

基于CNN和LSTM融合特征提取的车内声品质评价模型研究

杨礼强¹, 王攀¹, 王杰²

(1. 重庆大学机械与运载工程学院, 重庆 400044; 2. 作业帮教育科技有限公司(北京)有限公司, 北京 100085)

摘要: 基于深度学习方法建立的车内声品质评价模型不需要高度依赖声学理论和经验知识, 可以有效提取深层次特征, 客观高效地获得符合主观感受的评价结果。为获取噪声中符合人耳对声音感受的频率信息, 便于在深度学习中进行特征提取, 采用对数梅尔频谱和时频遮掩相结合的方法对采集到的噪声样本进行预处理。为有效提取车内噪声深层次特征, 融合卷积神经网络(Convolution Neural Network, CNN)和长短时记忆网络(Long Short-Term Memory Network, LSTM)各自的优点, 建立了融合特征提取层。使用全连接和Softmax输出单元组合构建了分类器模块。在合适的超参数下, 模型通过充足的训练获得了96.88%的训练准确度。使用大量样本对模型进行验证, 得到93.69%的验证准确度; 采用混淆矩阵对模型进一步验证, 总体的预测评价等级与真实评价等级偏差不大, 证明模型的预测结果与主观评价结果具有很好的一致性。

关键词: 车内声品质; 评价模型; 卷积神经网络; 长短时记忆网络

中图分类号: U467.1 文献标志码: A DOI: 10.3969/j.issn.2095-1469.2022.05.12

Research on Sound Quality Evaluation of Vehicle Interior Noise Based on Merged Features Extracted by CNN and LSTM

YANG Liqiang¹, WANG Pan¹, WANG Jie²

(1. College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing 400044, China;
2. Zuoyebang Education Technology (Beijing) Co., Ltd., Beijing 100085, China)

Abstract: The evaluation model of vehicle interior sound quality based on deep learning does not rely heavily on acoustic theory and empirical knowledge, can effectively extract deep-level features and efficiently obtains evaluation results in accordance with subjective feelings. Initially the collected noise samples was preprocessed by using a method combining logarithmic Mel spectrum and time-frequency masking to obtain the samples in the range of human frequency perception for feature extraction in deep learning. And then a fusion feature extraction layer combining CNN and LSTM was established to effectively extract the deep-level features of vehicle noise. A classifier module was built which consisted a fully connected layer and the Softmax output unit. With suitable hyperparameters, the model achieved a training accuracy of 96.88% after sufficient training. Finally a large number of test samples were used to verify the model and the verification accuracy was 93.69%. The confusion matrix was used for further validation. And there is little difference between the overall predicted evaluation levels and the real evaluation levels; i.e., the prediction results of vehicle interior sound quality are in good agreement with the subjective evaluation results.

收稿日期: 2022-05-23 改稿日期: 2022-07-28

基金项目: 重庆市技术创新与应用发展专项资助项目(cstc2019jscx-zdztzxX0024)

参考文献引用格式:

杨礼强, 王攀, 王杰. 基于CNN和LSTM融合特征提取的车内声品质评价模型研究[J]. 汽车工程学报, 2022, 12(5): 662-669.

YANG Liqiang, WANG Pan, WANG Jie. Research on Sound Quality Evaluation of Vehicle Interior Noise Based on Merged Features Extracted by CNN and LSTM[J]. Chinese Journal of Automotive Engineering, 2022, 12(5): 662-669. (in Chinese)



Keywords: vehicle interior sound quality; evaluation model; convolutional neural networks; long short-term memory network

声品质是一款汽车脱颖而出、吸引消费者的重要因素之一,其科学高效的评价是汽车噪声、振动与声振粗糙度(Noise、Vibration、Harshness, NVH)性能开发设计的关键。通过建立声品质客观评价模型来替代专家评审团做出符合人类主观感受的评价和分类,可以克服听音试验评价的缺点,有效节省成本和时间。

