

DOI:10.19651/j.cnki.emt.2211871

基于一维卷积神经网络的列车异响识别系统研究^{*}付孟新¹ 郭世伟¹ 王泽兴¹ 丁建明²

(1.西南交通大学机械工程学院 成都 610031; 2.西南交通大学牵引动力国家重点实验室 成都 610031)

摘要: 在列车行驶过程中,车内异响可作为反映车辆设备状态的信息源。为此提出一种基于1D-CNN的识别模型,对车辆异响进行识别,并设计列车异响识别系统。首先构建音频数据的试验样本库,然后利用MFCC提取异响数据样本的特征信息。针对列车噪声特征与车辆状态类型间的映射关系复杂、难解耦的问题,构建一种基于MFCC输入的1D-MCNN对异响所蕴含的故障信息进行识别分类。最后对识别模型进行实验与优化,确定MFCC阶数、学习率与批尺寸等模型参数,采用t-SNE算法、混淆矩阵进行模型特征提取的分析评价。试验结果表明该方法对列车异响识别诊断效果较好,准确率达98.38%。

关键词: 高速列车;异响;卷积神经网络;诊断识别;MFCC-1D-MCNN模型

中图分类号: TP 277 **文献标识码:** A **国家标准学科分类代码:** 510.40

Train abnormal sound recognition system based on 1D-CNN

Fu Mengxin¹ Guo Shiwei¹ Wang Zexing¹ Ding Jianming²

(1. School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, China;

2. State Key Laboratory of Traction Power, Southwest Jiaotong University, Chengdu 610031, China)

Abstract: The abnormal sound in the trains running can be used as information source to reflect the status of the vehicle equipment. For the reason that, a recognition model based on 1D-CNN was proposed to identify the abnormal sound of trains, and a set of recognition system for abnormal sound of trains was designed. Firstly, the experimental sample library of audio data was constructed. Then MFCC was used to extract the characteristic information of abnormal sound data samples. Aiming at the complex mapping relationship between train noise features and vehicle state types, a 1D-MCNN based on MFCC input was constructed to identify and classify the fault information contained in abnormal sound. Finally, the model parameters such as MFCC order, learning rate and batch size are determined by experiments and optimization. The t-SNE algorithm and confusion matrix were used to analyze the model feature extraction ability. The results show that the method is effective for the identification and diagnosis of abnormal sound of trains and its accuracy rate reaches 98.38%.

Keywords: high-speed train; abnormal sound; convolutional neural network; diagnosis and recognition; MFCC-1D-MCNN model

0 引言

自2007年中国开通并运行第一列车组列车以来,历经多年的发展壮大,运行里程已达3万多公里,但随着运营时间和里程的增加,车辆设备的状态异常和故障发生几率也在增加,导致列车多次停车检查,严重影响了列车的有序运行^[1-2]。目前,列车在运行途中出现的异响仅依靠随车机械师的人耳(也称为“顺风耳”)和轨旁检测设备来监听判

断,对于可以随车检测智能装备的研究尚浅。列车在运行过程中,其异响往往与运行噪声混合且与运行环境、运行速度、车辆零部件振动相关。人耳的疲劳特性与对声音的感知与辨识程度存在较大的个体差异的问题,易使异响检测出现误判,具有较高的随机性与不确定性,可靠性较低^[3]。对于轨旁检测设备,存在不能跟车随时监测和连续监测的问题。

目前国内外对于列车异响检测识别方法的研究,延用

收稿日期:2022-10-28

* 基金项目:四川科技厅重点研发项目(2020YFG0124)、成都科技局重点研发项目(2019-YF05-01823-SN)、中国博士后科学基金(2020M682506)项目资助

了 1997 年 Choe 等^[4]提出的方法框架,即首先通过声音传感器采集声学信号,然后利用一系列特征提取方法获得信号特征,最后输入分类器进行识别。在轨旁声学信号采集方面,美国铁道协会(American Association of Railroads, AAR)早在 20 世纪 60 年代开始研究^[5-6]。接着轨旁声学检测系统^[7](trackside acoustic detection system, TADS)迎来了高速发展。澳大利亚公司 VIPAC 利用声学信号与列车轴承故障异响的映射关系,开发了一套铁路轴承声学检测系统(Rail BAM)^[8]。Eronen 等^[9]采用多种特征提取与分类方法,包括了短时平均能量(short-time average energy)、过零率(zero-crossing rate, ZCR)、梅尔倒谱系数(Mel-scale frequency cepstral coefficients, MFCC)等 11 种方法,将提取到的特征经过融合降维后作为声音信号识别的特征输入,训练了低阶的隐马尔可夫模型(hidden Markov model, HMM)与高斯混合模型(Gaussian mixed model, GMM)^[10],并且提出将系统应用于便携式设备。孙永奎^[11]通过声学信息对列车塞拉门故障进行检测,采用经验模态分解(empirical mode decomposition, EMD)获得本征模态函数,并提取对应的小波包能量熵和多尺度排列熵,作为声音信号的特征提取方法,最后利用混合支持向量机(support vector machine, SVM)进行识别,准确率达 93.62%。邹瑛珂^[12]提出一种基于变分模态分解(variational mode decomposition, VMD)分解的 MFCC+伽玛通频率倒谱系数(Gammatone frequency cepstrum coefficient, GFCC)无人机噪音混合特征提取方法。狄晓栋^[13]提出了一种基于混合特征 MGCC 的干式变压器故障诊断模型最后送入长短期记忆网(long short-term memory, LSTM)分类模型进行模式识别。

