

自适应粒子群优化匹配追踪声音事件识别算法

苏映新*

辽东学院信息工程学院, 辽宁 丹东 118000

摘要 针对公共环境中的声音事件识别问题, 提出基于自适应粒子群优化(PSO)匹配追踪(MP)稀疏分解的声音事件识别算法。该算法在分析 MP 稀疏分解的基础上, 先基于适应度函数改进 PSO 算法相关参数的自适应设置, 再基于自适应 PSO 算法构建优化 MP 稀疏分解的目标函数及信号重构函数, 实现自适应 PSO 算法优化 MP 稀疏分解, 从而借助连续 Gabor 超完备集来提高最优原子的匹配程度, 增强声音信号, 提高特征的分类性能, 最后使用优化的支持向量机(SVM)和复合特征实现公共环境中的声音事件准确识别。实验结果表明, 与已有算法相比, 所提识别算法显著降低了计算量, 并取得了最优的声音识别率, 且具有较好的识别鲁棒性。

关键词 机器视觉; 声音事件识别; 自适应粒子群算法; 匹配追踪; 稀疏分解; 支持向量机

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/LOP57.101502

Sound Event Recognition Based on Adaptive Particle Swarm Optimized Matching Tracking

Su Yingxin*

College of Information Engineering, Eastern Liaoning University, Dandong, Liaoning 118000, China

Abstract A new algorithm based on optimized match pursuit (MP) sparse decomposition using adaptive particle swarm optimization (PSO) is proposed to address the recognition problem of sound events in public environment. Based on MP sparse decomposition analysis, the fitness function was used to improve the adaptive setting of parameters related to PSO algorithm; then, the objective function and signal reconstruction function were constructed for optimizing sparse decomposition, thus realizing adaptive PSO algorithm optimized MP sparse decomposition. Moreover, the continuous Gabor super complete set was used to improve the matching degree of the optimal atom, which enhanced the sound signal and improved the classification performance of the feature. Finally, optimized support vector machine (SVM) and composite features were used to achieve accurate recognition of sound events in public environments. Experimental results show that the proposed algorithm significantly reduces computational complexity, achieves optimal recognition rate, and demonstrates better robustness compared with existing algorithms.

Key words machine vision; sound event recognition; adaptive particle swarm optimization algorithm; match pursuit; sparse decomposition; support vector machine

OCIS codes 150.1135; 070.1060; 120.0280; 100.2000; 060.2630

1 引言

随着经济的发展和人民生活水平的提高, 安全监控受到越来越多的重视。由于突发事件通常伴有事件相关的异常声音, 因此, 公共场所声音事件的精确识别可作为公共安全监控的重要辅助和关键组

成, 具有重要的研究意义^[1-2], 已成为当前公共安全领域监控的重要发展方向。

李艳雄等^[3]提取声音信号的梅尔频率倒谱系数(MFCC)特征和能量谱特征, 利用混合高斯模型对组合特征进行分类识别, 但模型阶数难以合理设置。冯陈定等^[4]采用相关系数和过零率等多特征对

收稿日期: 2019-08-19; 修回日期: 2019-10-10; 录用日期: 2019-10-18

基金项目: 辽宁省教育厅科学研究项目(ldxy2017008)

* E-mail: su_yingx_555@163.com

四类声信号进行支持向量机(SVM)分类,对于类别数不多的异常声音具有较好的识别效果。Zhao等^[5]以 α 分布分析枪声等异常声音特征,并进行非线性变换以增强特征的抗噪性,通过深度置信网络(DBN)实现异常声音的识别,对几种典型枪声具有较好的识别效果。李伟红等^[6]分析了公共环境背景噪声的 T 分布特性,以极点对称模态分解排列熵作为分类判定准则,通过SVM实现异常声音的识别,提高了识别率并缩短了信号分解时间。李应等^[7]提出基于Gammatone滤波器组和Zigzag扫描的多频带能量声音检测算法,在低信噪比下取得较好的检测效果。McFee等^[8]采用稀疏合成非负矩阵分解提取声音事件的特征字典,通过多类平均SVM进行特征识别。

