

基于特征迁移的多物种鸟声识别方法

刘昊天¹ 姜海燕^{1,2} 舒欣¹ 徐彦¹ 伍艳莲^{1,2} 郭小清¹

(1. 南京农业大学信息科技学院, 南京, 210095; 2. 南京农业大学国家信息农业工程技术中心, 南京, 210095)

摘要: 针对多物种鸟声识别中多物种鸟声样本不足的问题, 尝试采用单物种鸟声样本训练多物种鸟声识别模型, 并提出一种基于特征迁移的多物种鸟声识别方法。该方法引入特征迁移学习算法, 利用最大均值差异(Maximum mean discrepancy, MMD)度量鸟声样本特征分布差异, 将不同分布的单物种鸟声和多物种鸟声的音频特征映射为同分布的潜在音频特征, 再基于同分布的音频特征构造识别模型。使得单物种鸟声样本训练的识别模型也能够适用于多物种鸟声识别。在自然形成的多物种鸟声数据集上, 算法在4项多标记评价指标上都取得了较好的识别效果; 在人工构造的多物种鸟声数据集上对比试验表明, 基于特征迁移的识别算法在单个物种上的正确识别率相较于对比算法最高提升了20%。

关键词: 鸟声识别; 多物种; 特征迁移; 迁移学习

中图分类号: TN912.34 **文献标志码:** A

Recognition of Multiple Bird Species in Audio Recordings Based on Feature Transfer

Liu Haotian¹, Jiang Haiyan^{1,2}, Shu Xin¹, Xu Yan¹, Wu Yanlian^{1,2}, Guo Xiaoqing¹

(1. College of Information Science and Technology, Nanjing Agricultural University, Nanjing, 210095, China; 2. National Engineering and Technology Center for Information Agriculture, Nanjing Agricultural University, Nanjing, 210095, China)

Abstract: To deal with the problem of inadequate sample in multiple bird species recognition, a new recognition method of multiple bird species in audio recordings is proposed based on feature transfer, which uses bird sounds of single species to train a multiple bird species recognition model. Maximum mean discrepancy (MMD) is used to measure the feature distributions difference of bird sounds, which maps audio feature of single-species bird sounds and multiple-species bird sounds into a new latent feature with the same distribution. Then single-species bird sounds with latent feature can be used to train a model of multiple-species bird sounds. The experimental result shows that method can achieve good recognition performance in a natural multiple-species bird sounds dataset based on four multi-label metrics. The recognition rate of proposed method increases by 20% compared with other methods in an artificial multiple-species bird sounds dataset.

Key words: bird sounds recognition; multiple bird species; feature transfer; transfer learning

引 言

鸟类的鸣声和形态特征一样,不同物种间具有独特性,是鸟类物种识别的重要依据^[1]。自然环境下鸟声识别技术能够帮助统计鸟类物种分布状态,从而反映区域生态系统中动植物资源的覆盖度和丰富度。鸟声识别技术研究起步较晚,但近几年来随着环保问题日益受到关注,对该技术的研究也在增加。在鸟声识别研究初期,绝大多数研究都只针对少数目标物种的鸟声构造识别模型。薛宇志^[2]针对台湾地区常见的 10 个鸟类物种提出一种二阶层式识别架构,但由于该方法需要人工对鸟声的鸣唱和鸣叫类型进行判断,因此自动化程度不高;中国科学院动物研究所^[3]基于鸟声音频的梅尔频率倒谱系数(Mel frequency cepstral coefficients, MFCC)用高斯混合模型(Gaussian mixture model, GMM)进行 4 种雀形目鸟类的种内个体识别,该方法仅在种内进行个体识别,未考虑不同物种间的识别。近年来一些研究则关注于实际应用中可能出现的噪音干扰以及识别效率问题:魏静明^[4]为了降低环境噪音对鸟声识别的干扰,提出了一种针对鸟声的抗噪纹理特征提取方法,该方法首先利用短时谱估计对鸟声做增强处理,然后提取增强后功率谱的纹理特征,该方法提取的抗噪纹理特征不但具有良好的识别率还减少了纹理特征提取环节的耗时。Dan^[5]则针对 MFCC 音频特征在抗噪能力上的不足,提出了一种无监督的音频特征,在提升了抗噪能力的同时提高了特征提取的效率。

