文章编号: 1003-0530(2017) 07-0943-10

基于鲁棒音阶特征和测度学习 SVM 的音乐和弦识别

王蒙蒙 关 欣 李 锵
(天津大学电子信息工程学院, 天津 300072)

摘 要: 和弦识别是音乐信息检索领域重要的研究内容之一, 在信息处理、音乐结构分析以及推荐系统等方面具有重要的作用。为了降低人声对和弦进程的影响且恢复和弦所对应的谐波信息, 文章分别对频谱中和弦所对应的谐波信息和人声信息进行建模, 构建双目标优化问题, 对和弦所对应的谐波信息进行有效重建, 同时去除人声; 其次, 对谐波信息进行降维处理得到鲁棒性的音阶轮廓特征; 最后为了提高支持向量机性能, 文章采用测度学习的方法得到马氏距离, 并使用马氏距离替换支持向量机的高斯核函数的欧氏距离, 使得支持向量机的判别函数包含有数据的空间分布信息。最终实验结果表明, 同基于现今流行的和弦识别算法相比, 提出的和弦识别算法识别正确率提高 3.5%~12.2%。

关键词: 和弦识别; 音阶轮廓特征; 核范数; 测度学习; 支持向量机
中图分类号: TP391.4 **文献标识码**: A **DOI**: 10.16798/j.issn.1003-0530.2017.07.006

Musical Chord Recognition Based on Robust Pitch Class Profile and Metric Learning Support Vector Machine

WANG Meng-meng GUAN Xin LI Qiang
(School of Electronic Information Engineer, Tianjin University, Tianjin 300072, China)

Abstract: Chord recognition is an important aspect of Music Information Retrieval, which plays an important role in information processing, musical structure analysis and recommender system. In order to reduce the influence of voice on chord progression and recovery harmonic information of chord, harmonic and voice component were modeled and a two-target optimal problem was constructed. Solving the optimal problem, harmonic structure was reconstructed and sparse voice was removed. Then, through performing a pitch mapping step, robust pitch class profile was obtained. At last, a Mahalanobis Metric was obtained from feature space of train samples through metric learning, then Euclidean Metric was replaced by Mahalanobis Metric in Radius Basis Function of SVM. Mahalanobis Metric contains distribution information of specified real dataset, so the classification result is more robust. Compared with currently popular chord estimation algorithm, Results show proposed system improves the accuracy ratio of 3.5%~12.2% on chord recognition.

Key words: chord recognition; pitch class profile; nuclear norm; metric learning; support vector machine

1 引言

和弦识别是音乐信号处理的重要研究问题之一,它在歌曲翻唱识别、音频匹配以及音乐推荐系统等领域都有重要作用[1]。

和弦特征设计和机器学习模型的选取一直是和弦识别领域最为活跃的研究内容。和弦内容不仅是对创作人情感的表达,也是乐曲频谱结构的外在表示。因此,音乐和弦识别离不开信号处理等各方面的知识。

收稿日期: 2016-12-22; 修回日期: 2017-02-10
基金项目: 国家自然科学基金资助项目(60802049, 61471263) ; 天津市自然科学基金重点项目(16JCZDJC31100)

文献[2]依据音乐谐波信息和音乐理论知识首次将信号频谱能量压缩到乐曲的 12 个音阶上,提出使用 12 维的音级轮廓特征(Pitch Class Profile, PCP)作为和弦特征来对乐曲的和弦进行转录。文献[3]提出了谐波音阶轮廓特征(Harmonic Pitch Class Profile, HPCP)特征,并将其用于和弦识别系统中。实验结果表明,HPCP 特征能够有效减弱乐器的类别对和弦的影响。文献[4]应用谐波积光谱与 PCP 特征结合的增强音阶轮廓特征(Enhanced Pitch Class Profile, EPCP)特征。EPCP 通过乘积而非求和的方式对谐波能量进行叠加映射,从而使得简单和弦间的区分更加明显,而对于相对复杂的和弦模式效果较差,因此很少使用。

不仅如此,近年来,学者进一步将矩阵分析等思想引入到和弦特征设计中。例如,文献[5]于 2012 年提出的基于非负最小二乘(Non-Negative Least Square, NNLS)解法的新型 NNLS-PCP 特征。该特征在进行频谱到 PCP 向量映射之前,通过先验知识获得 84 个音符所对应的频率值,并以此生成字典 $\boldsymbol{Q}$。字典中的每一个元素代表着一个音符所对应的基频及其谐波频率。假设 $\boldsymbol{Y}$ 为音乐信号原始频谱,通过使用最小二乘法解决问题 $\min \| \boldsymbol{Y}-\boldsymbol{Q}\boldsymbol{x} \|$,获得激活因子向量 $\boldsymbol{x}$。接着使用获得频谱的近似 $\boldsymbol{Q}\boldsymbol{x}$,并进行近似频谱到 PCP 特征之间的映射。这种方法通过先验知识,能够有效解决音符基频及谐波频率在进行映射时的串扰,增强了所提取特征的鲁棒性,并取得了较为广泛的应用[6]。另外,音乐人进行音乐创作的所使用乐器多种多样,使得同一和弦在使用不同乐器演奏时会有所不同。为了增强 PCP 特征对不同音色乐器所演奏和弦的鲁棒性,文献[7]基于离散余弦变换和美尔频率,提出了一种美尔对数 PCP 特征(discrete Cosine transform-Reduced Pitch, CRP)。这种特征结合美尔频率对音色的鲁棒性,提升了 PCP 特征性能,并在和弦识别系统中有所应用[8]。

