

引用格式: 李大鹏, 周晓彦, 叶如, 等. 基于特征选择和 GWO-KELM 的鸟声识别算法[J]. 声学技术, 2022, 41(5): 782-788. [LI Dapeng, ZHOU Xiaoyan, YE Ru, et al. Bird sound recognition algorithm based on feature selection and GWO-KELM[J]. Technical Acoustics, 2022, 41(5): 782-788.] DOI: 10.16300/j.cnki.1000-3630.2022.05.022

基于特征选择和 GWO-KELM 的鸟声识别算法

李大鹏, 周晓彦, 叶如, 夏煜, 徐华南

(南京信息工程大学电子与信息工程学院, 江苏南京 210044)

摘要: 针对鸟声识别算法中提取特征单一、分类准确率低等问题, 提出一种基于混合特征选择和灰狼算法优化核极限学习机的鸟声识别方法。首先从鸟声数据中提取大规模声学特征集 ComParE, 其次计算每个特征的 F_{score} 并进行排序, 然后以广义顺序向前浮动搜索(Generalized Sequential Forward Floating Search, GSFFS)为搜索策略, 特征子集在核极限学习机(Kernel Limit Learning Machine, KELM)上十折交叉验证的正确率, 作为特征选择标准进行特征选择, 得到适用于鸟声识别的特征子集, 最后通过灰狼算法(Grey Wolf Optimizer, GWO)选择最优 KELM 参数识别鸟声。在柏林自然科学博物馆鸟声数据库中进行实验, 该方法在 60 类鸟声识别平均正确率和 F1-score 达到 94.45% 和 92.29%。结果表明, 该方法相较于传统自行设计提取的单一特征集具有更高的识别精度, GWO-KELM 模型比网格搜索方式更易找到全局最优值。

关键词: 核极限学习机; 特征选择; 鸟声识别; 灰狼算法

中图分类号: TN912.34

文献标志码: A

文章编号: 1000-3630(2022)-05-0782-07

Bird sound recognition algorithm based on feature selection and GWO-KELM

LI Dapeng, ZHOU Xiaoyan, YE Ru, XIA Yu, XU Huanan

(College of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China)

Abstract: To address the problems of single feature extracted and low classification accuracy in bird sound recognition algorithms, a bird sound recognition method based on hybrid feature selection and Gray Wolf algorithm-optimized kernel limit learning machine is proposed. Firstly, the large-scale acoustic feature set ComParE is extracted from bird sound data, secondly, the F_{score} of each feature is calculated and ranked, then the generalized sequential forward floating search (GSFFS) is used as the search strategy, and the correct rate of feature subsets on the kernel limit learning machine (KELM) with ten-fold cross-validation is used as the feature selection criterion to select the features applicable to bird sound recognition subset, and finally the optimal KELM parameters are selected by the Grey Wolf Optimizer (GWO) to recognize bird sounds. In the experiments conducted in the bird sound database of the Berlin Museum of Natural Sciences, the average correct rate and F1-score of the method reaches 94.45% and 92.29% for 60 types of bird sounds. The results show that the method has higher recognition accuracy than the traditional self-designed single feature set, and the GWO-KELM model is easier to find the global optimal value than the grid search method.

Key words: kernel limit learning machine(KELM); feature selection; bird song recognition; Grey Wolf Optimizer (GWO)

0 引言

鸟类作为生态系统的重要组成部分, 对鸟类活动和分布的监测, 为了解一个地区的生物多样性变化和气候变化提供了重要的依据^[1-2], 因此对鸟类的监测与分类识别具有重要意义。鸟鸣声和形态特征

是区分鸟类的重要特征, 也是目前鸟类物种识别普遍采用的方式, 在实际监测中鸟鸣声相较于形态特征更加便于监测。虽然目前国内外对于鸟声识别的研究并不多, 但也取得了一定的成果。通过调查研究发现, 目前国内的鸟声识别技术主要通过改进鸟声特征提取算法, 提取各种鸟声特征然后使用机器学习算法构建分类器进行识别^[3]。

目前, 鸟声识别的分类方法可以分为 3 种: (1) 基于模板匹配的分类方法。最常见的是动态时间重整(Dynamic Time Warping, DTW)算法, 如徐淑正使用基于音长、梅尔倒谱系数(Mel-Frequency Ceps-

