

引用格式: 韩雪, 慕昱, 盛桂敏. CCPSO 优化支持向量机的鸟声识别技术研究[J]. 声学技术, 2023, 42(1): 118-126. [HAN Xue, MU Yu, SHENG Guimin. Research on the CCPSO optimized SVM based bird sound recognition technology[J]. Technical Acoustics, 2023, 42(1): 118-126.] DOI: 10.16300/j.cnki.1000-3630.2023.01.020

CCPSO 优化支持向量机的鸟声识别技术研究

韩雪^{1,2}, 慕昱¹, 盛桂敏³

(1. 黑龙江大学物理科学与技术学院, 黑龙江哈尔滨 150080; 2. 华南理工大学物理与光电学院, 广东广州 510640;
3. 绥化学院电气工程学院, 黑龙江绥化 152061)

摘要: 鸟类是生态系统中的重要组成部分, 鸟类物种的多样性对生态环境有重要作用。所以, 通过鸟声信号来识别鸟类从而对其进行保护有现实意义。文章对鸟声信号采用双参数的双门限法进行分段, 从鸟声信号中寻找出声音的起始点和终止点的具体帧, 进一步进行特征提取, 提取每段鸟声信号中的短时能量和短时平均幅度, 短时语谱图中的平均值、对比度、熵, 共 5 种特征, 采用优化参数的支持向量机进行鸟类物种分类。结果表明, 基于混沌云粒子群优化(Chaos Cloud Particle Swarm Optimization, CCPSO)的支持向量机比普通支持向量机的分类准确度得到提升, 可有效地识别鸟类。利用该方法实现鸟类物种保护和生态系统管理的目的。

关键词: 鸟声识别; 音节分割; 时频特征提取; 支持向量机; 混沌云粒子群优化(CCPSO)

中图分类号: TN911.7

文献标志码: A

文章编号: 1000-3630(2023)-01-0118-09

Research on the CCPSO optimized SVM based bird sound recognition technology

HAN Xue^{1,2}, MU Yu¹, SHENG Guimin³

(1. School of Physical Science and Technology, Heilongjiang University, Harbin 150080, Heilongjiang, China;
2. School of Physics and Optoelectronics, South China University of Technology, Guangzhou 510640, Guangdong, China;
3. College of Electrical Engineering, Suihua University, Suihua 152061, Heilongjiang, China)

Abstract: Birds are an important member of the ecosystem and the diversity of bird species has a self-evident effect on the ecological environment. Therefore, it is of great practical significance to protection of birds by identifying them through the sound of bird calls. In this paper, the bird sound signal is segmented by double-gate detection method of double parameters and used to find the specific frame of the sound starting point and ending point. A total of 5 features such as the short-term energy and short-term average amplitude in each segment of bird sound signal and the average value, contrast, and entropy in the short-time spectrogram are extracted, and then the support vector machine (SVM) with optimized parameters is used for birds species classification. The results show that compared with the ordinary support vector machine, the classification accuracy based on the support vector machine optimized by chaos cloud particle swarm optimization (CCPSO) is greatly improved. This method can effectively identify the birds and achieves the purposes of bird species protection, scientific research and ecosystem management.

Key words: bird recognition; syllable segmentation; time frequency feature extraction; support vector machine (SVM); chaos cloud particle swarm optimization (CCPSO)

0 引言

根据国际自然保护联盟的红色数据清单(IUCN2014), 1373 种(13% 以上)的世界鸟类物种濒临灭绝^[1]。相比形态特征, 鸟鸣声含有更多的信息, 可以最大程度表现出物种的特异性。用这种方

法对物种分类, 能评估生态系统中物种的存在、丰度, 并说明给定生物群落的总体平衡。

影响鸟声分类准确度有多方面因素。徐淑正等^[2]提取音节长度、梅尔频率倒谱系数(Mel Frequency Cepstrum Coefficient, MFCC)、基于线性预测编码系数(Linear Predictive Coding Coefficient, LPCC)的动态时间规整(Dynamic Time Warping, DTW)模板和时频纹理特征(Time-frequency texture feature, TFTF), 将这些特征和不同的参数优化后的分类器结合, 得到不同的分类结果。在对 11 种鸟声信号分类时, 准确率可以达到 92%。由于鸟声信号是非平稳信号, 程龙等^[3]提出一种基于经验模态

收稿日期: 2021-07-10; 修回日期: 2021-08-05

基金项目: 黑龙江省教育厅基本科研业务费(2017-KYYWF-0723)。

作者简介: 韩雪(1999—), 女, 黑龙江绥化人, 硕士研究生, 研究方向为声信号处理与机器学习。

通信作者: 慕昱, E-mail: muyuffff@126.com

分解(Empirical Mode Decomposition, EMD)改进的梅尔频率倒谱系数(MFCC),建立了瞬时频率和基本模式分量的概念,再采用高斯混合模型进行鸟鸣声识别,实验结果表明,改进的MFCC方法识别率达到70.09%,与未改进的MFCC识别率相比提高了3.42个百分点。陈莎莎^[4]将鸟声转换为时频图,对其音节进行分割,提取相应的纹理特征,为减少计算量,应用一维和差统计法,并且在一定程度上这种方法降低了背景噪声对精度的干扰,引入随机森林分类器对20种鸟类进行分类,平均准确度可以达到95.35%。Xie等^[5]提出MFCC-HMM鸟鸣识别技术,对6种鸟类进行分类,准确率达到90.47%。Hakim等^[6]提出一种计算工具来聚类 and 识别鸟声,在特征选择上采用主成分分析(Principal Component Analysis, PCA),采用余弦相似性进行计算,对11种鸟类数据样本进行三倍交叉验证,得到的识别率为98.5%±2.55%。Anderson等^[7]采用动态时间规整模板(DTW)结合时频图,比较待测鸟声的时频图与训练鸟声时频图的构成内容及边框,达到识别的目的。谢将剑等^[8]引入线性调频小波变换去计算鸟鸣声语图,并输入到深度卷积神经网络VGG16模型中进行分类,以18种鸟类为研究对象,利用Chirplet变换、短时傅里叶变换(Short Time Fourier Transform, STFT)和梅尔频率倒谱变换(Mel Frequency Cepstrum Transform, MFCT)计算得到3个不同的语图样本集。采用不同的语图样本集作为输入对比鸟类物种识别模型的性能,结果表明在以Chirplet语图作为输入时,测试集的平均识别准确率最高,可达到98.71%。Fagerlund^[9]提取两种不同的参数(Mel倒谱参数与低电平信号参数)作为特征,采用决策树支持向量机(Support Vector Machine, SVM)进行识别分类。以上都是对单一鸟类进行识别分类,在现实生活中,音频中可能会出现不同鸟声混叠的情况,这给分类带来不小的难题。刘昊天^[10]提出基于“问题转化”和“算法适应”两种策略的多标记迁移学习方法,对多物种鸟声进行识别。Briggs等^[11]提取频谱图每段音特征,应用随机森林分类器结合多示例多标记算法进行多物种分类。