上述的鸟声识别方法只能识别单一物种的鸟声音频,而在鸟类分布较为密集的野生湿地或森林中可能出现不同物种的鸟同时出现的情况,这时录制的音频中可能同时包含两种以上的鸟声。因此,近几年来多物种鸟声识别技术研究开始受到关注。2012 年,Briggs^[6]首次提出一种基于多实例多标记学习的多物种鸟声识别方法,该方法首先利用随机森林模型将样本音频的频谱图中鸟声部分以音节的形状从背景中分割出来,再提取每个音节的频谱形状特征,一段音频中的所有音节组成一个包,最后采用 MIML-KNN 等传统的多实例多标记算法构建分类模型。2013 年电气和电子工程师协会(Institute of electrical and electronics engineers, IEEE)组织的“信号处理中的机器学习技术”(Machine learning for signal processing, MLSP)研讨会^[7]以及同年的“生物声学神经信息处理”(Neural information processing scaled for bioacoustics, NIPS4B)研讨会都将多物种鸟声识别作为会议竞赛题,因此涌现出了不少针对多物种鸟声识别的研究成果。然而这个阶段以来,所有的鸟声识别方法仍然以多实例多标记学习为基础,这类方法假设的前提是训练样本与测试样本音节的频谱形状类似,但由于自然环境中多重鸟声相互叠加是不可控的,很难保证训练样本与测试样本有类似的频谱形状。2015 年 Jancovic^[8]将 HMM 由单一物种的鸟声识别问题扩展到多物种鸟声识别问题,提出了一种基于最大似然惩罚(Penalized maximum likelihood)的多物种鸟声识别算法,该方法将每一种鸟声描述为一组频率变化轨迹,再利用 HMM 估计测试样本与各轨迹间的相似程度。该方法在鸟声相互叠加较少的场景下可以发挥很好的效果。

上述的多物种鸟声识别算法均需要大量已标记的多物种鸟声数据用于训练识别模型。通过对现有鸟声数据库的统计发现(见表 1),目前绝大多数鸟声数据库中已标记的多物种鸟声样本很少,绝大多数音频样本都是单物种鸟声音频。然而,单物种鸟声音频并不能直接用于训练多物种鸟声识别模型,这是因为多物种鸟声相较于单物种鸟声更复杂。多重鸟声的相互叠加影响,造成了多物种鸟声与单物种鸟声的音频特征分布不同,单物种鸟声样本训练得到的识别模型无法适应多物种鸟声的特征分布。

特征迁移是一种常用的迁移学习方法,它通过构造训练样本与测试样本间的公共潜在特征表示,减小训练与测试样本的特征分布差异,从而实现不同分布样本间的知识迁移。特征迁移放宽了训练样本与测试样本的同分布要求,目前已经在图像分类、文本分类甚至疾病的早期诊断中取得了不错的应用效果。因此,为了解决多物种鸟声识别中训练样本不足的问题,可以尝试采用特征迁移方法,用单物种鸟声样本训练多物种鸟声识别模型。

表 1 常用鸟声数据集
Tab. 1 Common bird sounds data sets

| 样本类型 | 数据集名称 | 样本数 | 数据来源 |
|------|-----------------------------------|---------|--|
| 单物种 | Borror laboratory of bioacoustics | 26 689 | blb. osu. edu |
| | Macaulay Library | 4 929 | www. macaulaylibrary. org |
| | xeno-canto | 296 483 | www. xeno-canto. org |
| | LifeCLEF 2015 | 33 203 | www. imageclef. org/lifeclef/2015/bird |
| 多物种 | British Library Sound Archive | 869 | sounds. bl. uk |
| | NIPS4B | 687 | sabiod. univ-tln. fr/nips4b/challengel. html |