国内外研究者根据不同类型的汽车噪声,如车内噪声、汽车关门声、发动机噪声、汽车加速噪声等,从多元线性回归^[1-2]、支持向量机^[3-4]、BP神经网络^[2, 5-6]和小波神经网络^[7-9]等方法中选用一种来构建声品质评价模型,这些方法都用到了物理声学指标和客观心理声学参数,高度依赖于大量的、复杂的声学理论和经验知识。前3种方法通常在响度、尖锐度、粗糙度、波动度、A计权声压级、AI指数、主冲击时间、低频延续时间、抖动度、峰值频率、语音清晰度和言语干扰级等声学参数中,选用多个参数对噪声样本做预处理。基于小波神经网络法,有的研究者引用维格纳-威尔分布的时频分析方法,建立声品质参量SQP-RW,以此参量输入小波神经网络来预测汽车声品质;有的研究者为加快声品质评价模型的计算速度,使用噪声信号的能量、均值和标准差对响度、尖锐度、粗糙度、声调做出预测,采用这4个参数对噪声样本做预处理。

采用深度学习法建立车内声品质评价模型不需要高度依赖于复杂的声学理论和经验知识,还可以提取某些可能的未知深层次特征,使最终的声品质评价模型具有理想的预测准确度。首先使用对数梅尔频谱和时频遮掩相结合的方法对噪声样本做预处理;然后建立CNN和LSTM相融合的特征提取模块,以及使用全连接和Softmax输出单元组合搭建分类器模块;最后借助混合输入得到大量样本对所建立的评价模型进行训练,使其具备理想的精度。

1 车内噪声测试与主观评价

1.1 车内噪声数据采集

汽车行驶过程中会产生发动机噪声、轮胎噪声和风振噪声等,各种噪声经过不同的途径传递到车内,在较为封闭的空间内互相叠加和反射形成了车内噪声。车辆型号、车速以及乘坐位置等因素都会影响驾乘者坐在车内的声音舒适性,这些因素在设计车内噪声采集试验时起到了指导性作用。

为保证车内噪声能够被真实有效地记录到声音样本中,在进行噪声采集试验前对整车的各个系统进行了严格的检查,确保了各系统都处于正常工作状态且无异响。车内噪声采集试验场地是平滑干燥、无杂物、往来车辆少的硬地路面,周围没有高层建筑物,试验当天气温为16~22℃,沿测试路线于1.2 m高度测得风速为1.7~2.0 m/s,满足GB/T 18697—2002《声学-汽车车内噪声测量方法》规定的测试环境。依据国标对传声器的安装要求,将传声器安装在座椅头枕靠近驾驶员和后排乘员左右耳的位置,且传声器以最大灵敏度的方向水平指向行驶方向,调节驾驶员座椅的靠背,使其处于垂直位置。传声器在车内的安装位置如图1所示。

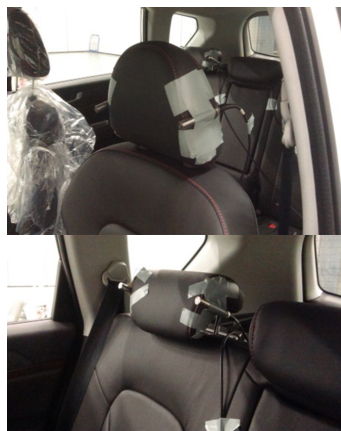


图1 传声器的车内安装位置

当车辆按照预设的速度匀速稳定行驶时开始采集车内噪声信号,且此时变速器挡位均处于最高挡

位。最终获得5辆不同品牌乘用车在60、80、100、120 km/h车速下的车内不同位置噪声样本。

1.2 车内噪声主观评价

主观评价试验组织了25位来自振动噪声领域的研究者进行听音试验,其年龄分布在22~45周岁之间,平均年龄为28岁。使用类别判断法作为主观评价方法,并对评审团进行声品质的知识培训以及正式试验前的听音训练。使用烦躁度作为评价试验的声品质指标,参考韩国现代公司提出的等级划分法,将评价指标由低到高分为很差(0~0.2)、差(0.2~0.4)、合格(0.4~0.6)、良好(0.6~0.8)和很好(0.8~1)5个等级供评审员选择。评价试验在具有良好隔声效果的实验室内进行,回放设备采用高保真解码器与某品牌高保真耳机组合,回放软件使用Simcenter Testlab软件下的Jury Testing模块。评分结束后使用皮尔逊相关分析法^[2]对评价分值进行检验,剔除相关系数低于0.6的3位评价者的主观评分,最终获得37个合格的噪声样本主观评价。