收稿日期: 2021-04-22; 修回日期: 2021-06-22

基金项目: 国家自然科学基金资助项目(61902064)。

作者简介: 李大鹏(1997—), 男, 江苏南京人, 硕士研究生, 研究方向为鸟声识别。

通信作者: 周晓彦, E-mail: xiaoyan_zhou@nuist.edu.cn

tral Coefficients, MFCC)、线性预测系数(Linear Prediction Coefficient, LPCC)系数和时频域纹理特征的动态时间规整算法可以达到90%的准确率^[4]。(2)通过人工算法提取特征,选择合适的分类模型。孙悦华等提取了4种鸟类鸣声的MFCC,然后使用高斯混合模型对其进行分类识别,正确率可以达到89.1%~92.5%^[5]。陈莎莎等为了降低环境噪声对鸟声识别的干扰,使用灰度共生法提取纹理特征,然后利用随机森林(Random Forest, RF)模型对20种鸟声进行识别,平均正确率可以达到95.35%^[6]。钱坤等首次使用OpenSmile提取大规模声学特征,使用ReliefF算法降低特征维度,极限学习机(Extreme Learning Machine, ELM)作为分类器实现鸟声的分类识别,该方法在10种、30种和54种鸟类的识别正确率分别为94.71%、89.56%和86.57%^[7]。张赛花等提取了一种梅尔子带参数化特征,使用支持向量机(Support Vector Machine, SVM)对野外11种鸟鸣声进行分类识别,结果表明该方法对11类鸟声查全率、查准率和F1-score均高于89%^[8]。(3)基于深度学习的方法。EmreCakir等基于提出了卷积递归神经网络(Convolutional Recurrent Neural Networks, CRNN)的方法实现鸣声的高维特征及短时帧间的相关性特征提取,对Freesound数据库中的鸟鸣声进行分类实验,正确率达到88.5%^[9]。冯郁茜提出了基于双模态特征融合的鸟类物种分类算法,融合卷积网络提取的语图特征和长短时记忆结构提取的鸣声时序序列特征,自适应完成基于鸣叫或者鸣唱的物种识别^[10]。Bold等利用卷积神经网络(Convolutional Neural Networks, CNN)提取语图特征并且提出跨模态结合特征,提高了分类识别的性能^[11]。

由于文献[5-7]所提取的鸟声特征较为单一,分类参数采用网格搜索的方式容易错过最优值,无法达到分类器的最好性能。本文受文献[7]的启发,将广泛应用于语音情感识别的ComParE特征集^[12]应用于鸟声识别领域,同时为了降低特征冗余度,采用适合高维样本的特征排序结合浮动搜索策略的混合特征选择方法,以核极限学习机(Kernel Extreme Learning Machine, KELM)^[13]十折交叉验证的正确率来评判特征的优劣,进行特征选择,得到适用于鸟声的特征子集,最后通过灰狼算法^[14]对KELM参数进行寻优,得到最优参数提高模型识别正确率。

1 基于特征选择和GWO-KELM的鸟声识别算法

本文所提出的鸟声识别系统总体框架如图1所

示主要分成两个部分:

(1) 基于KELM和 F_{score} 的混合特征选择

首先对获取的鸟声进行预处理,并通过OpenSmile提取ComParE特征集,然后计算每个特征的 F_{score} ,对特征的区分能力进行评价并排序。最后以KELM十折交叉验证正确率作为特征选择的标准,采用广义顺序向前浮动搜索(Generalized Sequential Forward Floating Search, GSFFS)^[15]作为搜索策略进行特征选择得到最终适用于鸟声识别的特征子集。

(2) 基于GWO的KELM识别算法

为了提高KELM模型在鸟声分类识别上的效果及稳定性,将特征子集在KELM模型上十折交叉验证的正确率,作为灰狼优化算法的适应度,迭代寻找最优的正则化参数 c 和核函数参数 σ 。最后在该参数上对KELM模型进行训练,得到识别结果。

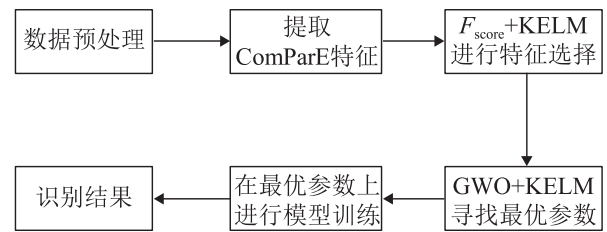


图1 鸟声识别系统框图

Fig.1 Block diagram of bird sound recognition system

1.1 核极限学习机

极限学习机^[16]由黄广斌在2006年提出,是一种只有一个隐藏层的神经网络结构,它的隐藏层节点参数 w 和 b 通过随机生成的方式产生,与训练集无关^[17],其输出函数为

$$f(x) = h(x)\beta \quad (1)$$

其中: $h(x)$ 为隐藏层的特征映射,隐藏层输出权重 β 根据广义逆矩阵原理和最小二乘法计算得到:

$$\beta = H^T(HH^T)^{-1}T \quad (2)$$

其中: H 是隐藏层的输出矩阵, T 是目标矩阵。

Huang等^[13]通过对比极限学习机与SVM的原理,结合核函数方法用核映射代替ELM中的随机映射,提出了核极限学习机(KELM),利用Mercer构造核矩阵 Ω_{ELM} 取代广义逆矩阵中的 HH^T 。

$$HH^T = \Omega_{ELM} = \begin{bmatrix} K(x_1, x_1) & \dots & K(x_1, x_N) \\ \vdots & & \vdots \\ K(x_N, x_1) & \dots & K(x_N, x_N) \end{bmatrix} \quad (3)$$

