

云模型和集成分类结合的故障数据不平衡学习*

马森财, 赵荣珍, 吴耀春
(兰州理工大学机电工程学院 兰州, 730050)

摘要 针对故障数据集不平衡而导致的误分类问题,在分析了不平衡数据对传统分类器影响的基础上,提出了一种基于高斯云模型正向、逆向云发生算法的样本再生成技术。首先,针对样本较少的类别,以现有样本特征值为逆向云算法的输入,计算出特征云模型的期望 E_e 、熵 E_e 和超熵 H_e 这3个指标;其次,以 E_e 、 E_e 和 H_e 为正向云发生算法的输入,衍生出数据量远大于原有样本的云滴 (x_i, y_i) ,采集若干云滴的 x 值作为新的样本特征值,补充了样本数量较少的类,在数据层面解决了不平衡问题;然后,借助集成极限学习机(ensemble extreme learning machine,简称E-ELM)对补充后的平衡数据集进行分类学习,在算法层面提高了最终的分类精度;最后,在一个滚动轴承故障数据集上验证了所提方法的有效性。

关键词 滚动轴承;分类;不平衡学习;集成学习
中图分类号 TH165+.3;TP391;TP18

引言

大数据时代的来临为旋转机械故障数据的处理、存储和利用带来了新的机遇和挑战^[1]。一直以来,分类问题是故障诊断甚至是机器学习研究领域的重要组成部分^[2-4]。然而,在采集的旋转机械状态数据中,某些类别的数据数量远少于其他一些类别的数据数量,但这些样本数量较少的类别往往又十分重要,不可忽略。此现象造成了不平衡数据集的产生,传统分类器在这种类间分布不平衡的数据集上训练时常常会出现分类面偏移,导致故障误分类,使模型辨识精度降低。因此,对不平衡数据分析技术进行深入研究,是工业大数据挖掘的重要前提之一,对旋转机械智能故障诊断技术的发展具有积极的促进作用。

针对不平衡数据的学习,相关研究主要从数据层面、分类算法层面来找寻和探讨其解决方法。文献[5]提出了一种不平衡分类与深度学习相结合的疾病自动检测方法,首先使用Borderline-SMOTE算法对训练集做了平衡处理,然后设计了一种一维深度卷积神经网络,并利用平衡处理后的训练集进行训练,有效避免了由于多数类样本较多而造成的过拟合。文献[6]提出了基于密度聚类与多工序制造特征的质检数据过采样方法,在少数类簇中进行

了数据生成,使用多数类簇剔除了无效数据,解决了汽车零部件质检数据存在的合格与不合格产品数量不平衡问题,提高了汽车零部件的检测效率。文献[7]提出了一种混合方法来解决不平衡数据学习问题,利用聚类方法指导了新的样本生成,并构建了集成模型以进一步提升算法的性能。上述研究成果都有效解决了不平衡数据带来的困扰,且经归纳后发现,结合集成学习的不平衡学习算法在解决类别分布不平衡的问题上卓有成效。大量研究表明^[8-11],在数据层面和分类算法层面共同处理数据不平衡问题将成为主要趋势。

笔者基于上述分析,首先,通过实验分析不同程度的类别不平衡数据作为训练集时对传统分类器的影响;其次,提出一种基于云模型的样本再生技术,并将其和集成ELM分类模型结合,在样本层面和分类器层面上解决训练样本类别不平衡对故障模式辨识精度的影响;最后,将该方法运用在滚动轴承故障数据集中,通过实验验证了方法的可行性。

1 相关理论简介

1.1 云模型

假设 $U=\{x\}$ 为一个精确数值组成的论域, C 为与 U 对应的定性概念, U 中的元素 x_i 对于定性概念

* 国家自然科学基金资助项目(51675253);兰州理工大学一流学科建设资助项目

收稿日期:2021-07-02;修回日期:2021-11-01

C 的确定度为 y_i , 是一个在稳定范围内的随机数, 则 x 在论域 U 上的分布称为云, 每一个 x_i 称为云滴。云模型用期望 E_x 、熵 E_n 和超熵 H_e 这 3 个特征来描述云滴群的整体特性。

