

基于子带谱特征的助听器背景噪声场景分类 算法



靳韡贇^{1,2}, 詹毅^{1,2*}, 樊晓华^{1,2,3}

(1. 中国科学院微电子研究所 北京 朝阳区 100029; 2. 中国科学院大学集成电路学院 北京 石景山区 100049;
3. 江苏集萃智能集成电路设计技术研究所有限公司 江苏 无锡 214115)

【摘要】针对助听器应用中背景噪声场景分类算法需同时具备低延时性和高分类准确率的问题,提出一种基于 LightGBM 集成学习模型的助听器场景分类算法以减少分类过程的计算时间,给出一种新的子带谱相关性特征并联合子带谱熵特征构成分类特征来提高助听器场景分类的准确率,使用双耳差分信号提取子带谱特征减少计算过程中的内存占用率以及模型离线训练工作量,提高计算效率。对双耳助听器声学环境识别数据集中的安静室内、交通环境、风噪声、音乐、鸡尾酒会、汽车噪声 6 种场景下的背景声音进行测试,实验结果表明,相对于基于随机森林模型和子带信号周期性特征、子带信号熵特征的场景分类算法,该算法在实时性和分类准确率方面的性能均有显著改善。

关键词 双耳差分信号; 子带谱特征; 背景噪声分类; 集成学习
中图分类号 TN911.7 文献标志码 A doi:10.12178/1001-0548.2021249

Background Noise Classification Algorithm for Hearing Aids Based on the Band Spectral Features

JIN Weiyun^{1,2}, ZHAN Yi^{1,2*}, and FAN Xiaohua^{1,2,3}

(1. Institute of Microelectronics of Chinese Academy of Sciences Chaoyang Beijing 100029;
2. School of Integrated Circuits, University of Chinese Academy of Sciences Shijingshan Beijing 100049;
3. Jiangsu JSIC Inc. Wuxi Jiangsu 214115)

Abstract Aiming at the challenge of real-time implementation and high classification accuracy for hearing aids, a background noise classification algorithm based on the LightGBM ensemble learning is proposed to reduce the computational time in the process. A newly proposed band spectral correlation feature concatenated with the band spectral entropy feature is also presented. This new acoustic feature is formed to improve the noise classification accuracy. Binaural differential signal is used to extract the band spectral features for reducing memory occupation and offline training workload, so as to improve the computational efficiency. Six commonly encountered noise environments of quiet indoors, traffic, wind turbulence, music, cocktail party and vehicle noise from hearing aid research dataset for acoustic environment recognition are considered. The experimental results show that our proposed algorithm significantly improves the performance of the background noise classification in real-time implementation and the accuracy compared with the algorithms based on the random forest model and band features.

Key words binaural differential signal; band spectral features; background noise classification; ensemble learning

助听器使用者的使用环境通常不是单一固定的环境,环境的变换会导致声音场景的变化,不同的声音场景会给助听器带来不同影响从而导致助听器性能偏差,通过调整和改变不同环境下助听器的参数可以改善助听器的性能。因此在使用助听器的过

程中,助听器需要持续检测使用者当前所属的环境,通过对环境中的背景噪声进行分类和识别,来调整和选取合适的参数及算法,以提高助听器产品的整体性能^[1]。噪声场景分类问题的本质是模式识别,主要由特征提取和分类两个过程组成,噪声场

收稿日期: 2021-09-17; 修回日期: 2022-05-14

作者简介: 靳韡贇(1991-),女,博士生,主要从事智能音频信号处理方面的研究。

*通信作者: 詹毅, E-mail: yizhan@ime.ac.cn

景分类采用的特征主要包括时域特征、频域特征以及倒谱域特征等, 分类过程中使用的模型主要包括 K 近邻 (K-nearest neighbor, KNN) 模型^[2], 高斯混合模型 (Gaussian mixed model, GMM)^[3-5]、隐马尔科夫模型^[6](hidden Markov model, HMM)、人工神经网络模型^[7-8](artificial neural network, ANN)、支持向量机^[9-11](support vector machine, SVM) 等。文献 [12] 基于随机森林集成学习算法和子带特征进行背景噪声场景识别, 在满足系统实时性要求的同时实现了高分类准确率。目前的噪声场景分类过程中使用的音频信号特征大部分是单通道音频信号特征, 而文献 [13-14] 使用双通道音频信号特征可以有效地进行音频场景分类, 其中包含从双通道差分信号中提取出的特征。2020 年, 德国听力系统能力中心提出了一个双耳助听器声学环境识别数据集, 适用于助听器的环境分类与识别需求, 并且基于深度神经网络验证了所提供数据集的有效性和可分离性^[15]。