上述方法均取得较好的声音识别效果,但由于声音事件的多样性和公共场所背景噪声的复杂性,真实环境噪声对特征的识别形成较强干扰。匹配追踪(MP)利用信号的稀疏性对信号进行分解并重构,适用于各种非平稳噪声环境下的声音事件识别。陈秋菊等^[9]通过优化的MP稀疏分解提取最优匹配原子的时频参数特征,通过改进的DBN进行声音事件识别,但其参数设置存在一定的局限性。Nivitha等^[10]在优化MP稀疏分解重构声音事件信号的基础上,利用短时谱估计对MP剩余信号进行再次重构,以消除背景噪声干扰,算法对多种声音事件识别都取得较好的识别率。Davila-Chacon等^[11]提取MP特征和MFCC特征,利用 K 最近邻算法和高斯混合模型对声音进行分类。韦娟等^[12]对声音信号进行MP稀疏分解后,利用主成分分解和隐含狄利克雷分布对最优匹配原子组进行降维,通过SVM识别降维特征。

本文在已有研究的基础上,提出基于自适应粒子群优化(PSO)算法优化MP稀疏分解的声音识别算法。该算法设计新的目标函数实现由改进的自适应PSO算法优化MP稀疏分解,从而借助连续Gabor超完备集来提高最优原子的匹配程度,最后使用优化的SVM和复合特征实现公共环境中的声音事件准确识别,实验结果表明,该算法有效提高了声音事件的识别准确度,并对环境噪声具有较好的鲁棒性。

2 MP信号稀疏分解

设集合 $\mathbf{D} = \{\mathbf{g}_i, i = 1, 2, \dots, Q\}$ 为张成空间 $\mathbf{H} = \mathbf{R}^N$ 所需的超完备库, N 为空间维数,且有 $Q \gg$

N ,则对于任意信号 $\mathbf{f} \in \mathbf{H}$ 都可由 \mathbf{D} 中若干匹配原子的线性组合表示^[13],即 $\mathbf{f} = \mathbf{a} \cdot \mathbf{g}$,其中 $\mathbf{a} = \{\alpha_1, \alpha_2, \dots, \alpha_m\}$ 为原子集组合表示所需的展开系数, $\mathbf{g} = \{\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_m\}$ 为信号稀疏表示所得的最优原子集,通常情况下最优原子数远小于原子集中的原子数,即 $m \ll Q$ 。

MP稀疏分解原理:从 \mathbf{D} 中选取原子 \mathbf{g}_i^1 ,使其满足

$$|\langle \mathbf{f}, \mathbf{g}_i^1 \rangle| = \sup_{\mathbf{g}_i \in \mathbf{D}} |\langle \mathbf{f}, \mathbf{g}_i \rangle|, \quad (1)$$

式中: \mathbf{g}_i^1 为第一次匹配追踪到的最优原子,即与信号 \mathbf{f} 最接近的原子; $\langle \mathbf{f}, \mathbf{g}_i^1 \rangle$ 表示内积; $\sup(\cdot)$ 为取极大值函数。由于 $m \ll Q$,因此(1)式存在多种表示形式,稀疏分解即为在所有 \mathbf{f} 的线性组合表达式中,取 m 值最小的表达式。对第 k 次迭代后的剩余信号 $R^k \mathbf{f}$ 进行重复分解,可得到

$$R^k \mathbf{f} = \langle R^k \mathbf{f}, \mathbf{g}_i^{k+1} \rangle \mathbf{g}_i^{k+1} + R^{k+1} \mathbf{f}, \quad (2)$$

式中: \mathbf{g}_i^{k+1} 满足 $|\langle R^k \mathbf{f}, \mathbf{g}_i^{k+1} \rangle| = \sup_{\mathbf{g}_i \in \mathbf{D}} |\langle R^k \mathbf{f}, \mathbf{g}_i \rangle|$ 。

令 $R^0 \mathbf{f} = \mathbf{f}$,经实验发现随着 k 值的增加, $R^k \mathbf{f}$ 逐渐趋向于零^[12],经 k 次匹配追踪后,此时信号 \mathbf{f} 可稀疏分解为

$$\mathbf{f} \approx \sum_{k=1}^K \langle R^{k-1} \mathbf{f}, \mathbf{g}_i^k \rangle \mathbf{g}_i^k. \quad (3)$$

由于MP算法在稀疏分解过程中遍历超完备原子集 \mathbf{D} 中的所有原子 \mathbf{g}_i ,因而要求 \mathbf{D} 中的原子数有限^[5]。Gabor字典集由函数 $g(t) = \exp(-\pi t^2)$,其中 t 为时间,通过调整伸缩因子 s 、频率相位因子 v, ω 及平移因子 u 得到,即

$$g_\gamma(t) = \frac{1}{\sqrt{s}} g\left(\frac{t-u}{s}\right) \cos(vt + \omega), \quad (4)$$