随着深度学习的发展与兴起,将深度学习模型应用于声音事件分类领域,已经逐渐替代传统的 GMM、HMM、SVM 等方法。2015 年, Zhang 等^[14]初次提出将卷积神经网络(convolutional neural networks, CNN)应用于声音事件分类领域。2020 年,王若平等^[15]在采用 MFCC 图谱方法提取汽车类异响特征时,将 MFCC 提取步骤之一的离散余弦变换(discrete cosine transform, DCT)摘除^[16],然后对各帧对数能量谱拼接得到(filterbank, Fbank)图谱,利用小波包分解的时频变换特性,对图谱进行改进融合得到新特征,送入 CNN 进行分类识别。最终该模型在其自测数据集上获得了 90.2% 的识别准确率。

因此本文采用依靠人耳听觉特性所产生的 MFCC 来提取列车异响特征信息。并将其作为卷积神经网络的输入,加入一维多尺度卷积层(1-dimensional multiscale convolution layer, 1D-MCL),成功构建一种基于 MFCC 与一维多尺度卷积神经网络的识别模型(1-dimensional multiscale convolutional neural network with MFCC input, MFCC-1D-MCNN)来完成列车异响的识别诊断任务。并采用硬件加软件的方式设计一套列车异响识别系统,拓宽

了列车异响检测的手段。

1 MFCC-1D-MCNN 识别算法的构建

1.1 MFCC 特征提取

根据人耳的听觉感知研究,在不同音频信号频率下人耳的听觉灵敏度有所不同。MFCC 就是基于人耳的听觉特性所产生的特征提取方法。在本文的声音识别事件中,随车机械师可以通过人耳判断一部分的列车异响事件,说明异响特征信息能够被人耳捕捉。因此采用贴近人耳听觉特性的 MFCC 特征提取方法是合理的。其提取过程如图 1 所示,具体提取步骤如下:

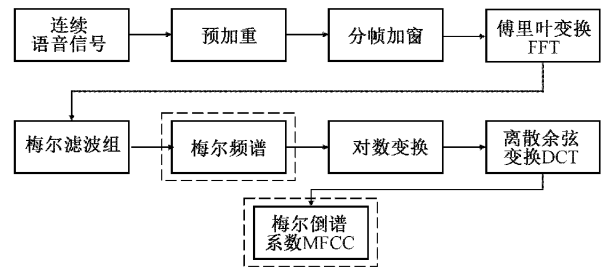


图 1 MFCC 的提取过程示意图

1) 预加重。为了弥补声音信号传输端的损失,尽量还原真实声音信号,需要对受损的高频信号进行如式(1)所示的补偿。式中 μ 为滤波器系数,取值 0.93~0.97,本文取 0.97。

$$H(z) = 1 - \mu z^{-1} \quad (1)$$

2) 分帧。音频信号为准稳态信号。将一定数量的采样点合并为一个观测单位,称为帧。在每一帧中即可看作为稳态信号,方便后续处理。选择将 1 024 个采样点合为一帧,涵盖的时间约为 23 ms,帧移取 512 个采样点,约 11 ms,重叠一帧的 1/2 即 512 个采样点。

3) 加窗。为了保证每帧音频信号边缘的平滑,通常使用一个平滑的窗函数与之相乘,从而使每帧信号两端平滑的衰减为 0。选择加入汉明窗,如式(2)所示,若规定分帧后的信号为 $S(n)$, N 为帧长,则加窗后的信号为 $S'(n)$ 如式(3)所示。参数 a 为常数取 0.46。

$$W(n, a) = (1 - a) - a \times \cos\left(\frac{2\pi n}{N-1}\right), n \in [0, N-1] \quad (2)$$

$$S'(n) = S(n) \times W(n, a) \quad (3)$$

4) 快速傅里叶变换(fast Fourier transform, FFT)。通过时域信号通常很难判断信号的特性,通常将其转换到频域上的能量分布来进行分析。进行 FFT 的计算过程如式(4)所示,其中 $S'(n)$ 为输入的语音信号, k 为第 k 个基信号, N 为帧长。

$$X(k) = \sum_{n=0}^{N-1} S'(n) e^{-j\frac{2\pi kn}{N}}, k \in [0, N-1] \quad (4)$$

5) 通过梅尔滤波组。在获得声音的能量谱之后,需要

将其转化为梅尔刻度,模拟人耳存在对低频更具辨识性的感知特点。定义 M 为三角滤波器的个数,中心频率为 $f(m), m=1,2,3,\dots,M$ 。 M 通常取 22~128 的整数,三角滤波器的频率响应的定义如式(5)所示。

$$H_m(k) = \begin{cases} 0, & k < f(m-1) \\ \frac{2(k-f(m-1))}{(f(m+1)-f(m-1))(f(m)-f(m-1))}, & f(m-1) \leq k \leq f(m) \\ \frac{2(f(m+1)-k)}{(f(m+1)-f(m-1))(f(m)-f(m-1))}, & f(m) \leq k \leq f(m+1) \\ 0, & k \geq f(m+1) \end{cases} \quad (5)$$

式中: $\sum_{m=0}^{M-1} H_m(k) = 1$

滤波后对信号取对数,每个滤波器组输出的对数能量如式(6)所示。

$$s(m) = \ln\left(\sum_{k=0}^{N-1} |X(k)|^2 H_m(k)\right), 0 \leq m \leq M \quad (6)$$

6) 提取梅尔倒谱系数。采用离散余弦变换(discrete cosine transform, DCT)对信号数据进行降维,将冗余的数据分开来。变换后大部分信号数据将集中在低频区域,如式(7)所示。

$$C(n) = \sum_{m=0}^{N-1} s(m) \cos\left(\frac{\pi n(m-0.5)}{M}\right), n = 1, 2, \dots, L \quad (7)$$

