

基于子带谱熵法和 PSO-GA-SVM 的汽车鸣笛识别

余凌浩¹, 陆铁文¹, 李 晨¹, 曾毓敏¹, 袁 芳²

(1. 南京师范大学计算机与电子信息学院/人工智能学院, 江苏 南京 210023)

(2. 杭州爱华智能科技有限公司, 浙江 杭州 311121)

[摘要] 针对鸣笛抓拍系统会产生误判的问题, 提出了一种基于子带谱熵法和支持向量机的汽车鸣笛识别算法. 首先, 使用子带谱熵法对声音样本进行初判, 将子带谱熵高于阈值的样本直接判定为非鸣笛样本. 然后, 对初判为鸣笛的样本中的疑似鸣笛部分进行分割, 并提取 Mel 频率倒谱系数作为声音的特征. 最后, 使用支持向量机对分割结果进行进一步分类, 并使用粒子群算法与遗传算法的融合来优化支持向量机的参数. 仿真结果表明, 该算法具有较好的鲁棒性. 在对实际采集样本的鸣笛识别中, 该算法也取得了较高的准确率.

[关键词] 汽车鸣笛识别, 子带谱熵法, 支持向量机, 粒子群优化, 遗传算法

[中图分类号] TN912.34 **[文献标志码]** A **[文章编号]** 1672-1292(2021)02-0027-07

Car Whistle Recognition Based on Sub-Band Spectral Entropy Method and PSO-GA-SVM

Yu Linghao¹, Lu Tiewen¹, Li Chen¹, Zeng Yumin¹, Yuan Fang²

(1. School of Computer and Electronic Information/School of Artificial Intelligence, Nanjing Normal University, Nanjing 210023, China)

(2. Hangzhou Aihua Intelligent Technology Co., Ltd., Hangzhou 311121, China)

Abstract: In order to solve the problem of misjudgment in car whistle capture system, an algorithm based on sub-band spectral entropy and support vector machine is proposed in this paper. Firstly, sub-band spectral entropy method is used to preliminarily judge the sound samples. The samples whose sub-band spectral entropies are higher than a threshold value are directly determined as non-whistle samples. Then, the suspected whistle parts are segmented in the samples which are initially judged as whistle, and Mel frequency cepstrum coefficient is extracted as sound feature. Finally, the segmentation results are classified by support vector machine, and the parameters of support vector machine are optimized by the combination of particle swarm optimization and genetic algorithm. Simulation results show that the algorithm has good robustness. In the whistle recognition of actual collected samples, the algorithm also achieves a higher accuracy.

Key words: car whistle recognition, sub-band spectral entropy method, support vector machine, particle swarm optimization, genetic algorithm

近年来, 交通噪声的关注度日益提升. 鸣笛噪声是交通噪声的主要组成成分, 汽车在禁鸣区违章鸣笛, 会严重影响他人的生活、学习与工作. 为了遏制违章鸣笛, 鸣笛抓拍系统应运而生. 鸣笛抓拍系统会实时检测当前的声压级, 当声压级超过一定阈值时, 系统通过远场波束形成技术定位声源的位置并进行音视频记录^[1]. 然而交通环境中声源情况非常复杂, 有很多非鸣笛声也能触发鸣笛抓拍系统的阈值, 只依靠声压级触发门限会出现较多误判的情况. 这也是目前鸣笛抓拍系统应用较少的主要原因. 因此, 对鸣笛抓拍系统采集的样本进行鸣笛与非鸣笛的分类是非常有必要的.

在汽车鸣笛识别方面, 目前已有许多学者进行了研究^[2-8], 较为常用的方法主要分为以下两类: (1) 阈值判断, 如文献[8]中使用的短时能量和短时过零率法等. 该方法将特征值高于或低于一定门限的声音样本直接进行分类, 其分类效果主要取决于特征和阈值的选取. (2) 机器学习, 如文献[2]中使用的卷积神经网络、文献[6]中使用的隐马尔科夫模型等. 该方法首先对样本提取特征, 通过训练一定数量的样本得到

收稿日期: 2020-09-23.

基金项目: 国家重点研发计划项目(2017YFB0503500)、江苏省自然科学基金资助项目(BK20171031).

