

基于领域覆盖算法的音乐情感识别

魏华珍¹, 戴安娜¹, 赵 姝^{1,2}

(1. 安徽大学 计算机科学与技术学院, 安徽 合肥 230601;

2. 安徽大学 智能计算与信号处理教育部重点实验室, 安徽 合肥 230039)

摘 要:音乐中具备很多情感的信息。文中通过分析音乐特征并用领域覆盖算法对音乐情感分类进行研究。音乐情感分类主要包括两个阶段:特征提取和分类。首先,通过 Matlab 语言提取音乐的特征,将提取到的特征值构建训练样本,然后使用训练样本训练领域覆盖算法分类器,得到音乐情感分类器,从而实现音乐的情感自动分类。文中借鉴 Weiner、Graham 的情感分类方法,将音乐分为开心和悲伤两类,并尝试用多种不同的音乐特征组合训练领域覆盖分类器,分析基于领域覆盖算法的音乐情感识别效果。

关键词:领域覆盖算法;音乐情感识别;音乐特征提取;情感分类

中图分类号:TP181

文献标识码:A

文章编号:1673-629X(2014)07-0072-05

doi:10.3969/j.issn.1673-629X.2014.07.018

Music Emotion Recognition Based on Neighborhood Covering Algorithm

WEI Hua-zhen¹, DAI An-na¹, ZHAO Shu^{1,2}

(1. School of Computer Science and Technology, Anhui University, Hefei 230601, China;

2. Key Lab of Intelligent Computing & Signal Processing of Ministry of Education, Anhui University, Hefei 230039, China)

Abstract: Music is strongly associated with emotions. In this paper, neighborhood covering algorithm is used to analyze musical genre so that music can be classified into different categories. Musical genre classification task falls into two major stages: feature extraction and classification. First of all, Matlab is used to extract the characteristics of the music, the characteristic values are applied to construct the training sample, then adopt the training sample to train the neighborhood covering algorithm classifier, obtaining the music emotion classifier and realizing the emotional automatic classification of music. According to emotion classification method of Weiner and Graham, divide the Chinese popular music into two categories: happy and sad, then try to use different music feature combinations to train neighborhood covering algorithm classifier as a method for music emotion recognition.

Key words: neighborhood coverage algorithm; music emotion recognition; music feature extraction; emotion classification

0 引言

音乐在创作时就表达了作曲者和作词者的情感,音乐与情感密切相关,传达着一种难以量化的信息。随着近些年互联网的发展,音乐在人们生活中扮演着重要的角色,人们开始越来越注重音乐的情感特性,音乐情感也开始应用于音乐检索和音乐推荐。

近些年国内越来越注重音乐情感的研究和音乐的个性化推荐,在计算机领域,人们将 Hevner 定义的情感环模型^[1]、Thayer 的二维情感模型^[2]等情感心理学理论用于音乐的情感分类;关于音乐情感的研究主要

使用文本分析^[3]、音频分析^[4-5]和两种分析方法结合^[6]的研究方式,使用 HMM^[4-5]、SVM^[6]等方法实现音乐情感自动识别。Juslin^[7]通过多种声学特征建立模型合成情感类别。D. Liu 等人^[8]利用模糊分类方法将约翰斯特劳斯圆舞曲的 MIDI 格式分成 5 类情感。H. Katayose 等人^[9-10]将声学信号转变为相应的数字符号对日本流行音乐提出了一种启发式的音乐情感分析系统。Lu Lie 等人^[11]对西方古典音乐采用 GMM 模型提取多种声学特征进行层次化分类取得了很多成果,但是关于领域覆盖算法^[12]在音乐情感识别

收稿日期:2013-10-11

修回日期:2014-01-15

网络出版时间:2014-04-24

基金项目:国家级大学生创新创业训练计划项目专项经费(cxcy2012066)

作者简介:魏华珍(1992-),女,江西新余人,硕士研究生,研究方向为机器学习;赵 姝,博士,硕士生导师,副教授,研究方向为智能计算、机器学习、神经网络、粒计算等。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20140424.0844.101.html>