作为整个和弦识别系统的第二个阶段,和弦的模式识别主要任务是对提取到的能够表征和弦的特征向量进行分类。由于 PCP 特征的广泛使用,模式识别阶段成为区别各个和弦识别系统的关键步骤。现今,模式识别即和弦识别分类阶段所使用的方法可以分为两大类:手动标注模板法和由和弦特征向量训练形成的机器学习模型。作为模板法的经典方法,Fujishima[2] 于 1999 年提出二进制模板法。这种方法主要是基于理想和弦的特点,把和弦所在的主音反映在 PCP 向量的对应维,即在对应的分量置 1。例如 C 大调和弦的模板:[1 0 0 0 1 0 0 1 0 0 0 0]。这种方法由于简单易于实施,从而被广泛使用[4]。同时,随着人工智能领域的快速发展和大量手动和弦标注数据的涌现,更为复杂的和弦模型越来越流行,并取得了一定的效果。其中,文献[9]采用了在语音信号中广泛使用的隐马尔科夫模型(Hidden Markov Model, HMM)。在隐马尔科夫模型中,采用 PCP 特征向量作为观测值,而真实的和弦标签作为隐藏状态,并经过训练获得状态转移概率分布,然后通过维特比译码来获得最可能的和弦序列。文献[10]采用动态贝叶斯网络(Dynamic Bayesian Network, DBN)进行和弦序列转录。动态贝叶斯网络是基于贝叶斯概率的推断网络,需要借助音乐乐曲中的其他信息。最常用的协助信息是调式。通过训练数据获得调式和和弦之间的条件概率分布,并通过最终的贝叶斯推断获得最可能的和弦序列。而文献[11]采用条件随机场(Condition Random Field, CRF)。不同于隐马尔科夫模型,条件随机场的当前时刻的和弦标签不仅依赖于当前观测 PCP 向量,还和完整的 PCP 序列有关。因此条件随机场模型的训练需要大量的带标签数据,训练速度较慢。

支持向量机(Support Vector Machine, SVM)是一种最大边界分类器,对离群点具有较好的鲁棒性。SVM 通过严格的映射关系将特征向量逐一映射为带标签的输出。而一种好的特征不仅能够更好的表征乐曲中包含的和弦信息,还会对模式识别效果产生较大影响。因此,本文选用一种鲁棒音阶轮廓特征(Robust Pitch Class Profile, RPCP)特征作为音频信号的和弦特征。RPCP 提取主要涉及频谱的谐波信息重建。RPCP 特征获取的关键步骤在于对音乐信号频谱中和弦所对应的谐波成分以及稀疏大噪声所对应的非谐波成分进行建模,得到一个优化问题,并通过增广拉格朗日乘子

法进行求解,并最终得到性能更加优越的 RPCP 特征。这种特征能够去除信号中大而稀疏的噪声,并重构音乐信号中的谐波信息,从而 RPCP 特征包含更加稳定而纯净的谐波信息。另外,由于一些和弦类型所对应的特征向量具有一定的相似度,所以不同和弦所对应的特征向量极大可能具有很小的欧氏距离,使得支持向量机的分类准确率不高。因此,本文利用测度学习的方法,根据和弦特征的特点,从问题本身的先验知识中有监督地学习到一个马氏距离测度矩阵。通过该距离测度,使原始特征空间投影到一个类别区分度更高的空间,使得在投影后的特征空间中,具有相同标签的特征向量分布更加紧凑,具有不同标签的特征向量间间隔更大。并进一步使用马氏距离改进原始 SVM 高斯核函数中的欧氏距离,使得改进后得到的基于测度学习的 SVM(Metric Learning based Support Vector Machine,mlSVM) 具有更好的类别区分度。最后,针对 12 类大调和弦和 12 类小调和弦的分类问题,本文使用新的 RPCP 特征输入多分类 mlSVM 分类器中完成和弦的识别,并完成同流行和弦识别系统的比较。

2　基于鲁棒 PCP 特征和测度学习 SVM 的和弦识别算法

如图 1 所示,本文所设计的和弦识别系统包括四个主要步骤: 音频输入、频谱预处理、RPCP 特征计算以及测度学习支持向量机分类。其中频谱预处理主要采用核范数凸优化技术对原始音频中的谐波成分进行重建,并加入一范数惩罚项,增强计算出的频谱对稀疏噪声的鲁棒性; 接着通过音阶映射提取基于谐波信息的鲁棒性 PCP 特征,最后将这些特征输入到基于测度学习的支持向量机中,输出和弦标签,进而完成和弦的识别。

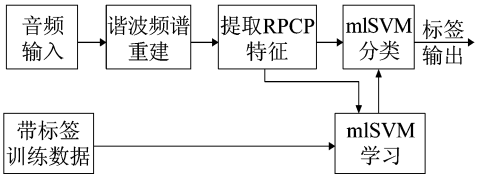

图 1　和弦识别系统结构图

Fig.1　Structure model of chord recognition

2.1　鲁棒 PCP 特征提取

乐曲的演奏过程中出现的音符掩蔽、音符缺失等现象会对最终的和弦识别与转录产生一定的影响。因此,鲁棒 PCP 特征的提取分为两步: 重建信号频谱中的谐波信息和音阶映射。

2.1.1　重建信号频谱中谐波信息

首先对音乐片段进行采样处理,采样率为 11025 Hz,生成一个长度为 I 的音频信号 $x(n)$。对于这个长度为 I 的音频信号 $x(n)$ 使用窗函数 $W(n)$ $(0<n\leqslant N)$ 对音频信号进行预处理,且窗移长度为窗宽的 50%,得到加窗后的待处理的信号矩阵 $\boldsymbol{M}_{N\times K}$,其中 $K=2I/N$。然后,将信号矩阵乘傅里叶变换矩阵 $\boldsymbol{\Omega}$,$\boldsymbol{\Omega}$ 是一个 $N\times N$ 的方阵,矩阵的行是标准正交基,这样得到分帧信号的频谱矩阵 $\boldsymbol{S}=\boldsymbol{\Omega}\cdot\boldsymbol{M}$。

假设 $\boldsymbol{A}$ 表示频谱中和弦所对应音阶的基频及其谐波成分所表征的矩阵,$\boldsymbol{E}$ 代表非谐波成分,例如人声等信息构成的矩阵。假设谐波成分和非谐波成分是相互独立的,那么就有: $\boldsymbol{S}=\boldsymbol{A}+\boldsymbol{E}$。也就是说,对于音乐信号而言,其频谱矩阵 $\boldsymbol{S}$ 可以看作是受外界噪声 $\boldsymbol{E}$ 影响的谐波矩阵 $\boldsymbol{A}$。

由信号理论可以知道,信号的谐波成分在谐波矩阵 $\boldsymbol{A}$ 上只分布在分散的数个频率成分上,且在时间尺度上和弦具有明显的重复性,因此矩阵 $\boldsymbol{A}$ 具有潜在的低秩特性,即和弦谐波信息分布在一个低维的子空间之内; 而非谐波成分 $\boldsymbol{E}$ 矩阵主要包含的是人声等噪声,且人声变化更加频繁,具有一定的稀疏特性[12]。因此,可以通过秩函数和零范数分别对谐波成分和稀疏噪声进行建模。此时对低秩矩阵的恢复是一个双目标优化问题:

$$\min_{\boldsymbol{A},\boldsymbol{E}}(\operatorname{rank}(\boldsymbol{A}),\|\boldsymbol{E}\|_0)$$
$$\text{st. } \boldsymbol{S}=\boldsymbol{A}+\boldsymbol{E} \tag{1}$$