针对双耳佩戴数字助听器接收到的双通道环境声音信号相对于单通道声音信号包含更多的环境声音信息, 同时, 助听器场景分类算法需具备实时性和高分类准确率, 提出基于 LightGBM 集成学习算法实现助听器的背景噪声场景分类, 并使用基于双耳差分信号子带谱联合特征进行信号表征, 充分利用不同环境中双耳信号差异信息完成背景噪声场景的识别。

1 模型构建

1.1 子带谱特征提取

不同种类的噪声场景信号在频域中不同的频带范围内具有不同的分布特性, 如白噪声是功率谱密度在整个频域内均匀分布的噪声, 所有频率具有相同的能量密度, 而粉红噪声的功率谱密度则与频率成反比。因此, 使用信号子带谱特征可以更全面地表达信号在各个频带上所具有的特性, 信号子带谱特征提取过程如图 1 所示。其中, N 表示频带划分数目; m 表示子带特征种类数目。在提取过程中, 首先对输入音频信号进行分帧加窗等预处理, 然后对信号进行傅里叶变换得到相应的频谱信号, 将频谱信号划分为 N 个互不重叠的等带宽子带后, 分别对每个子带进行 m 种类的特征提取, 最后将不同子带提取出的子带谱特征进行特征融合, 得到用于场景分类的信号特征。

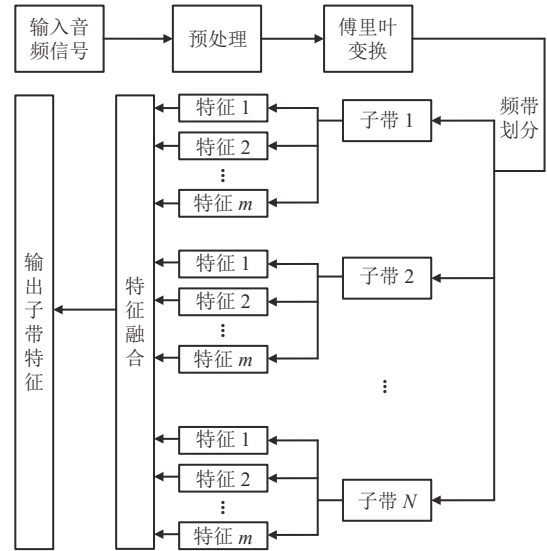


图 1 信号子带特征提取过程

1.2 LightGBM 模型

采用 LightGBM (light gradient boosting machine, LightGBM) 集成学习算法模型进行场景分类与识别, LightGBM 是一种基于决策树算法的梯度提升集成学习框架, 由于 LightGBM 基于梯度的单侧采样 (gradient based one side sampling, GOSS) 移除了梯度较小的数据实例, 保留了在信息增益的计算中起着更重要作用的梯度较大的数据实例, 同时利用特征捆绑方法 (exclusive feature bundling, EFB) 捆绑互斥的特征, 所以模型可以在较小的数据量下获得准确的信息增益估计并且降低模型分裂过程中的复杂度, 减少样本和特征数量, 具有训练速度快和内存占用率低的特点^[16]。图 2 为基于 LightGBM 的背景噪声分类框架, 对于多种背景噪声的场景识别系统, 分为模型训练和场景识别两个过程, 在模型训练阶段, 首先对背景噪声音频信号进行特征提取, 构建特征数据集。然后使用数据集中的数据对 LightGBM 模型进行训练。在场景识别阶段, 对需要分类与识别的音频信号提取相应的特征, 并且使用训练好的 LightGBM 模型进行场景分类与识别。

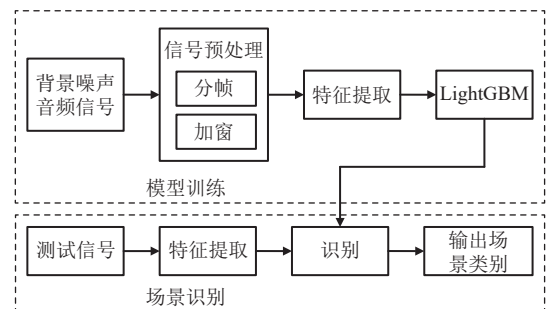


图 2 基于 LightGBM 的背景噪声分类框架

2 双耳信号子带谱特征

子带特征能较好地描述音频信号频域的局部特性, 由于不同种类噪声的音频信号频谱特性具有差异性, 因此提取噪声信号子带特征可以反映信号在不同频域范围内的细节特性。图 3 给出了 Noisex-92 噪音数据集中 3 种噪声信号 babble 噪声、volvo 噪声和 white 噪声的语谱图, 由图中可以看出不同种类噪声在不同频率范围内的频谱分布具有明显差异。