本文针对多物种鸟声训练样本缺少的问题,提出一种基于特征迁移的多物种鸟声识别方法(Recognition of multiple bird species in audio recordings based on feature transfer, FT-RMBS)。该方法以基于 MFCC 和 GMM 的鸟声识别框架为基础,引入最大均值差异嵌入算法^[9](Maximum mean discrepancy embedding, MMDE)来减少单物种与多物种鸟声的特征分布差异,构造单物种鸟声与多物种鸟声的共享潜在音频特征子空间,以期实现利用已有的大量单物种鸟声样本训练多物种鸟声识别模型的目的。

1 基于特征迁移的多物种鸟声识别方法

FT-RMBS 由“潜在特征提取”和“多物种分类”2 个关键步骤组成。“潜在特征提取”通过 MMDE 特征迁移算法提取训练样本与测试样本的共享潜在特征,着重解决单物种鸟声与多物种鸟声特征分布不一致问题;“多物种分类”以样本的潜在音频特征为输入,采用多标记学习算法训练多物种鸟声识别模型。FT-RMBS 的基本流程如图 1 所示。

1.1 潜在特征提取

1.1.1 提取 MFCC

MFCC 是最常用的音频特征之一,最初常用于人类语音识别^[10],但近年来也被应用于昆虫叫声和鸟声^[10]的识别中。MFCC 将人耳听觉感知特性与声音的产生相结合,利用人听觉的临界带效应,在声音频谱范围内设置若干个带通滤波器,每个滤波器具有三角形或正弦形滤波特性。然后将声音能量谱通过该滤波器组,求各个滤波器输出,对其取对数,并作离散余弦变换(Discrete cosine transform, DCT),即可得到 MFCC 系数。MFCC 系数的具体计算过程如图 2 所示。

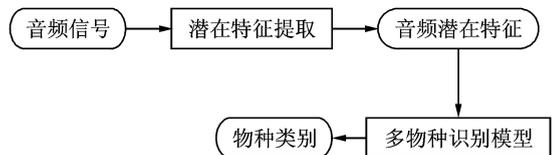


图 1 基于特征迁移的多物种鸟声识别流程

Fig. 1 Flow diagram of multiple bird species recognition based on feature transfer

(1)分帧加窗:鸟类声音信号具有短时平稳的特性,因此在分析鸟声时,为了便于处理,可以将长

时音频信号划分为多段短时帧。为了减小后续傅里叶变换中吉布斯效应^[10]的影响,增加帧左右两端的连续性。本文对音频信号加汉明窗(Hamming)交叠分帧,设置一帧的长度为 512 个采样点,帧间的偏移取为帧长的 1/2,即 256 个采样点。

(2)快速傅里叶变换:由于信号在时域上变化迅速,没有办法作有效的观察。为了方便分析信号的特性,通常会采用快速离散傅里叶变换(Fast Fourier transformation, FFT)将它转换到频域上的能量分布来观察。

(3)Mel 滤波:由于人耳听觉存在遮蔽效应,即听觉神经不只接受单一特定频率刺激,会被互相临近的频率影响。因此可以将 FFT 得到的频谱系数通过 M 个三角带通滤波器处理,让信号在频率上平均分布来模拟人耳听觉特性。

(4)对数转换:将 Mel 滤波得到的能量系数进行对数运算得到相应的音频功率。

(5)DCT:将不同频带的能量取对数变换后,再经由 DCT 变换回时域,得到原始 MFCC 系数。使用 DCT 主要用于减少维度间的关系,以增加识别的效率。

(6)差分变换:MFCC 特征参数是一种静态音频特征,无法反映音频信号随时间的动态变化。为了反映音频信号的动态特性常常在 MFCC 特征的基础上将一阶差分 MFCC(Δ MFCC)和二阶差分 MFCC(Δ^2 MFCC)作补充特征。

1.1.2 潜在特征映射

由于本研究问题中训练样本由单物种鸟声数据组成,而测试样本则是多物种鸟声数据。测试音频样本相较于训练音频样本包含了更多物种的鸟声,不同鸟声间的相互叠加影响使得单物种鸟声与多物种鸟声的音频特征具有不同的分布,而识别模型需要通过拟合样本的特征分布预测样本所属的类别,单物种鸟声与多物种鸟声的特征分布差异会使得训练阶段构造的识别模型无法正确拟合测试样本的特征分布。为此,本文引入 MMDE 算法,提取多物种鸟声与单物种鸟声的同分布的潜在特征。