方面的应用较少。

领域覆盖算法基于 M-P 神经元模型,使用领域覆盖算法构造的分类器学习速度快、复杂度低,能有效地解决有导师学习问题,可实现两类和多类分类的功能。

文中尝试采用 Weiner、Graham 的情感分类方法^[13],将中文流行歌曲分为开心和悲伤两类,并且将领域覆盖算法应用于音乐情感分类。文中使用音频分析的研究方式,尝试多种不同特征的组合,分析音乐特征对音乐情感识别的影响,取得了较好的音乐情感识别效果。

1 相关知识简介

1.1 音乐特征

音频信号有时域、频域、倒谱域等特征,表现了音乐的声学层和旋律层等特性,可用于识别音乐的情感,文中使用的音频特征简介如下:

1.1.1 时域特征

1) 短时能量:度量音频信号幅度值变化的函数,对高电平敏感。第 n 帧的音频信号的短时能量定义^[5]为:

$$E_n = \sum_{m=1}^N s_n^2(m) \quad (1)$$

2) 短时过零率:一帧音频的时域信号波形穿过零电平的次数,高频段过零率较高,低频段过零率较低,可用于区分清音与浊音。第 n 帧的短时过零率定义^[5]为:

$$Z_n = \frac{1}{2} \sum_{m=1}^N |\operatorname{sgn}[s_n(m)] - \operatorname{sgn}[s_n(m-1)]| \quad (2)$$

3) 短时平均幅度:度量音频信号幅度值变化的函数,与短时能量的区别是没有对音频的时域信号取平方,故不会对高频敏感,第 n 帧音频信号的短时平均幅度定义^[5]为:

$$M_n = \sum_{m=1}^N |s_n(m)| \quad (3)$$

4) 短时自相关函数(auto-correlation function):短时自相关函数是在信号的第 m 个样本点附近用短时窗截取一段信号,做自相关计算所得的结果。第 n 帧短时自相关函数定义^[5]为:

$$R_n(k) = \sum_{m=1}^{N-k+1} s_n(m) s_n(m+k-1) \quad (4)$$

自相关函数提供了一种获取周期信号周期的方法。在周期信号周期的整数倍上,它的自相关函数可以达到最大值,因此可以不考虑起始时间,而从自相关函数的第一个最大值的位置估计出信号的基音周期。

1.1.2 频域特征

把复杂的时间历程波形,经过傅里叶变换分解为

若干单一的谐波分量来研究,以获得信号的频率结构以及各谐波和相位信息,即将时域信号变换至频域,得到其幅度谱、相位谱、功率谱、对数谱,进而实现对信号的定量分析。频谱分析的相关定义如下:

傅里叶变换(FFT)^[14]:

$$X(k) = \sum_{j=1}^N x(j) \omega_N^{-(j-1)(k-1)} \quad (5)$$

其中, $\omega_N = e^{(-2\pi i)/N}$; $x(j)$ 为时域信号。

幅度谱:

$$B_{1w,n} = |X(k)| \quad (6)$$

相位谱:

$$B_{2w,n} = \angle X(k) \quad (7)$$

功率谱:

$$B_{3w,n} = |X(k)|^2 \quad (8)$$

对数谱:

$$B_{4w,n} = \log(|X(k)|) \quad (9)$$

1.1.3 倒谱域特征

1) 复倒谱变换(CCEPS)的定义^[15]:时域信号的傅里叶变换的对数的傅里叶反变换。

$$C_x(n) = \operatorname{IDFT}\{\ln[\operatorname{DFT}(x(n))]\} \quad (10)$$

2) 倒谱变换(RCEPS)的定义:复倒谱的实部,时域信号的傅里叶变换的幅值求自然对数,然后再做傅里叶逆变换。

$$\operatorname{RC}_x(n) = \operatorname{IDFT}\{\ln[|\operatorname{DFT}(x(n))|]\} \quad (11)$$

