

2022 第二届天府杯全国大学生数学建模竞赛论文

题 目 基于图像识别与多模型建模的故障检测技术

摘 要

通过数据收集器对故障数据进行收集,并依据故障数据进一步进行仪器设备的故障诊断是一种常见的技术收段。如何利用故障数据进行故障类型的识别是故障诊断中的一项重要工作。本文主要根据提供的仪器设备故障数据,对数据进行分析、去噪、特征提取,并进一步分别建立基于无监督和基于有监督的故障类型识别模型,具体如下:

针对问题一,我们首先对数据进行了初步的观察分析,并构建了数据矩阵便于后续研究;然后通过滑动平均法、Savitzky-Golay 滤波、小波变换(包括固定阈值,软阈值和硬阈值三种策略)对故障数据进行去噪,我们在给定的均方误差(MSE)、和方差(SSE)、均方根(RMSE)的基础上,又选取了信噪比(SNR)、平滑度(R)、互相关系数(ρ)作为去噪效果的评价指标。综合比较各方法的去噪效果,我们最终采用 Savitzky-Golay 滤波方法进行去噪。

针对问题二,我们选定了 10 类时域特征判据,分别为均值、标准差、方根幅值、均方根、偏度、峭度、峰值因子、裕度因子、波形因子、脉冲指数。通过提取的特征,我们首先分别对附件一和附件二中的 A, B 两类故障数据的各种特征分布以及特征间相关性进行了分析;最后我们将全部数据对应的特征判据进行了汇总。

针对问题三,我们选取了 3 种无监督分类算法进行实验,分别是 K-means、DBSCAN、Spectral 算法;基于准确率、召回率、耗时三项评价指标,我们另外选取了轮廓系数、兰德系数、调整兰德系数作为无监督预测结果的评价指标;除此之外,我们还对输入数据来源(原始数据或提取的特征数据)对预测结果的影响进行了实验分析。最终得出:在 K-means 模型上,采用我们提取的特征数据作为输入可以实现无监督条件下故障类型识别的最优解。

针对问题四,我们选取了 8 种有监督分类算法进行实验,分别是支持向量机(SVM)、Xgboost、随机森林(RF)、多层感知机(MLP)、决策树(DT)、高斯贝叶斯(GaussianNB)、K 最近邻(KNN)以及 Adaboost;基于准确率、召回率、耗时三项评价指标,我们另外选取了精准率(Precision)、F1-score、AUC 作为有监督预测结果的评价指标;我们不仅进行了多种算法建模实验,并且针对训练数据占比、训练数据来源(原始数据或提取的特征数据)对实验结果的影响进行了实验分析。最终得出:在 MLP 模型上,采用我们提取的特征作为输入,设置 80%训练样本可以实现有监督条件下故障类型识别的最优解。

针对问题五,我们选择了第三问,首先我们将一维信号数据转换为二维图像形式,并可视化,结果表明两类故障数据在视觉层面具有显著差异;其次,我们通过知识迁移技术(微调策略)结合深度预训练模型(AlexNet 和 ResNet-18)进行图像分类实验。

关 键 词: 时域特征挖掘, 多模型建模, 迁移学习, 图像分类

一、问题重述

问题一：针对附件一和附件二中的原始数据进行信号去噪处理，并利用全面恰当的评价指标（已给定 3 项评级指标，需自行添加至少 3 项评价指标）对去噪后数据进行去噪效果评价；

问题二：针对附件一和附件二中的全部数据进行信号的特征提取（特征判据数量不得少于 10），并将特征值进行汇总；

问题三：基于无监督或半监督方法，对附件一和附件二中数据进行二分类，并利用全面恰当的评价指标（已给定 3 项评级指标，需自行添加至少 3 项评价指标）对二分类的预测结果进行评价，保证预测结果的准确率均值在 90%以上，准确率标准差在 10 以内；

问题四：基于有监督方法，对附件一和附件二中数据进行二分类，并利用全面恰当的评价指标（已给定 3 项评级指标，需自行添加至少 3 项评价指标）对二分类的预测结果进行评价，保证预测结果的准确率均值在 95%以上，准确率标准差在 5 以内；

问题五：本题共有三个选择题，题 1 要求对比分析信号去噪前后的信号特征在信号分类中的区别和影响；题 2 要求讨论特征的数量以及质量对特征识别的影响，以及给出解决办法；题 3 要求引入图像识别技术解决问题三或问题四。

二、问题分析

（一）问题 1 的分析

在仪器设备故障诊断中，通常采用数据收集器收集到的信号数据进行故障的诊断。但是采集到的数据中往往掺杂有杂声，噪声的存在严重干扰了真实信号，且对于后续的故障诊断早成很大的困扰。因此，对于采集到的原始数据进行去噪是进行后续一系列实验的重要前提。问题 1 首先要求我们对于附件中的数据去噪并对去噪效果进行评价。信号数据的去噪方法以及去噪效果评价指标有多种，因此选择恰当的去噪方法以及去噪效果评价指标是去噪问题中的关键。我们可以结合故障信号数据的特点，选择多种去噪方法分别对数据进行去噪，并选用多种评价指标恰当全面的对去噪效果进行评价。

（二）问题 2 的分析

数据收集器采集到的信号数据中包含了大量的统计参数，而这大量的数据信息无法直接用于机器学习模型中，需要通过特征提取技术提取原始数据中的重要特征信息。通过特征提取能够将原始的数据整合成新的特征，既降低了数据维度避免数据冗余又能够保留数据的重要信息。因此，进行特征提取有利于提升后续故障检测的精度和效率。问题 2 要求我们选择恰当的特征判据并对附件中的数据去噪。我们可以根据故障信号数据一维、震荡等特点，选择恰当全面的特征判据，然后进行特征的提取。对提取到的特征进行特征分析，进一步确定提取到的特征的有效性。

（三）问题 3 的分析

无监督的分类算法即在样本数据没有任何标签（监督信息）的情况下，通过学习发现样本数据内在的特征，并自动对样本数据进行分类。经典的基于无监督的分类算法主要是聚类算法，我们可以通过选取多个不同的聚类算法进行无监督建模；其次，我们需要选择恰当全面的评价指标对无监督实验结果进行评价分析。

（四）问题 4 的分析

有监督的分类算法即在有标签的训练数据中构建分类模型，利用构建的模型对新的数据进行分类。有监督分类算法的准确性主要有三个关键：1）算法的有效性；2）数据的有效性；3）评价指标的合理性。因此，我们围绕以上三点展开实验。首先，我们需要选择多种有监督分类算法进行实验；其次我们需要验证训练数据对实验的影响（包括训

练数据比例、训练数据来源等);最后我们需要选择恰当全面的评价指标对实验结果进行评价分析。

(五) 问题 5 的分析

我们选择了问题 5 中的第三个问题:引入图像识别技术解决第四问。首先,我们将附件提供的一维数据可视化二维图像;其次,采用多种基于深度学习的神经网络解决图像识别问题。

三、模型假设

1. 假设题目所给的数据真实可靠;
2. 假设各类型故障数据之间差异明显;
3. 假设同一故障类别内的数据分布一致,且没有较多噪声较大的离群点;
4. 假设故障数据为均匀采样,对原故障类型分布具有较好的表征性。