通讯作者: 李晨, 博士, 讲师, 研究方向: 语音信号处理. E-mail: lichen@njnu.edu.cn

一个模型,再通过采集的样本是否与模型相匹配来进行分类。

进行阈值判断时,阈值的设置会直接影响分类结果的准确性,然而交通环境中声源情况非常复杂,会给阈值设置带来较大阻碍。机器学习虽然准确率较高,但对模型有比较高的要求。如果实测的鸣笛样本中含有较多噪声,也会对分类准确率产生很大影响。因此,我们需要结合实际交通环境中的鸣笛噪声,研究具有较高准确率的鸣笛识别算法。

由于鸣笛声具有丰富的谐波特征,其子带谱熵值相对较低。相比较于传统的短时能量和短时过零率特征,子带谱熵具有更高的区分度^[9-10],阈值的设置具有普适性。机器学习主要分为传统机器学习和深度学习两大类。深度学习非常适合处理大数据下的特征,但是在较小的数据量下,深度学习并没有表现出很好的效果,并且由于其计算量过大,很难满足实时性的需求。而支持向量机(support vector machine, SVM)作为传统机器学习的方法,在进行小样本集分类时效果较好^[11]。

SVM 参数优化的传统方法是网格搜索法,该方法通过遍历参数的所有组合来选取最优结果,运行速度较慢。我们使用粒子群优化(particle swarm optimization, PSO)和遗传算法(genetic algorithm, GA)的融合(PSO-GA)来优化 SVM 的参数^[12-13],在准确率没有降低的情况下,提高了选取最优参数的速度。

本文结合了阈值判断和机器学习的方法,提出了一种基于子带谱熵法和 PSO-GA-SVM 的汽车鸣笛识别算法。仿真结果表明,该算法具有较好的鲁棒性。在对实际采集样本的鸣笛识别中,本文算法也取得了较高的准确率。

1 算法介绍

1.1 算法总流程

本文首先使用子带谱熵法对鸣笛抓拍系统采集的声音样本进行初判,将子带谱熵高于阈值的样本直接判定为非鸣笛样本。为了使 SVM 训练出较好的模型,在子带谱熵法中再结合子带谱线能量门限,对初判为鸣笛的样本中的疑似鸣笛部分进行分割。然后使用 SVM 对分割结果进行进一步分类,并使用 PSO-GA 来优化 SVM 的参数。算法总流程如图 1 所示。

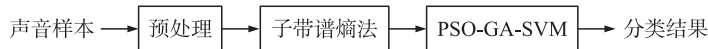


图 1 算法总流程

Fig. 1 Overall algorithm process

1.2 预处理

文中预处理的主要流程为:带通滤波、预加重、分帧、加窗。

一般汽车鸣笛的频率分布在 1 500~6 000 Hz。为了保留汽车鸣笛的主要特征并去除噪声的干扰,对声音样本进行 1 500~6 000 Hz 的带通滤波。

预加重是一个一阶高通滤波器,用一阶 FIR 滤波器表示为:

$$s'(n) = s(n) - as(n-1). \quad (1)$$

式中, a 为一常数, $s(n)$ 为原始信号, $s'(n)$ 为预加重后的信号。预加重可以对信号进行高频提升,使频谱变得更平坦,减少频谱的动态范围。一般来说, $0.9 < a < 1$ 。本文中 a 取 0.937 5。

分帧可以将一段较长声音信号分成多个小段,使每小段声音信号可以作为平稳信号来处理。为了使每一帧信号之间平滑地过渡,两个相邻帧之间需要有一些重叠区域。本文中帧长选为 50 ms,帧移长度选为帧长的 1/3。

加窗可以避免因信号两端不连续而导致的频谱泄漏。本文选用汉明窗作为窗函数 $w(n)$,若共有 N 个采样点,则第 n 个采样点的窗函数公式为:

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), 0 \leq n \leq N-1. \quad (2)$$

1.3 子带谱熵法

1.3.1 算法介绍

子带谱熵的思想是将每一帧的频谱分成若干个子带,这样就消除了每一条谱线受噪声影响的问题。

题. 设每个子带由 4 条谱线组成,共有 N_b 个子带. 第 i 帧中的第 n 个子带的谱线能量 $E(n,i)$ 为:

$$E(n,i) = \sum_{k=4(n-1)}^{4(n-1)+3} Y_i(k), 1 \leq n \leq N_b. \quad (3)$$

式中, $Y_i(k)$ 为傅里叶变换后的谱线能量. 子带谱线能量的概率分布 $p(n,i)$ 为:

$$p(n,i) = \frac{E(n,i)}{\sum_{k=1}^{N_b} E(n,i)}, 1 \leq n \leq N_b. \quad (4)$$