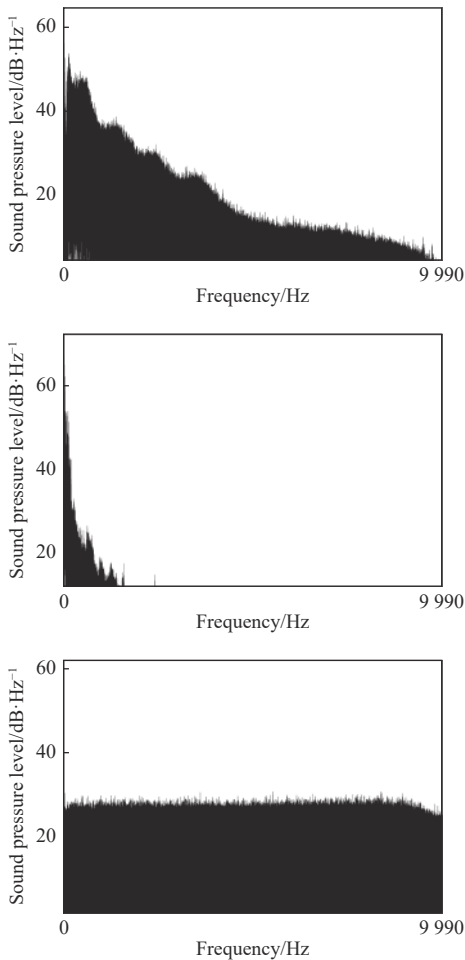


图 3 噪声信号语谱图

文献 [12] 已证明使用信号子带周期特征和信号子带熵特征可以有效地对背景噪声场景进行分类。信号子带周期特征可以根据信号每个子带中的周期性特征来区分不同场景的背景噪声, 这个特征可以反映信号中平稳的音频特性, 能够有效识别音乐信号。信号子带熵特征由不同子带的能量熵构成, 可以反映信号中非平稳的音频特性, 二者结合可以有效地对不同场景的音频信号进行表征从而实现场景分类。

为了充分利用音频信号频谱信息, 本文给出一种基于频谱子带信号的子带谱相关性特征, 并且结合子带谱熵特征形成联合特征来进行助听器的场景识别过程。其中, 子带谱相关性特征可以反映信号不同频率分量之间的相关程度, 而子带谱熵特征可以反映信号在频域内不同频率范围内的波动特性。

2.1 信号子带谱相关性特征

信号频谱相邻子带的相关性 (spectral correlation, SC) 使用归一化相关函数来计算。一帧音频信号频谱两个相邻频带之间的归一化相关函数为:

$$\text{Corr}_{\text{fr}}(b) = \frac{\sum_{l=1}^L |F_b(l)| |F_{b+1}(l)|}{\sqrt{\sum_{l=1}^L |F_b(l)|^2 \sum_{l=1}^L |F_{b+1}(l)|^2}} \quad (1)$$

式中, $\text{Corr}_{\text{fr}}(b)$ 表示一帧信号两个相邻子带谱之间的归一化相关函数; b 表示频带索引; fr 表示帧索引; $F(\cdot)$ 为输入信号的离散傅里叶变换; $F_b(\cdot)$ 和 $F_{b+1}(\cdot)$ 表示傅里叶变换后两个连续频带对应的子带信号; $|\cdot|$ 表示对应幅值; L 表示每个频带内所包含的频点数目; l 表示每个频带内的频点索引。噪声信号子带谱相关性特征计算如下:

$$\text{BC}_b = \frac{1}{N_f} \sum_{\text{fr}=1}^{N_f} \text{Corr}_{\text{fr}}(b) \quad (2)$$

式中, N_f 表示音频信号中包含的总帧数。信号子带谱相关性特征与信号子带周期特征的计算过程虽然都是基于归一化自相关函数, 但计算子带谱相关性特征不需要遍历所有延时点数并求取最大值。因此, 可以有效减少特征计算过程中的运算量, 降低计算时间, 对于帧长为 FL , 均匀划分为 N 个子带的音频信号, 假设傅里叶变换为 FL 点, 则每个频谱子带包含的频点数目为 $L = \text{FL}/N$, 计算一帧信号两个相邻频带间的相关性特征只需要计算一次 L 点的归一化自相关函数, N 个子带需要计算 $N-1$ 个谱相关性特征, 所以对于帧长 FL 的一帧信号只需要计算 $N-1$ 次 L 点的自相关函数。而对于信号子带周期性特征, 每个子带信号需要计算 FL 次 FL 点的归一化自相关函数并寻找最大值, 因此对于帧长 FL 的一帧信号, N 个子带信号需要计算 $N \times \text{FL}$ 次 FL 点的归一化自相关函数, 并且每个子带信号都需要找到归一化自相关函数的最大值。因此, 相