四、定义与符号说明

符号	含义
$x(\cdot)$	原始数据
$y(\cdot)$	去噪后数据
N	数据长度
$cov(\cdot, \cdot)$	协方差
$D(\cdot)$	数据方差
L	滑动窗口大小

五、模型的建立与求解

(一) 准备工作

➤ 数据的处理

通过对附件数据的观察,我们发现 A 类数据附件和 B 类数据附件的存储方式不同(A 数据按行存储,B 数据按列存储),为了便于后续的数据使用,我们首先将 txt 数据格式转换成矩阵方式存储,即 $100 * 4096$ 矩阵。

(二) 问题 1 的去噪模型以及评估指标

➤ 模型的建立与求解

对于问题一中去噪效果的评价指标,我们在给定的均方误差(MSE)、和方差(SSE)、均方根(RMSE)的基础上,又选取了信噪比(SNR)、平滑度(R)、互相关系数(ρ)评价指标。

(1) 信噪比(SNR)是指信号功率与噪声功率的比值:

$$SNR = 10 \log \frac{\sum_{n=1}^N x^2(n)}{\sum_{n=1}^N [x(n) - y(n)]^2},$$

其中, $x(n)$ 为原始信号数据, $y(n)$ 表示去噪后信号数据, N 表示信号数据长度。信噪比越大表示去噪效果越好。

(2) 平滑度(R)是指去噪后信号的差分数方差根与原始信号差分数方差根的比值:

$$R = \frac{\sum_{n=1}^{N-1} [y(n+1) - y(n)]^2}{\sum_{n=1}^{N-1} [x(n+1) - x(n)]^2},$$

其中, $x(n)$ 表示原始信号数据, $y(n)$ 表示去噪后信号数据, N 表示信号数据长度。平滑度越小表示去噪效果越好。

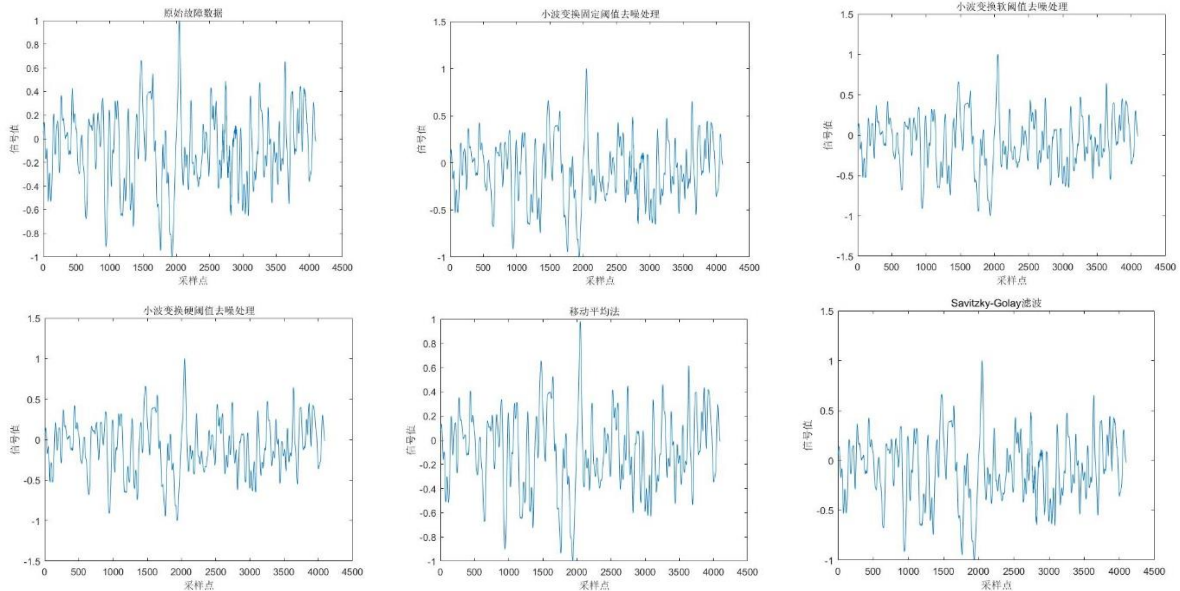
(3) 互相关系数(ρ)是指原始信号与去噪后信号之间的相关性:

$$\rho = \frac{cov(x, y)}{\sqrt{D(x)}\sqrt{D(y)}},$$

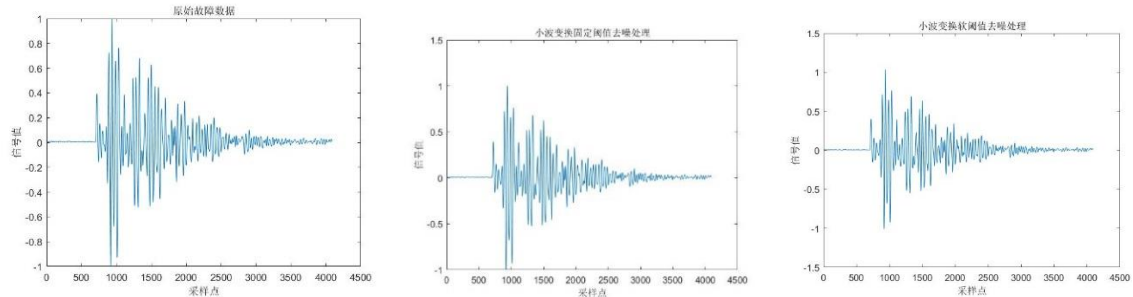
其中, x 为原始信号数据, y 为去噪后信号数据, $cov(x, y)$ 表示协方差, $D(x)$ 表示原始信号数据的方差, $D(y)$ 表示去噪后信号数据的方差。互相关系数越接近于 1 表示去噪效果越好。

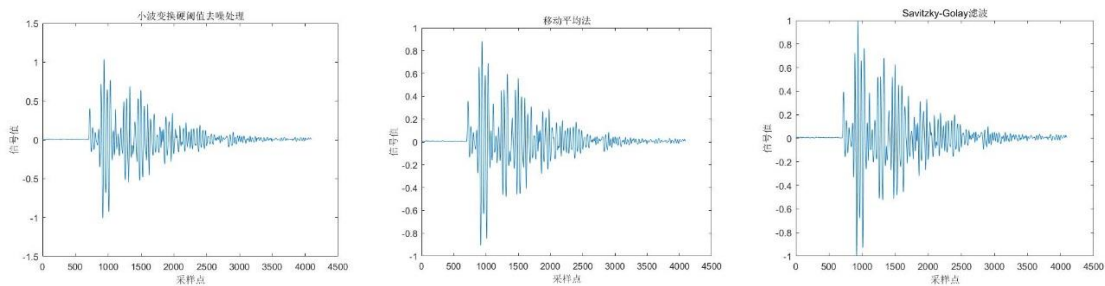
我们分别选取 A 类数据的第一个样本以及 B 类数据的第一个样本, 分别测试 (1) 滑动平均法; (2) Savitzky-Golay 滤波(简称 SG 滤波); (3) 小波变换(包括固定阈值, 软阈值和硬阈值三种策略)对故障数据进行去噪, 并通过六种评估指标进行对比。