本文针对11种鸟声信号进行处理,利用Matlab仿真实现鸟声信号的分段、特征提取、分类。鸟类识别流程如图1所示。具体步骤如下:鸟声信号分段与特征提取。鸟声信号在一般情况下都是不连续的,为了减少计算量,需要对鸟声信号进行分段,提取有语段中的信息;在特征提取上提出短时时域处理和短时频域处理方法。短时时域处理分析了短时平均能量和短时平均幅度;短时频域处

理方法对语谱图进行灰度分析,提取图片中包含的纹理特征。本文使用SVM算法进行分类,首先介绍线性支持向量机(SVM),考虑到鸟声识别涉及到多分类,扩展至SVM多分类算法,提出了两种构建原理。采用协同粒子群参数优化算法提高SVM分类准确度。引入传统声音特征参数Mel倒谱系数与本文提出的时频特征进行特征对比,引入决策数和随机森林分类算法与本文提出的SVM协同粒子群参数优化算法进行分类器对比。结果表明,SVM协同粒子群参数优化算法在分类方面具有较好的泛化性以及鲁棒性。

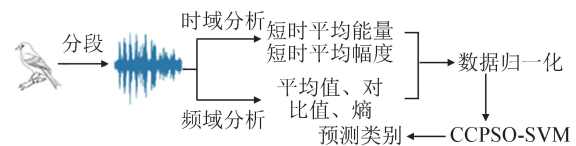


图1 鸟声识别流程图

Fig.1 Flow chart of bird sound recognition

1 鸟声信号的分段与特征提取

1.1 双参数的双门限检测法

为了识别录音中的鸟声信号,首先分割可能包含鸟鸣声的潜在声音片段。这有助于分析人员专注于长录音中的重要内容。当录音中包含很少的鸟鸣声时尤为重要。鸟声信号的检测和分段是整个分类过程中最复杂、难度最大的部分。

双参数的双门限检测法^[12]是基于短时平均能量和短时平均过零率而提出的。其思想是汉语的韵母中有元音,能量较大,所以可以从短时平均能量中找到韵母,而声母是辅音,频率较高,相应的短时平均过零率较大,据此能够确定完整的音节。

进行端点检测时,利用分帧函数对声音信号处理,逐帧计算短时平均能量和短时平均过零率,进而对每一帧根据阈值进行比较和判断。图2展示了对北扑翅鹭(Northern Flicker Call)鸟声信号进行端点检测的结果,图中的实线为鸟叫声的起始,虚线为鸟叫声的终止。

北扑翅鹭叫声总共有7623帧,分成68段。选取其中的3~4s的分段,在这段时间内鸟声信号被

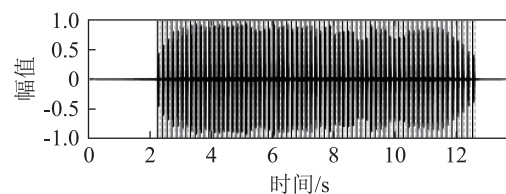


图2 北扑翅鹭鸟叫声的端点检测

Fig.2 Endpoint detection of Northern Flicker call

分为7段, 仿真结果如图3所示。

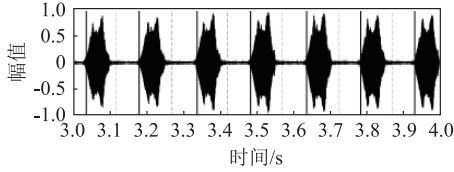


图3 北扑翅鸺鸟叫声在3~4 s内的端点检测
Fig.3 Endpoint detection of Northern Flicker call in the segment of 3-4 s

1.2 鸟声信号短时域处理

1.2.1 短时能量和短时平均幅度

设声音波形时域信号为 $x(n)$, 加窗并利用分帧函数处理, 获得第 i 帧声音信号为 $y_i(n)$:

$$y_i(n) = w(n) \times x[(i-1) \times q + n] \quad (1)$$

式中: $w(n)$ 为窗函数, $y_i(n)$ 是一帧的数值, $n=1, 2, \dots, L$, $i=1, 2, \dots, t$, L 为帧长; q 为帧移长度; t 为分帧的总帧数。

第 i 帧声音信号 $y_i(n)$ 的短时能量表达式为

$$E(i) = \sum_{n=0}^{L-1} y_i^2(n), \quad 1 \leq i \leq t \quad (2)$$

北扑翅鸺鸟叫声的短时能量图如图4所示。

鸟声信号的平均幅度定义为

$$M(i) = \sum_{n=0}^{L-1} |y_i(n)|, \quad 1 \leq i \leq t \quad (3)$$

读取北扑翅鸺鸟叫声数据, 计算短时平均幅度, 得到如图5所示的结果图。

$M(i)$ 也是一帧语音信号能量大小的表征, 它与

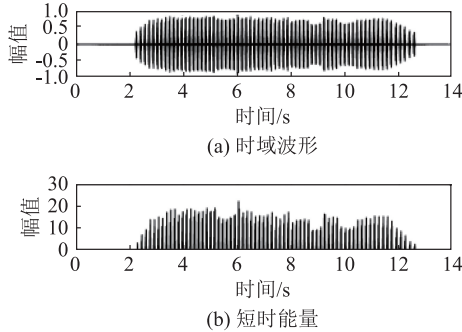


图4 北扑翅鸺鸟叫声的声信号波形及其短时能量
Fig.4 Sound signal waveform and its short-term energy map of Northern Flicker Call