其中 $\operatorname{rank}(\boldsymbol{A})$ 是矩阵 $\boldsymbol{A}$ 的秩函数,零范数 $\|\cdot\|_0$ 表示矩阵中非零元素的个数,即零范数表征矩阵的稀疏程度。

通过引入折中因子 $\lambda>0$,将双目标优化问题转换为如下单目标优化问题:

$$\min_{\boldsymbol{A},\boldsymbol{E}}\operatorname{rank}(\boldsymbol{A})+\lambda\|\boldsymbol{E}\|_0$$
$$\text{st. } \boldsymbol{S}=\boldsymbol{A}+\boldsymbol{E} \tag{2}$$

然而,式(2) 所表示的优化问题是 NP 难问题,

因此需要找到合适的能够代替秩函数和零范数的函数。为了保证整个优化问题存在全局最优解,因此要求替换函数是原函数的凸包络。而矩阵的核范数代表着矩阵中奇异值之和,是矩阵秩函数的凸包络;矩阵的 1 范数表示矩阵中非零元素之和,通常作为矩阵的稀疏算子,也是 0 范数的凸包。因此将问题(2)凸松弛到如下凸优化问题:

$$\min_{\boldsymbol{A},\boldsymbol{E}} \|\boldsymbol{A}\|_* + \lambda \|\boldsymbol{E}\|_1$$
$$\text{st. } \boldsymbol{S} = \boldsymbol{A} + \boldsymbol{E} \tag{3}$$

该优化问题可以通过增广拉格朗日乘子法[13]有效地进行解决,得到最优解谐波信息矩阵 $\boldsymbol{A}$,即 $\boldsymbol{A}$ 为对原始信号频谱进行重建后的谐波频谱。

2.1.2 音阶映射

通过使用增广拉格朗日乘子法对优化问题(3)进行求解,得到重建之后的谐波频谱矩阵。优化之后的频谱矩阵分离了人声等稀疏噪声,并重建了频谱的低秩信息,因此频谱矩阵 $\boldsymbol{A}$ 具有一定的鲁棒性。然而,乐曲的频谱矩阵 $\boldsymbol{A}$ 不仅包含和弦所对应音符的基频成分,还包含大量谐波成分,使得频谱矩阵 $\boldsymbol{A}$ 的维度仍然很大。直接使用高维度的频谱作为特征进行和弦的识别与转录不仅不直观,而且比较费时费力。故本文采用本节的音阶映射步骤将频谱矩阵 $\boldsymbol{A}$ 映射成为前文提到的 12 维 PCP 特征。为了得到整首乐曲的鲁棒 PCP 特征,即 RPCP 色度矩阵 $\boldsymbol{R}$,需要定义映射矩阵 $\boldsymbol{P}$,同时将谐波矩阵 $\boldsymbol{A}$ 的每一列映射成 12 维的 PCP 向量,即 $\boldsymbol{R}=\boldsymbol{P}\cdot\boldsymbol{A}$,其中映射矩阵 $\boldsymbol{P}$ 定义如下:

$$\boldsymbol{P} = \begin{bmatrix} \delta(2\pi\cdot\omega_0, f_1) & \cdots & \delta(2\pi\cdot\omega_{N-1}, f_1) \\ \vdots & \ddots & \vdots \\ \delta(2\pi\cdot\omega_0, f_{12}) & \cdots & \delta(2\pi\cdot\omega_{N-1}, f_{12}) \end{bmatrix}_{12\times N} \tag{4}$$

$\boldsymbol{P}$ 表示频谱矩阵 $\boldsymbol{A}$ 和由鲁棒 PCP 向量组成的色度矩阵之间的变换矩阵。$2\pi\omega_j(0\leqslant j\leqslant N-1)$ 表示进行频谱变换之后的频谱中各个频段的频率值。f_i $(1\leqslant i\leqslant 12)$ 表示音乐中 12 平均律所定义出来的 12 个音阶的基频。

国际上通常使用钢琴中央 C,即 C4 所对应的音符频率 $f_{C4}=261.626$ Hz,作为定义其他音符频率的基准频率,则各个音符对应频率 f_i 的计算公式如下:

$$f_i = f_{C4}\cdot 2^{\frac{B}{12}} \tag{5}$$

其中 B 为各个音符与中央 C 之间的音程差。

因此,映射矩阵中的映射函数 $\delta(x, f_i)$ 定义如下:

$$\delta(x, f_i) = \begin{cases} 0, \text{if } [12\cdot\log_2(2\pi\cdot x/f_i)]\%12 \neq 0 \\ 1, if\,[12\cdot log_2(2\pi\cdot x/f_i)]\%12 = 0 \end{cases} \tag{6}$$

其中 x 为对原始音频信号进行时频变换后得到的各个频率成分。

PCP 特征结合谐波信息,对信号频谱能量进行了有效地压缩,是一种广泛使用的和弦中级特征。本文将采用的 RPCP 特征矩阵同 HPCP,CRP 和 NNLS-PCP 进行对比,结果如图 2 所示。实验数据为从 Beatles 发行唱片中截取的一段持续时间为 3.5 s 的 C 大调和弦(音阶 C、E、G 叠加)片段。

从图 2(b)可以看出,在 0.8 s~1.3 s 之间,HPCP 所示的和弦特征在音阶 E 和 G 上存在明显的能量缺失现象。而图 2(e)所示的 RPCP 特征能够对特征矩阵进行结构化重建,对缺失的矩阵部分进行适当的填充,提高和弦特征的鲁棒性。同时,在 2.8 s~3.1 s 之间,图 2(b)所示的 HPCP 特征矩阵有一明显大而稀疏的噪声,即人声,这会使得在进行频谱映射后的主要音阶 E 上的能量有所降低,从而使和弦的识别出现偏差。而本文提出的 RPCP 特征矩阵,如图 2(e)所示,则会在削弱人声的同时,恢复出音阶 E 上的信息,即使音阶 E 上的能量有所提升。这是因为本文采用了核范数和 1 范数混合优化,能够约束矩阵的秩及其稀疏程度,从而能够从具有大而稀疏噪声的实际数据中较为完整的恢复出数据本身所具有的信息。同时,在 0.3 s~0.8 s 之间,NNLS-PCP 和 CRP 特征出现了明显的偏差,特征主要分布在音阶 D 和 F 之上,这会导致最终识别的错误。而 RPCP 特征不仅削弱了音阶 D 和 F 上的能量,还对音阶 C 和 G(C 大调和弦所包含的音阶)上的能量进行了一定的提升,提高了 PCP 特征的鲁棒性。总而言之,本文提出的 RPCP 特征在不改变和弦谐波信息的情况下,能够反映真实的和弦信息,同时抑制了人声等大而稀疏噪声的干扰,从而更加接近理想 PCP 特征,即图 2(a)。