引入一个正常量 C , 得到新的子带谱线能量的概率分布式:

$$p_b(n,i) = \frac{E(n,i) + C}{\sum_{k=1}^{N_b} (E(n,i) + C)}, 1 \leq n \leq N_b. \quad (5)$$

文献[9]中推导证明,在噪声环境下,引入正常量 C 后,分割的准确度能得到提高. 本文中 C 取 0.1. 子带谱熵 $H_b(i)$ 为:

$$H_b(i) = - \sum_{n=1}^{N_b} p_b(n,i) \ln p_b(n,i). \quad (6)$$

1.3.2 算法流程

子带谱熵法的主要流程如图 2 所示. 在信号的子带谱线能量包络线上选取一个较高的阈值 eth2 , 包络线与 eth2 相交于左右两个点 A 和 B . 认定高于 eth2 的部分一定是鸣笛, 而鸣笛的起止点应该在 AB 段之外. 再在信号子带谱线能量包络线上选取一个较低的阈值 eth1 , 从 A 点向左、从 B 点向右搜索, 分别找到包络线与 eth1 相交的左右两个点 C 和 D , 如图 3(a) 所示. CD 段即为基于子带谱线能量门限分割的鸣笛部分.

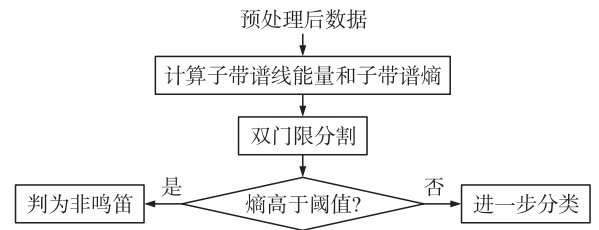


图 2 子带谱熵法流程

Fig. 2 Sub-band spectral entropy method process

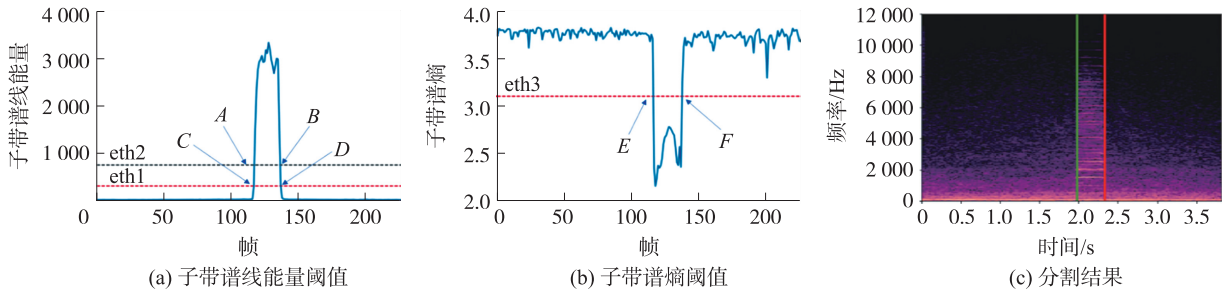


图 3 双门限分割示意图

Fig. 3 Schematic diagram of double threshold segmentation

在对应的子带谱熵包络线上,从 C 点向右、从 D 点向左进行更加细致的分割,找到子带谱熵低于阈值 eth3 的左右两个点 E 和 F ,如图 3(b) 所示. EF 段即为基于子带谱线能量和子带谱熵双门限分割的鸣笛部分. 图 3(c) 为声音样本的频谱图以及分割结果,左边的竖线为分割的起始点,右边的竖线为分割的停止点.

由于汽车鸣笛声具有丰富的谐波特征,其子带谱熵值相对较低,子带谱熵特征的区分度更高. 为了验证子带谱熵法的有效性,我们使用文献[8]的短时能量和短时过零率法与本文的子带谱熵法进行了对比实验. 实验结果如图 4 和图 5 所示.