A 类样本原始故障数据波形图与去噪后波形图如下:



B 类样本原始故障数据波形图与去噪后波形图如下:





同时，我们给出了不同去噪策略在各个评估指标下的性能，如下表：

	MSE	SSE	$RMSE$	SNR	R	ρ
固定阈值	1.05e-6	0.0043	0.0010	50.3012	1.0124	1.0000
软阈值	9.67e-5	0.3960	0.0098	30.6666	1.1853	0.9995
硬阈值	9.67e-5	0.3960	0.0098	30.6666	1.1853	0.9995
移动平均	2.01e-4	0.8243	0.0142	27.4824	1.3533	0.9991
SG 滤波	3.62e-7	0.0015	0.0006	54.9364	1.0220	1.0000

表 1 A 类数据第一个样本

	MSE	SSE	$RMSE$	SNR	R	ρ
固定阈值	4.85e-9	1.99e-5	6.96e-5	67.7095	1.0000	1.0000
软阈值	8.94e-5	0.3663	0.0095	25.0504	0.9588	0.9984
硬阈值	8.94e-5	0.3663	0.0095	25.0504	0.9588	0.9984
移动平均	3.14e-4	1.2877	0.0177	19.5901	1.2783	0.9987
SG 滤波	1.9e-10	7.78e-7	1.39e-5	81.7283	1.0001	1.0000

表 2 B 类数据第一个样本

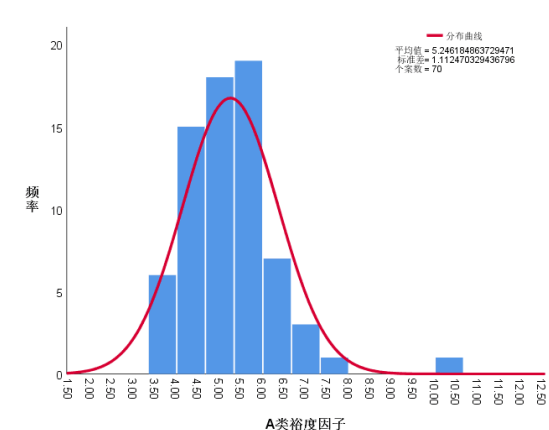
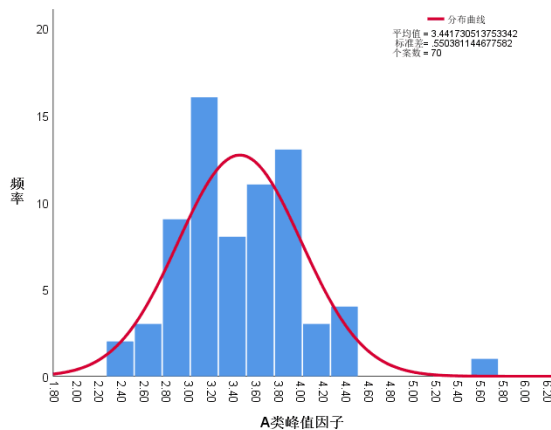
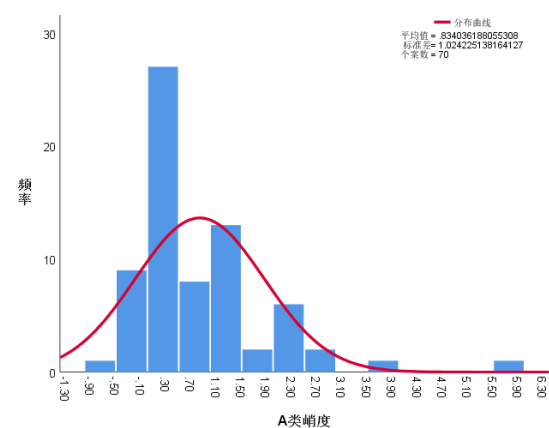
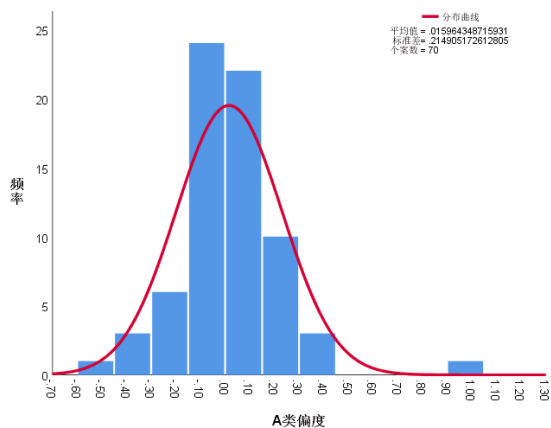
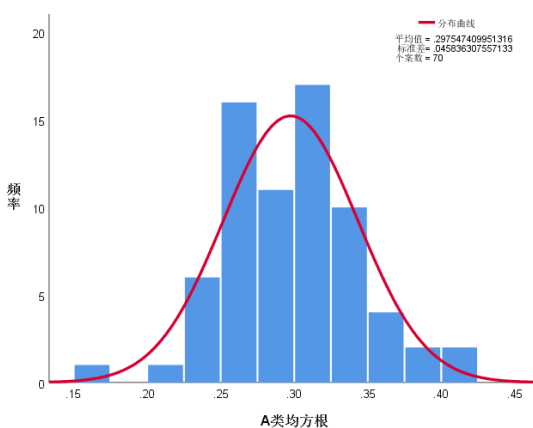
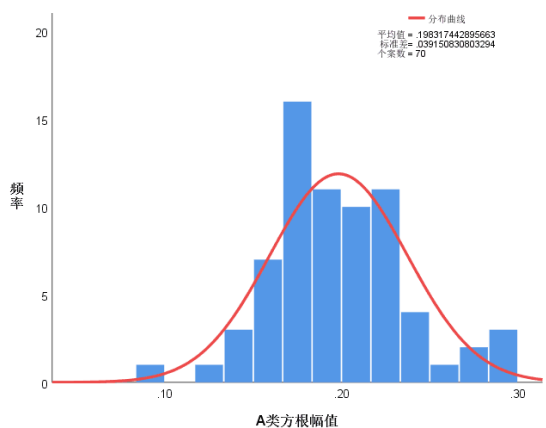
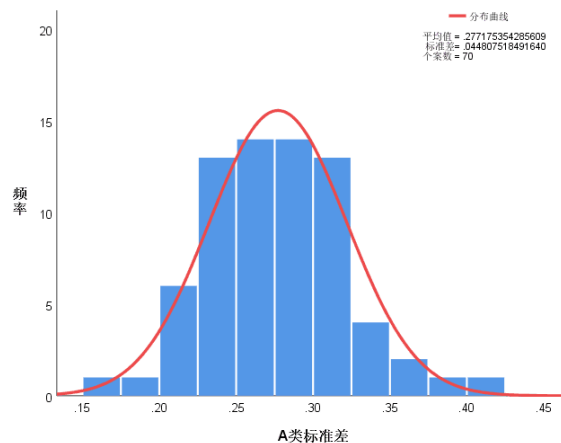
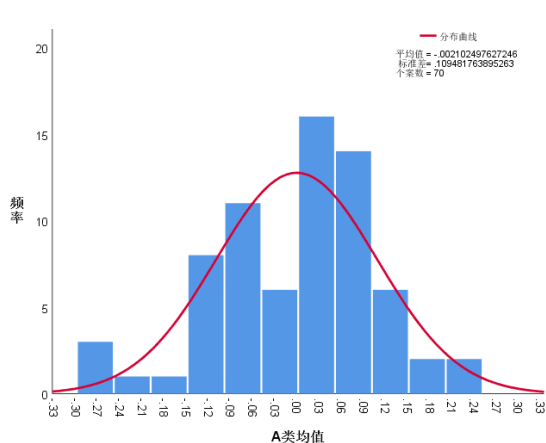
综上所述，我们采用 Savitzky-Golay 滤波对原始故障数据进行去噪处理，去噪后的数据结果参考附件 1 中表 1-1 和表 1-2。

