的区分度,而歌词在压力上有较好的区分度<sup>[14-15]</sup>,所以分别对音乐内容在能量上分为平静和充满活力,对歌词在压力上分类快乐和焦虑,然后结合二者的分类结果,得出最终分类结果。

具体结合方法如表 1 所示。

表 1 LFSM 融合法

能量	压力	情感
活力	焦虑	焦虑
活力	快乐	生机勃勃
平静	焦虑	沮丧
平静	快乐	令人满足

(3)改进的 LFSM 融合法。

文中在 LFSM 融合法的基础上,提出了改进的 LFSM 融合法。认为音乐内容不仅在能量有良好的区分度,它在压力上也有一定的区分度,但是在压力上的区分度比在能量上的区分度相对较弱(假设 1);并且认为歌词在压力上的区分度比音乐内容在压力上的区分度相对较强,即在压力维度上,如果音乐内容和歌词的判断不一致,则优先考虑歌词的判断(假设 2)。

具体做法是先将音乐根据音乐内容分为焦虑、生机勃勃、沮丧和令人满足 4 类,再根据歌词分为快乐和焦虑两类,最后根据以上假设重新确定音乐的情感类别。例如,一首歌曲根据音乐内容被分为生机勃勃(快乐,活力),而根据歌词却被分为焦虑类,则根据假设 2 将这首歌曲分为焦虑(焦虑,活力)。

具体结合方法如表 2 所示。

表 2 改进的 LFSM 融合法

音乐内容	歌词	情感
焦虑	焦虑/快乐	焦虑/生机勃勃
生机勃勃	焦虑/快乐	焦虑/生机勃勃
沮丧	焦虑/快乐	沮丧/令人满足
令人满足	焦虑/快乐	沮丧/令人满足

5 实验结果与分析

5.1 数据集

文中用于实验的数据集包括 4 个情感类别,分别是焦虑(Anxious)、生机勃勃(Exuberance)、沮丧(Depression)和令人满足(Contentment),且由统一格式的音乐片段(wav 格式,采样频率 16 kHz,单声道,时长 30 s)及其对应的歌词组成。这些音乐片段是根据 last.fm 音乐网站上对应情感标签下的英文歌曲列表免费下载,并截取其中最能代表整首歌情感的 30 s 片段所得。由于截取片段时有一定的主观性,所以该数据集也是经过多人独自确认的。然后,每个情感类别中随机选择 100 个音乐片段作为训练集,其余作为测试集。最终,训练集有 400 首,测试集有 122 首。

5.2 实验结果

实验 1:不同融合方法的比较。

首先验证文中提出的改进的 LFSM 特征融合方法的可行性和有效性,并比较不同融合方法的分类准确率。由于对不同的数据库,很难定量比较所提出的方法与现有方法的好坏,所以将 Audio-Only(仅基于音乐内容)和 Lyric-Only(仅基于歌词)两个单一模态作为基准线,然后比较不同融合方法的准确率。

实验中采用 SVM 分类器进行分类,分类器参数一致,且各个模块选定的参数也是一致的。其中,第一次降维后特征维数  $k$  选为 300,第二次降维后的特征维数  $p$  选为 30。

不同融合方法的准确率如表 3 所示。

表 3 不同融合方法的准确率 %

	准确率(4 类)	准确率(能量)	准确率(压力)
Audio-Only	70.49	97.54	69.67
Lyric-Only	58.20	68.03	80.33
LFLC <sub>0.5</sub>	73.77	96.71	79.51
LFSM	78.68		
改进的 LFSM	79.51		

从表 3 中可以看出:

(1)从表格 1,2 行可以看出 Audio-Only 对能量有高达 97.54% 的分类准确率,Text-Only 对压力有 80.33% 的分类准确率,然而 Audio-Only 对于压力和 Text-Only 对于能量的分类准确都很低,这说明声学特

征和歌词特征能够很好的互补,所以融合两种模态的分类方法是相当可行的;

(2)从表格整体来看,不论哪种融合方式的多模态的分类准确率都比单一模态的分类准确率高,其中经过改进的 LFSM 的分类精度达到最高(79.51%),这就验证了文中提出的改进 LFSM 方法的可行性。