图 4 的样本中没有鸣笛声,正确的结果应没有分割线,如图 4(a) 所示. 文献[8]的方法产生了分割线,而本文的方法没有产生分割线,并将其分类为非鸣笛样本. 图 5 的样本中含有鸣笛和噪声,正确的分割结果如图 5(a) 所示. 文献[8]的方法不仅分割出了鸣笛部分,还分割出了多个噪声片段. 而本文的方法则准确地分割出了鸣笛声,没有分割出噪声部分.

使用子带谱熵法可以将子带谱熵高于阈值的样本直接判为非鸣笛样本,如图 4(c) 所示. 然而在实际

交通环境中也存在一些非鸣笛声(如汽车转向提示声、公交车报站声、警笛声等),由于具有一定的谐波特征,产生了分割结果,会被子带谱熵法误判为鸣笛声,如图 6 所示. 因此,我们需要对子带谱熵法的分割结果进行进一步的分类,来消除此类误判.

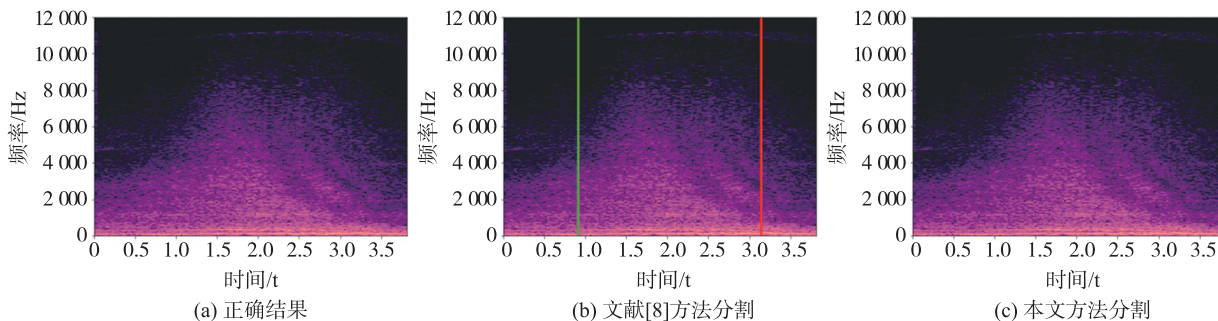


图 4 对一段非鸣笛声音样本的分割

Fig. 4 Segmentation of a non-whistle sound samples

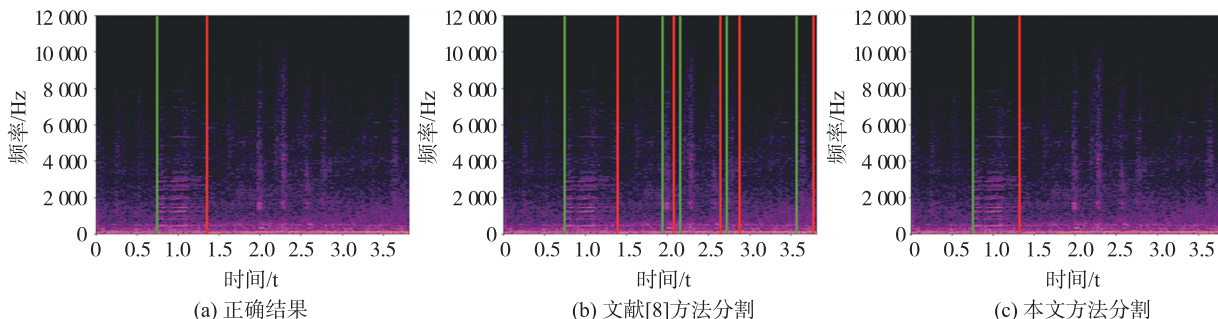


图 5 对一段含有鸣笛和噪声的声音样本的分割

Fig. 5 Segmentation of a sound samples with car whistle and noise

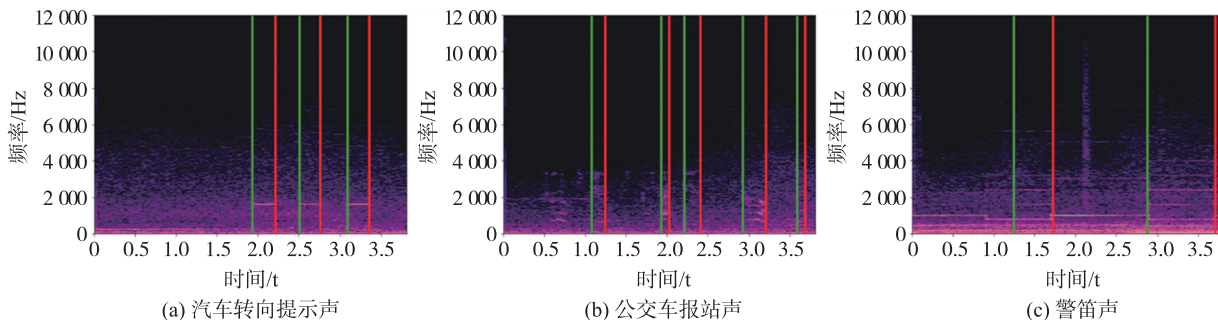


图 6 对部分非鸣笛样本的分割

Fig. 6 Segmentation of some non-whistle sound samples

1.4 PSO-GA-SVM

1.4.1 算法介绍

SVM 是一种二分类模型. 设样本集为 (x_i, y_i) , 其中 x_i 为样本特征, y_i 为类别符号. SVM 的目标是寻找一个分类面 $\omega^T x + b = 0$ 将样本正确地分成两类, 且使 $\|\omega\|^2$ 最小. 在样本集线性不可分时, 需要引入惩罚因子 c 和核函数 $K(x_i, x_j)$ 来进行分析. 核函数可以将低维向量内积的结果转化为高维向量内积的结果, 但不需要在高维中进行计算. 本文选用的核函数为径向基核函数,

$$K(x_i, x_j) = e^{-g \|x_i - x_j\|^2}. \quad (7)$$