MMDE 算法通过将训练样本与测试样本映射到同一个潜在特征空间中,使得训练样本和测试样本的特征分布相似。MMDE 采用最大均值差异(Maximum mean discrepancy, MMD)度量训练样本特征 $\mathbf{X}^{(s)} = [\mathbf{x}_1^{(s)}, \dots, \mathbf{x}_{n_s}^{(s)}]$ 与测试样本特征 $\mathbf{X}^{(t)} = [\mathbf{x}_1^{(t)}, \dots, \mathbf{x}_{n_t}^{(t)}]$ 的分布差异。令 H 表示再生核希尔伯特空间(Reproducing kernel Hilbert space, RKHS), $\varphi(\cdot): X \rightarrow H$ 表示原始特征空间映射到 RKHS 的映射函数,MMDE 的目的是找到一个理想的潜在特征映射函数 $\psi(\cdot)$,使得 $\tilde{\mathbf{X}}^{(s)} = \psi(\mathbf{X}^{(s)})$ 和 $\tilde{\mathbf{X}}^{(t)} = \psi(\mathbf{X}^{(t)})$ 在 RKHS 中的 MMD 距离尽可能小。MMDE 的目标函数为

$$\arg \min_{\psi} \text{mindist}_{\text{MMD}}^2(\mathbf{X}^{(s)}, \mathbf{X}^{(t)}) = \arg \min_{\psi} \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \varphi(\psi(\mathbf{x}_i^{(s)})) - \frac{1}{n_t} \sum_{j=1}^{n_t} \varphi(\psi(\mathbf{x}_j^{(t)})) \right\|_H^2 \quad (1)$$

该函数的详细求解方法可参考文献[9]。

1.2 多物种分类

1.2.1 GMM 估计

高斯混合模型(Gaussian mixture model, GMM)是利用高斯分布的概率密度函数的组合来描述特征

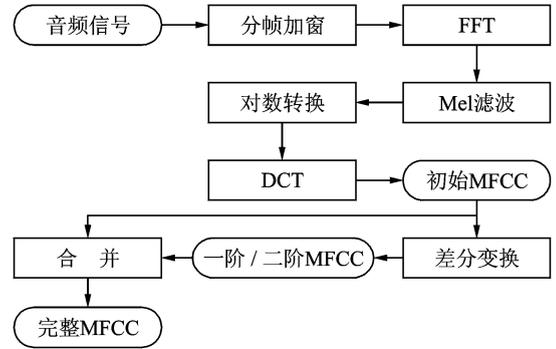


图 2 MFCC 计算过程

Fig. 2 MFCC calculation process

矢量在概率空间的分布状况的模型,可以很好地描述从声音信号中提取的特征参数这类不规则的数据,是当今最主流模型。一个 M 阶高斯混合模型的概率密度函数是由 M 个高斯概率密度函数加权求和得到的,表达为

$$P(\bar{\mathbf{x}} | \lambda) = \sum_{i=1}^M \omega_i b_i(\bar{\mathbf{x}}) \tag{2}$$

式中: ω_i 为混合权重,满足 $\sum_{i=1}^M \omega_i = 1$; λ 为 GMM 的参数集合, $\lambda = \{\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\}$, 其中 $\boldsymbol{\mu}$ 为均值矢量, $\boldsymbol{\Sigma}$ 为协方差矩阵; $b_i(\bar{\mathbf{x}})$ 为第 i 个 d 维高斯概率分部,可表示为

$$b_i(\bar{\mathbf{x}}) = \frac{1}{(2\pi)^{\frac{d}{2}} |\boldsymbol{\Sigma}_i|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\bar{\mathbf{x}} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_i)\right\} \tag{3}$$