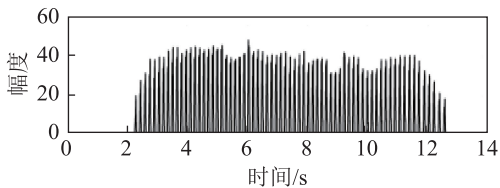


图5 北扑翅鸺鸟叫声的短时平均幅度
Fig.5 The short-term average amplitude map of Northern Flicker call

$E(i)$ 的区别在于计算时不论采样值的大小, 不会因取二次方而造成较大差异。短时能量和短时平均幅度函数的主要用于区分浊音段和清音段, 因为浊音时 $E(i)$ 值比清音时大得多。短时能量是信号绝对值的平方和, 短时幅度是信号绝对值的加权和, 解决了较大级别的信号能量值敏感的问题^[13-14]。

1.3 鸟声信号短时频域处理

1.3.1 语谱图

语谱图是把声音信号进行频谱分析, 横坐标为时间, 纵坐标为频率。像素的灰度值大小体现当前时刻与频率的信号能量密度^[15], 信号强则灰度深, 信号弱则灰度浅。

读取北扑翅鸺鸟叫声数据, 选取3~8 s画出其语谱图, 结果如图6所示。

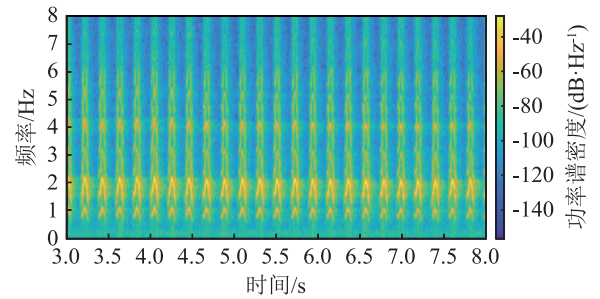


图6 北扑翅鸺鸟叫声的语谱图
Fig.6 The spectrogram of Northern Flicker call

本文主要分析语谱图的纹理特征^[16]。纹理是一种反映图像中同质现象的视觉特征, 体现了物体表面的具有缓慢变化或者周期性变化的表面结构组织排列属性。对于纹理特征本文采用灰度差分统计法。此方法依靠图像直方图提取如平均值、方差、能量和熵等来表示纹理特征。设 (x, y) 为图像中某点, 该点附近的点 $(x + \Delta x, y + \Delta y)$ 的灰度差值为 $\Delta g = g(x, y) - g(x + \Delta x, y + \Delta y)$, 也就是差分。

直方图定义为 $p(i) = h(i)$, $h(i)$ 是第 i 种颜色在整幅图像中的具体像素数, 由直方图可以得到 Δg 取值的概率 $p(i)$ 。直方图平坦, 表明纹精细^[17]。纹理特征中, 平均值、对比度、熵的计算公式分别为

$$g_{\text{Mean}} = \sum_i i p(i) / M \quad (4)$$

$$g_{\text{Co}} = i^2 p(i) \quad (5)$$

$$E_n = - \sum_i p(i) \log_2 [p(i)] \quad (6)$$

式中: M 为总帧数, g_{Mean} 表示图像的平均灰度; g_{Co} 代表图像局部灰度的总变化量, 像素的空间分布越不均匀, 表现为纹路越深, 反差越大, 视觉效果更明显; E_n 反映图像信息量, 它代表着图像复杂程度。若一个图像的随机性越大, 共生矩阵中的值之

间越相近,熵就越大。

1.4 鸟声信号的其他变换域处理

Mel 频率倒谱系数(MFCC)^[18]是基于人的听觉机理,依据人的听觉实验结果提取的声学特征,在音频领域被广泛使用。人的主观感知频域的划定是非线性的, Mel 刻度为

$$F_{\text{mel}} = 1125 \lg(1 + f/700) \quad (7)$$

式中: F_{mel} 是以梅尔(Mel)为单位的感知频率; f 是以 Hz 为单位的实际频率。MFCC 特征参数提取步骤包括:

(1) 预处理,将时域声音信号 $x(n)$ 进行预加重、分帧、加窗函数等加工,得到 $x_i(m)$, i 表示分帧后的第 i 帧。

(2) 快速傅里叶变换(FFT),对每一帧信号进行 FFT,将信号从时域数据转变为频域数据:

$$X(i, k) = \text{FFT}[x_i(m)] \quad (8)$$

(3) 计算谱线能量,对每一帧 FFT 后的数据计算谱线的能量:

$$E(i, k) = [X(i, k)]^2 \quad (9)$$

(4) 计算通过 Mel 滤波器的能量 $S(i, m)$,在频域中相当于把每帧的能量谱 $E(i, k)$ (其中 i 表示第 i 帧, k 表示频域中的第 k 条谱线)与 Mel 滤波器的频域响应 $H_m(k)$ 相乘并相加, $0 \leq m \leq M$, M 为滤波器的个数, $f(m)$ 为中心频率:

$$H_m(k) = \begin{cases} 0 & , k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)} & , f(m-1) \leq k \leq f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)} & , f(m) < k \leq f(m+1) \\ 0 & , k > f(m+1) \end{cases} \quad (10)$$

$$S(i, m) = \sum_{k=0}^{M-1} E(i, k) H_m(k), 0 \leq m < M \quad (11)$$

(5) 计算 DCT 倒谱,把 Mel 滤波器的能量取对数后计算 DCT:

$$E_{\text{MFCC}}(i, n) = \sqrt{\frac{2}{M}} \sum_{m=0}^{M-1} \lg[S(i, m)] \cos\left[\frac{\pi n(2m-1)}{2M}\right] \quad (12)$$