式中: $C(n)$ 为返回的梅尔信号,至此就得到了识别模型的特征输入, L 为 MFCC 的阶数,代表提取 MFCC 系数时会返回 L 阶的梅尔倒谱系数,这会直接影响特征输入的尺寸大小。后续将会对 L 的取值进行试验分析。

1.2 1D-MCL 的构建

Inception 模块可以相互堆叠对网络的深度和宽度进行高效的扩充,在提升网络性能的同时可以有效的防止过度拟合与梯度消失现象。因此,本文基于 Inception 模块原理设计出一种 1D-MCL,插入到传统的卷积神经网络中,其结构如图 2 所示。

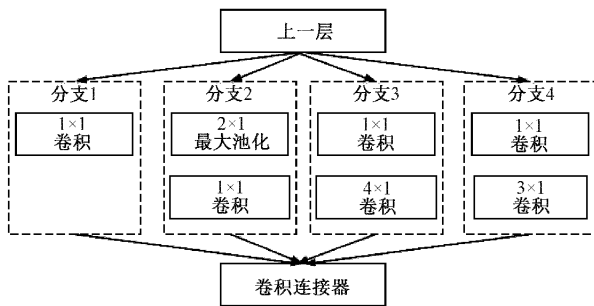


图 2 1D-MCL 结构

所设计的一维多尺度卷积层的核心模块由 4 个分支组成,使用 3 种不同尺寸的卷积和 1 个最大池化层分别进行卷积运算,然后将每个分支提取到的特征信息进行聚合,传递给下一层。有效的增加了卷积神经网络对不同尺度的自适应性。1D-MCL 具体参数如表 1 所示。

表 1 1D-MCL 具体参数

名称	卷积核数量	卷积核尺寸	步长
分支 1	64	1×1	1
分支 2	最大池化	2×1	1
	32	1×1	1
分支 3	16	1×1	1
	32	4×1	1
分支 4	64	1×1	1
	128	3×1	1

1.3 MFCC-1D-MCNN 的构建

声音信号作为典型的一维信号,二维矩阵不能凸显时变特性,也不能单独分析频带之间的相关性。为了适应样本的一维特性,提取到的 MFCC 特征采用一维排布法,每帧的信号特征首位拼接,纵向排列。将 MFCC 作为模型的输入,加入 1D-MCL 模块,设计一种基于 MFCC 输入的一维多尺度卷积神经网络模型 MFCC-1D-MCNN。其结构如图 3 所示。当 MFCC 阶数取 40 时,模型各层的具体参数如表 2 所示。提取的 MFCC 特征参数作为输入层的输入,将其转化为矩阵,使用卷积核对其进行卷积,从而得到卷积之后的特征。经过两层卷积后采取池化操作,有效的缩小参数矩阵的尺寸,从而减少最后连接层中的参数数量。再通过两次堆叠代替传统的一维卷积层,将学习到的多尺度特征信息融合输入到下一层网络。平均池化层可以视作一种稀疏的全连接层,可以有效减少参数量。Softmax 激活的全连接层在特征矩阵与最终的分类间转换显得更加简单自然。

2 仿真实验过程与结果分析

2.1 声音样本库的建立

列车异响的声音样本数据经随车机械师在途录制,由随后的停车检查车辆故障状态确定其对应的故障类型,并为异响样本打上标签。取 5 种典型故障类型作为研究对象:车轮廓形多边形故障异响(简称“多边形故障异响”)、钢轨波磨异响、轴承外圈剥离异响、转向架和车体低架零件松动悬吊异响(简称“零件松动异响”),再加上列车正常行驶状态的声音,一共 5 种音频数据样本类型,进行分析处理与诊断识别。

由于神经网络的训练需要大量的数据样本,否则神经网络容易出现过度拟合。因此本文采用滑动窗口数据采样法,进行数据样本扩增。具体方法如式(8)和图 4 所示。样本数量 N ,等于原始数据样本长度 L ,减去采样窗口长度 l 的差,除以采样窗口步长 s 的商。