云模型是一种能实现定性定量转换的双向认知模型, 其依靠云发生器来实现, 云发生器主要分为正向云发生器和逆向云发生器 2 种。正向云发生器根据云的 3 个数字特征 (E_x, E_n, H_e) 产生若干含有确定度 y_i 的云滴 (x_i, y_i) , 实现了定性到定量的映射。其中: 定量值 x_i 为论域 U 上的一次随机实现, 服从以 E_x 为期望、 E_n^2 为方差的高斯分布。同时, E_n^2 又是服从以 E_n 为期望、 H_e^2 为方差的高斯分布的一次随机实现。因此, x_i 对 U 的确定度 y_i 满足

$$y_i = \exp \left[-\frac{(x_i - E_x)^2}{2E_n^2} \right] \quad (1)$$

逆向云发生器可以将云滴群 (x, y) 转换为以数字特征 (E_x, E_n, H_e) 表示的云模型, 实现从定量值到定性概念的转换。云发生器模型如图 1 所示, 然而在实际工作中表示云滴确定度的 y 值很难获得, 因此图 1(b) 所示的逆向云发生器实用性并不高。针对此问题, 文献 [12] 提到了一种无需确定度 y 的逆向云发生器算法, 根据此改进算法, 可以在实际样本的基础上计算滚动轴承各状态下特征的高斯云分布模型指标, 即期望 E_x 、熵 E_n 和超熵 H_e 。具体计算公式如下

$$E_x = \bar{X} \quad (2)$$

$$E_n = \sqrt{\frac{\pi}{2}} \frac{1}{N} \sum_{i=1}^N |x_i - E_x| \quad (3)$$

$$H_e = \sqrt{S^2 - E_n^2} \quad (4)$$

其中: \bar{X} 为样本均值; S^2 为样本方差。

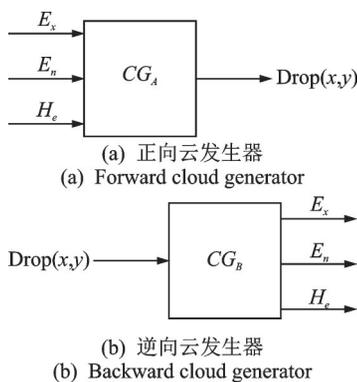


图 1 云发生器模型

Fig.1 The model of cloud generator

1.2 装袋式集成学习算法

装袋式 (Bagging) 集成学习是并行式集成学习方法的代表, 在学习过程中要构造多个学习器。若构造的学习器是同一类, 则称为同质集成, 反之则称之为异质集成^[13]。其在学习时, 先从包含 n 个样本的原始训练集中随机选择一个样本, 将其复制到训练子集中, 此时该样本仍存在于初始训练集中, 在下次采样时仍有被选中的可能。这样, 经过 n 次随机采样, 可得到一个内含 n 个样本的训练子集 (该训练子集中可含有重复的样本)。照此方法, 重复 t 轮可得到 t 个训练子集, 然后基于每个训练子集训练出一个子学习器。Bagging 集成学习的子学习器也被称作弱学习器, 常用的弱学习器包括人工神经网络和决策树等, 这些学习器在 m 类问题上的识别准确率一般不低于 $1/m$ 。

Bagging 集成学习结果通过综合 t 个子学习器的学习结果得到。对于回归预测问题, 通常将子学习器的结果经过简单平均后得到集成学习结果; 对于分类问题, 最终结果常用投票法得到。

综上所述, 集成学习综合多个分类器给出了最终的预测结果, 但多个分类器参与学习会消耗大量时间。为弥补这一缺陷, 需要选用效率较高、泛化性相对较好的子分类器。ELM 与其他分类器相比, 恰好具有计算复杂度低、泛化性强等特点^[14]。

1.3 定义的样本类别不平衡程度度量公式

在类别不平衡问题中, 将样本数较多的一类称为多数类, 反之被称为少数类。设多数类样本数为 k_{much} , 少数类样本数为 k_{least} , 则样本不平衡比率 (im-balance ratio, 简称 IR) 的计算公式为