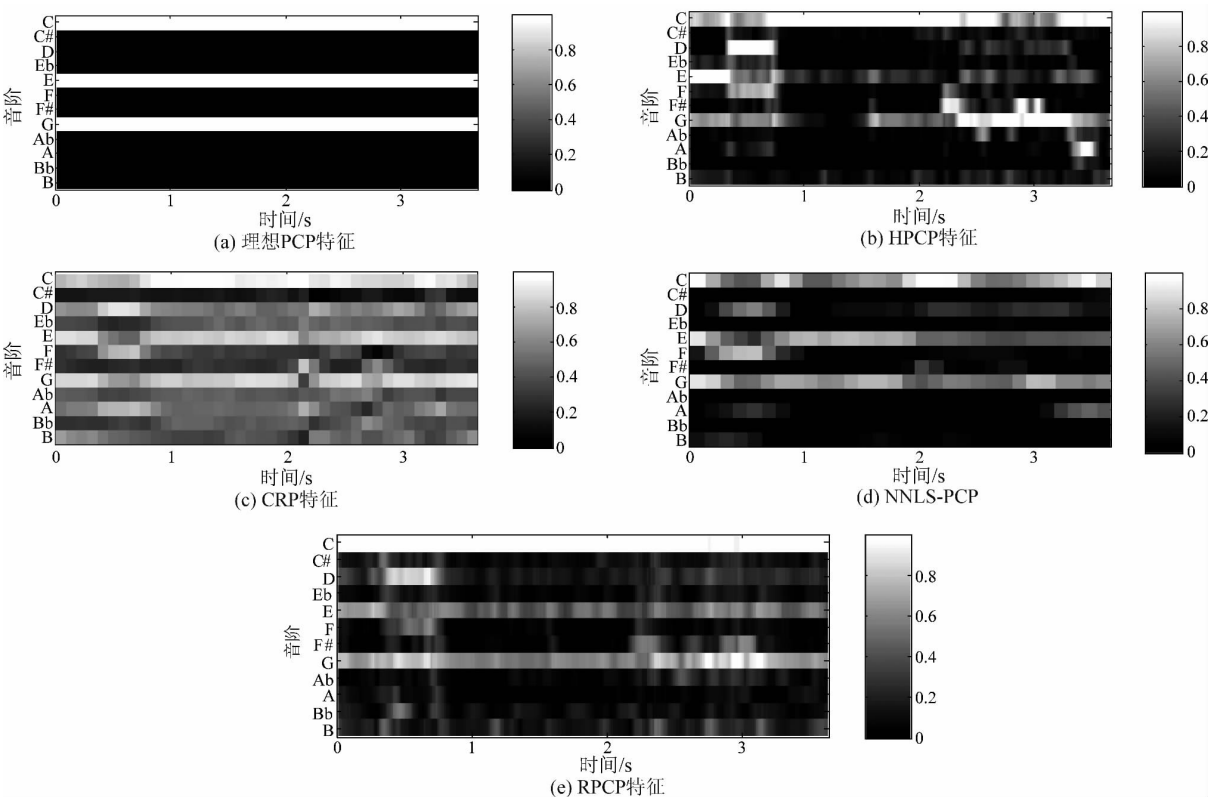

图 2　不同和弦特征比较

Fig.2　Comparison of different chord features

2.2　基于测度学习的支持向量机

支持向量机模型属于最大化间隔分类器(Largest Margin Classifier),对离群点有很好的鲁棒性。对和弦识别而言,SVM 是要建立一种多对一的映射,即将观测特征向量逐一映射到和弦标签集。

对于给定数据样本$\{(\boldsymbol{x}_i,\boldsymbol{y}_i),i=1,2,\cdots,n\}\in\mathbf{R}^{12}\times\mathbf{Y}$,其中$\boldsymbol{x}_i\in\mathbf{R}^{12}$为 12 维向量,表示 12 维的鲁棒 PCP 特征向量,$y_i\in\{-1,1\}$为样本标签,-1,+1 表示两个不同的和弦类别; n 为样本个数。那么基于高斯径向基核函数的支持向量机求解最优超平面的问题即为求解下列优化问题:

$$\max_{\alpha} f(\alpha)=\sum_{j=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j k_m(\boldsymbol{x}_i,\boldsymbol{x}_j)$$

$$\text{s.t.}\ \sum_{i=1}^{n}\alpha_i y_i=0$$

$$0\leqslant\alpha_i\leqslant C,\ i=1,\cdots,n \tag{7}$$

其中 k_m 为高斯径向基函数,$\boldsymbol{\alpha}$ 为拉格朗日系数,C 为惩罚参数。常规高斯径向基函数的表达式如下:

$$k_m(\boldsymbol{x}_i,\boldsymbol{x}_j)=\exp\left(-\frac{(\boldsymbol{x}_i-\boldsymbol{x}_j)^{\mathrm{T}}(\boldsymbol{x}_i-\boldsymbol{x}_j)}{2\sigma^2}\right) \tag{8}$$

通过解上述优化问题,得到支持向量机参数最优解$\boldsymbol{\alpha}^*=(\alpha_1^*,\cdots,\alpha_n^*)^{\mathrm{T}}$。接着,选择$\boldsymbol{\alpha}^*$的一个小于 C 的正分量$\alpha_k^*$,并据此计算支持向量机的另一个参数 b:

$$b^*=y_k-\sum_{i=1}^{n}y_i\alpha_i^*k_m(\boldsymbol{x}_i,\boldsymbol{x}_k) \tag{9}$$

从而,根据这两个参数构建和弦识别的分类决策函数:

$$f(\boldsymbol{x})=\operatorname{sgn}\left(\sum_{i=1}^{n}y_i\alpha_i^*k_m(\boldsymbol{x}_i,\boldsymbol{x})+b^*\right) \tag{10}$$