5类噪声样本的柱状图,如图2所示。由图可知,各类噪声样本的数量不一致,其中“很好”的噪声样本数量最少,为了平衡各类噪声样本的数量以及增加训练样本数量,对噪声样本的长度进行切割,每个训练样本长度为4 s。

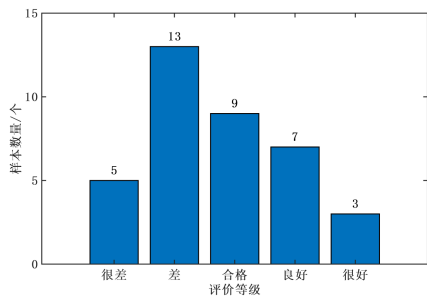


图2 各类噪声样本的数量

2 车内声品质评价模型的构建

基于CNN和LSTM融合特征提取的车内声品质评价模型的网络结构,如图3所示,由预处理层、CNN层、LSTM层和分类器组成,噪声样本首

先经过对数梅尔频谱和时频遮掩的预处理,其次进入标准卷积网络和空洞卷积网络,接着把得到的三维数组扁平展开成一维数组进入LSTM网络,之后进入分类器获得预测概率值序列,最终输出最大概率值所对应的噪声样本评价等级。

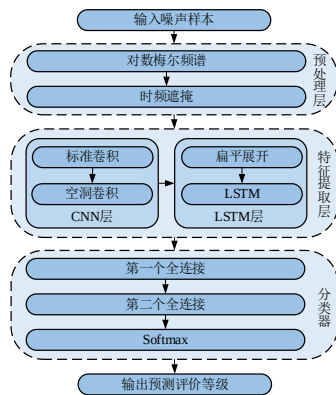
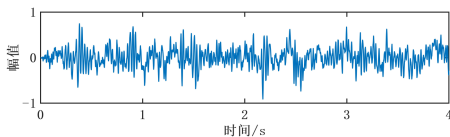


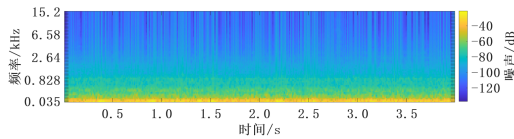
图3 车内声品质评价模型的网络结构

2.1 预处理层

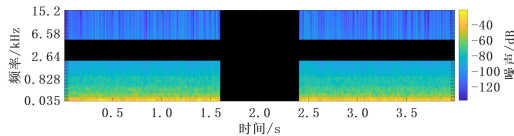
对噪声样本做预处理是为了把噪声样本的一维波形数据转换为高维数据,以及增强数据的特征,以助于CNN和LSTM提取其中深层次特征。本文采用对数梅尔频谱和时频遮掩相结合的方法对噪声样本做预处理,如图4所示。



(a) 噪声样本波形图



(b) 噪声样本的对数梅尔频谱



(c) 时频遮掩后的对数梅尔频谱

图4 车内声品质评价模型的预处理过程

2.1.1 对数梅尔频谱

梅尔频率倒谱系数^[10] (Mel Frequency Cepstrum Coefficient, MFCC) 能够获取音频数据中符合人耳

对声音感受的频率信息,有助于在深度学习中进行卷积操作和特征提取。MFCC的提取主要包括梅尔滤波和倒谱分析,后者由于删除了信息和破坏空间信息,不适用于深度学习。省略倒谱分析之后便得到对数梅尔频谱,其提取流程为:输入噪声样本,先进行预加重、分帧和加窗,然后做短时傅里叶变换得到功率谱,之后使用梅尔滤波器滤波,再取对数便得到对数梅尔频谱,图4b即为对数梅尔频谱。梅尔频率与物理频率的转换公式如式(1)所示。

$$f_m = 2595 \lg \left(1 + \frac{f}{700} \right) \quad (1)$$

式中: f_m 为梅尔频率, Hz; f 为物理频率, Hz。

2.1.2 时频遮掩

时频遮掩是一种数据增强的方法,通过故意使用受损数据来防止模型过拟合,有助于训练出更简单的网络,加快网络的收敛速度^[11]。它包括频率遮掩和时间遮掩两个部分,具体操作是把对数梅尔频谱指定的时间区间、频率区间内的数值变为零值,如图4c所示。