$$N = \frac{L-l}{s} \quad (8)$$

原始音频数据样本长度为 $L=1\ 396$,采样窗口长度设为 $l=1$,为了扩充样本数量,将窗口滑动长度设置为 $s=$

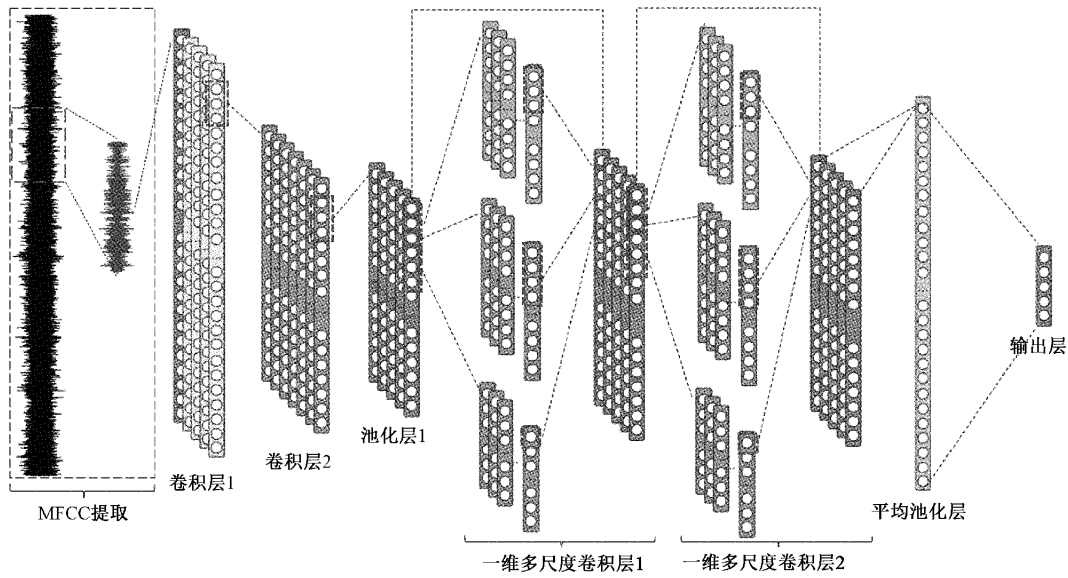


图 3 MFCC-1D-MCNN 模型结构图

表 2 MFCC-1D-MCNN 各层参数

层名称	卷积核			数据尺寸
	大小	步长	数目	宽×深
输入层	—	—	—	3 480×1
卷积层 1	8×1	2	96	1 740×96
卷积层 2	3×1	2	128	870×128
池化层 1	3×1	3	—	290×128
1D-MCL1	—	—	—	290×256
1D-MCL2	—	—	—	290×256
全局平均池化层	—	—	—	256×1
全连接输出层	—	—	—	5×1

表 3 扩增后的数据

声音类型	样本时间/s	采样频率/Hz
正常	1 020	44 100
多边形故障	1 288	44 100
钢轨波磨	1 088	44 100
轴承外圈剥离	1 048	44 100
零件松动	1 140	44 100
合计	5 584	44 100

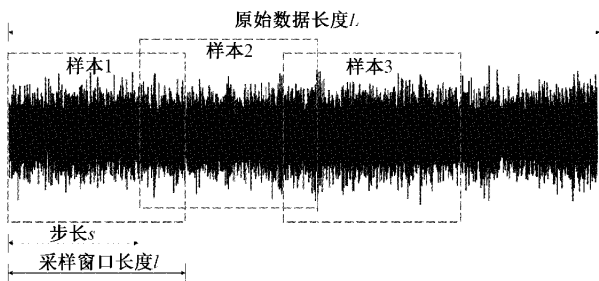


图 4 滑动窗口采样

0.25, 得到扩增后的样本 5 584 个。与原始音频样本数目相比, 数据扩增了 4 倍。数据扩增后得到由 5 584 个列车音频数据样本组成的样本库, 表 3 展示了扩增后音频样本的详情信息。

2.2 实验流程

使用表 3 数据进行实验验证。将 2/3 的样本作为训练集, 1/3 的样本作为验证集, 每条用于训练的音频尺寸为 1 s。所有模型训练初始迭代次数 (Epoch) 设置为 100, 当

出现 20 个回合准确率不上涨, 则将其确定为该次该种实验的最终准确率, 具体步骤如下:

- 1) 将采集到的原始音频数据进行预处理并切割为尺寸为 1 s 的音频样本文件, 以 2 : 1 的比例划分, 分别用于制作训练集与验证集;
- 2) 设置 MFCC 的相关参数, 对声音样本进行 MFCC 特征提取, 得到训练集与验证集;
- 3) 初始化网络模型参数, 并且设置学习率、迭代次数、批尺寸、激活函数等超参数;
- 4) 将训练集以批量的方式输入到模型中进行训练, 经过向前传播与反向传播后, 逐层更新网络参数;
- 5) 将验证集输入到训练好的 MFCC-1D-MCNN 的模型中, 输出分类结果, 验证其有效性;
- 6) 重复上述步骤 2)~5), 根据实验结果, 调整各参数直到获得满意的网络精度。

2.3 MFCC 阶数的选择

不同的阶数选择会导致网络模型输入尺寸的不同, 过小的阶数会降低特征提取的分辨率, 过大的阶数选择会导致引入过多的冗余特征信息。本小节分别选取阶数为 13、22、40、88、128 的 MFCC, 输入到网络模型进行验证实验,

以确定最适合列车异响声学特征的 MFCC 阶数。实验结果如表 4 所示。

表 4 MFCC 阶数选择实验结果

MFCC 阶数	准确率/%	训练用时/s	输入尺寸
13 阶	94.63	15	1 131×1
22 阶	97.12	36	1 914×1
40 阶	98.38	54	3 480×1
88 阶	98.57	104	7 656×1
128 阶	97.16	204	11 136×1

由表 4 可以看出以 40 阶与 88 阶的 MFCC 作为模型的输入,获得的准确率均较高。但是 40 阶的 MFCC,在仅仅比 88 阶的 MFCC 准确率低 0.19% 的情况下,训练用时减少了 48%。快速响应性优于 88 阶的 MFCC,因此综合评定 40 阶的 MFCC 更适合列车异响识别。

2.4 不同学习率与批尺寸对比实验

学习率会直接影响模型在学习训练中参数更新的幅度,过大与过小的学习率都会导致不好的模型训练结果。因此本文设置了 4 组不同的学习率进行实验比较,分别是 1.0×10^{-5} 、 3.0×10^{-5} 、 5.0×10^{-5} 和 1.0×10^{-4} ,训练结果如图 5 与表 5 所示。

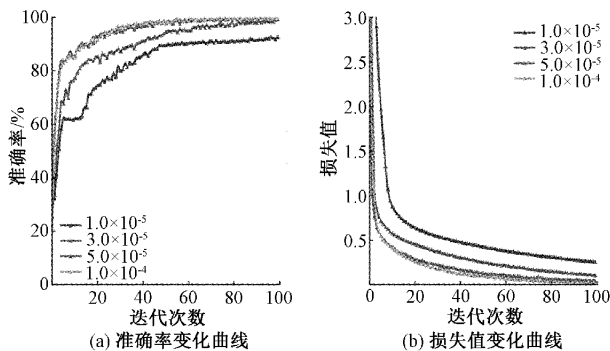


图 5 不同学习率所获得的准确率及损失值曲线图

表 5 不同学习率所获得的准确率

学习率	准确率/%
1.0×10^{-5}	90.12
3.0×10^{-5}	96.10
5.0×10^{-5}	98.38
1.0×10^{-4}	97.16