惩罚因子 c 和核函数参数 g 对 SVM 的分类性能有很大的影响. 本文使用 PSO-GA 来优化 c 和 g 参数.

PSO 将粒子 i 在搜索空间中的飞行速度表示为向量 v_i , 位置表示为向量 x_i . 在每一次迭代中, 计算每个粒子的适应度, 通过个体极值 p_{best} 和全局极值 g_{best} 来更新每一个粒子的速度和位置, 如下两式所示:

$$v_i = K[v_i + \varphi_1 r(p_{best} - x_i) + \varphi_2 r(g_{best} - x_i)], \quad (8)$$

$$\mathbf{x}_i = \mathbf{x}_i + \mathbf{v}_i. \quad (9)$$

式(8)中, r 为均匀分布在 $(0,1)$ 之间的随机数, K 由下式表示^[14]:

$$K = \frac{2}{|2 - \varphi - \sqrt{\varphi^2 - 4\varphi}|}, \varphi = \varphi_1 + \varphi_2, \varphi > 4. \quad (10)$$

本文中 φ 取 4.1, 则 K 取 0.730.

GA 将搜索空间中每一个可能的解编码为染色体, 随机选择一组染色体来形成初始群体. 在每一次迭代中, 计算每个粒子的适应度, 按照适应度大小的顺序来选择个体, 进行交叉、变异等操作. 反复执行遗传操作, 直到满足终止条件为止.

1.4.2 算法流程

PSO-GA-SVM 主要流程如图 7 所示, 其主要步骤如下:

步骤 1 对训练集进行 k 折交叉验证. 本文中 k 取 5.

步骤 2 将搜索空间中每一个 c 和 g 可能的解编码为染色体, 随机选择 N 个染色体来形成初始种群, 并初始化粒子速度. 计算每一个粒子的适应度, 求出 p_{best} 和 g_{best} . 本文中 N 取 20.

步骤 3 按照式(8)和(9)更新每一个粒子的速度和位置.

步骤 4 按照适应度选出 $N-1$ 个粒子, 对它们进行交叉、变异操作, 再与上一代的最优粒子组合成新的 N 个粒子. 计算每一个粒子的适应度, 更新 p_{best} 和 g_{best} .

步骤 5 转步骤 3, 直到满足一定迭代次数为止. 本文中迭代次数取 100.

步骤 6 输出最优 c 和 g .

步骤 7 将最优 c 和 g 代入 SVM 的 c 和 g 参数中, 对测试集进行分类.