2.2 CNN层

CNN是深度学习的一类网络结构,多用于图像、语音和视频等的分类和识别,它通常由标准卷积或空洞卷积、激活和池化构成,有时为了防止模型出现过拟合和加快训练速度,在卷积和激活之间会使用批量归一化,如图5所示。



图5 CNN层的网络结构

2.2.1 标准卷积与空洞卷积

卷积的过程是以时频遮掩后的对数梅尔频谱为输入,将卷积核在频谱上扫描,累加对应项相乘得到输出^[12],如式(2)所示。

$$y(t) = \int_{-\infty}^{+\infty} x(p)h(t-p)dp = x(t) \times h(t) \quad (2)$$

式中: $x(t)$ 为输入; $h(t)$ 为卷积核; $y(t)$ 为输出; p 为步长。

空洞卷积在标准卷积中添加零值来扩大卷积核尺寸,使其获得更大的感受野,从而更好地提取特征。空洞卷积的尺寸由膨胀系数(Dilation Factor, DF)来调整,膨胀系数为在卷积核相邻参数中填充(DF-1)个0。可将标准卷积看作是膨胀系数为1的特殊空洞卷积^[13]。空洞卷积的输出定义为:

$$y(i, j) = \sum_{h=1}^H \sum_{w=1}^W x(i + DF \times h, j + DF \times w) \times w(h, w) \quad (3)$$

式中: H 、 W 分别为输入特征图的长和宽; $x(i, j)$ 为该特征图上 (i, j) 位置的特征值; DF 为膨胀系数; $y(i, j)$ 为该特征图经过空洞卷积后的输出。本文的CNN采用一个标准卷积和一个膨胀系数为2的空洞卷积的组合。

2.2.2 批量归一化、激活函数和平均池化

批量归一化是数据预处理中的常用操作,可以统一各特征值的量纲,加快梯度的下降速度,从而缩短寻找最优值的时间^[14]。

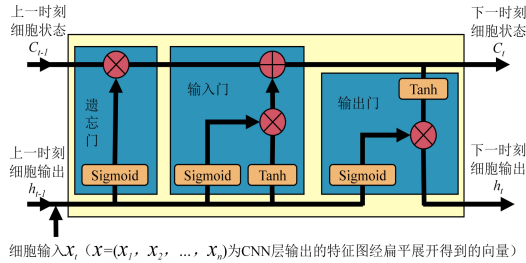
激活函数的作用是给网络引入非线性特性,增强模型的泛化能力。常用的激活函数有Sigmoid函数、Tanh函数和ReLU函数,本文模型中CNN和全连接的激活函数均使用ReLU函数,长短时记忆网络的细胞状态激活函数使用Tanh函数,门激活函数使用Sigmoid函数^[12]。

池化是对上一个特征图进行一次数据过滤以减少网络参数,包括平均池化、最大池化和全局池化,由于平均池化可以保留较多的信息,使提取的特征更完整,所以标准卷积和空洞卷积都采用平均池化,其运算过程为:首先,设置特征图上的窗口尺寸和步长,然后在特征图周围添加零值,这一步在深度学习中称为“padding”,接着计算特征图窗口内所有数值的平均值,这个平均值作为下一个特征图的数值,窗口从左到右、从上至下按预设的步长滑动,直至历遍整个特征图。

2.3 LSTM层

长短时记忆网络属于深度学习中循环神经网络(Recurrent Neural Network, RNN)的一种网络结构,多用于机器翻译、天气预测和音频识别等具有

时序特征的识别和分类任务。LSTM 使用 3 种门决定细胞状态中信息通过的比例,分别为遗忘门、输入门和输出门^[12],如图 6 所示。



细胞输入 x_t ($x = (x_1, x_2, \dots, x_n)$ 为 CNN 层输出的特征图经扁平展开得到的向量)

图 6 LSTM 层的网络结构

遗忘门决定细胞状态遗忘信息,其计算公式为:

$$f_t = \text{Sigmoid}(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (4)$$