式中: $g_\gamma(t)$ 为用于计算原子集中原子的函数,将其离散化即得到一个原子 $\mathbf{g}_i, \boldsymbol{\gamma} = \{\gamma_1, \gamma_2, \dots, \gamma_Q\}$ 为时频参数集。 v, ω 和 s 控制原子的波形, u 控制其中心位置,函数局部化特性保证了原子的非平稳性^[9]。根据(4)式,一个原子 \mathbf{g}_i 对应一组时频参数集,即 $\boldsymbol{\gamma}_i = \{s_i, u_i, v_i, \omega_i\}$,其描述了待分解信号的特征。

3 自适应PSO优化稀疏分解

3.1 自适应PSO算法

PSO算法^[13]的描述:种群粒子以一定的速度在 d 维空间飞行,根据粒子个体历史最优位置和种群历史最优位置更新当次迭代的粒子位置以获得最优解。但经典算法易陷入局部最优,且惯性因子的不合理设置也会造成PSO算法的“早熟”问题^[7]。

增强粒子群算法初期的全局搜索能加快其收敛速度,避免算法在接近收敛时陷入局部最优解。获得稳定的更优解,是粒子群算法改进的主要方向。基于此提出了区域划分自适应粒子群算法,每次迭代时,根据粒子与全局最优粒子的距离将种群划分为两部分:距离全局最优粒子较近的粒子,提高其局部搜索能力,以获得更好的最优解;较远的粒子,则增强其全局搜索能力,以期发现潜在的更好区域,避免算法陷入局部最优。粒子飞行轨迹会受到惯性因子和学习因子的影响,合理调整惯性因子有利于平衡粒子的全局和局部搜索性能,为此,每次迭代完成区域划分后,先根据适应度值计算自适应调整因子,表达式为

$$\delta(t) = \frac{1}{N} \sum_{i=1}^N f_{it}(p_i^t) / f_{it}(p_g^t), \quad (5)$$

式中: $\delta(t) \in (0, 1]$ 用于调整惯性因子; $f_{it}(t)$ 为适应度函数; p_i^t 和 p_g^t 分别为至 t 时刻粒子的历史最优位置和种群的历史最优位置。基于 $\delta(t)$ 的变化规律,可自适应调整惯性因子 $\omega(t)$,表达式为

$$\omega(t) = \exp[-\delta(t)]. \quad (6)$$

PSO算法中的学习因子 c_1 和 c_2 分别决定粒子个体和种群的历史最优对后续粒子速度的影响程度。由于初期进化很大可能仅形成局部最优,因而在初期时决定个体历史最优影响能力的 c_1 值应起主导作用,以增强粒子的自我进化形成更多潜在最优区域划分,随着进化的继续, c_2 的作用逐渐凸显,以增加最优区域的局部寻优能力,这样学习因子的自适应控制方式为

$$\begin{cases} c_1(t) = c_0[2 - \delta(t)] \\ c_2(t) = c_0[\delta(t) + 1] \end{cases}, \quad (7)$$

式中: c_0 为比例因子。根据(7)式, c_1 和 c_2 可在迭代进化过程中自适应地平衡个体和种群的历史最优位置对后续粒子速度的影响程度。

3.2 自适应粒子变异

基于3.1节因子的自适应调整改进,粒子群算法的收敛速度和避免局部最优的能力有较大提高,但仍存在局部最优的情况,为此,采用自适应粒子变异进一步避免算法陷入局部最优。

设某次迭代过程中全局最优粒子为 a ,则种群内粒子与最优粒子间的距离函数 $s(L)$ 为

$$s(L) = \sum_{j=1, j \neq a}^N (L_{aj} \pi)^{-1}, \quad (8)$$

式中: L_{aj} 为粒子 j 与 a 的欧氏距离。

如果当前最优粒子 a 在相邻 k 次迭代内无明显

改善, k 的取值范围为 $5\% \leq (k/k_{\max}) \leq 10\%$,其中 k_{\max} 为最大迭代次数,这表明种群粒子集中到一个极小的区域中,此时,需对粒子进行自适应变异,且变异概率取决于粒子与 a 的欧氏距离值,距离值越小,变异概率越大,其表达式为

$$P = d \left(\frac{L_{\max} - L_{aj}^k}{L_{\max} - L_{\min}} \right), \quad (9)$$