所以由式(2)、式(3)可得隐藏层输出权重:

$$\beta = H^T(\Omega_{ELM})^{-1}T \quad (4)$$

为了提高KELM的稳定性与泛化能力,引入正则化参数 $1/c$ 添加在对角矩阵 HH^T 对角线上。则输

出权重表示为

$$\beta = H^T \left(\frac{I}{c} + \Omega_{ELM} \right)^{-1} T \quad (5)$$

其中： I 为单位矩阵，由式(3)、(4)可得当特征映射 $h(x)$ 未知时KELM的输出函数可以表示为

$$f(x) = h(x)\beta = h(x)H^T \left(\frac{I}{c} + HH^T \right)^{-1} T \quad (6)$$

此时， $f(x)$ 的表达式为

$$f(x) = [k(x, x_1) \dots k(x, x_N)] \left(\frac{I}{c} + \Omega_{ELM} \right)^{-1} T \quad (7)$$

1.2 基于KELM和 F_{score} 的混合特征选择算法

由于ComParE特征集是为人类语音情感识别而设计，其中部分特征并不能反映鸟声的特点，为了更好地实现对鸟声的识别，降低特征冗余度，本文使用特征选择算法寻找合适的特征子集。

特征选择算法按照是否独立于分类器，可以分为Filter型和Wrapper型^[18]，Filter算法根据特定的准则评价特征重要性，独立于分类器，速度快，但特征分类能力较弱；Wrapper算法以分类器性能为评价标准，特征子集分类效果好，但算法速度慢，特征泛化能力较差^[19]。本文将这两种方式混合，把 F_{score} 作为特征区分能力评价标准结合KELM分类器，以广义顺序向前浮动搜索(GSFFS)^[15]为搜索策略寻找最佳特征子集。

1.2.1 F_{score} 特征区分能力评价标准

传统的 F_{score} 只能用来计算两类特征的区分能力，谢娟英等^[20]及杨勇^[21]对其进行了扩展，提出了改进的 F_{score} ，可以衡量多类特征的区分能力。为了满足对于多种鸟声识别的需求，选择改进的 F_{score} 作为鸟声特征区分能力评价标准对于提取到的鸟声特征样本 $\{(x_k, y_k) | x_k \in \mathbf{R}^m, y_k \in \{1, 2, \dots, l\}, l > 2, k = 1, 2, \dots, n\}$ ，第 i 个特征的 F_{score} 计算公式为

$$F_i = \frac{\sum_{j=1}^l (x_{Avg,i}^{(j)} - x_{Avg,i})^2}{\sum_{j=1}^l \frac{1}{n_j - 1} \sum_{k=1}^{n_j} (x_{k,i}^{(j)} - x_{Avg,i}^{(j)})^2} \quad (8)$$

其中： $x_{Avg,i}$ 、 $x_{Avg,i}^{(j)}$ 分别表示第 i 个特征在全部鸟声特征上的平均值、在第 j 类鸟声特征的平均值； $x_{k,i}^{(j)}$ 表示第 j 类第 k 个鸟声特征在第 i 个特征的值。式(8)中分子表示不同类别鸟声样本间的距离之和，分母表示相同类别样本间的紧密程度， F_i 越大则该特征的区分能力越强。

1.2.2 混合特征选择

为了提高搜索效率，本文首先对计算得到鸟声

特征的 F_{score} 按照大小进行降序排序，选择前20%特征作为初始特征集，删除得分最低的20%特征，剩下60%特征作为预选特征集。然后，依次不断加入预选特征集中 F_{score} 得分最高的特征，以KELM十折交叉验证正确率为特征区分能力评价标准，如果正确率没有提高就从特征子集中删除该特征，继续依次搜索剩下的预选特征集。为了解决严格地按照正确率上升作为特征选择标准造成特征子集无法更新的问题，引入惩罚参数 λ ，即KELM正确率下降不超过 λ 就不会删除该特征，混合特征选择算法步骤如下：