1.2 领域覆盖算法

领域覆盖算法是张铃教授等人基于 M-P 神经元模型的几何意义提出的一种求(次优)最小覆盖的方法与“球形领域”法相结合的神经网络学习算法^[12]。

设样本集 K 为: $K = \{s_i = (x_i, y_i), t = 0, 1, 2, \dots, n\}$, 神经网络的“记忆”功能是当输入为 x_i 时网络的输出为 y_i , 记住 x_i 与 y_i 之间的对应关系;“联想”功能,就是当输入为 $x_i + \$_i$ 时,其输出仍为 y_i (其中 $\$_i$ 是噪声或微扰)。有上面两种功能的网络,就可以认为是一个“联想记忆器”。

当输入样本为 $x_i + \$_i$ 时,其输出仍为 y_i , 可以理解为输入落入 x_i 附近时,输出为 y_i 。从几何意义角度,可以看作是以 x_i 为中心, $\$_i$ 为半径的一个球形领域,当输入的样本落在 x_i 的球形领域时,其输出为 y_i , 否则输出不为 y_i 。而该球形领域对应的神经元就可以实现分类的功能。对每一类用一组球形领域将其覆盖住,且不覆盖不属于该类的 x_i , 于是不同类的输入被不同组的球形领域覆盖,然后使用或门将同组的神经元集中输出,由此实现分类功能。上述的设计方法称之为领域覆盖算法。

1.2.1 领域覆盖算法球形领域的构造方法

第 K 类样本 X_i 的“球形领域”的构造方法如下:

取任一尚未被覆盖的点 a_i , 按公式(12)

$$\begin{cases} d_1(k) = \max_{x \in X_i} \{ \langle a_i, x \rangle \} \\ d_2(k) = \min_{x \in X_i} \{ \langle a_i, x \rangle \mid \langle a_i, x \rangle > d_1(k) \} \\ d(k) = (d_1(k) + d_2(k)) / 2 \end{cases} \quad (12)$$

计算覆盖半径 $r = d(k)$, 其中 $\langle x, y \rangle$ 表示 x 和 y 的内积, $d_1(k)$ 是样本 a_i 与异类样本 x 的最小距离, $d_2(k)$ 是 a_i 与同类样本 x 的最大距离, 其中 $d_2(k) > d_1(k)$ 。 r 为半径的覆盖 C_i 通过调整领域中心和半径, 可以覆盖更多的同类样本, 从而实现识别同类样本。

1.2.2 领域覆盖算法的应用过程

领域覆盖算法的应用过程分为训练和识别两个阶段, 具体如下:

1) 训练阶段。

将同类的训练样本用一组球形领域覆盖住, 该领域不覆盖其他类的样本, 于是不同组的球形领域覆盖不同类的样本, 然后再将同组球形领域对应的神经元的输出用或门集中起来。

2) 识别(测试)阶段。

对于某一测试样本 x , 如果落在第 y 组的球形领域, 则认为 x 属于第 y 类; 如果没有落入任何一组的球形领域, 则认为 x 为拒识样本, 使用“就近原则”将其归为距离其最近的类别。

2 音乐情感识别

文中通过对音乐进行特征提取, 基于领域覆盖算法实现对音乐进行音乐情感分类, 尝试多种不同音乐特征的组合对音乐情感进行识别, 分析音乐特征对音乐情感识别的影响, 以选出识别效果较好的特征组合。

2.1 音乐情感识别流程

文中基于领域覆盖算法对提取到的音乐特征进行情感识别, 主要对流行音乐进行情感分类, 并且使用 Weiner、Graham 的情感分类方法^[13], 将部分中文流行音乐分为开心和悲伤两类。文中将 mp3 格式音乐转换成 wav 格式用于音乐特征提取, 并将提取到的音乐特征数据构建训练样本用于训练基于领域覆盖算法的分类器, 然后将训练好的音乐情感分类器用于歌曲的情感分类。音乐情感识别的具体流程如图 1 所示。