式中: L_{\max} 和 L_{\min} 分别为欧氏距离的最大值和最小值; L_{aj}^k 为第 k 次迭代的 j 与 a 之间的距离; $d \in [0, 1]$ 为调节概率范围的常量。

对于粒子 j ,如果随机数 $r \in [0, 1]$ 满足 $r \leq p$,则需进行变异,在上述小区域外随机选取粒子 o ,则变异操作为

$$\mathbf{x}_j^{k, \text{new}} = \mathbf{x}_j^k + r(\mathbf{x}_j^k - \mathbf{x}_o^k), \quad (10)$$

式中: \mathbf{x}_j^k 和 \mathbf{x}_o^k 分别为第 k 次迭代中,粒子 j 和粒子 o 的位置矢量; $\mathbf{x}_j^{k, \text{new}}$ 为变异后的粒子位置矢量。

3.3 自适应稀疏分解构建

设含噪声音信号为 $\mathbf{f} = \mathbf{f}_s + \mathbf{f}_n$,其中 \mathbf{f}_s 为纯净声音信号, \mathbf{f}_n 为噪声信号。由于 \mathbf{D} 中原子结构与 \mathbf{f}_s 相关^[9],且字典集为连续超完备,因此 \mathbf{D} 中必包含与 \mathbf{f}_s 匹配最优的原子,MP稀疏分解就是寻找该最优原子。在内积定义^[12]的基础上,相关性表达式为

$$J(\mathbf{f}, \mathbf{g}_i) = \frac{\langle \mathbf{f}, \mathbf{g}_i \rangle}{\|\mathbf{f}\|}, \quad (11)$$

式中: $\|\cdot\|$ 为计算信号的幅值; $J(\mathbf{f}, \mathbf{g}_i) \in [0, 1]$ 为声音信号与原子的相关程度, $J(\mathbf{f}, \mathbf{g}_i) = 1$ 说明完全相关, $J(\mathbf{f}, \mathbf{g}_i) = 0$ 说明不相关,则自适应稀疏分解的目标函数 $\Gamma(\cdot)$ 及其重构函数 $\mathbf{f}_R(\cdot)$ 分别为

$$\Gamma = \max J(\mathbf{f}, \mathbf{g}_i), \quad (12)$$

$$\mathbf{f}_R = \|\mathbf{f}\| \Gamma \mathbf{g}_{\text{best}}, \quad (13)$$

式中: $\mathbf{g}_{\text{best}} = \{\mathbf{g}_i | \max J(\mathbf{f}, \mathbf{g}_i)\}$ 为(12)式迭代得到的最优匹配原子,其对应一组 $\boldsymbol{\gamma}_{\text{best}} = \{s, u, v, \omega\}$, $\boldsymbol{\gamma}_{\text{best}}$ 也反映了声音事件信号的特征。

4 特征提取与声音事件识别

4.1 稀疏分解复合特征

自适应稀疏分解经迭代运算后得到一组最优匹配原子集,其对应的 $\boldsymbol{\gamma}_{\text{best}}$ 反映了声音事件信号的特征,取 s 和 v 的均值和标准差,构成声音事件检测的特征。

设含噪信号分帧为 M 帧,对信号进行自适应稀疏分解和重构后,得到时频参数组 \mathbf{s} 和 \mathbf{v} ,则第 λ 帧信号的检测特征为 $\mathbf{F}_{\text{mp}}(\lambda) = \{u(s_\lambda, v_\lambda), \sigma(s_\lambda, v_\lambda)\}$ 。

对重构信号进行离散余弦变换和 Zigzag 扫描降维, 计算得到由多频带能量特征(MBP)^[9]、基于听觉感知的 MFCC 及声音基频特征(PITCH)与 \mathbf{F}_{mp} 组成的复合识别特征(CRF), 即

$$\mathbf{F}_{CRF}(\lambda) = \{\mathbf{F}_{mp}(\lambda), \mathbf{F}_{MBP}(\lambda), \mathbf{F}_{MFCC}(\lambda), \mathbf{F}_{PITCH}(\lambda)\}. \quad (14)$$