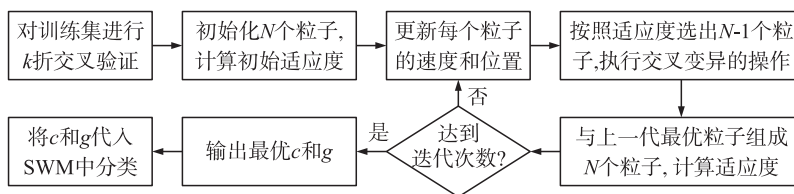


图 7 PSO-GA-SVM 流程

Fig. 7 PSO-GA-SVM process

2 实验及结果分析

本文使用 CPU 为 Intel i7-8750H、内存为 8G 的计算机作为硬件平台. 本文中所有音频的格式为 wav 格式, 位深度为 32 位. 所有程序均在 Python 3.7.0 下编写. 每次实验从数据集中随机选取 80% 的样本作为训练集, 剩下 20% 的样本作为测试集, 计算 100 次实验的平均准确率. 本文选用 Mel 频率倒谱系数作为分类器输入的声音特征^[15], 其中 Mel 滤波器的个数选为 96 个, 离散余弦变换后取第 2~65 维, 再对每一帧信号求平均, 最终形成一个 64 维的向量.

2.1 仿真实验

本节中对不同底噪、不同信噪比的声音样本进行仿真实验. 鸣笛声取自杭州爱华智能科技有限公司提供的纯净鸣笛声样本, 包含奥迪 A3、大众高尔夫、吉利等 13 种不同车型的鸣笛声, 共 120 个. 非鸣笛声取自 UrbanSound8K 数据集中的城市声音样本, 包含空调外机声、小孩玩耍声、警笛声等 8 种类型的声音, 共 211 个. 本节中音频的时长为 1 s, 采样频率为 48 kHz. 为了模拟鸣笛抓拍系统的采集方式, 每个声音样本中只有 0.25 s 存在待识别的声音, 并且随机设置该声音在音频中的起始位置. 为了验证本文算法的鲁棒性, 在上述声音样本的基础上, 分别添加信噪比为 20 dB、10 dB、5 dB、0 dB 的白噪声和粉噪声作为底噪进行仿真实验. 白噪声和粉噪声取自 NOISE92 数据集. 仿真实验结果如表 1 所示.

分析表 1 可知, 在信噪比较高的情况下, 本文

表 1 不同底噪不同信噪比下的平均准确率

Table 1 Average accuracy with different SNR under different noise

信噪比/dB	20	10	5	0
白噪声下的准确率/%	100.0	99.9	87.8	75.6
粉噪声下的准确率/%	100.0	99.9	99.7	86.5

算法对仿真样本的识别准确率可以达到 99% 以上. 随着信噪比不断降低, 本文算法的识别效果虽然也有所下降, 但依然维持在相对较高的准确率上. 在极差的信噪比环境下(信噪比为 0 dB 时), 本文算法依然能够有 75% 以上的准确率. 说明本文算法具有较好的鲁棒性.

2.2 结果分析

本节中对杭州爱华智能科技有限公司的鸣笛抓拍系统实际采集的 1 067 个声音样本进行鸣笛识别, 其中包括 683 个不含有鸣笛声的样本, 包含卡车发动机声、汽车转向提示声、公交车报站声、警笛声等超过 10 种城市环境音. 包括 384 个含有鸣笛声的样本, 包含各种车型的鸣笛声, 背景噪声中也存在多种上述城市环境音. 本节中音频的时长为 3.8 s, 采样频率为 24 kHz. 我们选择文献[2]算法、文献[7]算法与本文算法进行了对比实验. 实验结果如表 2 所示.

从表 2 可见出, 本文算法相比文献[2]算法准确率提高了 21.3%, 相比文献[7]算法准确率提高了 14.4%. 文献[2]和文献[7]只使用深度学习的方法对声音样本进行分类, 而本文首先使用子带谱熵法对声音样本进行初判和分割, 再使用分类器对分割结果进行进一步分类. 同时, 文献[2]和文献[7]中深度学习的方法在样本数量较少时分类效果并不理想, 而本文使用的 PSO-GA-SVM 更加适合小样本集的分类. 因此, 本文算法在汽车鸣笛识别上具有更高的准确率.

我们使用子带谱熵法结合文献[7]使用的反向传播神经网络(backpropagation neural network, BPNN)和本文使用的 PSO-GA-SVM 进一步进行了对比试验. 实验结果如表 3 所示.