2.2 音乐特征提取

音乐特征的分析与提取对实现音乐的自动识别起着至关重要的作用, 其提取对象的科学性、提取方法的可行性及其提取结果的准确性直接影响自动识别的效果。对于文中来说, 提取结果将进一步影响对 wav 格式的音频信号的分析, 及其后续判别、比较、评定工作。因此, 特征提取环节是心情音乐的计算机识别系统的重要研究讨论的内容, 也是实现系统功能必不可少的

组成部分。

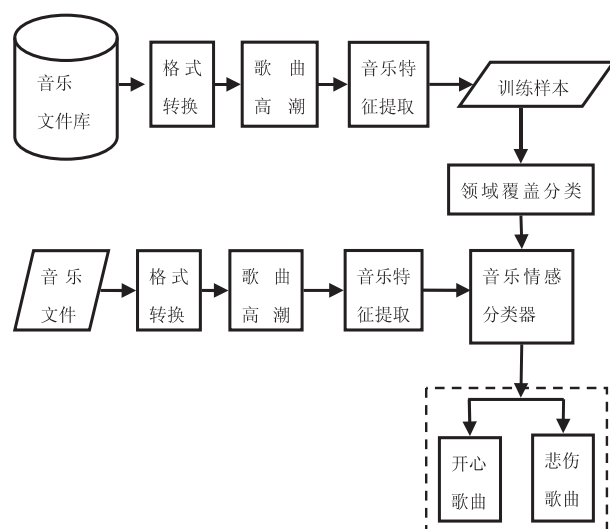


图 1 音乐情感识别流程图

2.2.1 歌曲高潮截取

歌曲的高潮部分情感特性较为明显, 所以文中获取歌曲总时长 t , 通过截取歌曲 $1/4t$ 到 $1/2t$ 的音频段获取歌曲的高潮部分, 用于音乐情感分析。

2.2.2 特征提取

1) 时域特征分析。

语音是一个时变信号, 具有短时平稳特性, 文中在提取歌曲高潮的时域特征前, 对音频进行分帧处理, 帧长为 $\text{fix}(\text{FS}/220)$, $\text{FS} = 44\ 100\ \text{Hz}$, 帧移与帧长长度相同。文中分别对歌曲高潮部分的音频提取了短时能量、短时过零率、短时平均幅度、短时自相关函数四种时域特征, 其中通过提取短时自相关函数求得音频的基音周期, 具体处理如表 1 所示。

2) 倒谱域特征分析。

有关倒谱域特征提取具体内容如表 1 所示。

3) 频域特征分析。

音频的频谱分析主要基于傅里叶变换, 文中通过傅里叶变换获取歌曲高潮的频谱、功率谱、对数谱, 并获取其幅值和相位, 具体处理如表 1 所示。

2.3 音乐情感识别的具体步骤

Step1: 将歌曲从 mp3 格式转换成 wav 格式;

Step2: 截取 wav 格式歌曲的高潮, 即截取歌曲 $t/4$ 到 $t/2$ 的音频段, t 为歌曲总时长;

Step3: 对截取的音频段提取各个音乐特征。提取时域特征前需对音频段进行分帧处理, 帧长为 $\text{fix}(\text{FS}/220)$, $\text{FS} = 44\ 100\ \text{Hz}$, 帧移与帧长长度相同。

Step3.1: 提取短时能量;

Step3.2: 提取短时过零率;

Step3.3: 提取短时平均幅度;

Step3.4: 提取短时自相关函数, 求基音周期 T , 并求出去除野点后的基音周期 $T_1 = \text{medfilt1}(T, 5)$;

Step3.5:求频域特征:频谱 $X = \text{fft}(x, 4096)$, 幅度谱 $\text{mag}X = \text{abs}(X)$, 相位谱 $\text{ang}X = \text{angle}(X)$, 功率谱 $\text{power} = \text{mag}X.^2$, 对数谱 $\ln = \log(\text{mag}X)$;

Step3.6:求复倒谱 $z = \text{ccep}(x(4000:4999))$, 倒谱 $rz = \text{rcep}(x(4000:4999))$ 。

Step4:对求得的特征求和、均值、方差、中值的特征值,具体内容见表 1;

Step5:使用提取好的特征值构建训练和测试样本;