- (1) 计算每个鸟声特征的 F_{score} ，并降序排序；
- (2) 选择前20%特征作为初始特征集，20%~80%特征作为预选特征集，个数为 n ，初始化 $i=0, R_{Acc}=0$ ；
- (3) 令 $i=i+1$ ，将第 i 个特征加入特征子集；
- (4) 计算特征子集的KELM十折交叉验证正确率为 R_{Acc1} ；
- (5) 若 $R_{Acc1} + \lambda > Acc$ ，更新特征子集和正确率令 $R_{Acc}=R_{Acc1}$ ，否则从特征子集中删除该特征；
- (6) 若 $i < n$ 返回步骤(3)，否则输出被选特征子集。

1.3 基于GWO优化的KELM分类模型

对于分类模型参数的选取，传统的方法一般使用网格搜索的方式进行，但该方法在较大范围使用小步长进行搜索时，需要较高的计算资源和较长的时间，而使用较大步长搜索往往可能错过全局最优值。因此本文采用随机搜索的方式寻找KELM模型参数的全局最优值。

1.3.1 灰狼算法

灰狼优化(Grey Wolf Optimizer, GWO)算法是由澳大利亚学者Mirjalili等人在2014年提出的一种元启发式群体智能优化算法，该算法通过模拟自然界中灰狼的社会等级制度和狩猎机制，将灰狼按照等级从高到低分为 α 狼、 β 狼、 δ 狼和 ω 狼四种狼，实现对于猎物的寻找、包围和攻击^[14]。在GWO算法中，群体适应度最优的三个解依次对应为 α 狼、 β 狼、 δ 狼，其余的解为 ω 狼，通过 α 狼、 β 狼、 δ 狼引导 ω 狼包围、攻击猎物，从而获得最优值。

GWO算法数学模型为

$$D = |C \odot X_p(t) - X(t)| \quad (9)$$

$$X(t+1) = X_p(t) - A \odot D \quad (10)$$

其中： D 表示灰狼个体与猎物间的距离， $X(t+1)$ 表示灰狼更新后的位置，公式中 $\odot$ 均为矩阵的哈达玛

积, $|\cdot|$ 为取矩阵中元素的绝对值。 t 表示当前迭代次数, X_p 和 X 分别表示猎物和灰狼的位置向量。 A 和 C 为系数向量, 计算公式为

$$A = 2a \odot r_1 - a \quad (11)$$

$$C = 2r_2 \quad (12)$$

其中: a 随迭代次数 t 线性递减, r_1 、 r_2 为 $[0, -1]$ 的随机向量。 C 是系数向量, 用于增加($C > 1$)或减少($C < 1$)灰狼与猎物之间的距离, 提高全局搜索能力。

假设狼 α 、 β 、 δ 知道猎物潜在位置, 则灰狼更新公式为

$$\begin{cases} D_\alpha = |C_1 \odot X_\alpha - X|, & X_1 = X_\alpha - A_1 \odot (D_\alpha) \\ D_\beta = |C_2 \odot X_\beta - X|, & X_2 = X_\beta - A_2 \odot (D_\beta) \\ D_\delta = |C_3 \odot X_\delta - X|, & X_3 = X_\delta - A_3 \odot (D_\delta) \end{cases} \quad (13)$$

$$X(t+1) = \frac{X_1 + X_2 + X_3}{3} \quad (14)$$

其中: D_α 、 D_β 、 D_δ 分别表示狼 α 、 β 、 δ 与其他个体的距离, X_α 、 X_β 、 X_δ 表示狼 α 、 β 、 δ 的当前位置。

1.3.2 GWO-KELM算法

常用的KELM核函数有高斯核函数, 多项式核函数, 拉普拉斯核函数和sigmoid核函数等^[13], 实验中KELM核函数选取高斯核函数, 其计算公式为

$$k(x, x') = e^{-\frac{\|x - x'\|^2}{2\sigma^2}} \quad (15)$$

由式(7)、式(15)可知, KELM性能与正则化参数 c 和核函数参数 σ 密切相关, 为了提高KELM模型分类识别效果及稳定性, 本文通过灰狼算法优化KELM的参数 c 和 σ , 把KELM模型十折交叉验证正确率作为灰狼优化算法的适应度。GWO优化KELM分类识别算法步骤如下:

- (1) 初始化GWO算法参数: 狼群数量、最大迭代次数 t 、参数 c 和 σ 上下界;
- (2) 随机初始化狼群位置, 位置维度为2;
- (3) 根据狼群位置即 c 和 σ , 计算KELM交叉验证正确率作为灰狼个体适应度;
- (4) 选择适应度最好的三个狼, 记录其位置为 X_α 、 X_β 、 X_δ ;
- (5) 根据式(13)、式(14)更新狼 ω 位置;
- (6) 根据式(11)、式(12)参数 A 和 C ;
- (7) 计算位置更新后的适应度, 并与上次迭代的最优适应度对比, 更新狼 α 、 β 、 δ 适应度和位置;
- (8) 若迭代次数超过最大次数, 输出全局最优适应度及其对应位置 X_α , 否则执行步骤(5)。