且美国东北大学 Keskar 等^[17]证明,使用更小的批尺寸训练神经网络,可以增强模型的泛化能力。为探究批尺寸对模型准确率的影响,本文设置 5 种批尺寸 20、40、80、160、320,在学习率均为 5.0×10^{-5} 下进行对比实验,结果如图 6 和表 6 所示。

由图 6 和表 6 可知,批尺寸为 160 和 320 的实验组在 100 次迭代后准确率仍呈上升趋势,其准确率仍然较低,若继续增加迭代轮次可以获得更高的准确率,训练用时也会

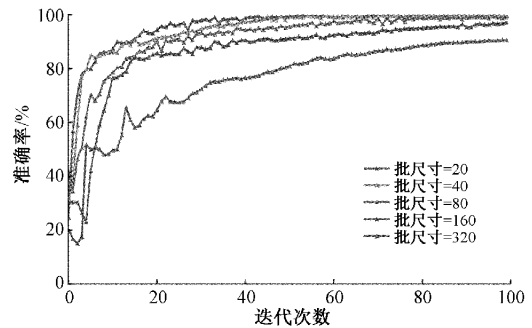


图 6 设置不同批尺寸所获得的准确率曲线图

表 6 不同批尺寸所获得的准确率和训练用时

批尺寸	准确率/%	训练用时/s
20	97.12	112
40	98.38	113
80	98.23	112
160	91.98	103
320	90.11	97

增加。批尺寸为 20、40、80 的实验组准确率表现较好,用时相差无几,因此取准确率最高所对应的批尺寸 40 为模型最合适的批尺寸。

2.5 消融实验

消融实验是验证神经网络模型各子模块作用的有效方法。通过删除系统个别子模块,来观察网络表现,进而确定该子模块的作用。

本文进行消融实验的具体方法为:首先将 MFCC 提取模块、多尺度卷积层、最大池化层分别抽离,然后补齐对应模块保证神经网络正常运行,观察模型性能的变化。设计出的 3 种消融网络实验组分别命名为消融网络 A、B、C。如图 7,红色虚线框内为删除部分,实验过程中使用相同的数据集与计算平台。

1) 消融网络 A,如图 7(a)所示,去除 MFCC 提取模块后,将原始音频完成预处理后直接输入网络,训练中迭代次数与准确度曲线如图 8(a)所示。

2) 消融网络 B,如图 7(b)所示,去除多尺度卷积层模块后,不添加其余模块,训练中迭代次数与准确度曲线如图 8(b)所示。

3) 消融网络 C,如图 7(c)所示,去除最大池化层模块,在多尺度卷积层后将多维输入展平,用传统的全连接层融合学习到的深度特征,训练中迭代次数与准确度曲线如图 8(c)所示。

4) 完整模型,训练中迭代次数与准确度曲线如图 8(d)所示。

可以发现随着不同网络模块的删减,训练过程中模型的收敛速度与准确率都出现了下降,表 7 详细展示了消融实验对比结果。

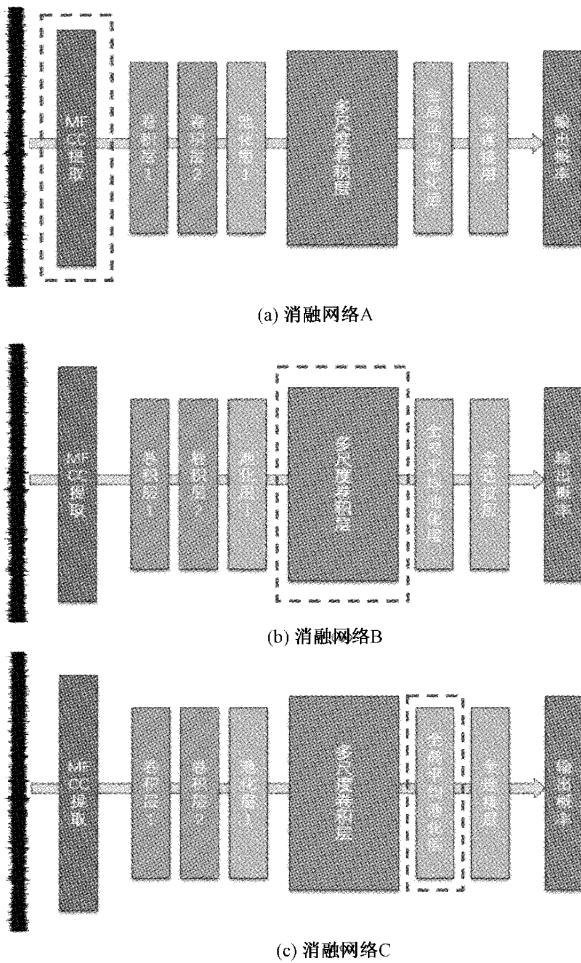


图 7 3 种消融网络结构图

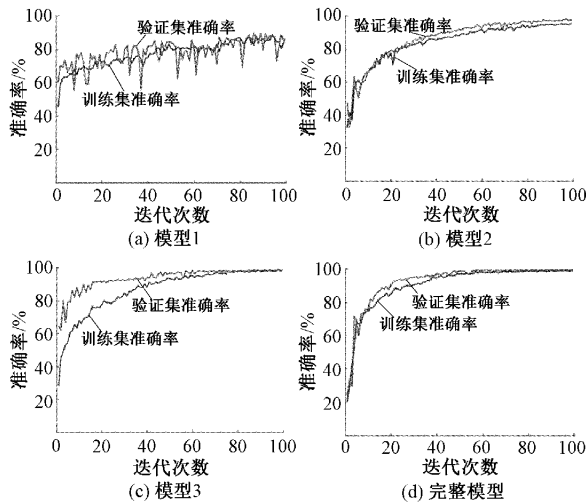


图 8 各网络结构性能表现

在准确率方面,4 组网络中消融网络 A、消融网络 B 相较于其余两组表现较差,证明 MFCC 提取模块和多尺度卷积层模块的缺失对模型准确率影响很大。消融网络 C 采用全连接层融合特征,导致参数量增加,在准确率比完全