GMM 的参数集合 λ 可以采用最大期望(Expectation maximization, EM)算法迭代求得。

1.2.2 多标记分类

在单物种鸟声识别中,当目标物种类别多于 2 种时可以针对每个物种的鸟声建立独立的 GMM 识别模型。最终以分布概率最高的 GMM 对应的物种类别标记样本^[11]。但该方法的识别结果只能是单一物种类别,无法同时标记样本中的多个物种类别。为了实现多物种分类,本文在 GMM 输出的分布概率基础上,增加 RankSVM 多标记分类算法进一步分类。

RankSVM 是一种基于传统支持向量机(Support vector machine, SVM)的多标记排序模型^[12]。算法将多标记分类问题中的样本标记看作 rank 值 $\mathbf{y} = [r_1, r_2, \dots, r_m]$, 并假设 rank 值间存在一个顺序关系 $r_1 > r_2 > \dots > r_m$, 其中“>”表示参考关系,理想情况下对于某个训练样本中任意一对标记 $\{y_1 = 1, y_2 = 0\}$ 都有 $\{r_1 > r_2\}$ 。SVM 的任务是找到最优的函数 f , 最小化样本对应 rank 值的代价函数。

本文将 GMM 输出的分布概率作为 RankSVM 的输入,构造多物种分类模型,该模型的基本架构如图 3。

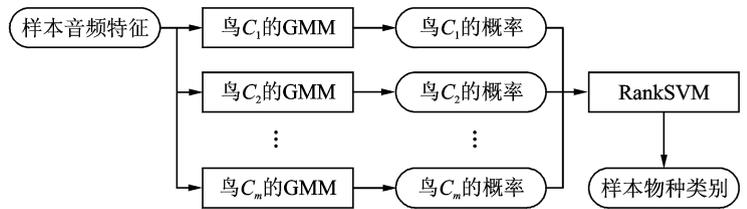


图 3 多物种鸟声识别模型基本架构

Fig. 3 Architecture of multiple bird species recognition model

2 试验结果与分析

2.1 实验设置

2.1.1 数据集构造

为了验证所提方法的识别能力,本文采用一个人工合成的多物种鸟声数据集和一个天然多物种鸟声数据集作为实验数据。

(1) 人工多物种鸟声数据集(Arti)

Arti 是本实验中人工合成的多物种鸟声实验数据集,它的原始数据来自 Xeno-canto 网站上的鸟声数据。考虑到地域合理性和样本多寡,Arti 选用美国纽约常见的 4 种鸟类(比氏夜鸫、紫朱雀、白翅交嘴雀和黄腹吸汁啄木鸟)的鸣叫音频作为单物种鸟声训练样本,并通过对单物种鸟声进行随机混合(如图 4),构造 1 000 段多物种鸟声作为测试样本。

(2) 天然多物种鸟声数据集(Natu)

Natu 是本实验中的天然多物种鸟声实验数据集,它的原始数据来自 NIPS4B 鸟声数据库。NIPS4B 是“生物的神神经信息处理”挑战赛(Neural information processing scaled for bioacoustics)公开的

鸟声数据库,其中收录了来自法国普罗旺斯的将近 700 段单物种和多物种鸟声数据。Natu 提取其中的 245 段单物种的鸟声作为训练样本,并提取与单物种鸟声存在相同物种的 51 段多物种的鸟声作为测试样本。

2.1.2 对比算法

为验证 FT-RMBS 的有效性,实验中将基于 MFCC 和 GMM 的鸟声识别算法作为对比算法,但为了使对比算法也具有多标记分类能力,实验中为对比算法增加了 RankSVM 分类器,并用 RMBS 表示对比算法。对比算法的组成结构见表 2。

表 2 对比算法组成

Tab. 2 Composition of algorithms

| 算法名称 | 特征映射算法 | 识别模型 | 多标记分类器 |
|---------|--------|------|---------|
| RMBS | | GMM | RankSVM |
| FT-RMBS | MMDE | GMM | RankSVM |