2 实验

2.1 鸟声数据库及预处理

为了验证算法的有效性, 本文采用国内国外两种数据库进行实验。一是德国柏林自然科学博物馆数据库, 该数据库由专业的鸟类学家在自然野外环境中采集的鸟鸣声数据组成。为了保证足够的训练、测试数据, 实验中删除数据库中鸟声音频文件数少于25个的鸟类, 采用了60种鸟类、共计4 468个鸟鸣声音频文件, 最短时长约为0.3 s, 最长为36 s。同时将这4 468个长度不等的鸟声数据统一为单声道、采样率44.1 kHz、32 bit的WAV格式音频。二是由北京百鸟数据科技有限责任公司发布的中国常见20种鸟类鸣声合集, 包含灰雁、大天鹅、绿头鸭、绿翅鸭、灰山鹑、西鹌鹑、雉鸡、红喉潜鸟、苍鹭、普通鸬鹚、苍鹰、欧亚鸺、西方秧鸡、骨顶鸡、黑翅长脚鹬、凤头麦鸡、白腰草鹬、红脚鹬、林鹬、麻雀, 共14 311个长度均为2 s的WAV格式音频。

2.2 特征提取

使用在语音情感识别中广泛使用的开源工具OpenSmile提取ComParE特征集。该特征集在2013年InterSpeech中被提出, 包含了大量手工设计的低水平特征LLDs和在LLDs基础上统计得到的高级统计特征HSFs, 共包含6 373维度特征, 包括MFCC、时间信号的过零率(Zero-Crossing Rate, ZCR)、帧能量均方根(Root Mean Square, RMS)、基音频率(F_0)等特征, 具体信息可参见文献[12]。

2.3 分类模型评价标准

混淆矩阵是一种直观地评价分类模型结果的指标, 如图2所示是二分类结果混淆矩阵。

	一类 (预测)	二类 (预测)
一类 (真实)	N_{TP}	N_{FN}
二类 (真实)	N_{FP}	N_{TN}

图2 二分类结果混淆矩阵

Fig.2 The confusion matrix of dichotomy results

图2中, N_{TP} 为样本的真实类别是正例且模型预测的结果也是正例的样本个数; N_{TN} 为真实类别是负例, 且预测成为负例的样本个数; N_{FP} 为真实类别是负例, 但预测成为正例的样本个数; N_{FN} 为真实类别是正例, 但预测成为负例的样本个数。

对混淆矩阵的数据进行简单计算,可以延伸得到查准率、查全率、正确率和F1-score等常用评价指标。对于每种鸟类其计算公式如下:

查准率(精准率):

$$R_{Pre} = \frac{N_{TP}}{N_{TP} + N_{FP}} \tag{16}$$

查全率(召回率):

$$R_{Recall} = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{17}$$

正确率(准确率):

$$R_{Acc} = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{FP} + N_{TN} + N_{FN}} \tag{18}$$

F1-score:

$$F_{F1-score} = \frac{2R_{Recall} \times R_{Pre}}{R_{Recall} + R_{Pre}} \tag{19}$$

本文选择正确率和兼顾查准率和查全率的F1-score作为分类模型评价指标。

2.4 实验与分析

为了验证本文算法的有效性,实验共分为三个部分,其中前两部分实验均在德国柏林自然科学博物馆数据库上实现。首先对比原始 ComParE 特征集在不同分类器上的表现;其次对比选择后的特征子集与原始 ComParE 特征集在不同分类器上的识别精度,并对比采用网格搜索方式和 GWO 随机搜索方式所得参数识别结果;最后对比了近年来相关论文所提算法和主流的深度学习算法。实验的特征提取部分,通过基于python的 OpenSmile3.0 工具包实现提取鸟声音频的 ComParE 特征,每个音频文件可以得到 6 373×1 维特征向量,循环遍历数据库中的所有音频提取其特征并保存为 CSV 文件。然后以 Matlab 2018b 软件为算法实验平台,采用十折交叉验证方式作为实验协议即将数据集分成十份,轮流将其中 9 份作为训练数据、1 份作为测试数据,进行试验。