4.2 SVM 声音事件识别

SVM 在少量样本情况下仍能取得较好的分类识别结果, 适用于公共场所声音事件检测环境。基于决策导向无环图(DDAG)的 SVM 扩展策略(记为 D-SVM), 在不增加计算量的情况下能够缓解不对称样本分类问题, 且无分类盲区。但 DDAG 存在多种节点排列结构可供选择, 为此, 结合已有改进^[14]方法, 优化 D-SVM 算法的步骤如下。

1) 将 y 类声音事件的训练样本数据成对组合, 生成 $y(y-1)/2$ 个二分 SVM 分类器 C_{ij} , 其中 i, j 为训练样本序号, 集合 $S = \{1, 2, \dots, y\}$ 描述 y 类声音事件。

2) 从训练样本中随机选取部分数据对生成的各类 SVM 进行识别率评估, 定义识别率为

$$p = (U_1/U) \times 100\%, \quad (15)$$

式中: U 和 U_1 分别为评估样本总数和准确分类的样本数。

3) 对 C_{ij} 的识别率值进行排序, 以识别率最大值的 C_{ij} 作为 DDAG 的根节点, 构建决策, 即

$$f_{i,j}(x) = \text{sign} \left[\sum_{V_C} y_m \alpha_m k(x - x_m) - b \right], \quad (16)$$

表 1 算法计算量及重构性能评价

Table 1 Evaluation of algorithm calculation and reconstruction performance

Algorithm	Ns	Nip	Time /s	MSE	NCC	SNR
OMP	218355	4459471	21705.4	0.0050	0.9241	15.9125
Pso-MP	23698	491720	1626.9	0.0255	0.7695	9.3298
APso-MP	897	3720	175.4	0.0047	0.9550	16.3921

表 1 中 Ns、Nip 及运行时间反映算法的计算量, 可以看到 Pso-MP 和 APso-MP 的计算量明显低于原始 OMP 算法, 说明采用 PSO 算法优化 MP 稀疏分解可有效降低稀疏分解的计算量, 而 APso-MP 算法进一步降低了计算量, 原因在于使用连续超完备集可实时产生所需最优匹配原子, 节省完备集的生成及原子搜索所需的时间开销。MSE、NCC 及 SNR 反映信号的重构性能, 可以看到, Pso-MP 算法的性能最低, 因为其参数设置不适用于所有声音信

式中: $\text{sign}(\cdot)$ 为符号函数; V_C 为节点集; α_m 为第 m 个拉格朗日乘子; y_m 为第 m 个类标志; x_m 为第 m 个样本; $k(x - x_m)$ 为决策核函数; b 为调节标量。 $f_{i,j}(x) = 1$ 说明样本 x 不属于类 j , 即 $x \notin j$, 从 S 中去除类 j ; $f_{i,j}(x) = -1$, 则有 $x \notin i$, 从 S 中去除类 i 。

优化的 D-SVM 算法减少了根节点随机选择导致的分类误差累积, 提高了分类识别率, 且对于不同实验数据和类别, 只需对 C_{ij} 值进行排序, 并以识别率最大值的 C_{ij} 作为根节。

5 实验分析

以 Freesound^[15] 库的动物叫声和 Webmaster 网站的爆炸声等 5 种异常声音作为实验数据, 分析自适应稀疏分解声音事件识别算法(简记为 APso-MP)的性能。声信号样本的采样率调整为 $f_s = 22.05$ Hz, 选取每类声音的 500 个样本作为训练数据, 50 个样本作为测试数据, 并以雨天环境噪声、办公室环境噪声、Noisex92 噪声库中的 babble 噪声和高斯白噪声共四种噪声按不同的信噪比与测试数据混合进行实验。

5.1 自适应稀疏分解性能分析

随机选取动物叫声作为实验样本, 以 -5 dB 分段信噪比与四种噪声随机混合作为测试数据, 采用所提算法、传统 PSO 优化的 MP 稀疏分解算法(记为 Pso-MP)^[11] 及 OMP 算法进行实测信号的稀疏分解和重构, 以搜索次数(Ns)、内积计算量(Nip)、均方误差(MSE)、波形相似性(NCC)^[11] 及信噪比(SNR)作为评价指标, 多次实验结果的平均值如表 1 所示, 信号重构效果如图 1 所示。

号, 而 APso-MP 算法的性能最优, 一方面因为算法对不同声音信号可自适应选择相关参数, 另一方面连续原子集计算的最优匹配原子, 不会因为参数离散化而带来精度损失。