式中: $S(i, m)$ 是式求出的 Mel 滤波器能量; m 是指第 m 个 Mel 滤波器; i 是第 i 帧; n 是 DCT 后的谱线。

2 样本数据集

2.1 类不平衡

本文研究对象为 11 种鸟声信号,音频文件来源于 Macaulay Library^[19], 11 种鸟类分别为: 绒啄木鸟、北扑翅鸢、黑顶山雀、白胸鹇、北美红雀、美洲家朱雀、松金朱雀、红腹啄木鸟、斯特勒鸟、

黄昏雀、冠蓝鸭。对所有鸟声的音频分段,仿真得到分段数如表 1 所示。

表 1 11 种鸟类鸟鸣声分段数
Table 1 Segmented numbers of bird call for 11 bird species

种类	分段数	种类	分段数
绒啄木鸟	75	松金朱雀	75
北扑翅鸢	72	红腹啄木鸟	24
黑顶山雀	43	斯特勒鸟	43
白胸鹇	44	黄昏雀	44
北美红雀	66	冠蓝鸭	21
美洲家朱雀	81		

提取每段音频信号的短时能量、短时平均幅度、平均值、对比度、熵作为特征。由于每种鸟类的发出叫声数量不同,造成各类鸟声的样本数量有很大差异,即存在类不平衡的情况,这导致分类模型对样本数较少的鸟类识别准确率差。常规解决类不平衡的方法有扩大数据集、过采样法、欠采样法、综合采样法,对小类错分进行加权惩罚等^[20]。

扩大数据集即增加样本量,其中含有较多的少数类样本。过采样法是随机复制少数类样本点,使它尽可能与多数类样本数量维持平衡。欠采样法是删除部分多数类样本使它和少数类样本平衡。综合采样法首先进行过采样,删除边界样本或预测结果错误的样本。对小类错分进行加权惩罚是增加少数类样本权重,使分类器更关注少数类样本^[21]。

在鸟声识别过程中,鸟声样本的数量是分类的关键。本文采用过采样法,生成本文需要的鸟声识别的数据集。

2.2 数据归一化

由于上述数据集差异很大,因此需要进一步进行归一化,这样可以缩小量值之间落差,提高计算精度^[22]。常见的归一化方法主要有线性归一化、标准分数归一化、中值归一化。本文采用第一种方法,把数据转换到 $[0, 1]$ 的范围内,转换公式为

$$X' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (13)$$

式中: x_{\max} 为样本的最大值, x_{\min} 为样本的最小值。对于本文的数据集是一个矩阵,归一化可以补偿特征向量的统计属性不匹配产生的影响,消除属性值之间的差别。

3 混沌云粒子群优化 SVM 参数

3.1 支持向量机 SVM 与决策树

二分类 SVM 是定义在特征空间上的间隔最大的线性分类器。在一个二维平面,有不同的两种数

据, 需要找到一条直线将其隔开, 在高维空间中此直线被称作超平面。

设分离超平面为

$$\mathbf{w}^T \mathbf{x} + b = 0 \quad (14)$$

式中: \mathbf{x} 是输入空间的样本, \mathbf{w} 和 b 分别表示平面的法向量和截距, 简记为 (\mathbf{w}, b) , 样本空间某点 x_i 到其距离可写为

$$r = \frac{|\mathbf{w}\mathbf{x}_i + b|}{\|\mathbf{w}\|} \quad (15)$$

假设超平面 (\mathbf{w}, b) 能将训练样本正确分类, 即对于 $(x_i, y_i) \in \mathbf{D}$, 若 $y_i = +1$, 则有 $\mathbf{w} \cdot x_i + b > 0$; 若 $y_i = -1$, 则有 $\mathbf{w} \cdot x_i + b < 0$ 。令

$$\begin{cases} \mathbf{w} \cdot x_i + b \geq +1, & y_i = +1 \\ \mathbf{w} \cdot x_i + b \leq -1, & y_i = -1 \end{cases} \quad (16)$$

两个不同类支持向量到 (\mathbf{w}, b) 的距离和为 $\gamma = \frac{2}{\|\mathbf{w}\|}$, r 被称为间隔(margin), 间隔最大的超平面即为所求, 因此支持向量在求解分离超平面过程中至关重要, 学者们把它称作支持向量机^[23]。

SVM 分类算法起初仅具有解决二分类任务的能力。后来随着需求向多分类方面扩展。现构建多分类 SVM 有两种方法: (1) 考虑全部类别。(2) 多个二分类模型集成。第一种方法的主要思想对公式改进并考虑全部类别的样本点, Weston 提出的“K-Class 多分类算法”就是此方法。它在典型存在的 SVM 原理的本源上, 从头架构多分类模型, 并且研究多种类别, 转换成一个解决二次规划(Quadratic Programming, QP) 问题, 达到多分类的目的^[24]。

第二种方法的原理是将多个二分类器集成, 普遍的集成方法有一对一和一对多两种^[25]。一对一对 n 类训样本两两组合, 构建 $\frac{1}{2}n(n-1)$ 个支持向量机, 每个分类模型训练两种不同类别的数据, 最终类别由“投票”决定。一对多构造相应类别数的支持向量机, 每个支持向量机来区分本类数据和非本类数据。每个支持向量机构造一个超平面, 最终结果依靠输出离与分界面 (\mathbf{w}, b) 距离最大的那个支持向量机。本文采用 libsvm 建立分类模型对鸟声信号进行分类, 该模型应用的是一对一方法^[26]。

决策树(decision tree)是一种类似流程图的树形结构, 主要用在分类问题上。决策树包含一个根节点、有向边(directed edge)、若干个内部节点和若干个叶子节点, 内部节点表示一个属性或特征, 有向边代表一个测试输出, 叶子节点表示一个类, 根节点包含样本全集。从根节点开始, 对样本的某一特

征进行测试, 根据测试结果, 将样本分配到其内部节点; 这时, 每一个内部节点对应着该特征的一个取值, 如此递归地对样本进行测试分配, 直至到达叶节点。最后将样本分到叶节点的类中^[27]。