表 7 消融实验对比结果

网络名称	准确率/%	训练用时/s
消融网络 A	88.12	482
消融网络 B	93.67	94
消融网络 C	96.13	519
完整网络	98.38	113

网络低的同时,训练时间也大量增加,证明在小样本的模型训练情况下,用全局平均池化代替全连接层可以在准确率与训练效率上有效提升网络性能。

本次消融实验证明了在设计神经网络中,各个模块均发挥了对网络性能提升的作用,不存在网络结构冗余的情况。

2.6 与不同识别模型对比实验

实验使用的模型除本章所提出的 MFCC-1D-MCNN 外,还加入与传统 SVM 分类器和将原始的音频经过 FFT 得到的一维数据输入到 1D-CNN 中进行识别的诊断方法 (FFT-1D-CNN) 对比实验,结果如表 8 所示。

表 8 不同模型识别实验对比结果

模型	准确率/%
SVM	78.84
FFT-1D-CNN	86.35
MFCC-1D-MCNN	98.38

可以看出,相较于传统的机器学习算法对列车异响的识别,本文构建的 MFCC-1D-MCNN 与 SVM、FFT-1D-CNN 在相同数据集下分别进行实验,在识别率上分别有 19.54%,12.03% 的提升,因此 MFCC 与卷积神经网络相组合的列车异响识别方法对数据的学习能力和泛化能力上的优异。

2.7 模型特征提取能力分析

根据上述实验结果,确定能发挥模型最佳性能相关参数如表 9 所示。

表 9 模型最佳性能参数

名称	最佳参数
MFCC 提取阶数	40 阶
学习率	5.0×10^{-5}
批尺寸	40

自适应加权的 t 分布随机近邻嵌入算法 (t-distributed stochastic neighbor embedding, t-SNE) 算法^[18]可以将高维数据降低到二维,通过视觉直观验证算法有效性。根据确定的最佳模型参数得到的训练模型,采用 t-SNE 可视化算法,对多层特征提取能力进行分析,不同分类结果用不同颜色的数字标记,如图 9 所示。

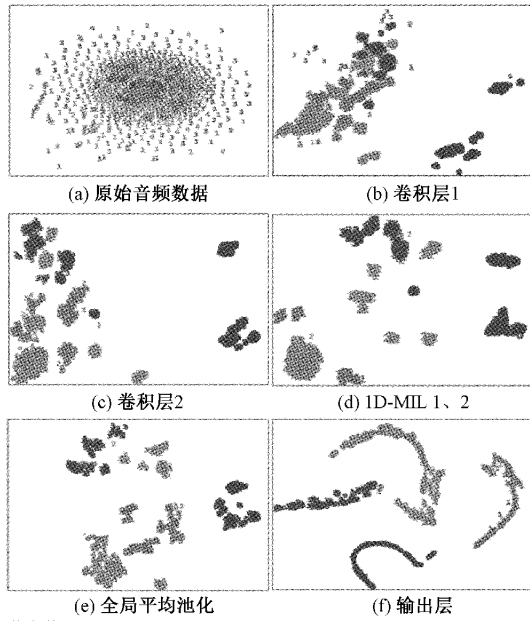


图 9 各层特征输出 t-SNE 可视化

图 9(a)展示了原始音频数据的 t-SNE 降维结果,可见各类故障信息难以区分。

图 9(b),(c)所示的是分别经过第一层卷积层与第二层卷积层后,各类故障信息开始聚集,不再是最初的无序状态。用红色数字 0 所代表的正常音频信号,和故障音频信号产生了一定的间隔。但各群仍然有一小部分粘连。

图 9(d)所示为再经过两层多尺度卷积层后,各群已经明显分离,并且同种故障类型聚集性增强。

图 9(e)表示又经全局平均池化层融合学习到的特征。图 9(f)为再经分类层的分类结果,可见 5 类音频信号已经完全分离聚集,各类故障的特征表达线性可分,展示了模型具有良好的特征提取能力。

混淆矩阵是精度评价的一种标准格式,引入混淆矩阵来分析本章所建模型 MFCC-1D-MCNN 的诊断效果。如图 10 所示,混淆矩阵用 5 行 5 列的矩阵形式来表示,纵坐标与横坐标的 0、1、2、3、4 分别对应类型:正常声音、车轮多

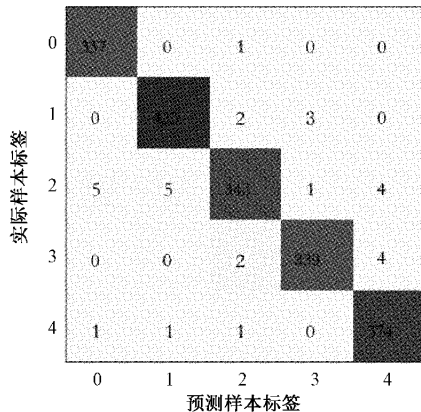


图 10 混淆矩阵

边形故障异响、钢轨波磨异响、轴承外圈剥离异响和零件松动异响。

从图 10 中可以看到,各类型均出现少量误分情况。其中,有 15 个标签为 2 的样本被误分为其余类型。说明钢轨波磨异响在本模型的验证中相对容易出现混淆。在 1 847 个验证样本中,共计 30 个样本被误分,仅占验证样本的 1.62%。在经过 t-SNE 与混淆矩阵两种方法的可视化后,说明模型特征提取能力较好,诊断效果理想。