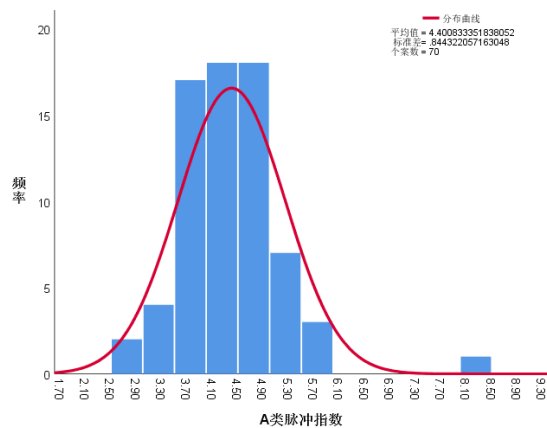
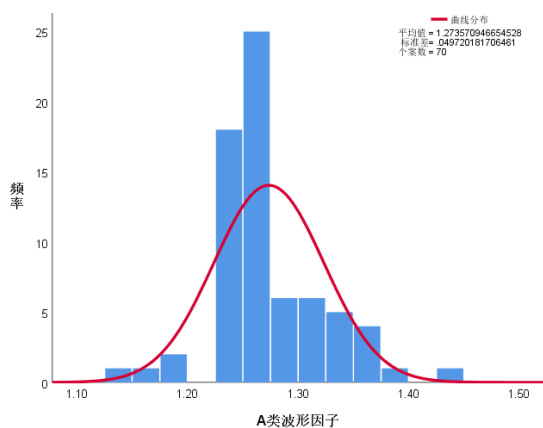
（三）问题 2 的特征提取模型以及特征值

➤ 模型的建立与求解

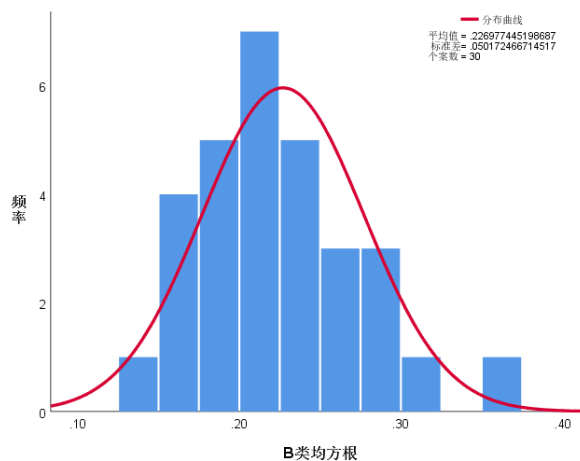
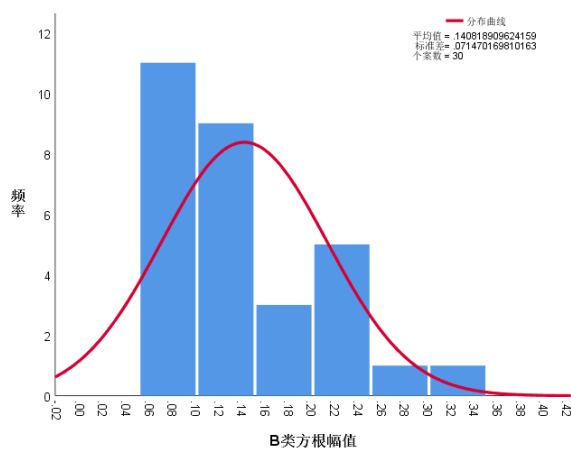
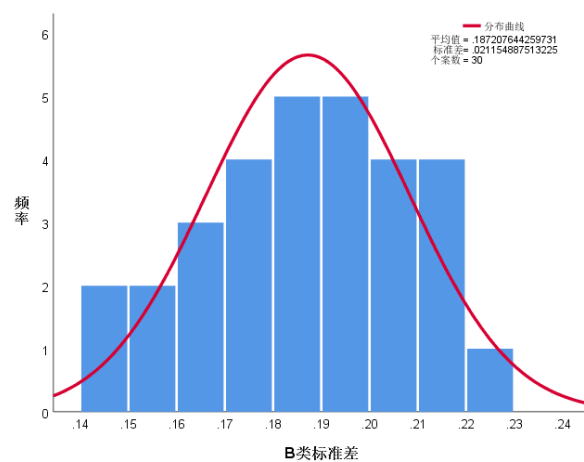
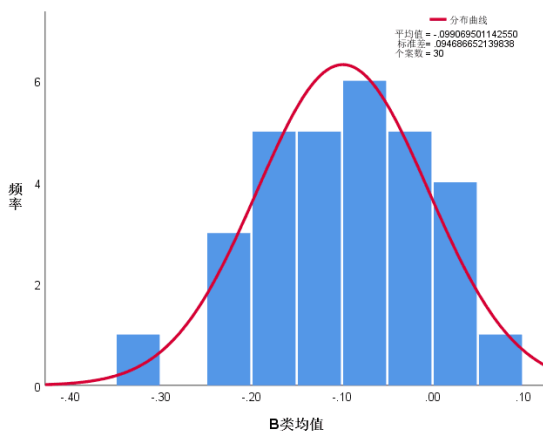
在对附件一和附件二中的数据进行特征提取时，我们选定了 10 类特征判据，分别为均值、标准差、方根幅值、均方根、偏度、峭度、峰值因子、裕度因子、波形因子、脉冲指数。通过提取的特征，我们首先分别对附件一中的 A 类故障数据和附件二中的 B 类故障数据的各种特征分布进行了分析，具体分析情况如下：