3.2 SVM 混沌云粒子群参数优化与随机森林

SVM 模型包含两个至关重要的参数 C 和 γ ^[28]。 C 是惩罚因子, 即对误差的宽容度。 C 越高说明越不能容忍出现误差, 容易过拟合。 C 越小容易欠拟合。 C 过大或过小, 泛化能力差。 γ 是选择径向基核函数作为内核后, 本身带有的参数。暗示着数据映射到新的特征空间后的分布, γ 与支持向量成反比, 所以 C 和 γ 的数值对准确度至关重要。

本文采用混沌云粒子群优化(CCP SO)算法选取 SVM 中最佳的参数 C 和 γ 。传统的粒子群优化算法倾向陷入局部最优, 不适用高维参数优化。协同混沌云粒子群优化(CCP SO)算法是一种基于群体智能理论的优化算法, 对适应值函数进行降维处理, 很大程度改善粒子群算法在高维空间效果不理想的问题, 在参数选择中可以进行全局寻找, 避免陷入局部最优, 影响实验准确率^[29]。

混沌云粒子群优化 SVM 参数的基本步骤如下^[30]:

(1) 将 n 维空间按维度分解, 均匀分成 k 个子空间。

(2) 使 k 个子种群在不同子空间内按 $x_{ik} = x_{ik} + v_{ik}$ ($k=1, 2, \dots, k$) 进行搜索。 x_{ik} 表示第 k 个子群中、第 i 个粒子的位置, v_{ik} 表示其速度。

(3) 计算每个种群的适应度值为 $g_{\text{best}, k}$, 全局最优解的构成为 $\mathbf{G}_{\text{best}} = (g_{\text{best}, 1}, g_{\text{best}, 2}, \dots, g_{\text{best}, k}, \dots, g_{\text{best}, K})$, 通过比较适应度值选择最优解, 计算 x_{ik} 的适应度值为 $\mathbf{X}_{ik} = (g_{\text{best}, 1}, g_{\text{best}, 2}, \dots, x_{ik}, \dots, g_{\text{best}, k-1}, g_{\text{best}, k})$, 更新所有子群内粒子, 达到全局最优。

本文只需优化两个参数, 问题的维度只有 2 维, 即每个粒子的位置向量只有 2 维, 在协同粒子群优化算法中还需嵌入适应度函数来判断参数的优劣。本文选用的是十折交叉验证, 即使用某一组 c 和 g 的 SVM 对训练集进行十折交叉验证, 得到的评分结果, 作为适应度值, 适应度函数就是一次 libsvm 的训练, 所以 libsvmtrain 就是本文的适应度函数。

随机森林(Random Forest, RF)是 Bagging(并行式集成方法)的一个扩展变体^[31]。在以决策树为基学习器构建 Bagging 集成的基础上, 进一步在决策树的训练过程中引入了随机属性选择。具体来说, 首先利用重采样方法有放回地从原始训练集中随机

抽取 n 个样本, 并构建 n 个决策树; 然后假设在训练样本数据中有 m 个特征, 则每次分裂时选择最好的特征进行分裂, 每棵树都一直这样分裂下去, 直到该节点的所有训练样例都属于同一类; 接着让每棵决策树在不做任何修剪的前提下最大限度地生长; 最后将生成的多棵分类数组成随机森林。

3.3 交叉验证

交叉验证(Cross-validation)是在机器学习中常引入的概念, 通过“交叉”将原始数据拆分成多种不同数据组合的方式对模型进行“验证”^[32]。交叉验证的基本思想是对原始数据进行分组, 一部分作为训练集(包含训练集和验证集两部分)来训练模型, 剩余部分作为测试集来评价模型。常见的交叉验证类型有简单交叉验证、2折交叉验证、 K 折交叉验证。

简单交叉验证仅将原始数据集随机分为两组, 一组作为训练集, 一组作为验证集, 利用训练集训练分类器, 利用验证集验证模型, 但实质没有达到交叉的目的, 结果具有随机性。2折交叉验证将数据集分成大小相等的两个子集, 进行两次的分类器训练, 两次训练时将训练集与测试集交换。但由于训练集样本数量太少, 训练时得到的模型准确度不具有说服力。 K 折交叉验证是最常用且效果最好的分类器性能评价方法, 通常 K 取 10, 将数据集分成 10 份, 轮流将 9 份作为训练集, 1 份作为测试集, 经过 10 次测试, 将 10 次测试结果求平均, 作为此分类器的性能指标。

4 实验结果分析

4.1 SVM与决策树的分类结果

第 2 节所提取的 11 种鸟叫声、共 873 个样本, 提取特征包括短时平均能量、短时平均幅度、平均值、对比度、熵共 5 种, 将特征归一化后进行 10 折交叉验证。设定每种鸟类 60 个样本为训练集(包含训练集与验证集), 剩余 213 个样本为测试集^[33]。11 种鸟类样本划分如表 2 所示。

SVM 中初始 C 默认为 1, γ 默认 $1/k$, k 是指输入数据中的特征数, k 为 5。分类准确率为 42.723%, 在待测的 213 个样本, 分类准确无误的样本仅有 91 个。

采用决策树进行分类, 分类准确率为 79.81%, 在待测的 213 个样本, 分类准确无误的样本有 170 个。

仿真计算得到 SVM 分类混淆矩阵如表 3 所示。仿真计算得到决策树分类混淆矩阵如表 4 所示。

表 2 11 种鸟类样本划分
Table 2 Sample division for 11 bird species

鸟类种类	类标号	训练样本数	测试样本数
绒啄木鸟	1	60	15
北扑翅鸢	2	60	12
黑顶山雀	3	60	26
白胸鹇	4	60	28
北美红雀	5	60	6
美洲家朱雀	6	60	21
松金朱雀	7	60	15
红腹啄木鸟	8	60	12
斯特勒鸟	9	60	26
黄昏雀	10	60	28
冠蓝鸭	11	60	24