2.1.3 评价指标

为了能从不同方面验证算法的识别能力,试验基于一套局部评价指标和一套全局评价指标对算法的识别效果进行了评价。

(1) 局部评价指标

局部评价指标是指评价识别算法在某个单一物种类别上的识别能力的指标。本文采用以下 2 种局部评价指标评价算法在单一物种上的识别能力。

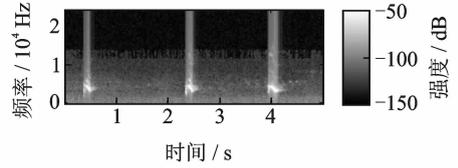
正确识别率:这是一种最典型的局部评价指标,它记录了某个类别上被识别到的正例样本占所有正例样本的比值,正确识别率越高说明预测函数的识别效果越好。以某一个类别为例,令 $E = \{(x, y) | y = 1\}$ 表示该类别的正例样本集,相应的 $\bar{E} = \{(x, y) | y = 0\}$ 表示负例样本集, h 表示该类别的预测函数。函数 h 的正确识别率为

$$R_{\text{true}}(h) = \frac{\sum_{x_i \in E} h(x_i)}{|E|} \quad (4)$$

式中: $| \cdot |$ 为集合中元素的个数。正确识别率只评价了预测函数识别正例样本的能力,但未考虑错误识别的情况,即负例样本也被识别为正例。因此本文提出一种称为错误识别率的指标。

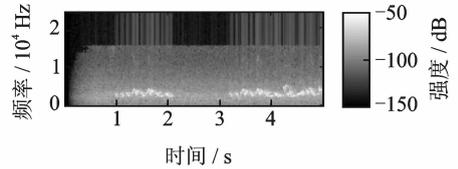
错误识别率:该指标记录了某个类别上被错误识别的负例占所有负例样本的比值,错误识别率越高说明预测函数的错误识别情况越严重。函数 h 的错误识别率为

$$R_{\text{false}}(h) = \frac{\sum_{x_i \in \bar{E}} h(x_i)}{|E|} \quad (5)$$



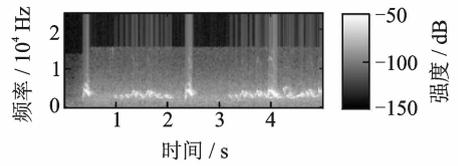
(a) 比氏夜鸫鸣叫声频谱图

(a) Audio spectrum of Bicknell's Thrush



(b) 紫朱雀鸣唱声频谱图

(b) Audio spectrum of Purple Finch



(c) 混合音频的频谱图

(c) Mixing audio spectrum

图 4 音频混合示例

Fig. 4 Audio mixing sample

(2)全局评价指标

全局评价指标用于评价识别算法在所有物种类别上的综合识别能力。由于多物种鸟声识别本质上是一个多标记分类问题^[13],因此本文采用4种目前被广泛采用的多标记学习评价指标:Hamming loss, Ranking loss, One error, Coverage从不同方面反映算法的损失程度,指标的值越小说明算法的识别错误率越低,上述4种指标的详细计算过程可以参考文献[13]。

2.2 算法在单个物种上的识别率比较

本文首先分析两种对比算法在Arti实验数据集上的识别效果。实验在Windows 7操作系统下采用Matlab R2012a编程实现算法。对比算法在Arti数据集上各运行10次,分别计算2个对比算法在各物种上的识别率平均值,结果如表3。其中最高的正确识别率和错误识别率被加粗显示。

表3 各物种识别率
Tab. 3 Recognition rate of each specie %

| 物种名称 | 正确识别率 | | 错误识别率 | |
|---------|-------|--------------|--------------|--------------|
| | RMBS | FT-RMBS | RMBS | FT-RMBS |
| 比氏夜鸫 | 34.48 | 53.56 | 21.59 | 47.96 |
| 绿头鸭 | 19.30 | 19.59 | 0.00 | 0.00 |
| 紫朱雀 | 54.59 | 61.48 | 24.67 | 24.01 |
| 白翅交嘴雀 | 45.12 | 64.42 | 13.68 | 54.91 |
| 黄腹吸汁啄木鸟 | 70.57 | 73.07 | 43.91 | 47.58 |