式中: x_t 为输入门; h_{t-1} 为上一时刻状态; f_t 为遗忘门; W_f 、 b_f 分别为遗忘门的权重和偏置项。

输入门决定增加信息到细胞状态,其公式为:

$$i_t = \text{Sigmoid}(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (5)$$

$$\tilde{C}_t = \text{Tanh}(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (6)$$

式中: i_t 为 t 时刻的输入门; $\tilde{C}_t$ 为候选向量; W_i 、 b_i 分别为输入门的权重和偏置项; W_c 、 b_c 分别为候选向量的权重和偏置项。

更新细胞状态:

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (7)$$

式中: C_{t-1} 为旧细胞状态,将 C_{t-1} 与 f_t 相乘以丢弃无用信息,将 $\tilde{C}_t$ 与 i_t 相乘以添加新信息。

输出门决定输出什么信息,其计算公式为:

$$o_t = \text{Sigmoid}(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (8)$$

$$h_t = o_t \times \text{Tanh}(C_t) \quad (9)$$

式中: h_t 为最终的输出状态信息; W_o 、 b_o 分别为输出门的权重和偏置项。

2.4 分类器

如图 7 所示,模型的分类器由两个全连接和 1 个 Softmax 输出单元构成,第 1 个和第 2 个全连接的神经元数量分别为 30 个和 5 个,分类器最后输出噪声样本被预测为 5 个评价等级的各个概率值,模型最终输出的是最大概率值所对应的评价等级。

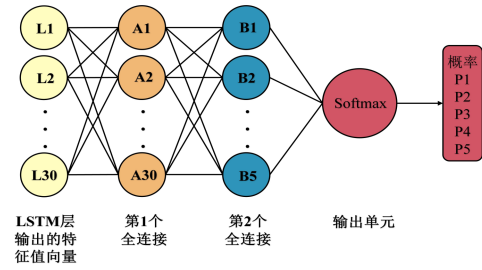


图 7 分类器的网络结构

2.4.1 全连接和随机失活

深度学习模型的分类器通常使用两个全连接,第 1 个全连接用于进一步提取输入数据的特征,第 2 个全连接则用于缩小最终的输出大小,其神经元数量与模型的分类数量相等。每个全连接后面也需要使用激活函数来引入非线性特性。

随机失活^[15]就是按设定的概率随机去掉网络当中的一些神经元,可降低模型对网络中某些神经元的依赖性,增强模型的泛化能力,同时可减少训练过程中的运算量,有效防止过拟合和加快训练速度。本文在两个全连接之间使用 1 个失活概率设置为 0.5 的随机失活。

2.4.2 损失函数

损失函数用于评估预测值与真实值之间的差距,网络反向传播计算的目标就是使损失函数达到全局最小值,使预测值最接近真实值。本文的声品质评价任务属于多分类任务,选用分类交叉熵作为损失函数,如式 (10) 所示。

$$J = - \sum_{i=1}^K y_i \lg(p_i) \quad (10)$$

式中: K 为种类数量; y 为样本评价等级向量,如果类别是 i ,则 $y_i = 1$,否则等于 0; p_i 为神经网络的输出,指预测类别为 i 的概率,由选定的输出单元计算得出。

2.4.3 Softmax 输出单元

Softmax 函数是一种常用于多分类任务的输出单元,能表示 N 个不同类别的概率分布,其计算公式为:

$$y_i = \frac{\exp(x_i)}{\sum_{j=1}^N \exp(x_j)} \quad (11)$$

式中: y_i 为车内噪声样本被分类为第 i 类的概率; x_i

和 x_j 为Softmax单元的输入,即第2个全连接的 N 个输出。由于上文将车内噪声分为5个评价等级,所以 N 值为5。

3 评价模型的训练与验证

本文的车内声品质评价模型是在Matlab Deep Network Designer环境下建立的,首先使用训练集寻找模型的最佳网络参数,然后应用验证集检验每次迭代训练后模型的预测准确度,训练全部结束后使用测试集评估模型的性能。深度学习模型的训练不仅需要大量的样本,而且超参数的选择对最终训练出来的模型性能产生重要的作用。混淆矩阵是深度学习中常用于测试模型性能的方法,可以直观地看出各类别和全部类别的预测结果。