对于子带周期特征, 子带谱相关性特征可以有效地减少特征提取时间, 提高计算效率。

2.2 信号子带谱熵特征

谱熵 (spectral entropy, SE) 特征可以分析信号的功率谱和熵率之间的关系。熵特征是对随机进行试验不确定性的一种度量, 事件概率分布的熵越大, 试验可能出现的结果确定性越小。子带谱熵特征提供了噪声信号每个子带谱的熵度量, 即:

$$BE_b = \frac{1}{N_f} \sum_{fr=1}^{N_f} H(fr) \quad (3)$$

式中, $H(fr)$ 表示第 fr 帧信号的子带谱熵特征。计算过程如下:

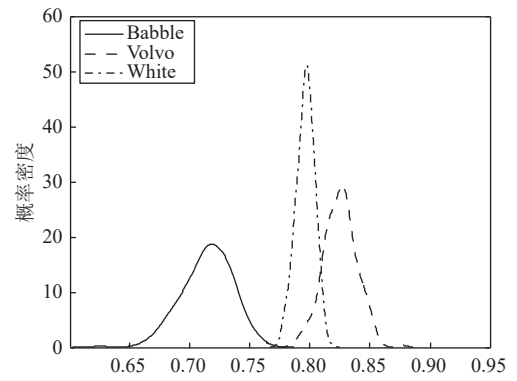
$$P_b(l) = \frac{|F_b(l)|^2}{\sum_{l=1}^L |F_b(l)|^2} \quad (4)$$

$$\text{band_entropy}(b) = - \sum_{l=1}^L P_b(l) \log_2 P_b(l) \quad (5)$$

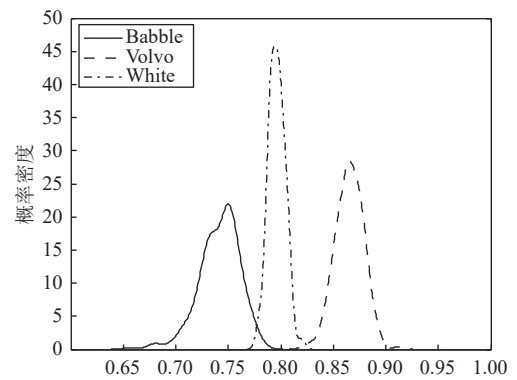
式中, $P_b(l)$ 表示相对功率谱概率, 最后对 band_entropy 进行归一化得到子带谱熵特征:

$$H(fr) = \frac{\text{band_entropy}(b)}{\log_2 L} \quad (6)$$

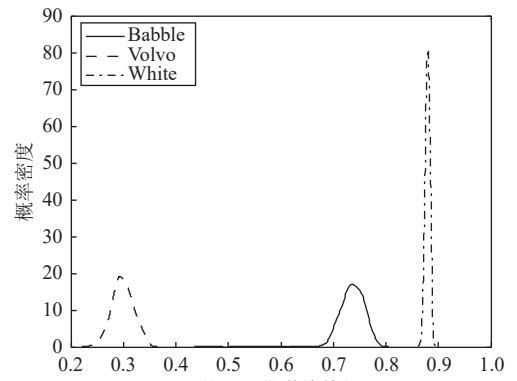
音频信号子带谱熵特征可以反映出信号在频域子带内的稳定特性。在频域均匀划分为 8 个子带时, Noisex-92 噪声数据集中 babble、volvo、white 这 3 类噪声频域子带间谱相关性特征和子带谱熵特征的概率密度差异如图 4 所示。不同场景中的声信号特征概率密度曲线分布具有明显的差别, 因此可以通过子带谱相关性特征和子带谱熵特征对声音场景信号进行分类。