由表3可知:(1)FT-RMBS相较于对比算法在各物种上的正确识别率均有提升。其中FT-RMBS对“比氏夜鸫”和“白翅交嘴雀”的正确识别率相较于RMBS提升了20%以上。仅从正确识别率上来看经过特征映射后的识别算法有更好的识别效果。(2)FT-RMBS在个别物种上的错误识别率比RMBS算法高,其中“比氏夜鸫”和“白翅交嘴雀”的错误识别率比RMBS高20%以上,说明FT-RMBS在提升了正确识别率的同时错误识别了部分本不属于该物种的样本。2个对比算法在“绿头鸭”上的错误识别率为0,说明该物种没有出现任何一例错误识别的情况,但同时该物种的正确识别率也不到20%,FT-RMBS算法对于该物种的识别率也仅提升了不到0.3%,识别效果不突出。而在“紫朱雀”这个物种上,FT-RMBS的正确识别率提升了7%,同时错误识别率降低了0.6%,识别效果较好。

综上,本文提出的FT-RMBS在提升了正确识别率的同时可能会出现错误识别的情况。为了进一步考量算法的综合识别精度,需要对算法的综合识别精度进行评价。

2.3 算法综合识别精度比较

为比较FT-RMBS和RMBS的综合识别效果,本次实验选用Natu实验数据集,基于4项多标记学习评价指标比较2种算法对多物种鸟声识别的精度。算法在原始MFCC(以下简称MFCC)和经过差分变换的完整MFCC(以下简称 Δ MFCC)下的各项指标数值见表4,其中每个指标的最小值被加粗表示。

由表4可知:(1)在基于原始MFCC特征的鸟声识别中,FT-RMBS的Ranking loss, One error和

表 4 算法综合识别精度

Tab. 4 Algorithm synthesis recognition accuracy

| 评价指标 | MFCC | | Δ MFCC | |
|--------------|----------------|-----------------|----------------|-----------------|
| | RMBS | FT-RMBS | RMBS | FT-RMBS |
| Hamming loss | 0.075 2 | 0.090 4 | 0.170 5 | 0.102 4 |
| Ranking loss | 0.400 6 | 0.388 6 | 0.374 8 | 0.333 5 |
| One error | 0.843 1 | 0.823 5 | 0.725 5 | 0.862 7 |
| Coverage | 20.960 8 | 18.725 5 | 17.647 1 | 16.568 6 |

Coverage 相较于 RMBS 都有不同程度的降低,其中 Coverage 降低了 2.2,而 Ranking loss 和 One error 降低 0.02。但新算法 Hamming loss 轻微升高了 0.015。综合来看 4 个综合评价指标中有 3 个得到了提升,说明本文提出的 FT-RMBS 在原始 MFCC 特征上相比于 RMBS 算法有更高的识别精度。

(2)在基于完整 MFCC 特征的鸟声识别中,FT-RMBS 相较于 RMBS 在 One error 上升高了 0.14,但 Hamming loss,Ranking loss 和 Coverage 上分别降低了 0.07,0.04 和 1.1。与原始 MFCC 的识别结果类似,实验中 4 个综合评价指标有 3 个得到了提升,说明在完整 MFCC 特征下 FT-RMBS 相较于 RMBS 同样具有更高的识别精度。

综上,基于特征迁移的 FT-RMBS 算法相较于直接采用原始音频特征进行识别的 RMBS 算法有更高的识别精度。

3 结束语

为解决多物种鸟声识别中训练样本不足的问题,本文尝试采用单物种鸟声样本训练多物种鸟声识别模型,并提出了 FT-RMBS,该方法通过引入 MMDE 算法提取单物种鸟声与多物种鸟声的同分布潜在音频特征,从而使单物种鸟声训练的识别模型也能够用于识别多物种鸟声。在多物种鸟声识别实验中,FT-RMBS 算法相较于对比算法最高提升了 20% 的识别率。新算法在多物种鸟声识别问题中有更高的精度,为多物种鸟声识别中样本不足的问题提供了一种解决方案。然而本文提出的识别算法是针对没有背景噪声的理想情况,并没有考虑实际应用中噪音的干扰,也没有进一步对目标鸟类的鸣声进行特征选取和识别模型的优化。在实际应用中不仅需要音频进行降噪或增强等预处理,还需根据目标物种选取能够反映其鸣声特点的音频特征,并对识别模型进行优化。除了对识别算法进行优化外,鸟声识别还可以通过引入先验知识帮助提高识别率,例如同一段鸟声中出现的不同鸟类物种肯定处于相同的地域环境,并且具有类似的生活习性,这样才能保证它们在同一地点同时鸣叫。因此,引入鸟类的生活习性等先验知识或许能有助于提高识别模型的精度。这也是多物种鸟声识别技术未来发展的一个可行方向。