3.1 评价模型的训练

3.1.1 混合输入和超参数

混合输入^[16]通过对同一数据集的两个样本和目标值进行插值获得更多的样本,从而克服因样本数量少而导致训练过程中出现过拟合的缺点,理论上通过这种方法可获得无穷多个样本。本文使用混合输入获取训练集和验证集,图4a为通过混合输入获得的噪声样本波形图。

优化器、学习率、L2正则化系数、最小批次数量、训练集和验证集的样本数量等超参数对模型最终的预测准确度具有非常重要的影响。可供选择的优化器有SGDM、RMSProp和Adam,由于Adam训练速度较快、收敛性更好,所以选择Adam作为优化器,与之相对应的学习率通常使用0.001。L2正则化是一种有效防止过拟合的方法,它在权重的损失函数后面添加1个惩罚项,惩罚项前面的系数就是L2正则化系数,也称为权重衰减系数。最小批次是训练集的子集,用于每次迭代中评估损失函数的梯度并更新权重。训练集用于更新模型的网络参数,训练集的样本数量太少,容易导致训练出来的模型过于简单而出现欠拟合现象,训练集的样本数量太多,又会使模型过于复杂而出现过拟合现象,所以需要选择合适的训练集样本数量。验证集用于评估不同网络参数下模型的性能,根据模型在

验证集上的效果选择是否停止训练,它的数量需要与训练集保持一定的比例,一般选择3:7或2:8的比例,本文选择后者的比例。表1列举了超参数的选取情况。

表1 超参数的选择

优化器	学习率	L2正则化系数	最小批次数量	训练集的样本数量	验证集的样本数量
Adam	0.001	0.000 1	32	2 160	555

3.1.2 模型的训练结果

如图8所示,在共12轮的训练过程中,训练损失曲线除了在第11轮时有所回升,总体上稳定下降,最终模型的训练损失下降至0.728;训练准确度曲线在1~3轮快速上升,3~6轮经过大幅下降后大幅回升,6~12轮以较小的波动幅度缓慢上升,训练准确度曲线总体上呈现波动上升的趋势,最终模型的训练准确度达到了96.88%。训练损失和训练准确度的最后结果说明评价模型使用训练集学习到了理想的网络参数,使模型预测的2 160个训练样本评价等级与真实评价等级总体上大致接近,同时也反映了CNN和LSTM共同提取到了噪声的深层次特征,使分类器能对大部分噪声样本做出正确的分类。

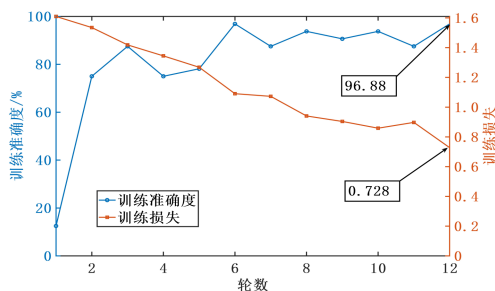


图8 训练准确度与训练损失曲线

3.2 评价模型的验证

对模型的验证分为两个部分,第1个部分使用555个噪声样本作为验证集对模型进行验证,考察其总体准确度;第2个部分基于混淆矩阵的方法,使用30个样本作为测试集查看模型对每一类样本的分类精度。