2.4.1 ComParE特征集在不同分类器上的表现

该实验在德国柏林自然科学博物馆数据库上进行,并随机选取了其中 10 类, 30 类和全部 60 类数据进行实验。对比了原始 ComParE 特征集在 RF、SVM、ELM 和 KELM 分类器上的表现结果如表 1 所示。实验设置如下,除 RF 模型以外在将特征向量输入分类器之前,首先对特征进行归一化处理,其中 RF 模型采用默认设置决策树数量为 500,特征数量为 79;由于特征参数维度很高,因此 SVM 模型采用线性核函数,正则化参数设置为 1; ELM 模型隐藏层数量通过循环搜索的方式遍历[1 000

2 000 ... 30 000]最终在 10 类, 30 类和 60 类实验中对应的设置为 10 000、16 000、20 000; KELM 模型核函数选择为高斯核函数,采用网格搜索的方式,高斯核函数参数 $\sigma \in [2^{-5} \ 2^{-4} \dots 2^{15}]$ 和正则化参数 $c \in [2^{-5} \ 2^{-4} \dots 2^{15}]$ 最终 σ 和 c 设置为 4 096 和 2 048。

从表 1 中可以看到 KELM 分类器在 10 类、30 类和 60 类鸟声识别十折交叉验证正确率为 96.67%、93.77% 和 93.23%, 相对于其他分类器均具有更高的正确率。结果表明 KELM 算法相较于其他算法在高维度鸟声特征分类识别中更具优势,体现了 KELM 分类器的优越性。

表 1 ComParE 特征集在分类器上的表现
Table 1 The performance of the ComParE feature set on the classifier

分类	10 类数据结果		30 类数据结果		60 类数据结果	
	R_{Acc}	$F_{F1-score}$	R_{Acc}	$F_{F1-score}$	R_{Acc}	$F_{F1-score}$
RF	88.51%	84.66%	85.82%	76.95%	83.88%	76.13%
SVM	95.67%	95.24%	91.93%	88.68%	89.41%	85.87%
ELM	94.81%	94.65%	90.59%	86.95%	87.24%	82.27%
KELM	96.67%	96.65%	93.77%	91.32%	92.23%	89.45%

2.4.2 基于特征选择和GWO-KELM算法实验

为了进一步提高鸟声识别的效果,本文对提取到的 6 373 维特征集采用 F_{score} +KELM 的方法在德国柏林自然科学博物馆全部 60 类数据库进行特征选择,减少冗余特征,实验惩罚参数 λ 设置为 0.001。为了更好地体现特征对于分类器性能的影响,将 KELM 参数 σ 和 c 固定为 4 096 和 2 048,最终选择出 2 710 维特征。

GWO 算法初始设置为初始化狼群数量为 10,最大迭代次数为 100,参数 c 和 σ 的上下界设置为 10 000 和 0.000 1,初始狼群位置在上下界范围内随机生成。如图 3 所示 GWO-KELM 在选择后的特征集上的迭代结果,最终选择的最优参数 c 和 σ 分别为 316、6 112。其他分类器的参数设置与 2.4.1 节相同。

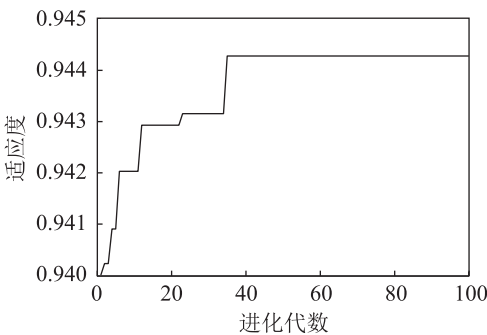


图3 GWO-KELM迭代结果
Fig.3 GWO-KELM iteration results

表2所示为选择后的特征子集与原始ComParE特征集在不同分类器和GWO-KELM算法的识别结果。从表2中可以看出选择后的特征子集在四个分类器上的识别正确率和F1-score均高于原始特征集,提升幅度约2%~5%。结果表明,基于 F_{score} 和KELM特征选择算法减少了冗余特征,所选特征集具有良好的分类能力,能够更好地反映鸟鸣声的特点。在GWO-KELM模型($c=316$, $\sigma=6112$)上识别正确率为94.45%,相比采用网格搜索的方式KELM算法提高0.5%左右。与传统的网格搜索方式相比,GWO搜索方式更容易找到全局最优值,证明了GWO-KELM模型的有效性。

表2 不同分类器对60类鸟声特征选择前后的特征集的识别结果
Table 2 Identification results of different classifiers for the feature sets before and after the selection of the 60-class bird acoustic features

分类器	特征选择前结果		特征选择后结果	
	R_{Acc}	$F_{F1-score}$	R_{Acc}	$F_{F1-score}$
RF	83.88%	76.13%	85.11%	78.41%
SVM	89.41%	85.87%	92.10%	89.85%
ELM	87.24%	82.27%	92.12%	89.45%
KELM	92.23%	89.45%	93.89%	91.78%
GWO-KELM	92.96%	90.16%	94.45%	92.29%