$$\text{IR} = k_{\text{much}} / k_{\text{least}} \quad (5)$$

根据 IR 值, 类别不平衡可划分为轻度不平衡和重度不平衡问题。前者的 IR 值较小, 对传统分类器的分类影响不大; 后者会在极端条件下令分类器完全失效。

2 基于云模型的样本生成模型和基于装袋法的集成 ELM 模型

2.1 滚动轴承状态信号采集和信号特征提取

本研究以无锡市厚德自动化仪表有限公司的一套双跨双转子综合故障模拟平台 (图 2) 来模拟滚动轴承故障。故障轴承安装端如图 3 所示, 信号的采

样通道共5个,其中:通道1~3连接1个三向加速度传感器,采集故障轴承的2个径向和1个轴向方向的振动加速度信号;通道4和5各自连接1个电涡流传感器,采集轴的径向振动信号(故障轴承直接安装在该轴上)。图4为具体的故障轴承,其中:1为内圈故障轴承;2为外圈故障轴承;3为滚动体故障轴承;4为保持架故障轴承。实验测得滚动轴承NSK6308在4种故障和正常共5种状态下的振动信号 $5 \times 80 = 400$ 组,选择 5×20 组作为测试样本。采样转速分别为2 600,2 800,3 000和3 200 r/min,采样频率为8 kHz。



图2 双跨双转子综合故障模拟平台

Fig.2 Double-span double-rotor comprehensive fault simulation platform

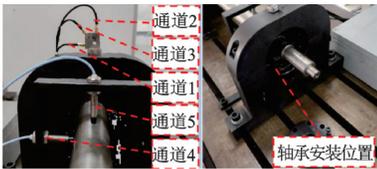


图3 故障轴承安装端

Fig.3 Fault bearing mounting end

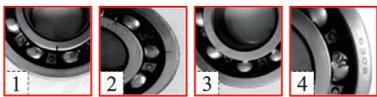


图4 故障轴承

Fig.4 Fault bearing

转速为3 000 r/min时轴承部分故障的振动信号如图5所示。将振动信号用一维小波进行消噪

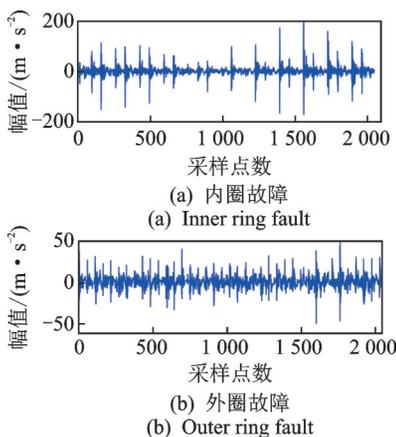


图5 3 000 r/min时轴承部分故障的振动信号

Fig.5 Vibration signal of partial failure when rolling bearing speed is 3 000 r/min

后,构建出 $5 \times 11 = 55$ 维的多域多通道特征数据集,为*i*通道构造的原始特征集如表1所示。为后续实验及分类考虑,将特征数据归一化至 $[-1, 1]$ 的区间上。

表1 为*i*通道构造的原始特征集($i=1-5$)

Tab.1 The original feature set for the channel $i(i=1-5)$

序号	特征	序号	特征
1	峰峰值	7	频谱二阶矩
2	波形指标	8	频率标准偏差
3	峰值指标	9	频率峭度
4	脉冲指标	10	均方根频率
5	裕度指标	11	中心频率
6	峭度指标		

2.2 基于云模型的滚动轴承特征样本再生技术

设滚动轴承的状态序号 $m=(1, 2, 3, 4, 5)$,分别表示滚动体故障、保持架故障、内圈故障、外圈故障和正常状态;特征序号为 $j=(1, 2, \dots, 55)$ 。首先,以滚动轴承在状态*m*下第*j*个特征的特征值为输入,利用1.1节所提的改进逆向云算法计算出云模型指标($E_{xm,j}, E_{nm,j}, H_{em,j}$);其次,将该指标作为正向云发生器的输入,可以获得带有确定度的云滴 (x_i, y_i) ,设云滴个数 $i=1\ 000$ 。高斯云由这些云滴汇聚而成,云滴的 x_i 值即可视为状态*m*下第*j*个特征的再生样本。

设样本数为60的特征值期望为标准值,以实际样本数为自变量、实际样本数目下的特征期望与标准期望的比值为因变量来绘图,不同数目实际样本的期望与标准期望的关系如图6所示。由图可知,当实际样本的数量极少时(最少为1),特征云模型会发生期望 E_x 偏移的现象,其中外圈故障特征偏移较为明显,但总体上样本特征的期望值趋于相对稳定,这说明基于云模型的样本再生技术在该故障数据集上具有可行性。

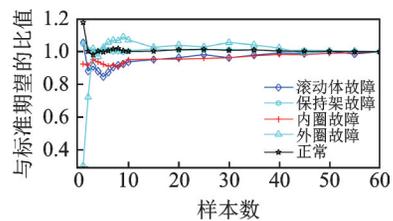


图6 不同数目实际样本的期望与标准期望的关系

Fig.6 Relationship between expectations of different numbers of true samples and standard expectations