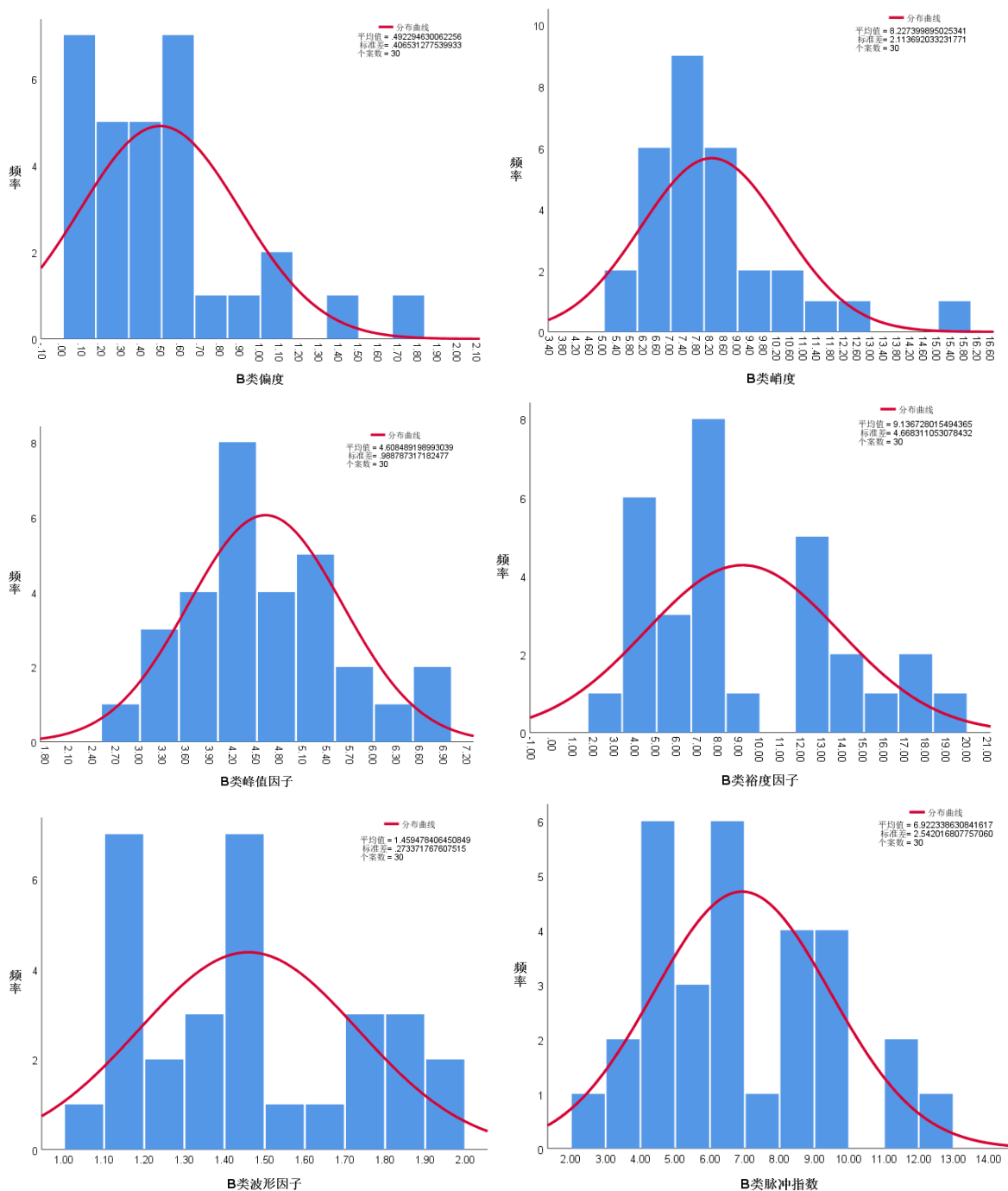
A 类数据：





B 类数据:





同时，我们对 A、B 两类数据各特征之间的相关性进行了分析，并将分析结果通过热力图进行了展示，如图所示：

A 类数据：



B 类数据:



最终，通过特征提取，我们获得了如下 10 种特征值（此处仅对每类数据各展示一条数据的特征值，详情请见附件 1 中表 2）：

数据 类型 \ 特征	均值	标准 差	方 根 幅 值	均 方 根	偏 度	峭 度	峰值 因子	裕 度 因子	波 形 因子	脉 冲 指 数
A 类 数据	-0.087	0.324	0.226	0.336	-0.123	0.084	2.979	4.424	1.261	3.757
B 类 数据	0.007	0.169	0.053	0.169	0.019	9.132	5.912	18.960	1.932	11.426

（四）问题 3 的无监督分类模型以及评估指标

➤ 模型的建立与求解

针对问题 3 中预测结果的评价指标，在给定的准确率、召回率、耗时的前提下，我们另外选取了轮廓系数、兰德系数、调整兰德系数评价指标。

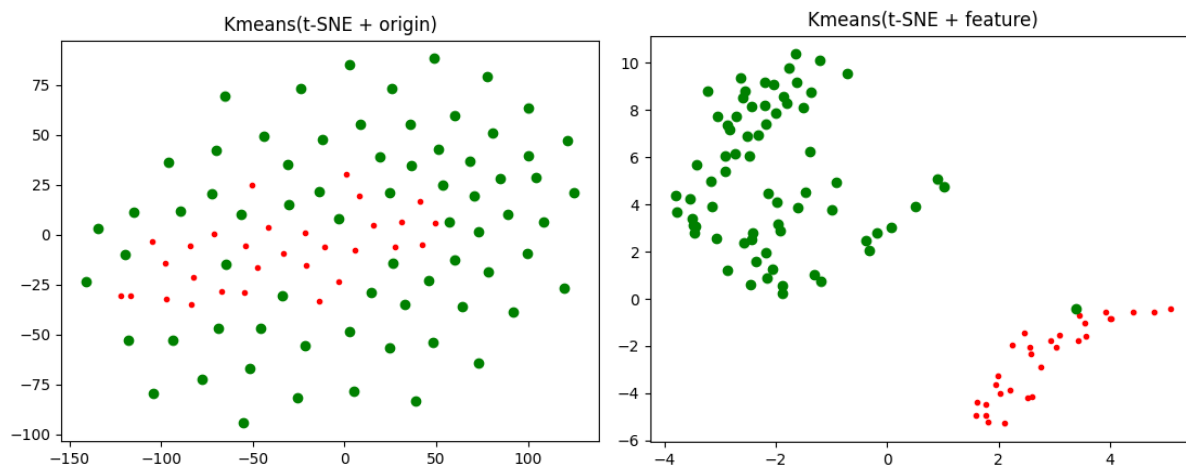
轮廓系数描述的是簇的内外差异，兰德系数表示的是分类结果与真实情况的吻合度，调整兰德系数衡量的是两个数据分布的吻合程度。

对于无监督预测模型，我们采用经典的无监督算法——聚类，其基本思想为：对于给定的数据样本，设置聚类（簇）的个数，初始化每个样本所属的类别，根据设定的规则不断地迭代并划分数据类别。我们采用了 K-means、DBSCAN、Spectral 算法进行无监