图 1(a) 为实验中包含 2 个有效音节的纯净动物叫声声谱, 图 1(b) 为该叫声与雨天环境噪声混合后信号的声谱, 图 1(c)~1(e) 为三种比较算法对叫声信号进行稀疏分解和重构后所得信号的声谱。从图 1 可以看到, Pso-MP 算法的重构信号含有较多

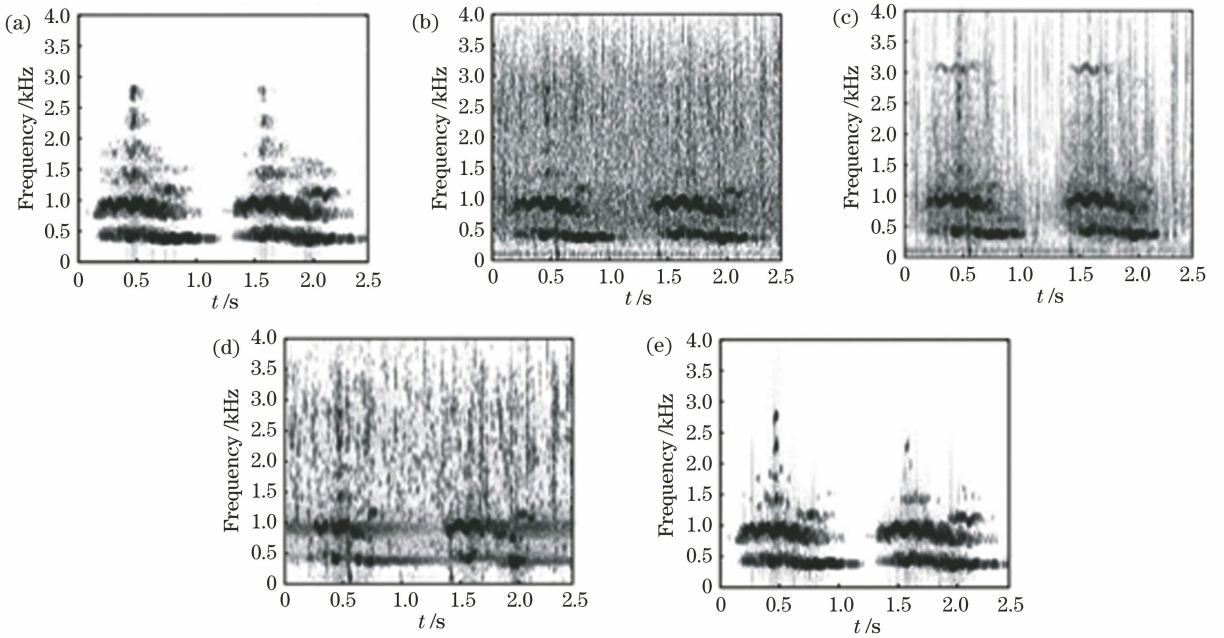


图1 动物叫声及不同算法重构信号的声谱图。(a)动物纯净叫声;(b)加入噪声的叫声;(c) Pso-MP 重构信号;(d) OMP 重构信号;(e) APso-MP 重构信号

Fig. 1 Animal calls and sound spectra of reconstructed signals by different algorithms. (a) Pure animal calls; (b) noise-added calls; (c) Pso-MP reconstructed signals; (d) OMP reconstructed signals; (e) APso-MP reconstructed signals

背景噪声,OMP 算法的重构信号对背景噪声有较好的抑制作用,但使用该算法对原始纯净信号造成较大损失,APso-MP 算法的重构信号在有效抑制背景噪声的同时,很好地保留原始纯净动物叫声,这与表 1 实验结果相符合。

实验结果表明,APso-MP 算法重构信号具有较好的准确性和鲁棒性,有效减少计算量,降低最优原子搜索的复杂度,有利于信号快速稀疏分解。

5.2 算法声音识别性能比较

为了验证所提算法的声音事件识别性能,以(16)式的识别率作为评价指标,以 APSVM^[11]、MPSVM^[15]、Pso-MPSRF^[9]、APso-MPSVM 和 APso-MPDSVM 作为比较算法。APSVM 算法提取声音信号的分类特征,以 SVM 作为分类器;MPSVM 算法在提取 MP 特征前先对信号进行主成分分解降维;APso-MPSVM 算法以 $\mathbf{F}_{CRF}(\lambda)$ 作为识别特征,以 SVM 作为分类器;Pso-MPSRF 以粒子群优化稀疏分解得到的复合特征作为特征,以随机森林(BF)算法作为分类器,决策树数为 500,使用训练样本训练各算法的分类器,再将测试样本与噪声以不同的分段信噪比进行混合。图 2 为在不同信噪比下,各算法对含噪测试样本的识别结果。