4.2 SVM 协同粒子群参数优化后和随机森林分类结果

将参数优化与交叉验证结合选取最优参数, 将 10 次训练得到的参数进行平均, 得到 $C=88.367\ 0$ 、 $\gamma=87.545\ 4$, 对测试集进行分类, 分类准确率达到 91.08%, 在预测的 213 个样本中, 正确分类的样本可达到 194 个。仿真计算得到分类混淆矩阵如表 5 所示。

采用随机森林进行分类, 分类准确率为 91.08%, 在预测的 213 个样本中, 正确分类大的样本可达到 194 个。

仿真计算得到随机森林分类混淆矩阵如表 6 所示。

4.3 MFCC 与 SVM 协同粒子群参数优化、决策树、随机森林

本文对 11 种鸟类提取其 Mel 倒谱系数, 对得到的数据集进行过采样, 归一化处理, 采用十折交叉

表 3 SVM 分类结果混淆矩阵
Table 3 Confusion matrix of SVM classification results

真实	预测										
	类1	类2	类3	类4	类5	类6	类7	类8	类9	类10	类11
类1	9	0	0	0	4	0	0	0	2	0	0
类2	0	0	0	0	12	0	0	0	0	0	0
类3	0	4	0	0	19	9	1	0	0	0	2
类4	0	0	0	24	0	0	1	0	3	0	0
类5	0	0	0	0	4	0	0	1	1	0	0
类6	4	0	0	3	0	12	1	0	1	0	0
类7	0	0	0	0	1	0	11	0	2	0	1
类8	0	2	0	0	4	0	3	0	3	0	0
类9	1	0	0	1	0	0	6	0	18	0	0
类10	8	6	0	0	3	1	4	1	5	0	0
类11	0	0	0	1	7	0	0	0	3	0	13

表4 决策树分类结果混淆矩阵

Table 4 Confusion matrix of decision tree classification results

真实	预测										
	类1	类2	类3	类4	类5	类6	类7	类8	类9	类10	类11
类1	7	2	0	0	1	0	0	2	0	3	0
类2	1	10	0	0	1	0	0	0	0	0	0
类3	2	3	21	0	0	0	0	0	0	0	0
类4	0	0	0	26	0	0	0	0	2	0	0
类5	0	1	1	0	4	0	0	0	0	0	0
类6	1	0	0	2	0	17	0	0	0	1	0
类7	0	0	0	0	2	0	10	1	1	0	1
类8	0	0	0	0	0	0	0	12	0	0	0
类9	0	0	0	0	0	0	2	0	24	0	0
类10	0	0	2	2	1	0	3	1	0	19	0
类11	0	2	0	0	0	0	0	0	2	0	20

表5 SVM优化分类结果混淆矩阵

Table 5 Confusion matrix of SVM optimized classification results

真实	预测										
	类1	类2	类3	类4	类5	类6	类7	类8	类9	类10	类11
类1	13	1	0	0	0	0	0	1	0	0	0
类2	0	12	0	0	0	0	0	0	0	0	0
类3	0	0	24	0	0	0	2	0	0	0	0
类4	0	0	0	26	0	0	0	0	2	0	0
类5	0	0	0	0	5	0	0	0	0	1	1
类6	0	0	0	0	0	21	0	0	0	0	0
类7	0	0	0	0	3	0	9	0	1	1	1
类8	0	0	0	0	0	0	0	12	0	0	0
类9	0	0	0	0	0	0	0	0	26	0	0
类10	0	0	0	2	0	0	4	0	0	22	0
类11	0	0	0	0	0	0	0	0	0	0	24

验证, 设定每种鸟类60个样本为训练集(包含训练集与验证集), 剩余213个样本为测试集。将10次训练得到参数进行平均, 得到 $C=96.919\ 2$ 、 $\gamma=63.712\ 6$, 对测试集进行分类, 分类准确率达到95.31%, 在预测的213个样本中, 正确分类的样本可达203个。

仿真计算得到分类混淆矩阵如表7所示。

采用决策树进行分类, 分类准确率为92.49%, 在待测的213个样本, 分类准确无误的样本有197个。

采用随机森林进行分类, 分类准确率为97.65%, 在待测的213个样本, 分类准确无误的样本有208个。

时频特征、MFCC特征参数分别与协同粒子群参数优化、决策树, 随机森林分类方法相结合, 得到的分类结果如表8所示。

表6 随机森林分类结果混淆矩阵

Table 6 Confusion matrix of random forest classification results

真实	预测										
	类1	类2	类3	类4	类5	类6	类7	类8	类9	类10	类11
类1	13	0	1	0	1	0	0	0	0	0	0
类2	0	12	0	0	0	0	0	0	0	0	0
类3	0	0	24	0	0	0	0	0	0	0	2
类4	0	0	0	28	0	0	0	0	0	0	0
类5	0	0	1	0	5	0	0	0	0	0	0
类6	1	0	0	2	0	17	0	0	0	1	0
类7	0	0	0	0	3	0	11	0	0	0	1
类8	0	0	0	0	0	0	0	12	0	0	0
类9	0	0	0	0	0	0	2	0	24	0	0
类10	0	0	0	0	0	0	2	2	0	24	0
类11	0	0	0	0	0	0	0	0	0	0	24

表7 MFCC特征下SVM优化分类结果混淆矩阵

Table 7 Confusion matrix of SVM optimized classification results under MFCC features

真实	预测										
	类1	类2	类3	类4	类5	类6	类7	类8	类9	类10	类11
类1	10	0	0	0	0	0	5	0	0	0	0
类2	0	12	0	0	0	0	0	0	0	0	0
类3	0	0	24	0	0	0	2	0	0	0	0
类4	0	0	0	28	0	0	0	0	0	0	0
类5	0	0	0	0	6	0	0	0	0	0	0
类6	0	0	0	0	1	18	2	0	0	0	0
类7	0	0	0	0	0	0	15	0	0	0	0
类8	0	0	0	0	0	0	0	12	0	0	0
类9	0	0	0	0	0	0	0	0	26	0	0
类10	0	0	0	0	0	0	0	0	0	28	0
类11	0	0	0	0	0	0	0	0	0	0	24