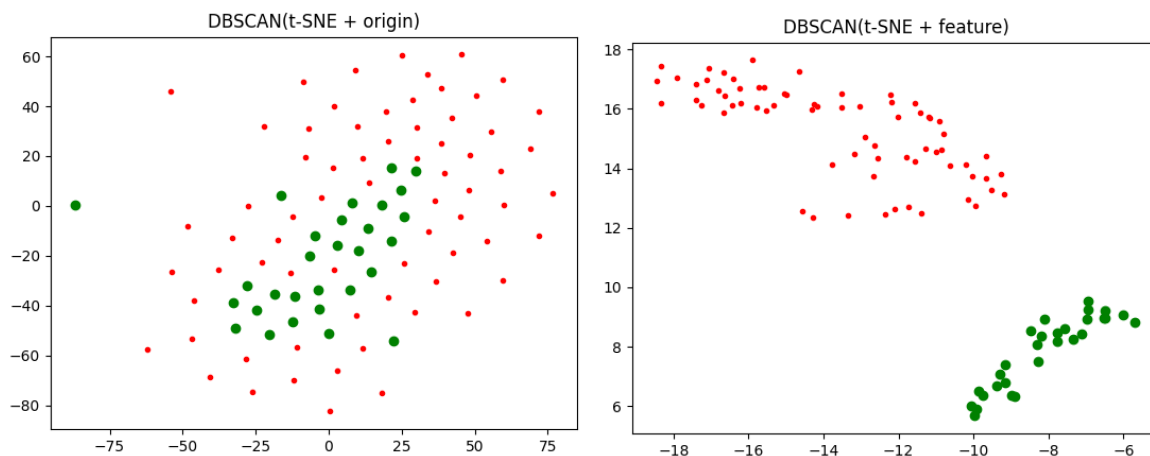
督模型建立。

为了研究原始输入数据和提取的特征数据对无监督模型预测效果的影响,我们将这两类数据分别作为输入进行实验,并利用 t-SNE 技术结合聚类输出的伪标签进行可视化,结果如下(详情见附件 1 中表 3):

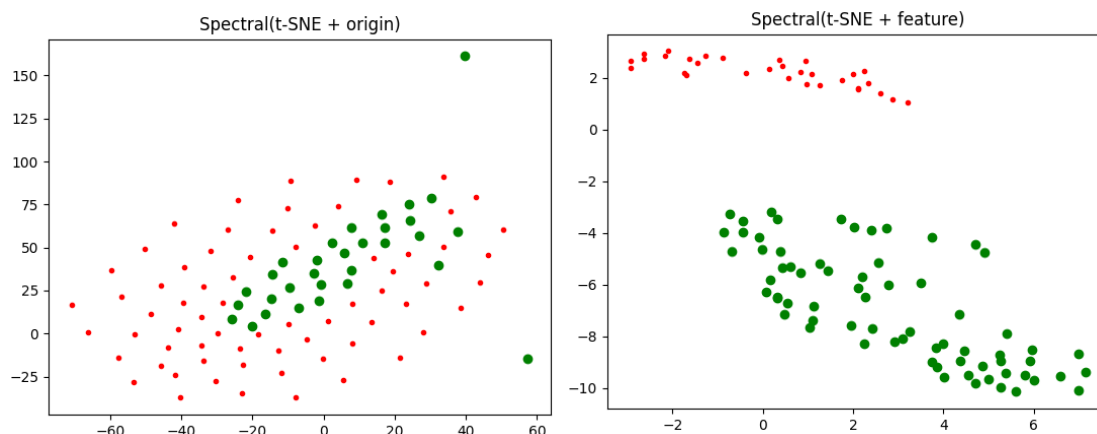
K-means:



DBSCAN:



Spectral:



通过上图可得，以提取的特征数据作为输入数据可以实现更好的聚类。这同时证明了我们提取的特征数据的有效性和必要性。我们选取 K-means 进行 A 和 B 类数据的聚类 and 预测，由于无监督方法无法输出与人类定义类别编号的对应，只能获取聚类后的类簇编号，因为我们设置与 A 类第一个样本最近的类簇中心作为 A 类伪标签，反之另外的为 B 类伪标签。我们进行了 100 次测试实验，实验结果见附件 1 中表 3。

（五）问题 4 的有监督分类模型以及评估指标

➤ 模型的建立与求解

针对问题 4 中预测结果的评价指标，在给定的准确率、召回率、耗时的前提下，我们另外选取了精准率（Precision）、F1 得分（F1-score）、AUC 评价指标。

精准率指的是预测结果中预测正确的占比，F1-score 是精准率和召回率的调和平均数，AUC 是 ROC 曲线与坐标轴形成的面积，而 ROC 曲线是反映算法特异性和敏感性关系的曲线。

为了最终能够选择出最佳的有监督预测方法，我们选取了 8 种有监督分类算法进行实验，分别是支持向量机（SVM）、Xgboost、随机森林（RF）、多层感知机（MLP）、决策树（DT）、高斯贝叶斯（GaussianNB）、K 最近邻（KNN）以及 Adaboost。

为了验证训练样本占比对预测结果的影响，我们利用问题 2 中提取到的特征，并分别采取了 20% 训练样本、50% 训练样本、80% 训练样本进行实验对比，实验结果如下表所示：

实验设置 1：提取的特征数据+20%训练样本

	准确率 (Acc)	召回率 (Recall)	耗时 (ms)	F1 得分	精准率 (Precision)	AUC 值
SVM	0.9875	1.0000	0.0000	0.9804	0.9615	0.9909
Xgboost	0.9875	1.0000	2.9995	0.9804	0.9615	0.9909
RF	0.9125	0.7600	2.9997	0.8444	0.9500	0.8709
MLP	1.0000	1.0000	1.0011	1.0000	1.0000	1.0000
DT	0.9875	1.0000	0.0000	0.9804	0.9615	0.9909
GaussianNB	0.9625	1.0000	0.0000	0.9434	0.8929	0.9727
KNN	0.9500	0.8400	1.9987	0.9130	1.0000	0.9200
Adaboost	0.9875	1.0000	1.0004	0.9804	0.9615	0.9909

实验设置 2: 提取的特征数据+50%训练样本

	准确率 (Acc)	召回率 (Recall)	耗时 (ms)	F1 得分	精准率 (Precision)	AUC 值
SVM	0.9800	1.0000	0.0000	0.9630	0.9286	0.9865
Xgboost	0.9800	1.0000	2.0008	0.9630	0.9286	0.9865
RF	0.9400	0.8462	2.0885	0.8800	0.9167	0.9096
MLP	0.9800	1.0000	0.0000	0.9630	0.9286	0.9865
DT	0.9800	1.0000	0.0000	0.9630	0.9286	0.9865
GaussianNB	0.9600	1.0000	0.0000	0.9286	0.8667	0.9730
KNN	0.9800	1.0000	2.0008	0.9630	0.9286	0.9865
Adaboost	0.9800	1.0000	1.0021	0.9630	0.9286	0.9865

实验设置 3: 提取的特征数据+80%训练样本

	准确率 (Acc)	召回率 (Recall)	耗时 (ms)	F1 得分	精准率 (Precision)	AUC 值
SVM	1.0000	1.0000	0.0000	1.0000	1.0000	1.0000
Xgboost	1.0000	1.0000	1.9989	1.0000	1.0000	1.0000
RF	1.0000	1.0000	1.9996	1.0000	1.0000	1.0000
MLP	1.0000	1.0000	0.0000	1.0000	1.0000	1.0000
DT	1.0000	1.0000	0.0000	1.0000	1.0000	1.0000
GaussianNB	1.0000	1.0000	0.0000	1.0000	1.0000	1.0000
KNN	1.0000	1.0000	1.0004	1.0000	1.0000	1.0000
Adaboost	1.0000	1.0000	6.9994	1.0000	1.0000	1.0000