由于高斯径向基函数能够将低维空间内的数据映射到高维空间乃至无穷维,从而使得在低维空间中线性不可分数据在高维空间内有很好的线性可分特性。从式(8)可以看出,高斯径向基函数的距离测度一般采用欧氏距离测度,即$(\boldsymbol{x}_i-\boldsymbol{x}_j)^{\mathrm{T}}(\boldsymbol{x}_i-\boldsymbol{x}_j)$。

从欧氏距离测度公式可以明显地看出,欧式距离测度没有包含任何真实数据的分布信息,即不包含任何真实数据的先验信息,因此对于特定的应用问题,并不总能取得较好的效果。

对于本文的研究对象而言,一部分和弦所对应

的 PCP 向量虽属于不同标签,但是却具有比较小的空间距离。对于真实的数据而言,就有可能出现数据混叠、分界面不明显的现象。因此,本文采用测度学习的方法,根据和弦特征的特点,从问题本身的先验知识中有监督的学习到一个距离方程,使原始特征空间投影到一个类别区分度更高的空间,即希望在投影后的特征空间中,具有相同标签的特征向量更相似,具有不同标签的特征向量间区分度更大[14]。进而使用从真实数据中学习得到的测度矩阵改进高斯径向基函数的距离测度。具体改进方法如下介绍。

给定 $\mathbf{R}^{12}$ 空间的 n 个数据点 $[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]$,为了得到测度矩阵,通常采用马氏距离进行距离度量,即要找到一个度量矩阵 $\boldsymbol{M}$ 来衡量样本对 $(\boldsymbol{x}_i,\boldsymbol{x}_j)$ 之间的距离:

$$d_{\boldsymbol{M}}(\boldsymbol{x}_i,\boldsymbol{x}_j)=\sqrt{(\boldsymbol{x}_i-\boldsymbol{x}_j)^{\mathrm{T}}\boldsymbol{M}(\boldsymbol{x}_i-\boldsymbol{x}_j)} \tag{11}$$

为了求得度量矩阵 $\boldsymbol{M}$,Xing 等人[15]基于样本对距离提出了一种经典的测度学习算法。对于给定的相似对集合 $(\boldsymbol{x}_i,\boldsymbol{x}_j)\in\mathbf{S}$ 和区分对集合 $(\boldsymbol{x}_i,\boldsymbol{x}_j)\in\mathbf{F}$,则可以用下面的优化问题对度量矩阵 $\boldsymbol{M}$ 进行求解:

$$\begin{aligned}&\min_{\boldsymbol{M}\succ 0}\sum_{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in\mathbf{S}}d^2_{\boldsymbol{M}}(\boldsymbol{x}_i,\boldsymbol{x}_j)\\&\text{st.}\sum_{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in\mathbf{F}}d_{\boldsymbol{M}}(\boldsymbol{x}_i,\boldsymbol{x}_j)\geqslant 1\end{aligned} \tag{12}$$

从优化问题可以看出,目标函数为相似对集 $\mathbf{S}$ 中各个相似对中两个数据点的马氏距离平方和,而约束条件为区分对集合中所有区分对中两个数据点的马氏距离平方和。也就是说,在使用测度矩阵 $\boldsymbol{M}$ 进行空间映射之后的新空间内,相似对之间的空间距离变小,而区分对之间的空间距离更大。可以看出,该优化问题是凸优化问题,从而可以使用梯度下降算法求得全局最优解。进而可以用求得的度量矩阵 $\boldsymbol{M}$ 的最优解对 SVM 的高斯径向基核函数进行优化,得到如下的基于测度学习的高斯径向基核函数:

$$k_m=\exp\left(-\frac{(\boldsymbol{x}_i-\boldsymbol{x}_j)^{\mathrm{T}}\boldsymbol{M}(\boldsymbol{x}_i-\boldsymbol{x}_j)}{2\sigma^2}\right) \tag{13}$$

综合式(10)和式(13)可以得到 mlSVM 的分类决策函数:

$$f(\boldsymbol{x})=\mathrm{sgn}\left(\sum_{i=1}^{n}y_i\alpha_i^*\exp\left(-\frac{(\boldsymbol{x}_i-\boldsymbol{x})^{\mathrm{T}}\boldsymbol{M}(\boldsymbol{x}_i-\boldsymbol{x})}{2\sigma^2}\right)+b^*\right) \tag{14}$$

3 实验与结果分析

3.1 实验数据

本文采用的数据集为国际音频检索评测大赛 MIREX(Music Information Retrieval Evaluation eXchange)所提供的数据源。该数据集包括 Beatles 乐队的 12 张专辑,共 180 首歌曲(采样率 44100 Hz,比特率 16 bits,单通道)。本文使用来自学者 Chris Harte[16]对这些歌曲手动标记的标签文件作为和弦序列的真实的标签文件,并且将这些标签文件作为测试数据的生成依据和最终和弦输出的对比依据。其中标签文件的格式如图 3 所示。其中包括起始时间、结束时间以及持续时间内的和弦类型。

```
19.725      21.281  Cmaj
21.281      22.848  Gmaj
22.848      23.650  Fmaj
23.650      27.516  Cmaj
27.516      29.396  Gmaj
……
```

图 3 标签文件格式

Fig.3 Format of annotation file

在本文中将和弦分为 25 类,12 个根音音级分别对应的大调和弦和小调和弦,再加上一个无和弦类型。并根据标签文件将 180 首歌曲中的 3/4 音乐文件进行分割,得到单独的和弦片段,进而得到测试数据集 $\mathbf{S}^{\mathrm{train}}$,并作为得到测度学习模型和机器学习模型的训练数据,而剩余的 1/4 则作为测试数据 $\mathbf{S}^{\mathrm{test}}$ 进行最终识别结果的计算。

3.2 不同特征的主要音阶在持续时间内的方差比较

3.2.1 实验数据以及实验结果

实验依据图 1 所示的和弦识别系统的结构图,对测试数据集 $\mathbf{S}^{\mathrm{train}}$ 中音乐和弦片段提取表征和弦的特征矩阵,例如图 2 所示的矩阵,并按照 24 类不同的和弦类型 Cmaj,Cmin,…,Bmaj,Bmin 进行分类。从每类和弦的特征数据集中分别取 200 个数据作为本实验的实验数据。