3 列车异响识别系统的设计

3.1 系统功能介绍

列车异响识别系统由软件和硬件结合组成,即便携式音频采集设备(硬件)与个人计算机(personal computer, PC)端列车异响识别软件(软件)组成。可以方便快捷地实现音频采集与识别等系统的全部操作。

设计便携的音频采集模块这一硬件设备,可录制得到单声道 44 100 与 8 000 Hz 的 WAV 文件,随车机械师使用硬件设备完成异响音频采集,通过读卡器将音频数据导入软件,利用 PC 端强大的算力进行异响识别。图 11 展示了系统主要组成及功能。

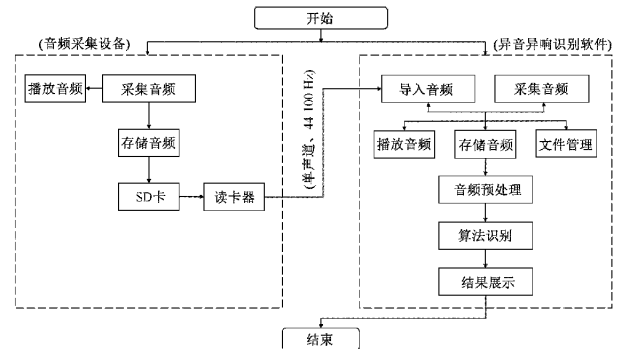


图 11 系统主要组成及功能图

3.2 异响音频采集硬件

硬件介绍:

- 1) 微控制单元:STM32F103ZET6。
- 2) 音频模块:VS1053,支持录音、解码播放功能,芯片采用 IIS 与 SPI 协议驱动。
- 3) 音频文件存储设备:SD 卡,采用 SPI 协议驱动。
- 4) 音频文件移植设备:读卡器。实物如图 12 所示。

异响音频采集设备设置了 3 个按键,分别为:KEY0、KEY1 与 KEY2,功能与按键各操作如下:

1) 连接电源,按下开关机键可以实现开关机,在开机后按下 KEY0 键可以开始录制音频,再次按下 KEY0 键停止或者继续录音。

2) 按下 KEY1 键结束录制,音频数据自动保存到 SD 卡指定目录下。

3) 按下 KEY2 键,进入回放模式,自动扫描 SD 卡中的 WAV 文件,播放最近一次保存的 WAV 文件。

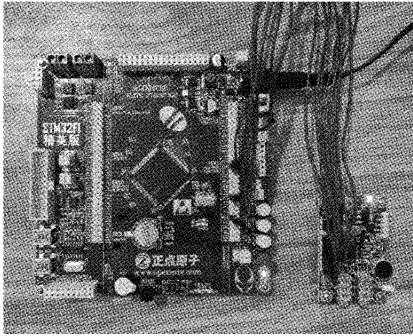


图 12 异响采集设备实物图

3.3 异响音频采集软件

基于深度学习的列车异响识别软件主要由 3 个模块组成,即音频采集、音频信号预处理和识别与结果展示。功能结构如图 13 所示。

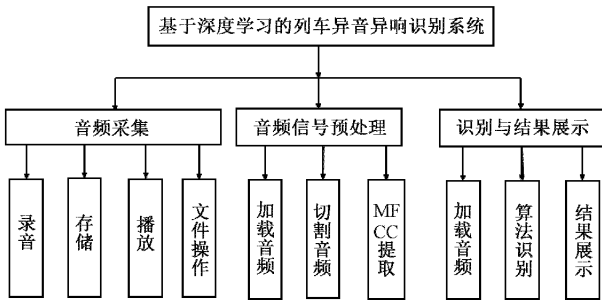


图 13 列车异响识别软件功能结构

基于深度学习的列车异响识别系统的软件界面如图 14 所示。

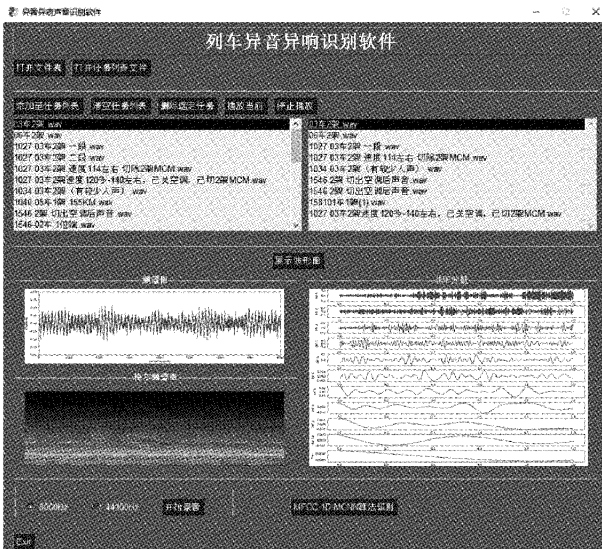


图 14 列车异响识别系统界面

具体功能设置如下:

1) 音频采集。利用 Python 的 pyaudio 库调用笔记本电脑上有自带的收音装置完成录音,采样频率为单声道

44 100 Hz。点击相关按钮可以实现文件管理,播放声音文件等功能。

2) 音频信号预处理。在将待检测数据送入识别算法前,系统会将原始音频信号进行预处理,为适应算法模型输入的数据尺寸,首先将加载音频获取音频长度,利用 Python 的 AudioSegment 库将其切割为 1 s 的样本数据,并提取 40 阶 MFCC 进行一维排布。

3) 展示波形图。选中文件操作区(右侧白色矩形框)的文件,点击“展示波形图”按钮,将展示选中声音文件的频谱图、梅尔频谱图和经过 EMD 分解的本征模态函数(Intrinsic mode function, IMF)分量图。

4) 识别与结果展示。在将文件从文件加载区,添加至文件操作区,点击“MFCC-1D-MCNN 算法识别”即可将音频文件进行预处理后,送入已经训练好的网络中进行识别分类,将最终结果进行展示。

4 结 论

提出了基于 MFCC 提取声音信号特征,并利用 1D-CNN 进行分类的列车异响识别诊断方法。首先建立了一个列车异响样本库。然后设计了一种 1D-MCL,结合 MFCC 的特征输入,构建了一种基于 MFCC 与一维多尺度卷积层的网络模型 MFCC-1D-MCNN 用于异响的识别。在实验与优化中,选取到了能发挥模型性能的最佳 MFCC 提取阶数、学习率和批尺寸。使用 t-SNE 算法与混淆矩阵两种方法对模型的特征提取能力进行可视化展示。结果表明,MFCC-1D-MCNN 模型每层都能自适应的提取特征信息,模型误分情况较少。模型的识别准确率达 98.38%。

采用软件加硬件的方式搭建了列车在途异响识别系统。介绍了该系统功能与操作流程。硬件上实现了数据采集、存储、处理与传输功能。软件上实现了简洁的系统操作界面,搭载了 MFCC-1D-MCNN 模型算法,可以实现列车在途异响识别诊断功能,使随车机械师可以方便快捷的进行异响识别工作。

参考文献

[1] 刘辉. 中国高速铁路的创新与发展[J]. 领导科学论坛, 2018(12):42-62.

[2] ZHAO P, BAI M, YANG J. Train safety detection technology based on audio analysis [P]. Communication and Information Systems, 2017: 399-403.

[3] 花伟,梅吉祥. 动车组运用的研究进展及现状[J]. 中国铁路,2017(4): 33-38.

[4] CHOE H, WAN Y, CHAN A. In neural pattern identification of railroad wheel-bearing faults from audible acoustic signals; Comparison of FFT, CWT, and DWT features [C]. Proceedings of the SPIE, Bellingham, 3 April; Bellingham, 1997: 480-496.

- [5] DYBALA J, RADKOWSKI S. Reduction of doppler effect for the needs of wayside condition monitoring system of railway vehicles[J]. *Journal of Mechanical Systems and Signal Processing*, 2013, 38(1): 125-136.
- [6] HE B, WANG J, HU F, et al. Wayside acoustic diagnosis of defective train bearings based on signal resampling and information enhancement[J]. *Journal of Sound and Vibration*, 2013, 332(21): 5635-5649.
- [7] 铁道车辆滚动轴承故障轨边声学诊断系统(TADS)探测设备: TB/T 3340-2013[S]. 北京: 中国标准出版社, 2013.
- [8] SNELL O D, NAIRNE I. Acoustic bearing monitoring-the future RCM 2008 [C]. 4th IET International Conference on Railway Condition Monitoring(RCM 2008).
- [9] ERONEN A J, PELTONEN V T, TUOMI J T, et al. Audio-based context recognition [J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2006, 14(1): 321-329.
- [10] ZAHEER M Z, KIM J Y, KIM H G, et al. A preliminary study on deep-learning based screaming sound detection[C]. *International Conference on IT Convergence and Security(ICITCS)*, 2015, 20(1): 1-4.
- [11] 孙永奎. 基于声音信号的列车塞拉门故障特征提取与诊断方法研究[D]. 北京: 北京交通大学, 2021.
- [12] 邹瑛珂, 李祖明, 刘晓宏, 等. 基于VMD分解的MFCC+GFCC无人机噪音混合特征提取方法[J]. *中国测试*, 2021, 47(11): 141-146.
- [13] 狄晓栋, 李震梅, 李宗哲, 等. 基于混合特征MGCC的干式变压器故障诊断[J]. *电子测量技术*, 2021, 44(12): 57-62.
- [14] ZHANG H, SONG Y. Robust sound event recognition using convolutional neural networks[J]. *Speech and Signal Processing, Brisbane, QLD*, 2015: 559-563.
- [15] 王若平, 房宇, 陈达亮, 等. 基于小波包FBank图谱与CNN的车内异响识别研究[J]. *重庆理工大学学报(自然科学)*, 2020, 34(7): 1-9.
- [16] BOUJELBEN O, BAHOURA M. Efficient FPGA-based architecture of an automatic wheeze detector using a combination of MFCC and SVM algorithms[J]. *Journal of Systems Architecture*, 2018, 88: 54-64.
- [17] KESKAR N S, MUDIGERE D, NOCEDAL J, et al. On large-batch training for deep learning: Generalization gap and sharp minima [J]. *CoRR*, 2016, abs/1609.04836.
- [18] VAN DER MAATEN L, HINTON G. Visualizing data using t-SNE[J]. *Journal of Machine Learning Research*, 2008, 9(11): 2579-2605.

作者简介

付孟新, 硕士研究生, 主要研究方向为信号处理、故障诊断。

E-mail: fmengxin@foxmail.com

郭世伟(通信作者), 副教授, 硕士生导师, 主要研究方向为测控技术、机械故障诊断、人工智能。

E-mail: guoswall@163.com