通过上述表格中展示的实验结果可知,当训练样本占比达到 80%时,预测结果最好,各个评价指标均能够达到最优结果。因此,我们选取实验设置 3 作为我们最终的实验设置。

同时,为了验证特征提取的有效性以及必要性,我们进行了实验设置 4,在该实验中,我们将 4096 维原始故障数据直接作为输入,采用 80%训练样本,实验结果如下表所示:

实验设置 4: 原始 4096 维的故障数据+80%训练样本

	准确率 (Acc)	召回率 (Recall)	耗时 (ms)	F1 得分	精准率 (Precision)	AUC 值
SVM	0.6000	0.8333	1.0111	0.5556	0.4167	0.6667
Xgboost	0.9000	0.8333	5.9988	0.8333	0.8333	0.8810
RF	0.9000	0.6667	2.9986	0.8000	1.0000	0.8333
MLP	0.8500	0.8333	0.9990	0.7692	0.7143	0.8452
DT	0.8000	0.8333	1.0009	0.7143	0.6250	0.8095
GaussianNB	1.0000	1.0000	0.9992	1.0000	1.0000	1.0000
KNN	0.3500	1.0000	2.0001	0.4800	0.3158	0.5357
Adaboost	1.0000	1.0000	11.9994	1.0000	1.0000	1.0000

通过上表,我们可以得出,在分类实验中,利用提取的特征作为输入比直接利用原始数据作为输入能够获得更好的实验效果。这既证明了特征提取的必要性,同时验证了我们在问题 2 中提取到的特征的有效性。

最终, 综合分析上述四个表格的实验结果, 证明采用我们提取的特征可以更加有效地对故障数据进行识别和区分, 并选取 MLP 作为有监督学习模型进行 100 次测试实验, 实验结果见附件 1 中表 4。

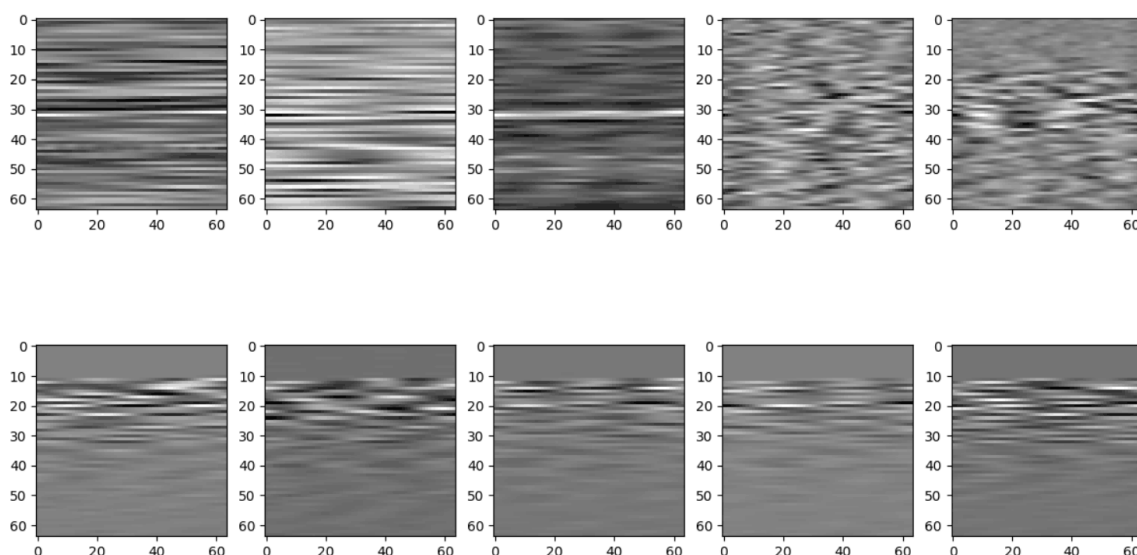
(六) 问题 5 的图像识别技术解决分类预测

➤ 模型的建立与求解

我们选择了问题 5 中的第三个问题: 引入图像识别技术解决第四问。

首先, 我们将附件提供的数据可视化二维图像, 具体重构过程为: 将长度为 4096 的一维原始数据进行 Reshape 操作得到为 64*64 的二维特征图, 由于原始数据的取值量纲为 $[-1, 1]$, 为了便于可视化, 将重构后的特征图数据量纲进行转换, 将 $[-1, 1]$ 平移缩放到 $[0, 1]$, 并乘以 255。

这里我们分别在 A 类数据和 B 类数据中分别选取了 5 条数据进行展示 (第一行是 A 类数据, 第二行是 B 类数据):

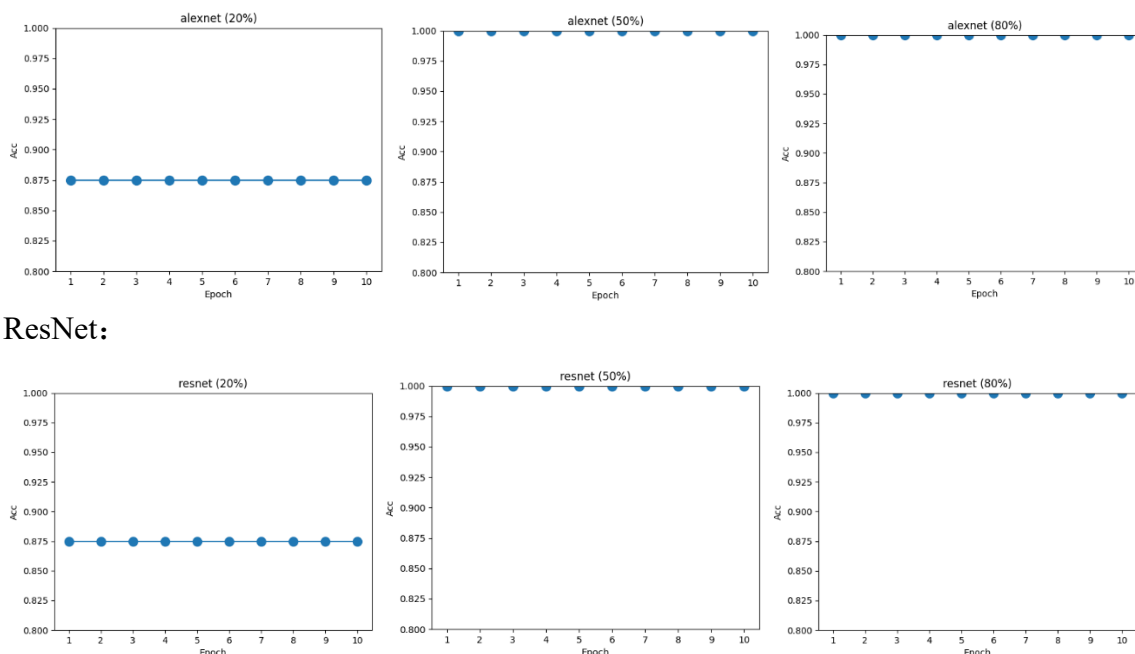


通过上图可知, A 类故障数据和 B 类故障数据在可视化图像上的区别明显, 这就为利用图像识别技术进行故障类型识别奠定了良好的基础。

其次, 我们采用预训练深度神经网络模型进行图像分类, 选择常见的 Alexnet 和 ResNet-18 网络进行建模。模型训练过程中, 利用 SGD 优化器更新参数, 采用了二分类交叉熵损失函数进行有监督训练, 训练和测试样本批次大小均设置为 16。为了提升模型效率和精度, 我们利用迁移学习技术, 即预训练加微调的策略进行网络训练。