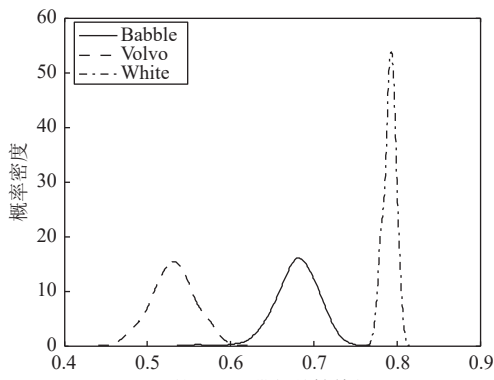
第 2, 3 子带相关性特征
b. 2, 3 子带相关性特征概率密度曲线



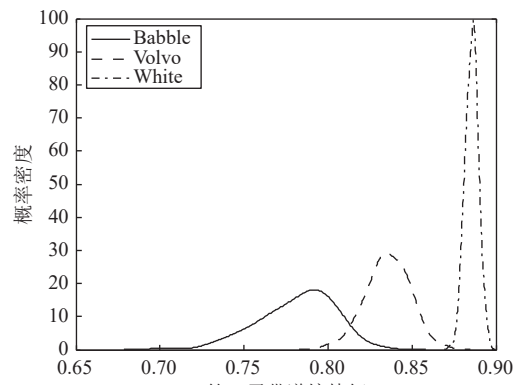
第 3, 4 子带相关性特征
c. 3, 4 子带相关性特征概率密度曲线



第 1 子带谱熵特征
d. 第 1 子带谱熵特征概率密度曲线



第 1, 2 子带相关性特征
a. 1, 2 子带相关性特征概率密度曲线



第 2 子带谱熵特征
e. 第 2 子带谱熵特征概率密度曲线

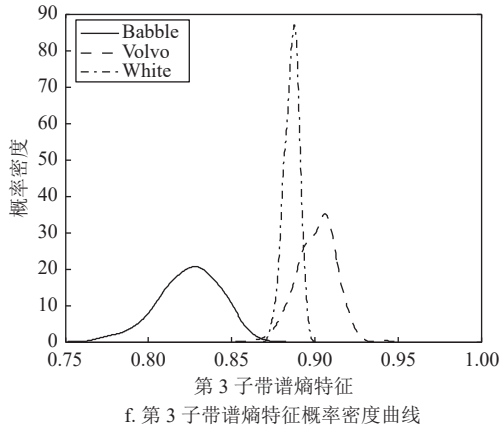


图4 3类噪声频域子带间谱相关性特征和子带谱熵特征的概率密度曲线

2.3 基于双耳信号子带特征

在双耳佩戴助听器时，双耳信号往往包含更多的环境声音信息，因此，对双耳接收到的声音信息进行整合与利用也是非常重要的一个环节。假设助听器左耳通道接收到的声音信号是 S_L ，右耳通道接收到的声音信号是 S_R ，对左右耳接收到的声音信号分别进行子带特征提取，假设提取出的子带特征维数为 d ，共包含 $d1$ 维子带谱相关性特征和 $d2$ 维子带谱熵特征。提取出的左耳接收信号子带特征表示为：

$$\mathbf{BF}_{\text{left}} = [\mathbf{BC}_{\text{bl}}(1), \mathbf{BC}_{\text{bl}}(2), \dots, \mathbf{BC}_{\text{bl}}(d1), \mathbf{BE}_{\text{bl}}(1), \mathbf{BE}_{\text{bl}}(2), \dots, \mathbf{BE}_{\text{bl}}(d2)] \quad (7)$$

右耳接收信号子带特征为：

$$\mathbf{BF}_{\text{right}} = [\mathbf{BC}_{\text{br}}(1), \mathbf{BC}_{\text{br}}(2), \dots, \mathbf{BC}_{\text{br}}(d1), \mathbf{BE}_{\text{br}}(1), \mathbf{BE}_{\text{br}}(2), \dots, \mathbf{BE}_{\text{br}}(d2)] \quad (8)$$

1) 双耳信号联合子带特征

文献 [17] 通过双耳信号特征互联的方式构成场景分类特征，基于双耳信号的联合子带特征同时保留左右耳接收信号子带特征的完整信息，并且对 $\mathbf{BF}_{\text{left}}$ 和 $\mathbf{BF}_{\text{right}}$ 按顺序依次进行拼接，联合特征的长度为单声道子带特征长度的 2 倍，因此 $2d$ 维双耳

信号联合子带特征表示为：

$$\mathbf{BF}_{\text{joint}} = [\mathbf{BF}_{\text{left}}, \mathbf{BF}_{\text{right}}] = [\mathbf{BC}_{\text{bl}}(1), \dots, \mathbf{BC}_{\text{bl}}(d1), \mathbf{BE}_{\text{bl}}(1), \dots, \mathbf{BE}_{\text{bl}}(d2), \mathbf{BC}_{\text{br}}(1), \dots, \mathbf{BC}_{\text{br}}(d1), \mathbf{BE}_{\text{br}}(1), \dots, \mathbf{BE}_{\text{br}}(d2)] \quad (9)$$

2) 双耳信号均值子带特征

文献 [15] 分别对双耳信号提取特征后，采用取算术平均值的方法构成助听器的场景分类特征，均值特征可以消除单个信号特征的误差给信号特征表示所带来的影响，表示为：

$$\mathbf{BF}_{\text{mean}} = \frac{\mathbf{BF}_{\text{left}} + \mathbf{BF}_{\text{right}}}{2} \quad (10)$$