然后,计算每类和弦的根音、三度音和五度音在和弦特征矩阵中对应行的方差,即和弦在持续时间内的变化频繁程度。实验结果如图 4 所示。

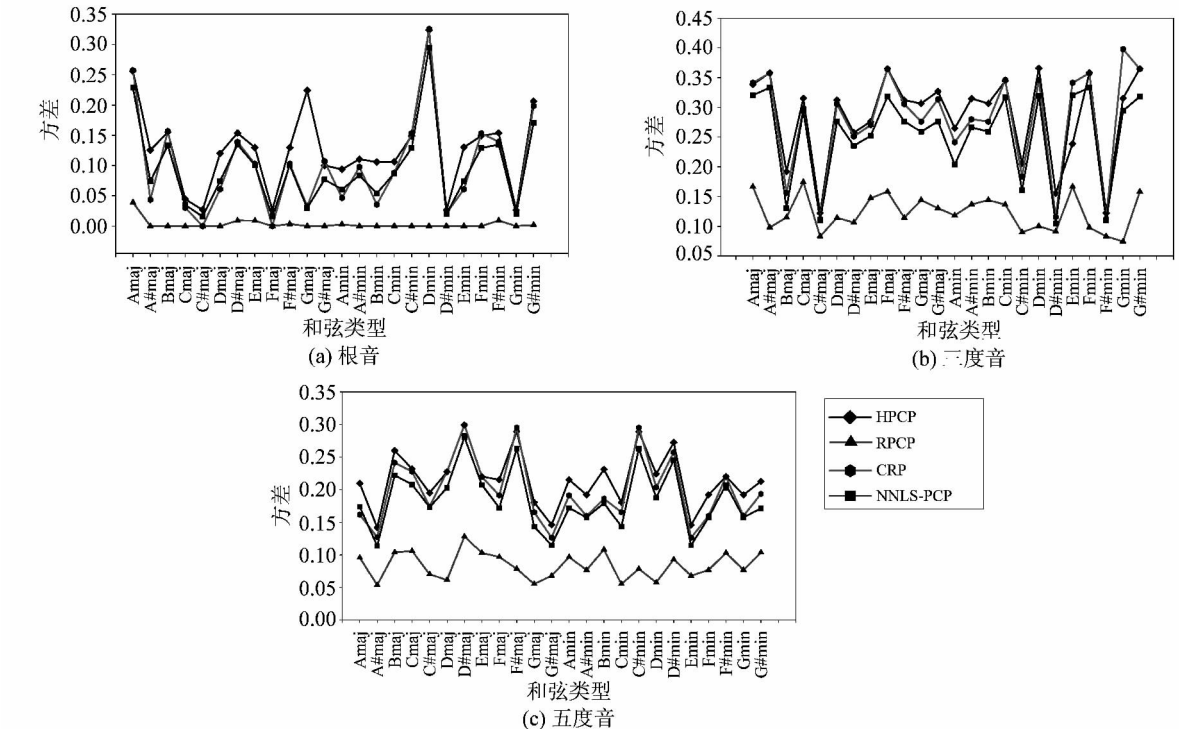

图 4　不同和弦特征的主要音阶在持续时间内的方差比较

Fig.4　Comparison of variance of main pitch during lasting time for different chord features

3.2.2　实验结果分析

从实验结果可以看出,在和弦的持续时间内,相比其他和弦特征,如 CRP、HPCP 和 NNLS-PCP,本文所采用的 RPCP 特征在根音、三度音和五度音之上均表现出较小的方差。这也说明了 RPCP 特征在和弦的持续时间内表现出更强的稳定性,即增强了 PCP 特征的鲁棒性。

在音乐演奏过程中,和弦的行进是一个复杂的过程,在这个过程中难以避免出现音符缺失、引入人声等情况,这会极大降低和弦的转录与识别的准确性。而本文所采用的 RPCP 特征在提取过程中,首先对原始音频频谱进行优化处理,分别对谐波成分和稀疏噪声成分进行建模,充分考虑了谐波成分的低秩特性和噪声成分的稀疏特性,并求解最优化问题得到优化后的具有鲁棒特性的频谱。优化后的频谱不仅分离了稀疏噪声,还对具有低秩特性的和弦谐波进行了重建,在一定程度上重建缺失频谱,使得和弦频谱在持续时间内具有更好的稳定性。随后对得到的频谱进行音阶映射,得到最终的 RPCP 特征。进而,RPCP 特征具有在持续时间内稳定的特性,同一和弦的不同音阶在帧与帧之间变化较小,即具有更小的方差。因此,从实验结果可以看出,相比其他和弦特征,本文所采用的 RPCP 特征具有更好的鲁棒性。

3.3　高斯距离与马氏距离比较

3.3.1　实验数据以及实验结果

实验依据图 1 所示的和弦识别系统的结构图,对测试数据集 $\mathbf{S}^{\text{test}}$ 中音乐和弦片段提取 RPCP 特征向量。为了得到相似对和区分对集合,本文从得到的 24 类特征集中各选取 100 个数据点。本文采用的是“一对余”(one-versus-rest) 多分类模型,因此需要设计 24 个相似对集合 Cmaj,Cmin,…,Bmaj,Bmin 和 24 个区分对集合 Cmaj/rest,Cmin/rest,…,Bmaj/rest,Bmin/rest,即每个相似对或者区分对集合对应一个和弦类型。相似对集设计如下: 对于每类和弦,从选取的 100 个数据点中随机组合构成 1000 个相似对作为训练相似对集合; 类似,同样另外选取 500 个相似对作为测试相似对集和。而区分对集合如下设计: 选定一和弦类型,并将剩余和弦类型的数据点进行混合构成混合数据集; 接着从选定的和弦类型所对应的特征向量集和混合数据集中随机选取数据点进行组合得到区分对集合; 最后,从区分对集合中选取 1000 个区分对作为训练数据集,选

取另外 500 个区分对作为测试数据集。

然后依据优化问题(12) 使用测试数据对马氏距离测度矩阵 $\boldsymbol{M}$ 进行求解,并计算相似对和区分对测试数据集中各个数据对的欧氏距离$(\boldsymbol{x}_i-\boldsymbol{x}_j)^{\mathrm{T}}(\boldsymbol{x}_i-\boldsymbol{x}_j)$ 以及马氏距离$(\boldsymbol{x}_i-\boldsymbol{x}_j)^{\mathrm{T}}\boldsymbol{M}(\boldsymbol{x}_i-\boldsymbol{x}_j)$,实验结果分别如图 5 所示。