按照问题 4 中的实验设置, 我们同样进行了 20% 训练数据、50% 训练数据和 80% 训练数据的划分, 两个网络的训练效果如图所示:

AlexNet:



ResNet:

六、模型评价与推广

（一）模型的优点

- (1) 本文充分挖掘了时域信号上的特征字段，并验证了其有效性（相对于原始信号输入）。
- (2) 本文采用了多种有监督和无监督模型进行建模并测试 6 种评估指标，全面地衡量不同模型的学习性能，并选择最优模型进行测试。
- (3) 本文将原始一维信号转换为二维矩阵，并利用图像识别技术，提高局部和全局信息提取的能力，并集合知识迁移策略测试二分类性能。

（二）模型的缺点

- (1) 本文对于数据的特征挖掘信息有待于进一步提升，需要充分理解频域和时域的有关理论，提出更具有隐式语义表达能力的特征字段。
- (2) 缺少对多种机器学习模型的复杂集成技术，单一检测模型可能无法捕获全面的潜在特征信息。
- (3) 基于知识迁移技术的预训练模型微调策略缺少对故障数据的特定性适配，缺少对局部纹理信号的挖掘和探索。

（三）模型的改进

本次建模所用的数据仅有两个类别，并且数据量较少，这两类故障的信号特征差异性较大，因此使用一些基础模型便可以得到十分可观的分类性能。然而在实际应用中，故障类型往往是复杂且多形态的，因此，1) 一些故障的信号特征可能存在一定的相似性；2) 故障信号与非故障信号差异很小。在此提出以下可能解决方法：

- 1、利用多种机器模型融合技术或某些多模型决策策略，提高分类鲁棒性。不同的模型对不同的特征感知能力是有差异的。
- 2、利用多种聚类等无监督策略进行数据建模。当故障类别很多时，不同类别空间分布的流形结构可能存在差异，并且可能存在样本不均衡的实际场景，直接进行简单的多分类可能导致一些重要但不常见的故障类别识别能力降低。

3、基于(5)的建模基础,将一维信号转换为二维形式,并采用图像识别技术进行检测。对于相似性较大的故障类别,应该更加关注低阶的局部纹理信息,才能更好地感知信号的局部差异。Alexnet 结构更加关注高阶语义信息,Resnet 虽然引入残差结构,将低阶信息信号传向高阶特征提取的卷积层,但是最终的输出带有较多的高阶语义信息。因此,可以利用 Transformer 结构,对信号区域进行分割得到 Patch 序列,并通过 Self-Attention 结构,挖掘 Patch 之间的关联信息。

4、利用 GCN 等图网络模型对故障数据进行结构建模,利用同类别的邻域样本增强特征表示,从而更好地挖掘本类的故障数据的原型表征,该原型表征可以更好地表示该类故障数据的潜在的一致信息。

八、参考文献

- [1] 苏成志,陈洪印,孟凡一,等. 新阈值二进小波去噪算法在齿轮信号中的应用[J]. 计算机工程与应用, 2014, 50(018):206-209.
- [2] 朱建峰,郭伟剑,刘俊邦,等. 小波变换在多层复合材料的超声波检测中的应用[J]. 兵器装备工程学报, 2014, 000(007):67-70.
- [3] 刘垚,楼光海,黄微.结合双树复小波变换和滑动平均滤波的心电信号去噪方法[J].电子测量技术,2018,41(19):112-117.DOI:10.19651/j.cnki.emt.1801706.
- [4] 张宝,周麟奉,杨涛.Savitzky-Golay 滤波与局部均值分解相结合的滚动轴承故障诊断方法[J].机械设计与制造,2017(03):192-196.DOI:10.19356/j.cnki.1001-3997.2017.03.052.
- [5] 董广凯. 基于改进小波变换的地震资料信噪比提升方法研究[D].西安石油大学,2021.DOI:10.27400/d.cnki.gxasc.2021.000842.
- [6] 李佩佩,崔凤英.基于二分 K-means 聚类的曲率分级点云数据精简优化算法研究[J/OL]. 电子测量技术:1-7[2022-03-21].
- [7] 陈文龙,时宏伟.基于 KD 树改进的 DBSCAN 聚类算法[J].计算机系统应用,2022,31(02):305-310.DOI:10.15888/j.cnki.csa.008310.
- [8] 孙伟伟,杨刚,彭江涛,孟祥超.鲁棒多特征谱聚类的高光谱影像波段选择[J].遥感学报,2022,26(02):397-405.
- [9] Cherkassky V, Ma Y. Practical selection of SVM parameters and noise estimation for SVM regression[J]. Neural networks, 2004, 17(1): 113-126.
- [10]Chen T, He T, Benesty M, et al. Xgboost: extreme gradient boosting[J]. R package version 0.4-2, 2015, 1(4): 1-4.
- [11]Liu K, Hu X, Zhou H, et al. Feature Analyses and Modeling of Lithium-Ion Battery Manufacturing Based on Random Forest Classification[J]. IEEE/ASME Transactions on Mechatronics, 2021, 26(6): 2944-2955.
- [12]Raju M, Gopi V P, Anitha V S. Multi-class Classification of Alzheimer's Disease using 3DCNN Features and Multilayer Perceptron[C]//2021 Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET). IEEE, 2021: 368-373.
- [13]Charbuty B, Abdulazeez A. Classification based on decision tree algorithm for machine learning[J]. Journal of Applied Science and Technology Trends, 2021, 2(01): 20-28.
- [14]Wang Q, Pan L, Shen J, et al. Power plant data filtering based on Gaussian naive Bayesian classification and prediction error method[C]//2019 Chinese Automation Congress (CAC). IEEE, 2019: 1490-1495.
- [15]Guo G, Wang H, Bell D, et al. KNN model-based approach in classification[C]//OTM Confederated International Conferences" On the Move to Meaningful Internet Systems".

Springer, Berlin, Heidelberg, 2003: 986-996.

[16]Schapire R E. Explaining adaboost[M]//Empirical inference. Springer, Berlin, Heidelberg, 2013: 37-52.

[17]Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25.

[18]Targ S, Almeida D, Lyman K. Resnet in resnet: Generalizing residual architectures[J]. arXiv preprint arXiv:1603.08029, 2016.

九、附录

1、支撑材料——附件 1

表 1-1, 表 1-2, 表 2, 表 3, 表 4

2、支撑材料——代码

实验代码