从图 2 可以看到,基于优化 MP 稀疏分解的算法得到复合特征的平均识别率整体高于 APSVM

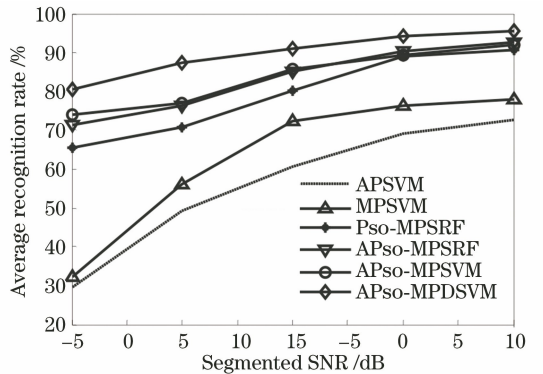


图2 不同算法在四种噪声环境的平均识别率
Fig. 2 Average recognition rate of different algorithms in four noisy environments

算法和 MPSVM 算法,尤其在低信噪比时效果更明显;在使用 BF 分类器时,APso-MPSRF 算法的平均识别率更高,说明改进 MP 稀疏分解和重构后得到信号的 $\mathbf{F}_{CRF}(\lambda)$ 特征具有更好的分类特性;对比 APso-MPSVM 和 APso-MPDSVM 的实验结果可以看到,在相同的识别特征下,APso-MPDSVM 的识别率更高,说明对 SVM 进行多分类扩展后,可有效降低各层累积误差对 SVM 分类性能的影响;在不同信噪比下,APso-MPDSVM 算法都取得最好的平均识别率,说明该算法对环境噪声具有较好的识别鲁棒性。图 3 为接收机工作特性曲线(ROC),进一步验证了所提算法的识别性能最优。

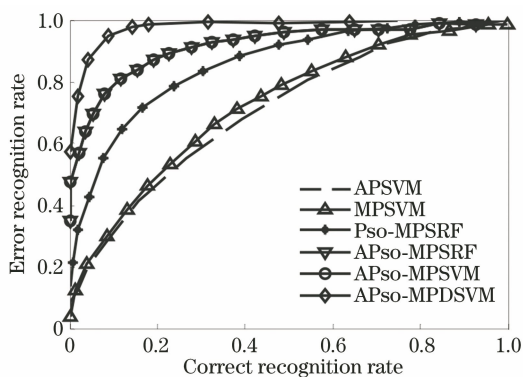


图3 各算法的 ROC 曲线

Fig. 3 ROC curve of each algorithm

6 结 论

针对公共环境中的声音事件识别问题,提出基于自适应 PSO 算法优化 MP 稀疏分解的声音识别算法。先基于适应度函数实现自适应 PSO 算法,再构建自适应 PSO 优化 MP 稀疏分解的目标函数及重构函数,将 PSO 算法的连续空间搜索能力引入 MP 稀疏分解中,提高了最优原子的匹配程度,最后使用优化的 SVM 分类器和复合特征实现公共环境中声音事件的准确识别。实验结果表明,与已有算法相比,所提识别算法显著降低了计算量,并取得了最优的声音识别率,且对环境背景噪声具有较好的鲁棒性。由于 PSO 和 MP 算法的识别效率不高,将进一步研究如何通过并行运行等方式提高算法的运行效率,以进一步解决识别效率问题。

参 考 文 献

- [1] Zhang X H, Hao R F, Li T Y. Hyperspectral abnormal target detection based on low rank and sparse matrix decomposition-sparse representation [J]. *Laser & Optoelectronics Progress*, 2019, 56(4): 042801.
张晓慧, 郝润芳, 李廷鱼. 基于低秩稀疏矩阵分解和稀疏字典表达的高光谱异常目标检测 [J]. *激光与光电子学进展*, 2019, 56(4): 042801.
- [2] Chen Y, Cheng Y N, Liu H L. Demodulation of light sensing overlapping spectral signal by improved particle swarm optimization algorithm [J]. *Chinese Journal of Lasers*, 2018, 45(7): 0710003.
陈勇, 程亚男, 刘焕淋. 利用改进粒子群优化算法解调光传感重叠光谱信号 [J]. *中国激光*, 2018, 45(7): 0710003.
- [3] Li Y X, Wang Q, Zhang X, et al. Audio events clustering based on agglomerative information

bottleneck[J]. *Acta Electronica Sinica*, 2017, 45(5): 1064-1071.