Step6:训练基于领域覆盖算法的音乐情感分类器;

Step7:使用训练好的分类器测试测试样本。

表 1 特征提取的具体内容

音频信号特征	音乐特征	维数	处理方法
时域	短时能量	4	对求得的时域特征进行求和、求均值、求方差、求中值的处理
	短时过零率	4	
	短时平均幅度	4	
	短时自相关函数	8	使用短时自相关函数求音频的基音周期和去除野点后的基音周期,并对基音周期进行求和、求均值、求方差、求中值的处理
倒谱域	复倒谱	16	对求得的倒谱域特征求幅值和求相位,然后对求得的特征值进行求和、求均值、求方差、求中值的处理
	倒谱	16	
频域	频谱	2	对求得的频谱,求方差
	幅度谱	8	对求得的频域特征求幅值和求
	相位谱	8	相位,然后对求得的特征值进行
	对数谱	16	求和、求均值、求方差、求中值的
	功率谱	16	处理

3 实验

实验歌曲主要来源于网络上各音乐网站上推荐的心情歌单,如酷我音乐盒、百度随心听等音乐网站,此次实验总共使用了 637 首歌曲,其中使用 445 首流行音乐(215 首开心音乐、230 首悲伤音乐)训练音乐情感分类器,使用 192 首流行音乐(93 首开心音乐、99 首悲伤音乐)作为测试歌曲,并使用提取到的音乐特征数据构建训练和测试样本。

为了找到音乐情感识别效果较好的特征值的组合,尝试将时域、频域、倒谱域的特征值组合成多种组合通过领域覆盖算法进行音乐情感识别,实验结果如图 2、图 3 所示。

1)按时域、频域、倒谱域特征值所属域不同,组合的六组特征组合的情感识别

由图 2 可以看出,时域特征对音乐的情感识别效果较好,对开心的歌曲识别的正确率为 80.65%,对悲伤的歌曲识别的正确率为 79.79%,总的正确率为

80.21%;时域特征值与倒谱域特征值组合的音乐情感识别效果也较好,开心歌曲的正确率为 81.72%,悲伤歌曲的正确率为 76.77%,总的正确率为 79.17%;但是其他四组特征值组合的歌曲情感识别效果都较差。

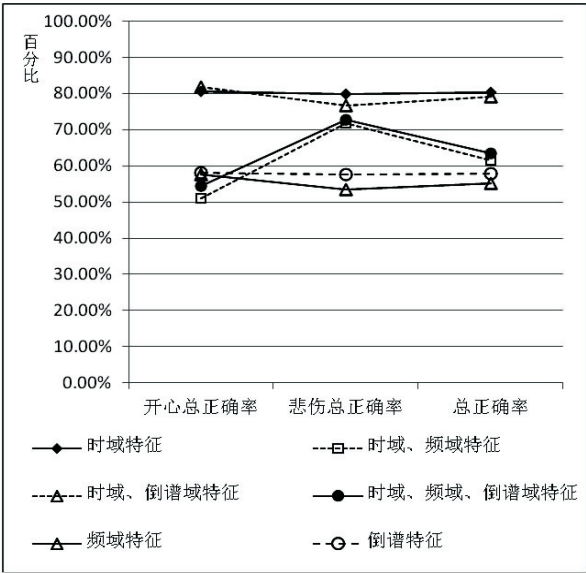


图 2 时域、频域、倒谱域特征组合的测试结果

2)由 1)的测试结果,可以发现时域特征的音乐情感识别效果较好,所以第二组特征值的组合主要以时域特征为主体,随机抽取频域、倒谱域的特征值与之组合,以发现音乐情感识别效果更好的特征组合。具体的测试结果如图 3 所示。