表8 不同特征与分类方法相结合的分类准确率

Table 8 Classification accuracies of different features combined with classification methods

分类方法	分类准确率/%	
	时频特征	MFCC特征
CCPSO-SVM	91.08	95.31
决策树	79.81	92.49
随机森林	91.08	97.65

由表8可知, SVM协同粒子群参数优化算法、随机森林算法与时频特征相结合对11种鸟类测试集分类结果相同, 分类准确率为91.08%, 在用MFCC参数作为特征时, SVM协同粒子群参数优化算法仅比随机森林分类准确率低2.34个百分点。随机森林是目前多分类最有效的方法, 可见交叉验证得到的SVM协同粒子群优化参数算法具有较高的泛化能力。在分类时, 数据测试集在SVM协同

粒子群参数优化、决策树以及随机森林3种分类器上的计算时间相差较小。

5 结论

本文提取了11种鸟类的短时平均能量、短时平均幅度、语谱图中平均值、对比度、熵,共5种特征,并对提取特征进行分类验证。在分类准确率相似的情况下,相比将特征与分类器相结合来找到最优的分类系统的分类方法,本文提出的算法更加简洁。在传统的MFCC特征下,可以看到SVM协同CCPSO的分类准确率可达95.31%,相比改进后MFCC特征参数的准确率提高的幅度大,在特征参数相同的情况下,本文提出的分类算法也比MFCC-HMM分类模型分类准确率高。

从仿真实验可以看出,在时频特征下参数优化后的SVM模型与随机森林分类精度相同。在MFCC特征参数下,三种分类器的分类精度都得到了提高。对于协同粒子群参数优化算法,未来将进一步研究是由于MFCC特征矩阵的空间维度增加对参数优化产生影响,从而导致的识别准确率提高,还是由于MFCC特征参数比时频特征更好地体现语音特性从而导致识别准确率提高。

参 考 文 献

- [1] KROODSMA D E, MILLER E H. Ecology and evolution of acoustic communication in birds[M]. Ithaca, N. Y.: Comstock Pub./Cornell University Press, 1996.
- [2] 徐淑正, 孙忆南, 皇甫丽英, 等. 基于MFCC和时频图等多种特征的综合鸟声识别分类器设计[J]. 实验室研究与探索, 2018, 37(9): 81-86, 91.
XU Shuzheng, SUN Yinan, HUANGFU Liying, et al. Design of synthesized bird sounds classifier based on multi feature extraction classifiers and time-frequency chat[J]. Research and Exploration in Laboratory, 2018, 37(9): 81-86, 91.
- [3] 程龙, 张华清. 基于改进MFCC的鸟鸣声识别方法研究[J]. 中国传媒大学学报(自然科学版), 2017, 24(3): 41-46.
CHENG Long, ZHANG Huaqing. Research of birdsong recognition method based on improved MFCC[J]. Journal of Communication University of China (Science and Technology), 2017, 24(3): 41-46.
- [4] 陈莎莎. 基于时频纹理和随机森林的鸟声识别[D]. 福州: 福州大学, 2013.
CHEN Shasha. Bird sounds recognition using time-frequency texture and random forests[D]. Fuzhou: Fuzhou University, 2013.
- [5] XIE S S, XU H F, LIU J, et al. Research on bird songs recognition based on MFCC-HMM[C]//2021 International Conference on Computer, Control and Robotics (ICCCR). Shanghai, China. IEEE, 2021: 262-266.
- [6] HAKIM A, MAHMOOD M T. Automated birdsong clustering and interactive visualization tool[J]. Pakistan Journal of Agricultural Sciences, 2021, 58(4): 1395-1403.
- [7] ANDERSON S E, DAVE A S, MARGOLIASH D. Template-based automatic recognition of birdsong syllables from continuous recordings[J]. The Journal of the Acoustical Society of America, 1996, 100(2 Pt 1): 1209-1219.
- [8] 谢将剑, 李文彬, 张军国, 等. 基于Chirplet语图特征和深度学习的鸟类物种识别方法[J]. 北京林业大学学报, 2018, 40(3): 122-127.
XIE Jiangjian, LI Wenbin, ZHANG Junguo, et al. Bird species recognition method based on Chirplet spectrogram feature and deep learning[J]. Journal of Beijing Forestry University, 2018, 40(3): 122-127.
- [9] FAGERLUND S. Bird species recognition using support vector machines[J]. EURASIP Journal on Advances in Signal Processing, 2007, 12(1): 1-8.
- [10] 刘昊天. 多标记迁移学习算法研究以及在鸟声识别中的应用[D]. 南京: 南京农业大学, 2016.
LIU Haotian. A study on multi-label transfer learning algorithm and application in the bird sounds recognition[D]. Nanjing: Nanjing Agricultural University, 2016.
- [11] BRIGGS F, HUANG Y H, RAICH R, et al. The 9th annual MLSP competition: new methods for acoustic classification of multiple simultaneous bird species in a noisy environment[C]//2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP). Southampton, UK. IEEE, 2013: 1-8.
- [12] 卜玉婷, 曾庆宁, 郑展恒. 一种低信噪比环境下的语音端点检测算法[J]. 声学技术, 2020, 39(5): 592-602.
BU Yuting, ZENG Qingning, ZHENG Zhanheng. A speech endpoint detection method in low SNR environment[J]. Technical Acoustics, 2020, 39(5): 592-602.
- [13] 武志翔, 王顺利, 李占锋, 等. 基于短时平均幅度和HMM的语音识别系统研究[J]. 化工自动化及仪表, 2013, 40(6): 779-782.
WU Zhixiang, WANG Shunli, LI Zhanfeng, et al. Study on speech recognition system based on short-time average amplitude and HMM[J]. Control and Instruments in Chemical Industry, 2013, 40(6): 779-782.
- [14] 潘泽波. 基于改进MFCC与短时能量特征的异常音频检测方法[J]. 电子世界, 2020(23): 139-141, 145.
PAN Zebo. Abnormal audio detection method based on improved MFCC and short-term energy characteristics[J]. Electronics World, 2020(23): 139-141, 145.
- [15] 白燕燕, 胡晓霞. 基于MATLAB语谱图的声乐研究[J]. 软件工程, 2019, 22(9): 1-4.
BAI Yanyan, HU Xiaoxia. Research on vocal music based on MATLAB spectrogram[J]. Software Engineering, 2019, 22(9): 1-4.
- [16] 赵书涛, 王紫薇, 陈志华, 等. 有载分接开关GLCM纹理特征及改进随机森林算法的故障诊断方法[J]. 高电压技术, 2022, 48(9): 3593-3601.
ZHAO Shutao, WANG Ziwei, CHEN Zhihua, et al. GLCM texture features of on-load tap changer and fault diagnosis method based on improved random forest algorithm[J]. High Voltage Engineering, 2022, 48(9): 3593-3601.
- [17] 薄华, 马缚龙, 焦李成. 图像纹理的灰度共生矩阵计算问题的分析[J]. 电子学报, 2006, 34(1): 155-158, 134.
BO Hua, MA Fulong, JIAO Licheng. Research on computation of GLCM of image texture[J]. Acta Electronica Sinica, 2006, 34(1): 155-158, 134.
- [18] 陈瀚翔, 邱志斌, 王海祥, 等. 基于MFCC特征与GMM的输电线路涉鸟故障相关鸟种智能识别[J]. 水电能源科学, 2021, 39(7): 171-174, 67.
CHEN Hanxiang, QIU Zhibin, WANG Haixiang, et al. Intelli-