3) 双耳差分信号子带特征

左右耳通道时域差分信号表示为：

$$S_{\text{Diff}} = S_L - S_R \quad (11)$$

或：

$$S_{\text{Diff}} = S_R - S_L \quad (12)$$

两者只存在相位差异，幅值相同，文中统一使用式 (11) 的差分信号形式作为左右耳通道时域差分信号。对左右耳差分信号进行子带特征提取，时域差分信号子带特征表示为：

$$\mathbf{BF}_{\text{diff}} = [\mathbf{BC}_{\text{bd}}(1), \mathbf{BC}_{\text{bd}}(2), \dots, \mathbf{BC}_{\text{bd}}(d1), \mathbf{BE}_{\text{bd}}(1), \mathbf{BE}_{\text{bd}}(2), \dots, \mathbf{BE}_{\text{bd}}(d2)] \quad (13)$$

双耳差分信号子带特征主要通过左右耳声道接收到的信号之间的差异来对场景特征进行表征，不需要分别提取左右耳接收到的信号特征。表 1 给出了使用不同层面双耳助听器信息在内存效率、计算效率和离线训练工作量方面的对比。其中决策层面信息结合表示左右耳助听器进行场景识别后，对双耳场景识别结果进行判断与决策。对比结果显示差分信号特征在内存占用率、计算效率以及离线工作量方面均表现优异。助听器设备对存储以及计算资源有一定限制，因此减少资源消耗有利于助听器信号处理过程中的算法与模型部署。

表 1 内存效率、计算效率、离线训练工作量对比

算法	内存效率	计算效率	离线训练工作量
决策层面信息结合 ^[17]	$2 \times d$ 输入特征 $2 \times$ 分类器参数	$2 \times$ 特征提取器 $2 \times$ 分类器计算	$2 \times$ 分类器训练
双耳信号联合特征 ^[17]	$1 \times 2d$ 输入特征 $1 \times$ 分类器参数	$2 \times$ 特征提取器 $1 \times$ 分类器计算	$1 \times$ 分类器训练
双耳信号均值特征 ^[15]	$1 \times d$ 输入特征 $1 \times$ 分类器参数	$2 \times$ 特征提取器 $1 \times$ 分类器计算	$1 \times$ 分类器训练
双耳差分信号特征	$1 \times d$ 输入特征 $1 \times$ 分类器参数	$1 \times$ 特征提取器 $1 \times$ 分类器计算	$1 \times$ 分类器训练

3 实验仿真和结果分析

3.1 实验设置

1) 双耳助听器声学环境识别数据集

实验数据来自德国听力系统能力中心给出的双耳助听器声学环境识别数据集，选取常见的安静室内、交通环境、风噪声、音乐、鸡尾酒会、汽车噪声场景中的背景声音信号，每一组背景声音信号分

别包含左右耳两个通道的音频数据信号, 信号采样率为 16 000 Hz, 每个信号片段持续时间为 10 s, 总计 4 241 组双耳接收音频信号, 共 8 482 个音频数据片段。对单个音频信号片段进行预加重、分帧和加窗的预处理, 帧长 25 ms, 帧重叠为 0, 对信号进行傅里叶变换, 并且划分为 8 个子带信号, 取 8 个子带中相邻子带的谱相关性特征和前 4 个子带的频带谱熵特征构成子带信号特征。分别对左耳接收信号、右耳所接收信号以及双耳差分信号进行子带特征提取, 并且构成基于单声道的信号子带特征数据集和基于双耳信号的均值特征数据集、联合子带特征数据集以及差分信号子带特征数据集。数据集中 80% 的数据用来训练 LightGBM 模型, 剩下 20% 的数据用于对训练好的模型进行测试。

2) 模型设置

实验仿真过程中基于随机森林的分类模型与基于 LightGBM 的分类模型均使用 50 个子估计器进行实验。并且, 在进行模型训练与测试前先对数据集进行缺失值与异常值过滤预处理去掉数据集中的异常特征向量。

3.2 仿真结果与分析

基于 LightGBM 模型进行单声道信号子带特征场景分类, 并与文献 [12] 中使用的子带特征与分类模型实验结果进行对比, 特征提取过程使用一组滤波器对输入音频信号进行滤波, 分别得到不同频率范围内的子带信号, 提取子带信号的周期性特征与熵特征, 选取前 6 个子带的周期性特征和前 4 个子带的熵特征构成场景分类特征。表 2 给出了使用子带周期性特征和子带熵特征时, 基于随机森林模型和基于 LightGBM 模型在整个单声道信号特征数

数据集上进行场景分类得到的测试集分类准确率与运行时间 (包括训练和预测过程) 对比。LightGBM 模型相对于 RF 准确率可以提高约 0.53%, 模型训练和预测时间可以减少约 40%。因此, 基于 LightGBM 模型进行助听器的背景噪声分类在维持场景分类准确率的情况下可以提高信号处理的实时性。