子带谱熵法对非鸣笛样本的误判率在 1% 以内, 但是对鸣笛样本的误判率在 34% 以上, 不能满足鸣笛抓拍系统对准确率的要求, 因此需要对子带谱熵法初判为鸣笛的样本作进一步的分类. 由表 3 可知, 使用子带谱熵法前, BPNN 和 PSO-GA-SVM 的准确率并不是非常理想, 而使用子带谱熵法后, 两种分类器的准确率均有了较大提高. 这是因为子带谱熵法可以分割出疑似鸣笛的部分, 大幅减少了噪声对分类器模型造成的影响. 同时从表 3 中还可以看出, 本文使用的 PSO-GA-SVM 相比较于 BPNN 具有更高的准确率, 更加适合小样本集的分类.

此外, 由于本文使用的声音特征维数较低, 仅有 64 维, 提取特征以及分类的速度较快, 在本文使用的计算机上处理一个样本仅需约 0.3 s, 可以很好地应用到实时检测中.

3 结论

针对现有的汽车鸣笛识别算法易受噪声影响、计算量较大的问题, 本文提出如下算法: 首先利用阈值判断对声音样本进行初判, 使用子带谱熵法分割出样本中疑似的鸣笛部分; 然后利用机器学习对疑似鸣笛样本作进一步分类, 使用 PSO-GA-SVM 训练模型, 最终分类出所有的鸣笛与非鸣笛样本. 实验表明, 本文算法准确率较高且计算量较小, 可以满足鸣笛抓拍系统实时检测的要求.

[参考文献] (References)

- [1] 袁芳, 闫建伟, 张勇, 等. 汽车鸣笛声实时抓拍的理论研究和系统实现[J]. 电声技术, 2018, 42(11): 13-15, 21.
- [2] 侍艳华, 刘菁原, 卞飞, 等. 基于 MFCC 和 CNN 的汽车鸣笛声识别算法[J]. 电声技术, 2020, 44(5): 30-33.
- [3] 郑皓, 赵庶旭, 屈睿涛. 一种用于城市交通的优化声音识别仿真[J]. 计算机技术与发展, 2019, 29(2): 60-64.
- [4] 祁兵兵, 刘金磊, 张焕胜, 等. 智能交通系统中声源端点识别算法[J]. 电声技术, 2018, 42(10): 8-13, 68.
- [5] 项彬, 马石磊, 陈建峰, 等. 一种基于支持向量机的汽车鸣笛分类方法; 201811405638.8[P]. 2019-04-12.
- [6] 蒋翠清, 邵宏波. 基于 MFCC 与改进 ACF 的汽车声音识别算法研究[J]. 计算机技术与发展, 2015, 25(2): 140-143.
- [7] 张铁成. 基于麦克风阵列的声音识别与定位算法研究[D]. 南京: 南京师范大学, 2016.

- [8] 刘建平,张一闻,刘颖. 基于麦克风阵列的汽车笛语识别及笛声定位方法[J]. 西安电子科技大学学报,2012,39(1): 163-167.
- [9] 王琳,李成荣. 一种基于自适应谱熵的端点检测改进方法[J]. 计算机仿真,2010,27(12):373-375,395.
- [10] 张毅,王可佳,席兵,等. 基于子带能熵比的语音端点检测算法[J]. 计算机科学,2017,44(5):304-307,319.
- [11] 刘灵,孙晨晨,徐寅林,等. 基于图像预处理分类分割的盲道分割算法[J]. 南京师范大学学报(工程技术版),2020, 20(1):42-48.
- [12] SETTLES M, SOULE T. Breeding swarms: a GA/PSO hybrid[C]//Proceedings of the 7th Annual Conference on Genetic and Evolutionary Computation Conference, CECCO 2005. Washington DC, USA: ACM, 2005:161-168.
- [13] ALOMARI F, LIU G H. Novel hybrid soft computing pattern recognition system SVM-GAPSO for classification of eight different hand motions[J]. Optik, 2015, 126(23):4757-4762.
- [14] CLERC M, KENNEDY J. The particle swarm—explosion, stability and convergence in a multidimensional complex space[J]. IEEE Transactions on Evolutionary Computation, 2002, 6(1):58-73.
- [15] 宋知用. MATLAB 语音信号分析与合成[M]. 2 版. 北京:北京航空航天大学出版社,2017.

[责任编辑:陈 庆]