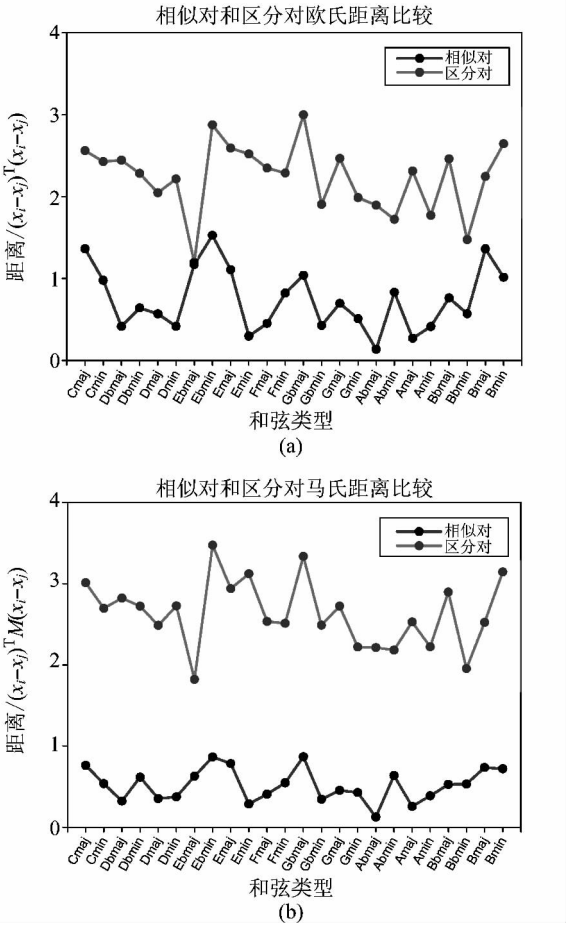

图 5 相似对集和区分对集的欧氏距离与马氏距离比较

Fig.5 Comparison of Euclid distance and Mahalanobis distance in similar pair set and dissimilar pair set

3.3.2 实验结果分析

从图 5(a) 和图 5(b) 对比可以看出,使用解优化问题(12) 得到的距离测度矩阵 $\boldsymbol{M}$ 计算得到的马氏距离能够使属于同一和弦的特征向量之间的距离,即类内间距变小; 同时使不属于同一和弦的特征向量之间的距离,即类间间距变大。总的来说,基于测度矩阵的马氏距离能够使得不同类之间的间隙变得更大。从支持向量机的判决函数,即式(14) 的角度而言,用马氏距离替换 SVM 高斯核函数中的欧氏距离能够使判决函数具有更强的鲁棒性。例如,假设一数据点 x 属于正类,即标签为+1,由于类内数据分布更加紧凑,那么对于正类支持向量就有

$$y_i\alpha_i^*\exp\left(-\frac{(\boldsymbol{x}_i-\boldsymbol{x})^{\mathrm{T}}\boldsymbol{M}(\boldsymbol{x}_i-\boldsymbol{x})}{2\sigma^2}\right)=$$
$$\alpha_i\exp\left(-\frac{(\boldsymbol{x}_{+sv}-\boldsymbol{x})^{\mathrm{T}}\boldsymbol{M}(\boldsymbol{x}_{+sv}-\boldsymbol{x})}{2\sigma^2}\right)>$$
$$\alpha_i\exp\left(-\frac{(\boldsymbol{x}_{+sv}-\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{x}_{+sv}-\boldsymbol{x})}{2\sigma^2}\right) \quad (15)$$

同时由于类间间距变大,那么对于负类支持向量有

$$y_j\alpha_j^*\exp\left(-\frac{(\boldsymbol{x}_j-\boldsymbol{x})^{\mathrm{T}}\boldsymbol{M}(\boldsymbol{x}_j-\boldsymbol{x})}{2\sigma^2}\right)=$$
$$-\alpha_j\exp\left(-\frac{(\boldsymbol{x}_{-sv}-\boldsymbol{x})^{\mathrm{T}}\boldsymbol{M}(\boldsymbol{x}_{-sv}-\boldsymbol{x})}{2\sigma^2}\right)>$$
$$-\alpha_j\exp\left(-\frac{(\boldsymbol{x}_{-sv}-\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{x}_{-sv}-\boldsymbol{x})}{2\sigma^2}\right) \quad (16)$$

其中 α_i,α_j 分别表示正类支持向量和负类支持向量所对应的拉格朗日乘子,$\boldsymbol{x}_{+sv}$、$\boldsymbol{x}_{-sv}$表示正类支持向量和负类支持向量。由于非支持向量的拉格朗日乘子为 0,所以判决函数式(14) 的计算实际就是对式(15) 和式(16) 进行求和。很明显可以看出

$$\sum_{i=1}^{n}y_i\alpha_i^*\exp\left(-\frac{(\boldsymbol{x}_i-\boldsymbol{x})^{\mathrm{T}}\boldsymbol{M}(\boldsymbol{x}_i-\boldsymbol{x})}{2\sigma^2}\right)>$$
$$\sum_{i=1}^{n}y_i\alpha_i^*\exp\left(-\frac{(\boldsymbol{x}_i-\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{x}_i-\boldsymbol{x})}{2\sigma^2}\right) \quad (17)$$

也就是说,对于改进之后的支持向量机,高斯核函数中包含着真实数据的分布信息,并且能使不同类别的数据点之间的间隙变大,从而使得测试数据点距离分类超平面的距离变得更远,即改进之后的支持向量机对测试数据具有更好的鲁棒性。

3.4 和弦识别结果与比较

3.4.1 实验数据以及实验结果

实验依据图 1 所示的和弦识别系统的结构图,对测试数据集 $\mathbf{S}^{\mathrm{train}}$中音乐和弦片段提取表征和弦的特征,并将这些特征作为多分类 mlSVM 的训练数据进行模型训练,从而获得 SVM 的训练参数 a^* 和 b^* 。对 mlSVM 的参数寻优采用网格寻优: 采用 5 折交叉验证,惩罚参数 C 和高斯径向基函数核半径 $g=1/\sigma^2$的寻优区间均为 $[10^{-5},10^5]$。并将之前计算好的测度矩阵 $\boldsymbol{M}$ 代入到改进后的高斯核函数中。接着,选取 $\mathbf{S}^{\mathrm{test}}$ 歌曲作为测试数据进行测试。为了说明本文所提特征与识别算法的有效性,本文选取

HPCP[3]，CRP[7]，NNLS-PCP[5] 特征作为对比特征，同时选取 HMM[9]，DBN[10]，CRF[11] 作为对比识别模式来进行实验对比。为了直观地说明识别效果，依据如下公式作为识别效果的评价指标:

$$\text{识别率} = \frac{\text{正确识别的帧数}}{\text{总帧数}} \times 100\% \quad (18)$$

具体实验结果如表 1 所示。

表 1　识别结果对比(%)