以转速为3 000 r/min时滚动体故障和轴承外圈故障情况下通道1信号的频谱二阶矩为例来说明样本的再生成过程。设少数类的实际样本数仅为8个,2种故障情况下通道1振动信号的频谱二阶矩云模型如图7所示,其横、纵坐标均为无量纲量。实际样本原本无对应的确定度值,故无法进行基于云模型的可视化表达。每个特征属于对应概念的程度,即确定度可按式(6)进行计算

$$y_{m,j} = \exp\left[\frac{-(F_{m,j}^k - E_{x_{m,j}})^2}{2(E_n'_{m,j})^2}\right] \quad (6)$$

其中: $y_{m,j}$ 为状态 m 下第 k 个实际样本 $N_{m,k}$ 的第 j 个特征与对应特征云的确定度; $F_{m,j}^k$ 为实际样本 $N_{m,k}$ 的第 j 个特征值; $E_{x_{m,j}}$ 为状态 m 中第 j 个特征的期望值;

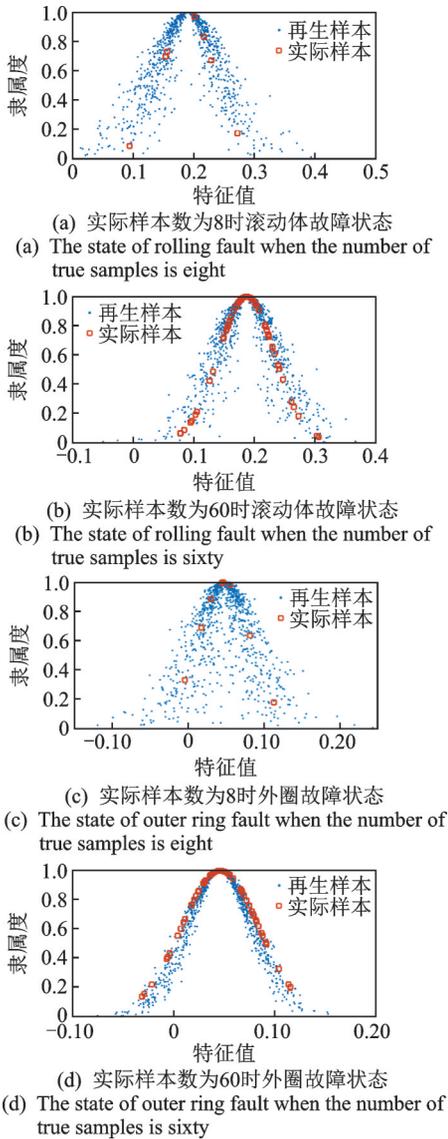


图7 2种故障情况下通道1振动信号的频谱二阶矩云模型
Fig.7 Spectrum second-order momen cloud model of the vibration signal of channel 1 under two fault conditions

$E_n'_{m,j}$ 为状态 m 中第 j 个特征的等效方差,是以特征的 E_n 为期望、 H_n 为标准差生成的一个正态随机数。

由图7(b,d)还可看出,特征值符合高斯分布,且大多数样本分布在高斯云的峰顶,从顶至尾逐渐减少。因此,在借助云模型进行样本生成时要进行分层抽样,本研究将确定度范围 $[0,1]$ 等分为5个区间,分别按照5:4:3:2:1的比例来进行抽样,以获得不同确定度区间内的再生特征样本。虚拟样本的采样分布如图8所示。

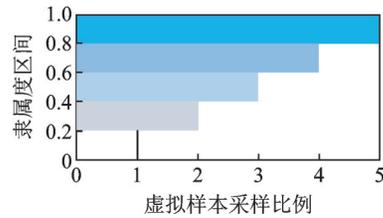


图8 虚拟样本的采样分布
Fig.8 Sampling distribution of virtual samples

云模型是根据现存的实际样本构造而出,同样,再生样本也必须根据已有的样本来衍生。通过基于固有样本的高斯云模型,可以衍生出滚动轴承在不同状态下各个特征的再生样本。

2.3 基于装袋法的集成极限学习机

在隐含层节点相同的情况下,相较于极限学习机,集成极限学习机具有更好的分类精度。本研究将ELM的隐含层节点数设定为10个,在确定了隐含层节点的情况下,以分类误差不小于95%为原则,确定参与集成学习的子分类器数目(即ELM的数目)。经验证,子分类器ELM的数目不小于30时满足实验要求,至此完成了集成极限学习机的2个超参设置。E-ELM学习流程如图9所示,其中输入至子分类器 ELM_i 的训练子集 T_i 由训练集通过自助采样得到,测试样本 U 的最终结果借助了相对多数投票法,综合考虑了各子分类器的结果。