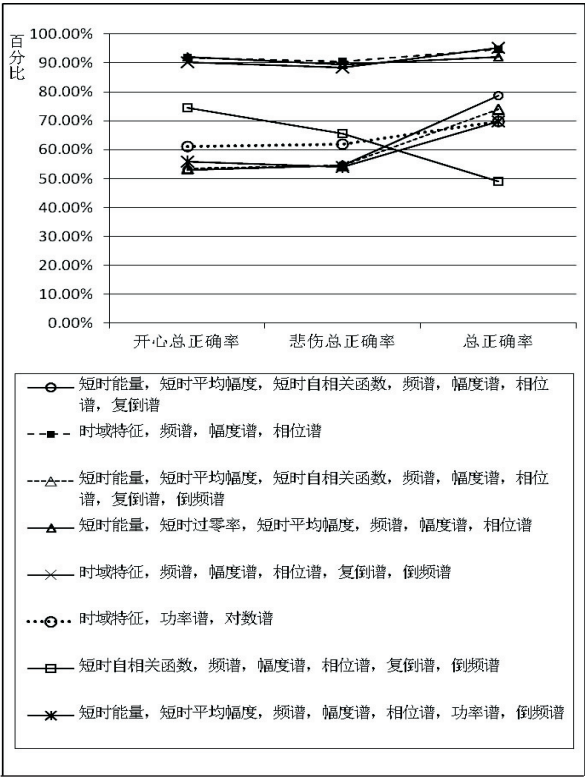


图 3 以时域特征为主体的各个特征组合测试结果
由图 3 可以看出第二、四、五种特征组合的情感识

别效果最好,正确率都在 90% 左右。三种特征组合的识别正确率见表 2。

表 2 三种识别效果最好的特征组合的识别正确率

领域覆盖测试结果	第二种 特征组合	第四种 特征组合	第五种 特征组合
开心总正确率/%	88.17	89.25	88.17
悲伤总正确率/%	92.93	89.89	88.89
总正确率/%	90.63	89.58	88.54

由 1) 和 2) 可以看出,基于领域覆盖算法的音乐情感识别实验识别效果最好的特征组合是由四个时域特征、频谱、幅度谱和相位谱组成的特征组合、由短时能量、短时平均幅度、短时过零率、频谱、幅度谱和相位谱组成的特征组合和由四个时域特征、频谱、幅度谱和相位谱组成的特征组合。从实验结果可以看出时域特征对音乐情感识别有较好的影响。

4 结束语

音乐的声学特征在音乐情感方面有一定的体现,基于音频自身特性,应用领域覆盖算法实现的音乐自动情感分类效果较好。

文中的音乐情感分类的研究样本是研究者根据研究需要建立的,样本选择的规模和实际情况有一定的差异,随着音乐情感分类研究的深入需要构建大规模的样本数据。文中只尝试对中文流行歌曲进行情感分析,关于其他语言的音乐情感识别还待进一步的研究。

参考文献:

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

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

(上接第 71 页)

[8] Muthumayil K,Rajamani V,Manikandan S. A novel cross layered energy based ad hoc on-demand routing protocol for MANETs[C]//Proc of 2011 third international conference on advanced computing. Chennai;IEEE,2011:276-281.

[9] 曹爱文,郭达伟,符可可,等. 一种避免路径丢失的按需多径路由协议[J]. 计算机技术与发展,2013,23(9):94-98.

[10] 嵇海进,邹修明. 基于链路状态的 AODV 路由协议改进[J]. 计算机工程,2010,36(24):81-83.

[3] 刘冬平. 面向中文歌词的音乐情感分类研究[D]. 广州:广东工业大学,2011.

[4] 张伟,谢湘. 基于 HMM 的音乐情感识别研究[C]//第九届全国人机语音通讯学术会议论文集. 黄山;出版者不详,2007:456-460.

[5] 朱菊霞. 基于 ACON/SVM/HMM 混合算法的情感识别研究[D]. 合肥:安徽大学,2011.

[6] 钟将,程一峰. 音乐和歌词融合的歌曲情感分类研究[J]. 计算机工程,2012,38(8):144-146.

[7] Juslin P N. Cue utilization in communication of emotion in music performance;relating performance to perception[J]. J Exp Psychol Hum Percept Perform,2000,26(6):1797-1813.