图 3 为不同的融合方法对应不同情感类别的分类准确率。

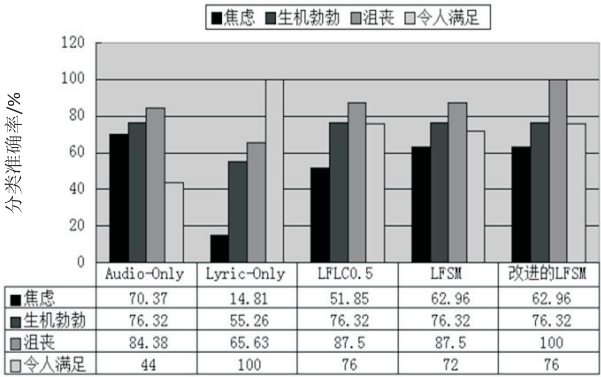


图 3 不同融合方法对每类情感的影响

从图中可以看出,单一模态的分类方法会使部分情感类别的准确率偏低,从而导致总的准确率不高。经过多模态融合之后,这种情况有明显好转。其中 LFSM 和改进的 LFSM 效果最为明显。

实验 2:一次降维和二次降维对实验结果的影响。

实验比较了仅使用 CHI 特征选择方法和使用 CHI 特征选择方法之后引入 LSA 进行二次降维的分类准确率。

当  $k = 300, p = 30$  的时候,仅使用 CHI 特征选择方法的分类准确率为 74.59%,而结合 CHI 和 LSA 进行特征降维的分类准确率为 79.51%。

实验结果表明,引入 LSA 进行二次降维后,其分类准确率有了一定程度的提高,这说明采用 LSA 可以进一步减少噪声的干扰,并且维数得到了很大程度的降低。

6 结束语

文中描述了一个将音乐内容和歌词相结合的多模态的音乐情感分类方法,应用统计自然语言处理技术来分析歌词,通过实验表明歌词确实能够对音乐信号进行语义信息的补充。对于多模态融合方法的问题,提出了改进的 LFSM 融合方法。在与传统的融合方法对比中发现,该方法的分类准确率有一定程度的提高,充分验证了该方法的可行性和有效性。今后,将把研究重点放在如何进一步提高系统分类准确率,特别是基于歌词的分类准确率上。



的变形和应力在允许范围之内。文中给出的分析方法和模型等具有通用性,对其他固体火箭发动机喷管设计具有指导意义。

文中在研究固体火箭发动机球形接头柔性喷管时,对问题进行了一定的简化,如:对发动机的燃烧室进行了简化处理,没有考虑球形接头处的密封问题,没有考虑新型的碳/碳复合材料。此外,实际情况中喷管是运动的,动态状态下喷管性能和喷管的优化设计,需要在以后的研究中深入探讨。

参考文献:

[1] 陈汝训. 固体火箭发动机设计与研究(下册)[M]. 北京:宇航出版社,1992.

[2] 王元友. 固体火箭发动机设计[M]. 北京:国防工业出版社,1984.

[3] 邢耀国,董可海,沈伟,等. 固体火箭发动机使用工程[M]. 北京:国防工业出版社,2010.

[4] 鲍福延,郭大庆,赵飞,等. 固体火箭发动机喷管集成设计分析技术研究[J]. 固体火箭技术,2004,27(3):169-172.

[5] 虞跨海,莫展,张亮,等. 固体火箭发动机特型喷管造型设计与优化[J]. 弹箭与制导学报,2012,32(4):137-138.

(上接第187页)

参考文献:

[1] 刘怡,高玥. 一种基于文本关键字模型的 Audio 音乐情感分类方法[C]//第四届人机环境联合学术会议论文集. 出版地不详:出版者不详,2008:1-7.

[2] 蒋盛益,李霞,李碧,等. 音乐情感自动分析研究[J]. 计算机工程与设计,2010,31(18):4112-4115.

[3] 甄超,宋爽,许洁萍,等. 多模态音乐流派分类研究[J]. 计算机科学与探索,2011,5(1):50-58.

[4] Yang Dan, Lee W S. Music emotion identification from lyrics[C]//Proc of the 11th IEEE international symposium on multimedia. Washington D C, USA: IEEE Computer Society, 2009:624-629.