第1部分验证的结果如图9所示,在0~12轮过程中验证损失稳定下降,最终降至0.681;验证准确度在第3轮之前快速上升,之后缓慢爬升,在

第8轮后波动变化,验证准确度基本保持不变,70明模型的性能基本稳定下来了,最终的验证准确度为93.69%。

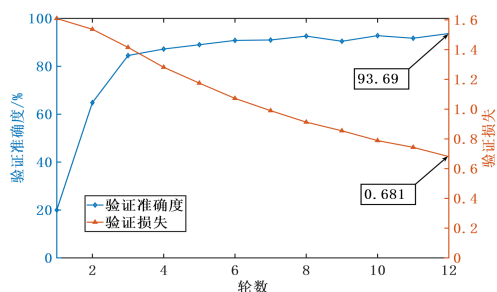


图9 验证准确度与验证损失曲线

第2部分验证的结果如图10所示,从长度为4s的真实噪声样本集中,每一类随机挑选6个,总共30个噪声样本作为测试集输入已训练好的评价模型,获得预测评价等级与真实评价等级组成的混淆矩阵。混淆矩阵的行代表预测评价等级,列代表真实评价等级,对角线上的数字和百分比为各类噪声样本预测正确的样本数量和预测准确度,非对角线上的数字和百分比则是分类错误的样本数量和预测偏差度。从混淆矩阵可以看出,有1个真实评价等级为“差”的噪声样本被分类为“很差”,有3个真实评价等级为“良好”的噪声样本被分类为“很好”,这4个噪声样本均被分类于相邻的评价等级,与真实评价等级差距不大,其余的噪声样本均被正确分类,表明模型对“很差”、“合格”和“很好”样本预测准确度最高。

总体来说,评价模型的预测结果与主观评价结果有着很高的吻合度,能够用于车内噪声的声品质评价。

参考文献 (References)

- [1] 张俊红,段超阳,林杰威,等.商用车柴油机加速声品质主客观评价研究[J].天津大学学报(自然科学与工程技术版),2019,52(2):150-156.
ZHANG Junhong, DUAN Chaoyang, LIN Jiwei, et al. Research on Commercial Vehicle Diesel Engine Sound Quality Subjective and Objective Evaluation During Acceleration [J]. Journal of Tianjin University (Science and Technology), 2019, 52(2): 150-156. (in Chinese)
- [2] 刘哲,高云凯,解馥荣.电动汽车关门声品质预测模型

真实评价等级	很差	6 20.0%	1 3.3%	0 0.0%	0 0.0%	0 0.0%	85.7% 14.3%
	差	0 0.0%	5 16.7%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	合格	0 0.0%	0 0.0%	6 20.0%	0 0.0%	0 0.0%	100% 0.0%
	良好	0 0.0%	0 0.0%	0 0.0%	3 10.0%	0 0.0%	100% 0.0%
	很好	0 0.0%	0 0.0%	0 0.0%	3 10.0%	6 20.0%	66.7% 33.3%
		100% 0.0%	83.3% 16.7%	100% 0.0%	50.0% 50.0%	100% 0.0%	86.7% 13.3%
		很差	差	合格	良好	很好	
		预测评价等级					

图10 评价模型的混淆矩阵

4 结论

本文基于深度学习法建立了车内声品质评价模型,该模型由预处理层、CNN层、LSTM层和分类器组成。通过研究得出以下结论:

(1) 使用对数梅尔频谱的方法把一维的波形噪声变换成二维的频谱,同时使用时频遮掩法增强数据的特征,为模型的CNN层、LSTM层和分类器提供更多、更强的特征信息。

(2) 评价模型在训练集的训练之下获得了96.88%的训练准确度,说明模型已获得理想的网络参数,使模型对大部分样本的评价等级预测正确。

(3) 在验证集的检验下,评价模型的验证准确度为93.69%,使用测试集对评价模型进行评估,发现评价模型的预测结果与主观评价结果具有很高的吻合度,证明基于CNN和LSTM融合特征提取的车内声品质模型具有足够的精度,可用于车内声品质的评价。

研究[J].汽车工程,2021,43(12):1858-1864.

LIU Zhe, GAO Yunkai, XIE Furong. Study on Predictive Models for Door Slamming Sound Quality of an Electric Vehicle [J]. Automotive Engineering, 2021, 43(12): 1858-1864. (in Chinese)

[3] 贺岩松,涂梨娥,徐中明,等.支持向量机在汽车加速车内声品质预测中的应用[J].汽车工程,2015,37(11):1328-1333.