表 2 背景噪声场景分类结果对比

参数	RF ^[12]	LightGBM
准确率/%	87.93	88.46
运行时间/s	0.897 5	0.532 1

表 3 给出了基于 LightGBM 算法对双耳信号子带特征进行场景分类, 对比子带周期与子带熵联合特征和在单声道特征提取数据集上使用信号子带谱联合特征在安静室内、交通环境、风噪声、音乐、鸡尾酒会、汽车噪声场景下以及整个测试集上的分类准确率。表 4 给出了基于子带谱联合特征使用双耳信号联合特征、双耳信号均值特征以及双耳差分信号子带特征进行分类的实验结果, 实验结果表明, 相对于采用子带周期与子带熵特征, 使用信号子带谱联合特征可以有效提高助听器的场景分类准确率, 测试集上的分类准确率可以提升约 9%。此外, 使用信号子带谱联合特征进行场景分类在 6 种背景噪声环境中分类准确率均有显著提升。使用双耳差分信号子带特征进行场景分类与联合特征和均值特征相比, 在部分场景中分类准确率有所下降, 如风噪声和汽车噪声场景, 而在部分场景中有所提升, 如音乐和鸡尾酒会场景, 但是, 在整个测试数据集上基本保持不变。

表 3 信号特征分类准确率对比

特征类别	安静室内	交通环境	风噪声	音乐	鸡尾酒会	汽车噪声	测试集
子带周期子带熵联合特征 ^[12]	82.24	77.34	83.01	96.03	88.63	89.22	88.47
子带谱联合特征	98.61	94.94	94.62	98.88	97.94	97.51	97.55

表 4 双耳信号特征分类准确率对比

特征联合方式	安静室内	交通环境	风噪声	音乐	鸡尾酒会	汽车噪声	测试集
双耳信号均值 ^[17]	97.61	96.78	95.12	98.05	97.12	98.65	97.44
双耳信号联合 ^[15]	97.82	96.83	95.76	98.00	97.58	98.63	97.59
双耳差分信号	96.48	96.26	92.98	99.67	97.80	95.80	97.35

为了验证基于 LightGBM 与子带谱联合特征声场景分类算法的有效性和普适性, 根据日常听觉场景调查^[2]给出的 20 种人们日常接触的环境声 (分别

是车站、公交车内、汽车内、马路上、卧室、办公室、会议室、教室、酒店内、酒吧、餐厅、电影院、超市、集市、公园内、体育场馆、工地、田野、

山林和车间)以及助听器常见噪声风声,构建了一个包含 8 种声音类别的环境声数据集,音频数据来自 NOISEX-92 噪声数据集、NOIZEUS 语音增强数据库^[18]、ESC-50 环境声数据集^[19]、UrbanSound8K 城市环境声分类公共数据集、AISHELL-2 中文语音数据库^[20],除 NOISEX-92 中数据进行了数据切分,其余数据均保持原时间长度和原始采样率。数

据集中的数据保留了原始数据的多样性,具有不同的采样率和数据长度。在验证过程中,数据集中 80% 的数据特征用来训练模型,剩下 20% 的数据特征用于对训练好的模型进行测试。对 8 种常见环境声音的分类结果如表 5 所示。实验结果显示基于 LightGBM 与子带谱联合特征声场景分类算法在日常生活环境声音分类中也表现良好。

表 5 8 种常见环境声音的分类结果

环境声音	空调发动机	嘈杂人声	汽车鸣笛声	纯净语音	工厂车间	火车	车内噪声	风声	测试集
准确率	100.00	98.75	75.56	96.87	99.00	84.31	98.75	77.69	92.03

4 结束语

针对助听器应用中背景噪声场景分类算法需同时具备低延时性和高分类准确率的问题,提出一种基于 LightGBM 集成学习模型的助听器场景分类算法以减少分类过程中的计算时间,给出一种新的子带谱相关性特征并且联合子带谱熵特征构成分类特征来提高助听器场景分类的准确率,使用双耳差分信号提取子带谱特征减少计算过程中的内存占用率以及模型离线训练工作量,提高计算效率。实验结果表明,与随机森林模型相比,基于 LightGBM 算法的场景分类可以在维持算法准确率的情况下减少约 40% 的程序运行时间,使用子带谱相关性特征联合子带谱熵特征进行场景分类可以进一步提高场景分类的准确率,与子带周期和子带熵特征相比,场景分类准确率在整个测试集上可以提高约 9%。通过对 8 种常见环境声分类,结果显示了算法具有一定的鲁棒性。与双耳信号均值子带特征以及双耳信号联合子带特征相比,采用双耳差分信号子带特征进行场景分类可以在维持高分类准确率的条件下减少内存与计算资源的占用。因此,基于 LightGBM 和双耳差分信号子带谱联合特征的场景分类算法更适用于对实时性、准确率要求高的应用场景。本文工作对助听器场景分类等相关研究具有意义,但研究工作还缺少实际数据的验证,未来将在此基础上做进一步的研究和开发,考虑基于 FPGA 平台通过硬件测试算法的有效性和实时性。