Tab.1　Comparison of recognition results(%)

特征	识别模式				
	HMM	SVM	DBN	CRF	mlSVM
HPCP	80.7	79.4	82.3	83.1	**85.5**
CRP	78.2	77.5	84.3	82.4	**84.8**
NNLS-PCP	82.9	82.0	85.1	88.3	**89.2**
RPCP	83.6	84.2	85.8	89.4	**92.9**

3.4.2　实验结果分析

本文基于信号处理、矩阵分析理论和凸优化相关理论提出了一种基于 RPCP 特征和测度学习支持向量机的音乐和弦识别算法。

PCP 特征基于音乐理论，压缩了信号的能量，但是当音乐出现人声时，此时对信号频谱能量进行压缩，有可能在特征矩阵引入较为明显的噪声，使得信号能量分散，不能集中于和弦所对应的音阶之上，使特征向量的空间分布产生偏差，从而导致最终分类错误。基于谐波成分的核范数约束和稀疏噪声的一范数约束优化后得到的 RPCP 特征不仅能够重建和弦的谐波信息，还表现出对稀疏噪声良好的鲁棒性。因此，从实验结果可以看出，相比采用 HPCP、NNLS-PCP 和 CRP 特征的和弦识别系统，采用 RPCP 作为和弦特征的识别系统的分类正确率平均提高了 1.3%~5.8%。

支持向量机的高斯核函数的距离测度通常采用欧氏距离。这种做法虽然简单易行，但是对于特定数据集的数据，欧氏距离测度并不能反映数据集的分布特点，在使用判别函数进行分类时，不能达到预期的效果。本文所采用的基于马氏距离的测度学习算法能够根据和弦特征空间中特征向量的分布得到一个测度矩阵。这个矩阵包含着特征空间中特征向量的分布信息，也就是说测度矩阵包含有特征分布的先验知识。因此，采用测度学习算法对高斯核函数进行改进，能够拉大不同类别和弦特征向量之间的距离，同时缩小相同和弦类型特征向量之间的距离，因而能够更加有效地改善和弦识别效果。从表 1 中可以看出，相比于比较流行的 HMM 模型、常规 SVM、DBN 网络和 CRF 模型，采用 mlSVM 作为识别模式的识别系统的分类正确率平均提高了 2.3%~6.8%。

4　结论

本文采用了一种有效地提取鲁棒性 PCP 特征的算法，该算法使用核范数和 1 范数结合的混和范数约束。该算法所得到的 RPCP 特征不仅能够削弱大而稀疏噪声，同时也能够对和弦特征矩阵的缺失部分进行必要的填充，使得所得到的和弦信息更加完整，更加鲁棒，为和弦识别算法打好基础。最终采用基于测度学习的支持向量机对和弦特征进行分类，同时同现今流行的和弦特征和识别模式进行双向对比，实验结果表明，本文所提出的 RPCP 特征和 mlSVM 算法能够有效地提高和弦的分类正确率。

参考文献

[1] McVicar M, Santos-Rodríguez R, Ni Y, et al. Automatic chord estimation from audio: A review of the state of the art [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2014, 22(2): 556-575.

[2] Fujishima T. Realtime chord recognition of musical sound: a system using Common Lisp Music [J].Proc Icmc, 1999: 464-467.

[3] Herrera P. Automatic Extraction of Tonal Metadata from Polyphonic Audio Recordings [C]// Audio Engineering Society Conference: 25th International Conference: Metadata for Audio. Audio Engineering Society, 2004: 74-80.

[4] Lee K. Automatic Chord Recognition from Audio Using Enhanced Pitch Class Profile [C] // ICMC, 2006: 306-313.

[5] De Haas B, Magalhães J P, Wiering F. Improving Audio Chord Transcription by Exploiting Harmonic and Metric Knowledge [C]//ISMIR, 2012: 295-300.

[6] Deng J, Kwok Y K. A Hybrid Gaussian-HMM-Deep-Learning Approach For Automatic Chord Estimation with Very Large Vocabulary [C]//ISMIR, 2016.

[7] Muller M, Ewert S. Towards timbre-invariant audio features for harmony-based music [J]. IEEE Transactions

on Audio, Speech, and Language Processing, 2010, 18 (3): 649-662.

[8] Korzeniowski F, Widmer G. Feature Learning for Chord Recognition: The Deep Chroma Extractor [C] // Proceedings of the 17th International Society for Music Information Retrieval Conference (ISMIR), 2016: 37-43.

[9] Maruo S, Yoshii K, Itoyama K, et al. A feedback framework for improved chord recognition based on NMF-based approximate note transcription [C] // 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2015: 196-200.

[10] Di Giorgi B, Zanoni M, Sarti A, et al. Automatic chord recognition based on the probabilistic modeling of diatonic modal harmony [C] // Multidimensional Systems (nDS), 2013. Proceedings of the 8th International Workshop on. VDE, 2013: 1-6.

[11] Wang F, Zhang X. Research on CRFs in music chord recognition algorithm [J]. J. Comput, 2013, 8: 1017.

[12] Huang P S, Chen S D, Smaragdis P, et al. Singing-voice separation from monaural recordings using robust principal component analysis [C] // Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference, 2012: 57-60.

[13] Lin Z, Chen M, Ma Y. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices [J]. arXiv Preprint arXiv: 1009.5055, 2010.

[14] Kulis B. Metric learning: A survey [J]. Foundations and Trends in Machine Learning, 2012, 5(4): 287-364.

[15] Xing E P, Ng A Y, Jordan M I, et al. Distance metric learning with application to clustering with side-information [J]. Advances in Neural Information Processing Systems, 2003, 15: 505-512.

[16] Harte C. Towards automatic extraction of harmony information from music signals [D]. Department of Electronic Engineering, Queen Mary, University of London, 2010.

作者简介

王蒙蒙 男,1991 年生,河南襄城县人。天津大学硕士研究生,主要研究方向为信号处理和模式识别。

E-mail: wangmengjunior@ tju.edu.cn

关 欣 女,1977 年生,河北石家庄人。2009 年获天津大学博士学位,现为天津大学讲师,主要研究方向为音乐信息检索。

E-mail: guanxin@ tju.edu.cn

李 锵 男,1974 年生,山西太原人。天津大学教授,博士生导师,主要研究方向为信息处理和滤波器设计。

E-mail: liqiang@ tju.edu.cn