HE Yansong, TU Li'e, XU Zhongming, et al. The

- Application of Support Vector Machine to the Prediction of Vehicle Interior Sound Quality During Acceleration [J]. *Automotive Engineering*, 2015, 37 (11): 1328–1333. (in Chinese)
- [4] FENG Tianpei, SUN Yuedong, WANG Yansong, et al. Sound Feature Space Effects on the Performance of Annoyance Evaluation Model Based on Support Vector Machine [J]. *Applied Acoustics*, 2019, 154: 99–113.
- [5] 杨易, 高骏, 谷正气, 等. 基于GA-BP的汽车风振噪声声品质预测模型 [J]. *机械工程学报*, 2021, 57(24): 241–249.
- YANG Yi, GAO Jun, GU Zhengqi, et al. Research on Sound Quality Prediction Model of Automobile Wind Buffeting Noise Based on GA-BP [J]. *Journal of Mechanical Engineering*, 2021, 57 (24): 241–249. (in Chinese)
- [6] CHEN Pushuang, XU Liangyuan, TANG Qiansheng, et al. Research on Prediction Model of Tractor Sound Quality Based on Genetic Algorithm [J]. *Applied Acoustics*, 2022, 185: 108411.
- [7] 曾发林, 孙苏民. 基于RNR-WVD与GA-小波的非稳态排气噪声声品质研究 [J]. *振动与冲击*, 2019, 38(12): 74–80, 104.
- ZENG Falin, SUN Sumin. Study on the Sound Quality of Unsteady Exhaust Noise Based on RNR-WVD Calculation and GA-Wavelet Network [J]. *Journal of Vibration and Shock*, 2019, 38(12): 74–80, 104. (in Chinese)
- [8] POURSEIEDREZAEI M, LOGHMANI A, KESHMIRI M. Prediction of Psychoacoustic Metrics Using Combination of Wavelet Packet Transform and an Optimized Artificial Neural Network [J]. *Archives of Acoustics*, 2019, 44(3): 561–573.
- [9] POURSEIEDREZAEI M, LOGHMANI A, KESHMIRI M. Development of a Sound Quality Evaluation Model Based on an Optimal Analytic Wavelet Transform and an Artificial Neural Network [J]. *Archives of Acoustics*, 2021, 46(3): 55–65.
- [10] PURWINS H, LI B, VIRTANEN T, et al. Deep Learning for Audio Signal Processing [J]. *IEEE Journal of Selected Topics in Signal Processing*, 2019, 13 (2): 206–219.
- [11] CHAKRABARTY S, HABETS E A P. Time-Frequency Masking Based Online Multi-Channel Speech Enhancement with Convolutional Recurrent Neural Networks [J]. *IEEE Journal of Selected Topics in Signal Processing*, 2019, 13(4): 787–799.
- [12] 韩瑞东. 基于卷积网络与长短时记忆网络的钢轨裂纹检测方法研究 [D]. 哈尔滨: 哈尔滨工业大学, 2019.
- HAN Ruidong. Research on Rail Crack Detection Method Based on CNN and LSTM [D]. Harbin: Harbin Institute of Technology, 2019. (in Chinese)
- [13] 候少麒, 梁杰, 殷康宁, 等. 基于空洞卷积金字塔的目标检测算法 [J]. *电子科技大学学报*, 2021, 50(6): 843–851.
- HOU Shaoqi, LIANG Jie, YIN Kangning, et al. Object Detection Algorithm Based on Atrous Convolutional Pyramid [J]. *Journal of University of Electronic Science and Technology of China*, 2021, 50 (6): 843–851. (in Chinese)
- [14] 刘建伟, 赵会丹, 罗雄麟, 等. 深度学习批归一化及其相关算法研究进展 [J]. *自动化学报*, 2020, 46(6): 1090–1120.
- LIU Jianwei, ZHAO Huidan, LUO Xionglin, et al. Research Progress on Batch Normalization of Deep Learning and Its Related Algorithms [J]. *ACTA Automatica Sinica*, 2020, 46(6): 1090–1120. (in Chinese)
- [15] LIU Lei, LUO Yuhao, SHEN Xu, et al. Beta-Dropout: A Unified Dropout [J]. *IEEE Access*, 2019, 7: 36140–36153.
- [16] LIANG Daojun, YANG Feng, ZHANG Tian, et al. Understanding Mixup Training Methods [J]. *IEEE Access*, 2018, 6: 58774–58783.

作者简介



杨礼强 (1998–), 男, 广西桂平人, 硕士研究生, 主要研究方向为汽车振动噪声控制。

Tel: 18269572523

E-mail: yangliqiang@cqu.edu.cn