参 考 文 献

[1] 梁瑞宇,王青云,邹采荣.数字助听器原理及核心技术[M].北京:电子工业出版社,2018.
LIANG R Y, WANG Q Y, ZOU C R. Digital hearing aid theory and key technology[M]. Beijing: Publishing House

of Electronics Industry, 2018.
[2] 陈克安.环境声的听觉感知与自动识别[M].北京:科学出版社,2014.
CHEN K A. Auditory perception and automatic recognition of environmental sound[M]. Beijing: Science Press, 2014.
[3] ALAVI Z, AZIMI B. Application of environment noise classification towards sound recognition for cochlear implant users[C]//The 6th International Conference on Electrical and Electronics Engineering. Istanbul: IEEE, 2019: 144-148.
[4] CHU S, NARAYANAN S, KUO C C J. Environmental sound recognition with time-frequency audio features[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2009, 17(6): 1142-1158.
[5] LOZANO H, HERNÁEZ I, PICÓN A, et al. Audio classification techniques in home environments for elderly/dependant people[C]//International Conference on Computers for Handicapped Persons. Berlin, Heidelberg: Springer, 2010: 320-323.
[6] BÜCHLER M, ALLEGRO S, LAUNER S, et al. Sound classification in hearing aids inspired by auditory scene analysis[J]. *EURASIP Journal on Advances in Signal Processing*, 2005, 2005(18): 1-12.
[7] BARKANA B D, SARICICEK I. Environmental noise source classification using neural networks[C]//The 7th International Conference on Information Technology. [S. l.]: IEEE, 2010: 259-263.
[8] TANWEER S, MOBIN A, ALAM A. Environmental noise classification using LDA, QDA and ANN methods[J]. *Indian Journal of Science and Technology*, 2016, 9(33): 1-8.
[9] LI Y, LI Y. Eco-environmental sound classification based on matching pursuit and support vector machine[C]//The 2nd International Conference on Information Engineering and Computer Science. Wuhan: IEEE, 2010: 1-4.
[10] WEI P, HE F, LI L, et al. Research on sound classification based on SVM[J]. *Neural Computing and Applications*, 2020, 32(6): 1593-1607.
[11] WANG J C, WANG J F, HE K W, et al. Environmental sound classification using hybrid SVM/KNN classifier and MPEG-7 audio low-level descriptor[C]//The 2006 IEEE

- International Joint Conference on Neural Network Proceedings. Vancouver: IEEE, 2006: 1731-1735.
- [12] SAKI F, KEHTARNAVAZ N. Background noise classification using random forest tree classifier for cochlear implant applications[C]//The 2014 IEEE International Conference on Acoustics, Speech and Signal Processing. Florence: IEEE, 2014: 3591-3595.
- [13] EGHBAL-ZADEH H, LEHNER B, DORFER M, et al. CP-JKU submissions for DCASE-2016: A hybrid approach using binaural i-vectors and deep convolutional neural networks[J]. IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events, 2016, 6: 5024-5028.
- [14] HAN Y, PARK J, LEE K. Convolutional neural networks with binaural representations and background subtraction for acoustic scene classification[C]//Detection and Classification of Acoustic Scenes and Events. Munich: [s. n.], 2017: 46-50.
- [15] HÜWEL A, ADILOĞLU K, BACH J H. Hearing aid research data set for acoustic environment recognition [C]//The 2020 IEEE International Conference on Acoustics, Speech and Signal Processing. Barcelona: IEEE, 2020: 706-710.
- [16] KE G, MENG Q, FINLEY T, et al. Lightgbm: A highly efficient gradient boosting decision tree[J]. Advances in Neural Information Processing Systems, 2017, 30: 3146-3154.
- [17] MIRZAHASANLOO T, KEHTARNAVAZ N. Real-time dual-microphone noise classification for environment-adaptive pipelines of cochlear implants[C]//The 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Osaka: IEEE, 2013: 5287-5290.
- [18] HU Y. Subjective evaluation and comparison of speech enhancement algorithms[J]. *Speech Communication*, 2007, 49: 588-601.
- [19] PICZAK K J. ESC: Dataset for environmental sound classification[C]//Proceedings of the 23rd ACM International Conference on Multimedia. Brisbane: ACM, 2015: 1015-1018.
- [20] DU J, NA X, LIU X, et al. Aishell-2: Transforming mandarin asr research into industrial scale[EB/OL]. [2021-01-20]. <https://arxiv.org/pdf/1808.10583.pdf>.

编辑 叶芳