参考文献:

- [1] 孙斌, 万鹏威, 陶达, 等. 基于自适应最优核时频分布的鸟类识别[J]. 数据采集与处理, 2015, 30(6): 1187-1195.
Sun Bin, Wan Pengwei, Tao Da, et al. Identification of birds based on adaptive optimal kernel time-frequency distribution [J]. *Journal of Acquisition and Processing*, 2015, 30(6): 1187-1195.
- [2] 薛宇志. 依照鸟类鸣叫与鸣唱声识别其种类[D]. 台北: 台北科技大学, 2010.
Xue Yuzhi. Automatic identification of bird species by their calls and songs[D]. Taipei, China: Taipei University of Technology, 2010.

- [3] Cheng J, Sun Y, Ji L. 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.
- [4] 魏静明, 李应. 利用抗噪纹理特征的快速鸟鸣声识别[J]. 电子学报, 2015(1):185-190.
Wei Jingming, Li Ying. Rapid bird sound recognition using anti-noise texture features[J]. Acta Electronica Sinica, 2015(1): 185-190.
- [5] Stowell D, Plumbley M D. Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning[J]. Peer J, 2014, 2: 488.
- [6] Briggs F, Lakshminarayanan B, Neal L, et al. Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach[J]. Journal of the Acoustical Society of America, 2012, 131(6):4640-4650.
- [7] Briggs F, Raich R, Eftaxias K, et al. The ninth annual MLSP competition: Overview[C]//IEEE International Workshop on Machine Learning for Signal Processing. Southampton, United Kingdom: IEEE, 2013: 22-25.
- [8] Jancovic P, Kokuer M. Acoustic recognition of multiple bird species based on penalized maximum likelihood[J]. IEEE Signal Processing Letters, 2015, 22(10):1585-1589.
- [9] Pan S J, Kwok J T, Yang Q. Transfer learning via dimensionality reduction[C]//Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence. [S. l.]: IEEE, 2008:677-682.
- [10] 李燕萍, 唐振民, 钱博, 等. 基于 PLAR 特征补偿的鲁棒性说话人识别仿真研究[J]. 系统仿真学报, 2009, 21(2):409-412.
Li Yanping, Tang Zhenmin, Qian Bo, et al. Robust speaker recognition based on plar features compensation transformation and its simulation study[J]. Journal of System Simulation, 2009, 21(2):409-412.
- [11] Hewitt E, Hewitt R E. The Gibbs-Wilbraham phenomenon: An episode in fourier analysis[J]. Archive for History of Exact Sciences, 1979, 21(2):129-160.
- [12] Elisseeff A, Weston J. A kernel method for multi-labelled classification[C]//Advances in Neural Information Processing Systems. Cambridge, Massachusetts, USA: MIT Press, 2001: 681-687.
- [13] Zhang M L, Zhou Z H. A review on multi-label learning algorithms[J]. Knowledge and Data Engineering, IEEE Transactions on, 2014, 26(8): 1819-1837.

作者简介:



刘昊天(1990-),男,硕士研究生,研究方向:迁移学习、多标记学习和并行计算, E-mail: liuhaotian1990@foxmail.com。



姜海燕(1967-),女,博士,教授,研究方向:人工智能与机器学习、智能计算与并行计算和农业信息技术。



舒欣(1984-),通信作者,男,博士,研究方向:机器学习、计算机视觉。



徐彦(1979-),男,博士,研究方向:神经网络学习、神经计算。



伍艳莲(1970-),女,博士,研究方向:图像处理、计算机图形学。



郭小清(1976-),女,博士,研究方向:分布式计算、作物病害诊断。