2.4.3 与其他论文方案的比较

由于德国柏林数据库语音长短差距较大,无法直接作为CNN模型的输入,所以为了与目前主流的深度学习模型进行对比,实验增加了北京百鸟数据库的实验。该部分实验主要与以下4种方法进行对比:

(1) ComParE+SVM: 使用OpenSmile工具包提取ComParE特征集,通过经典的SVM算法进行识别。

(2) 09IS+ELM^[7]: 通过p-centre方法实现对鸟鸣声的音节检测,然后使用OpenSmile工具包提取09IS特征集,并通过ReliefF算法减少特征维数,最后采用ELM进行识别。

(3) Logmel+CNN^[11]: 通过librosa提取鸟鸣声的梅尔谱图,输入CNN进行识别。

(4) Logmel+CRNN^[9]: 通过librosa提取鸟鸣声的梅尔谱图,输入CNN和两层循环神经网络(Recurrent Neural Network, RNN)进行识别。

(注: 09IS+ELM模型采用原文在德国柏林数据库上的实验结果, Logmel+CNN和Logmel+CRNN模型实验结果通过python 3.7和基于TensorFlow的kreas 2.3框架,采用原文参数复现所得)。

实验结果如表3所示,从表中可以看出相较于

手工特征加分类器的方式,端到端的深度学习模型在识别率上更具备优势,本文所提的基于特征选择和GWO-KELM的鸟声识别算法,在传统手工特征方法中具备明显优势,可以达到与目前常见的卷积神经网络(CNN)近似相同的正确率,与CNN加RNN的方式仍有一定差距。但深度学习网络往往需要更高的计算量和更多数据样本。同时本文的方法不受音频长短的影响,在音频长短一致的北京百鸟数据库和音频长短不一的德国柏林数据库都可以实现较好的分类效果。而深度学习网络需要固定的语音长度或者通过补0等方法实现特征补齐,对于语音长短差距较大数据库处理较为困难。因此本文方法对于数据要求更低,具有更好的适应性。

表3 与其他论文方案的对比
Table 3 Comparison with other paper schemes

方案	德国柏林数据库结果		北京百鸟数据库结果	
	R_{Acc}	$F_{F1-score}$	R_{Acc}	$F_{F1-score}$
ComParE+SVM	89.41%	85.87%	87.49%	86.04%
09IS+ELM ^[7]	86.57%	—	—	—
Logmel+CNN ^[11]	—	—	91.12%	88.47%
Logmel+CRNN ^[9]	—	—	92.89%	89.64%
GWO-KELM	94.45%	92.29%	91.16%	88.54%

3 结论

为了提高种类较多的鸟声识别正确率,解决目前鸟声识别算法中特征单一的问题,本文提出了一种基于混合特征选择和灰狼算法优化核极限学习机的鸟声识别算法。该算法将 F_{score} 特征区分能力标准与KELM分类器相结合对提取的ComParE特征集进行特征选择,然后通过GWO算法优化KELM进行分类识别。通过对比实验可以看出,相较于原始ComParE特征集,选择后的特征子集在不同分类器上的识别正确率均有提高,表明了基于KELM和 F_{score} 的混合特征选择算法所选特征子集具有良好的分类能力。该方法在德国柏林数据库和北京百鸟数据库的分类识别正确率分别可以达到94.45%和91.16%。虽识别正确率略低于目前计算量更大的深度学习模型,但由于深度学习模型需要更多的数据样本,本文所提方法在小样本鸟声数据的识别上仍具有一定优势。

参 考 文 献

- [1] 赵洪峰,雷富民. 鸟类用于环境监测的意义及研究进展[J]. 动物学杂志, 2002, 37(6): 74-78.
ZHAO Hongfeng, LEI Fumin. Birds as monitors of environmental change[J]. Chinese Journal of Zoology, 2002, 37(6):