[8] Liu D,Zhang N Y,Zhu H C. Form and mood recognition of Johann Strauss's waltz centos[J]. Chin J Electron,2003,12(4):587-593.

[9] Katayose H,Imai M,Inokuchi S. Sentiment extraction in music[C]//Proc of 9th international conference on pattern recognition. Rome;IEEE,1988:1083-1087.

[10] Katayose H,Fukuoka T,Takami K,et al. Expression extraction in virtuoso music performances[C]//Proc of 10th international conference on pattern recognition. Atlantic City,NJ;IEEE,1990:780-784.

[11] Liu D,Lu L,Zhang H J. Automatic music mood detection from acoustic music data[C]//Proc of international symposium on music information retrieval. [s.l.]:[s.n.],2003.

[12] 张铃,张钹. M-P 神经元模型的几何意义及其应用[J]. 软件学报,1998,9(5):334-338.

[13] Weiner B,Graham S. An attributional approach to emotional development[M]//Emotions, cognition, and behavior. New York;Cambridge University Press,1984.

[14] 许作辉. 基于信息熵的语音端点检测算法研究与实现[D]. 长春:吉林大学,2012.

[15] 颜青青,钱盛友. 基于混沌的复倒谱域双重音频水印算法[J]. 通信技术,2011,44(1):63-65.

[11] 肖克江,熊忠阳,张玉芳. 多径路由协议 AOMDV 的改进与性能分析[J]. 计算机工程与应用,2012,48(6):99-101.

[12] 夏利,刘宗奇,田东渭. 基于主备链路稳定性的节能多径 AOMDV-ES 路由协议[J]. 计算机应用与软件,2012,29(8):237-241.

[13] 卢先领,陈树,于继明,等. Ad hoc 网络能量均衡的跨层多径路由算法[J]. 信息与控制,2009,38(5):585-590.

[14] 杨吟君,汪学明. 基于 AOMDV 的多径路由协议改进及仿真分析[J]. 通信技术,2011,44(8):45-47.

作者: 魏华珍, 戴安娜, 赵姝, WEI Hua-zhen, DAI An-na, ZHAO Shu
作者单位: 魏华珍, 戴安娜, WEI Hua-zhen, DAI An-na(安徽大学 计算机科学与技术学院, 安徽 合肥 230601), 赵姝, ZHAO Shu(安徽大学 计算机科学与技术学院, 安徽 合肥 230601; 安徽大学 智能计算与信号处理教育部重点实验室, 安徽 合肥 230039)
刊名: 计算机技术与发展 **ISTIC**
英文刊名: Computer Technology and Development
年, 卷(期): 2014(7)

参考文献(15条)

1. Hevner K Expression in music: a discussion of experimental studies and theories 1935
2. Thayer R E The biopsychology of mood and arousal 1989
3. 刘冬平 面向中文歌词的音乐情感分类研究 2011
4. 张伟; 谢湘 基于HMM的音乐情感识别研究 2007
5. 朱菊霞 基于ACON/SVM/HMM混合算法的情感识别研究 2011
6. 钟将; 程一峰 音乐和歌词融合的歌曲情感分类研究 2012(08)
7. Juslin P N Cue utilization in communication of emotion in mu-sic performance: relating performance to perception 2000(06)
8. Liu D; Zhang N Y; Zhu H C Form and mood recognition of Jo-hann Strauss's waltz centos 2003(04)
9. Katayose H; Imai M; Inokuchi S Sentiment extraction in music 1988
10. Katayose H; Fukuoka T; Takami K Expression extraction in virtuoso music performances 1990
11. Liu D; Lu L; Zhang H J Automatic music mood detection from acoustic music data 2003
12. 张铃; 张钺 M-P神经元模型的几何意义及其应用 1998(05)
13. Weiner B; Graham S An attributional approach to emotional development 1984
14. 许作辉 基于信息熵的语音端点检测算法研究与实现 2012
15. 颜青青; 钱盛友 基于混沌的复倒谱域双重音频水印算法 2011(01)

本文链接: http://d.g.wanfangdata.com.cn/Periodical_wjtz201407018.aspx