- gent identification of bird species related to transmission line faults based on MFCC features and GMM[J]. *Water Resources and Power*, 2021, **39**(7): 171-174, 67.
- [19] Cornell Lab of Ornithology - Macaulay Library[DB/OL]. <http://birds.cornell.edu/MacaulayLibrary/>.
- [20] 王乐, 韩萌, 李小娟, 等. 不平衡数据集分类方法综述[J]. *计算机工程与应用*, 2021, **57**(22): 42-52.
WANG Le, HAN Meng, LI Xiaojuan, et al. Review of classification methods for unbalanced data sets[J]. *Computer Engineering and Applications*, 2021, **57**(22): 42-52.
- [21] 汤荣志, 段会川, 孙海涛. SVM训练数据归一化研究[J]. *山东师范大学学报(自然科学版)*, 2016, **31**(4): 60-65.
TANG Rongzhi, DUAN Huichuan, SUN Haitao. Research on data normalization for svm training[J]. *Journal of Shandong Normal University (Natural Science)*, 2016, **31**(4): 60-65.
- [22] 刘雅洁. 基于集成学习的不平衡数据集分类问题研究[D]. 大连: 大连理工大学, 2021.
LIU Yajie. Research on unbalanced data classification based on ensemble learning[D]. Dalian: Dalian University of Technology, 2021.
- [23] 王晓辉, 王小娟, 谷峥, 等. 基于优化支持向量机的实验设备故障诊断[J]. *实验技术与管理*, 2021, **38**(6): 254-257.
WANG Xiaohui, WANG Xiaojuan, GU Zheng, et al. Fault diagnosis of experimental equipment based on optimized support vector machine[J]. *Experimental Technology and Management*, 2021, **38**(6): 254-257.
- [24] WESTON J, WATKINS C. Support vector machines for multi-class pattern recognition[C]// *Proc. European Symposium on Artificial Neural Networks*, 1999.
- [25] 王冲, 马晓楠, 宋冬慧. SVM一对一多分类的图像反馈检索优化[J]. *计算机工程与设计*, 2019, **40**(1): 173-178.
WANG Chong, MA Xiaonan, SONG Donghui. Improved image feedback retrieval method based on one-versus-one multi-class algorithm of SVM[J]. *Computer Engineering and Design*, 2019, **40**(1): 173-178.
- [26] CHANG C C, LIN C J. LIBSVM: a library for support vector machines[J]. *ACM Transactions on Intelligent Systems and Technology*, 2011, **2**(3): 1-27.
- [27] SONG Y Y, LU Y. Decision tree methods: applications for classification and prediction[J]. *Shanghai Archives of Psychiatry*, 2015, **27**(2): 130-135.
- [28] 姜雯, 吴陈. 基于自适应粒子群的SVM参数优化研究[J]. *计算机与数字工程*, 2021, **49**(7): 1302-1304, 1309.
JIANG Wen, WU Chen. Research on SVM parameter optimization based on adaptive particle swarm optimization[J]. *Computer & Digital Engineering*, 2021, **49**(7): 1302-1304, 1309.
- [29] Yuhua, Li, . Competitive and cooperative particle swarm optimization with information sharing mechanism for global optimization problems[J]. *Information Sciences*, 2015, **293**: 370-382.
- [30] 满春涛, 盛桂敏. 改进的协同粒子群优化算法[J]. *哈尔滨理工大学学报*, 2010, **15**(6): 51-53, 58.
MAN Chuntao, SHENG Guimin. An improved algorithm based on cooperative particle swarm optimization[J]. *Journal of Harbin University of Science and Technology*, 2010, **15**(6): 51-53, 58.
- [31] PROBST P, BOULESTEIX A L. To tune or not to tune the number of trees in random forest? [EB/OL]. 2017: arXiv: 1705.05654. <https://arxiv.org/abs/1705.05654>.
- [32] REFAEILZADEH P, TANG L, LIU H. *Cross-validation* [M]// *Encyclopedia of Database Systems*. Boston, MA: Springer US, 2009: 532-538.
- [33] 陈德昊, 林建恒, 衣雪娟, 等. 三种鱼类发声信号频带能量特征提取与分类研究[J]. *声学技术*, 2021, **40**(2): 254-259.
CHEN Dehao, LIN Jianheng, YI Xuejuan, et al. Research on band energy extraction and classification of three kinds of fishes sound signals[J]. *Technical Acoustics*, 2021, **40**(2): 254-259.