- 74-78.
- [2] 吴伟伟, 徐海根, 吴军, 等. 气候变化对鸟类影响的研究进展[J]. 生物多样性, 2012, **20**(1): 108-115.
WU Weiwei, XU Haigen, WU Jun, et al. The impact of climate change on birds: a review[J]. Biodiversity Science, 2012, **20**(1): 108-115.
 - [3] 乔玉, 钱昆, 赵子平. 基于机器听觉的鸟声识别的中文研究综述[J]. 复旦学报(自然科学版), 2020, **59**(3): 375-380.
QIAO Yu, QIAN Kun, ZHAO Ziping. A survey on Chinese literature for bird sound recognition based on machine listening[J]. Journal of Fudan University (Natural Science), 2020, **59**(3): 375-380.
 - [4] 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, HUANGPU 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.
 - [5] CHENG J K, SUN Y H, JI L Q. A call-independent and automatic acoustic system for the individual recognition of animals: a novel model using four passerines[J]. Pattern Recognition, 2010, **43**(11): 3846-3852.
 - [6] 陈莎莎, 李应. 结合时-频纹理特征的随机森林分类器应用于鸟声识别[J]. 计算机应用与软件, 2014, **31**(1): 154-157, 161.
CHEN Shasha, LI Ying. Applying random forest classifier combined with time-frequency texture features to bird sounds recognition[J]. Computer Applications and Software, 2014, **31**(1): 154-157, 161.
 - [7] QIAN K, ZHANG Z X, RINGEVAL F, et al. Bird sounds classification by large scale acoustic features and extreme learning machine[C]//2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP). Orlando, FL, USA. IEEE, 2015: 1317-1321.
 - [8] 张赛花. 面向鸟声传感网的鸟鸣自动分类方法研究[D]. 南京: 南京理工大学, 2018.
 - [9] CAKIR E, ADAVANNE S, PARASCANDOLO G, et al. Convolutional recurrent neural networks for bird audio detection [C]//2017 25th European Signal Processing Conference (EUSIPCO). Kos, Greece. IEEE, 2017: 1744-1748.
 - [10] 冯郁茜. 基于深度学习的双模态特征融合鸟类物种识别算法[D]. 北京: 北京林业大学, 2019.
FENG (Yu)(Qian|Xi). Birds species identification algorithm for two-modal feature fusion based on deep learning[D]. Beijing: Beijing Forestry University, 2019.
 - [11] BOLD N, ZHANG C, AKASHI T. Cross-domain deep feature combination for bird species classification with audio-visual data[J]. IEICE Transactions on Information and Systems, 2019, E102.D(10): 2033-2042.
 - [12] SCHULLER B, STEIDL S, BATLINER A, et al. The INTER-SPEECH 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism[C]//Interspeech 2013. ISCA: ISCA, 2013: 148-152.
 - [13] HUANG G B, ZHOU H M, DING X J, et al. Extreme learning machine for regression and multiclass classification[J]. IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics: a Publication of the IEEE Systems, Man, and Cybernetics Society, 2012, **42**(2): 513-529.
 - [14] MIRJALILI S, MIRJALILI S M, LEWIS A. Grey wolf optimizer[J]. Advances in Engineering Software, 2014, **69**: 46-61.
 - [15] XIE J Y, LEI J H, XIE W X, et al. Two-stage hybrid feature selection algorithms for diagnosing erythemato-squamous diseases[J]. Health Information Science and Systems, 2013, **1**: 10.
 - [16] HUANG G B, CHEN L, SIEW C K. Universal approximation using incremental constructive feedforward networks with random hidden nodes[J]. IEEE Transactions on Neural Networks, 2006, **17**(4): 879-892.
 - [17] 徐睿, 梁循, 齐金山, 等. 极限学习机前沿进展与趋势[J]. 计算机学报, 2019, **42**(7): 1640-1670.
XU Rui, LIANG Xun, QI Jinshan, et al. Advances and trends in extreme learning machine[J]. Chinese Journal of Computers, 2019, **42**(7): 1640-1670.
 - [18] 李郅琴, 杜建强, 聂斌, 等. 特征选择方法综述[J]. 计算机工程与应用, 2019, **55**(24): 10-19.
LI Zhiqin, DU Jianqiang, NIE Bin, et al. Summary of feature selection methods[J]. Computer Engineering and Applications, 2019, **55**(24): 10-19.
 - [19] 谢娟英, 王春霞, 蒋帅, 等. 基于改进的F-score与支持向量机的特征选择方法[J]. 计算机应用, 2010, **30**(4): 993-996.
XIE Juanying, WANG Chunxia, JIANG Shuai, et al. Feature selection method combing improved F-score and support vector machine[J]. Journal of Computer Applications, 2010, **30**(4): 993-996.
 - [20] 谢娟英, 谢维信. 基于特征子集区分度与支持向量机的特征选择算法[J]. 计算机学报, 2014, **37**(8): 1704-1718.
XIE Juanying, XIE Weixin. Several feature selection algorithms based on the discernibility of a feature subset and support vector machines[J]. Chinese Journal of Computers, 2014, **37**(8): 1704-1718.
 - [21] 杨勇. 基于粒子群优化策略的核极限学习机及其应用研究[D]. 兰州: 兰州大学, 2019.
YANG Yong. Study on KELM based on particle swarm optimization strategy and its application[D]. Lanzhou: Lanzhou University, 2019.