李艳雄, 王琴, 张雪, 等. 基于凝聚信息瓶颈的音频事件聚类方法 [J]. *电子学报*, 2017, 45(5): 1064-1071.

- [4] Feng C D, Li S B, Yao Y, et al. Environmental sound recognition with improving convolutional neural networks and learning rate decay [J]. *Science Technology and Engineering*, 2019, 19(1): 177-182.
冯陈定, 李少波, 姚勇, 等. 基于改进卷积神经网络与动态衰减学习率的环境声音识别算法 [J]. *科学技术与工程*, 2019, 19(1): 177-182.
- [5] Zhao S, Zhang Y, Xu H F, et al. Ensemble classification based on feature selection for environmental sound recognition [J]. *Mathematical Problems in Engineering*, 2019, 2019: 4318463.
- [6] Li W H, Tian Z Z, Gong W G, et al. Developed ESMD for the feature extraction of abnormal sound in public places [J]. *Chinese Journal of Scientific Instrument*, 2016, 37(11): 2429-2437.
李伟红, 田真真, 龚卫国, 等. 改进的 ESMD 用于公共场所异常声音特征提取 [J]. *仪器仪表学报*, 2016, 37(11): 2429-2437.
- [7] Li Y, Wu L F. Detection of sound event under low SNR using multi-band power distribution [J]. *Journal of Electronics & Information Technology*, 2018, 40(12): 2905-2912.
李应, 吴灵菲. 用多频带能量分布检测低信噪比声音事件 [J]. *电子与信息学报*, 2018, 40(12): 2905-2912.
- [8] McFee B, Salamon J, Bello J P. Adaptive pooling operators for weakly labeled sound event detection [J]. *ACM Transactions on Audio, Speech, and Language Processing*, 2018, 26(11): 2180-2193.
- [9] Chen Q J, Li Y. Sound recognition based on optimized orthogonal matching pursuit and deep belief network [J]. *Journal of Computer Applications*, 2017, 37(2): 505-511, 516.
陈秋菊, 李应. 基于优化正交匹配追踪和深度置信网的声音识别 [J]. *计算机应用*, 2017, 37(2): 505-511, 516.
- [10] Nivitha V V, Ramachandran K I, Soman K P. Wavelet-based fundamental heart sound recognition method using morphological and interval features [J]. *Healthcare Technology Letters*, 2018, 5(3): 81-87.
- [11] Davila-Chacon J, Liu J D, Wermter S. Enhanced robot speech recognition using biomimetic binaural sound source localization [J]. *IEEE Transactions on*

- Neural Networks and Learning Systems, 2019, 30 (1): 138-150.
- [12] Wei J, Yue F L, Qiu P, et al. Abnormal sound multiclass recognition algorithm based on EEMD[J]. Journal of Huazhong University of Science and Technology(Natural Science Edition), 2018, 46(7): 117-121.
韦娟, 岳凤丽, 仇鹏, 等. 基于 EEMD 的异常声音多类识别算法[J]. 华中科技大学学报(自然科学版), 2018, 46(7): 117-121.
- [13] Wang D L, Hu S. Optimization method of camera calibration based on quantum-behaved particle swarm optimization algorithm[J]. Laser & Optoelectronics Progress, 2018, 55(12): 121502.
王道累, 胡松. 基于量子粒子群优化算法的摄像机标定优化方法[J]. 激光与光电子学进展, 2018, 55(12): 121502.
- [14] Mallat S G, Zhang Z F. Matching pursuits with time-frequency dictionaries [J]. IEEE Transactions on Signal Processing, 1993, 41(12): 3397-3415.
- [15] Kim H G, Kim J Y. Acoustic event detection in multichannel audio using gated recurrent neural networks with high-resolution spectral features [J]. ETRI Journal, 2017, 39(6): 832-840.