[5] Hu Xiao, Downie J S. Improving mood classification in music digital libraries by combining lyrics and audio[C]//Proc of the 10th annual joint conference on digital libraries. New York, USA: ACM Press, 2010:159-168.

[6] Taylor J G, Fellenz W A, Cowie R, et al. Towards a neural-based theory of emotional dispositions[C]//Proc of IMACS IEEECS'99. [s. l.]:[s. n.], 1999.

[7] Hevner K. Expression in music: a discussion of experimental studies and theories[J]. Psychological Review, 1935, 42:186-204.

[6] Zebbiche T, Youbi Z E. Supersonic two-dimensional minimum length nozzle design at high temperature: application for air[J]. Chinese Journal of Aeronautics, 2007, 20(1):29-39.

[7] 尤军峰, 校金友, 张铎, 等. 固体火箭发动机延伸喷管展开动力学分析[J]. 推进技术, 2008, 29(1):37-42.

[8] 刘勇琼, 汪亮. 固体火箭发动机柔性喷管摆动机构的结构可靠度分析[J]. 推进技术, 1997, 18(4):51-53.

[9] 刘勇琼, 尤军峰. 固体火箭发动机柔性接头拉伸载荷下强度分析[J]. 航空动力学报, 2003, 18(2):264-268.

[10] Kraiko A N, Myshenkov E V, P'yankov K S, et al. Effect of gas non-ideality on the performance of laval nozzles with an abrupt constriction[J]. Fluid Dynamics, 2002, 37(5):834-846.

[11] Río-Cidoncha M D, Martínez-Palacios J, Ortuño-Ortiz F. Task automation for modelling solids with Catia V5[J]. Aircraft Engineering and Aerospace Technology, 2010, 79(1):53-59.

[12] Zhang Xiaoya, You Junfeng, Zhang Duo. Application of ADAMS and ANSYS to mechanism analysis[J]. Journal of Solid Rocket Technology, 2010, 33(2):201-204.

[13] 《导弹与航天丛书》编委会. 固体火箭发动机设计与研究[M]. 北京:宇航出版社,1993.

[14] Sutton G P, Biblarz O. 火箭发动机基础[M]. 洪鑫, 张宝炯, 译. 第7版. 北京:科学出版社,2003.

[8] Salton G, Wong A, Yang C S. A vector space model for automatic indexing[J]. Communications of the ACM, 1975, 18(11):613-620.

[9] 程一峰. 基于 TF-IDF 的音频和歌词特征融合模型的音乐情感分析研究[D]. 重庆:重庆大学,2012.

[10] 张玉峰, 何超. 基于潜在语义分析和 HS-SVM 的文本分类模型研究[J]. 情报理论与实践, 2010, 33(7):104-107.

[11] 熊小梅, 刘永浪. 基于 LSA 的二次降维法在中文法律案情文本分类中的应用[J]. 电子测量技术, 2007, 30(10):111-114.


[12] 刘云峰. 基于潜在语义分析的中文概念检索研究[D]. 武汉:华中科技大学,2005.

[13] Yang Yihsuan, Lin Yuching, Cheng Hengtze, et al. Toward multi-modal music emotion classification[C]//Proceeding of pacific rim conference on multimedia. Tainan, Taiwan: [s. n.], 2008:70-79.

[14] Lu Lie, Liu Dan, Zhang Hongjiang. Automatic mood detection and tracking of music audio signals[J]. IEEE Trans on Audio, Speech and Language Processing, 2006, 14(1):5-18.

[15] Yang Y H, Lin Y C, Su Y F, et al. A regression approach to music emotion recognition[J]. IEEE Trans on Audio, Speech and Language Processing, 2008, 16(2):448-457.

# 基于音乐内容和歌词的音乐情感分类研究

作者：[邵曦](#)，[陶凯云](#)，[SHAO Xi](#)，[TAO Kai-yun](#)  
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](#)  
英文刊名：[Computer Technology and Development](#)  
年，卷(期)：2015(8)

引用本文格式：[邵曦](#). [陶凯云](#). [SHAO Xi](#). [TAO Kai-yun](#) [基于音乐内容和歌词的音乐情感分类研究](#)[期刊论文]-[计算机技术与发展](#) 2015(8)