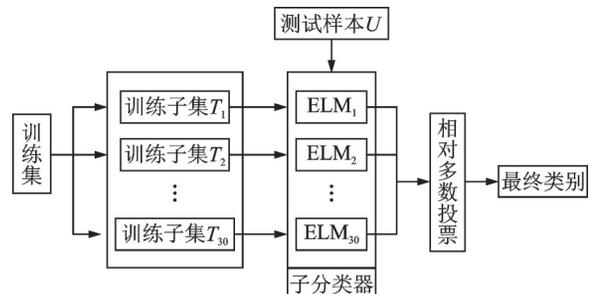


图9 E-ELM学习流程
Fig.9 Learning process of E-ELM

2.4 总体流程

设计的方法流程如下:

1) 计算训练集中各类的样本数 $k_m (m=1, 2, \dots, 5)$, $k_{\text{much}} = \max(k_m)$, 若 $50 < k_{\text{much}} < 70$, 则保留 k_{much} , 否则令 $k_{\text{much}} = 60$;

2) 按式(1)计算各类样本的IR值, 将 $IR > 1.4$ 的类视为少数类, 需要进行样本生成, 将该类的样本数扩充至60; 若 $IR < 0.857$, 则该类样本过多, 需要进行欠采样, 使样本数目至60;

3) 将平衡样本集输入至E-ELM中进行训练;

4) 利用测试集验证分类效果。

本研究中各类训练样本最大数量为60, 故无需进行欠采样操作。图10为不平衡数据学习流程图。

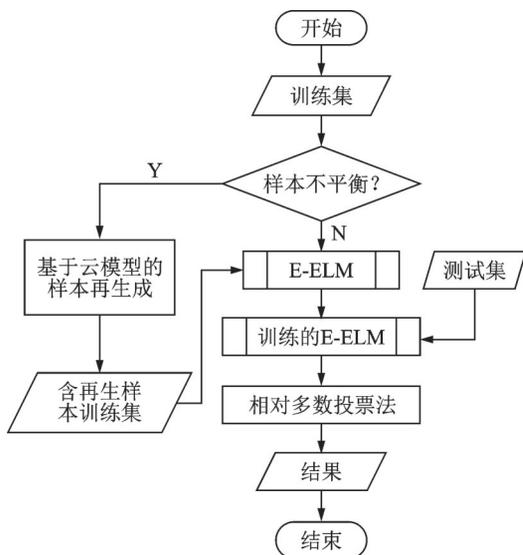


图10 不平衡数据的学习流程

Fig.10 The learning process of imbalanced data

3 实验分析及探讨

3.1 训练样本类别不平衡对传统分类器的影响

实验时, 少数类样本数目统一设为5, 即 $IR = 60/5 = 12$, 此时训练样本类别分布极度不平衡。分别用“G”, “B”, “N”, “W”和“Z”表示训练集中滚动体故障样本较少、保持架故障样本较少、内圈故障样本较少、外圈故障样本较少以及正常状态样本较少, 将其定义为一类不平衡情况; “GB”, “GN”, “GW”, “BN”和“BW”表示对应的两类样本数据量较少, 将其定义为二类不平衡情况。

本节验证了滚动轴承故障训练数据存在一类和

二类不平衡时, 传统分类器ELM、误差反向传播网络(back propagation neural networks, 简称BPNN)和K近邻(K-nearest neighbor, 简称KNN)分类器的识别精度。在实验时, ELM和E-ELM的隐层神经元数目均为10; BPNN的隐含层为1层, 隐层神经元数目设为12, 训练最小误差为0.1, 训练次数为1000次; KNN的近邻样本数目设置为6。

在2600和3000 r/min 2种转速下, 一类不平衡训练样本分类器的分类精度分别如表2, 3所示; 二类不平衡训练样本分类器的分类精度分别如表4, 5所示。按照组合 $C_5^2 = 10$, 即两类不平衡的情况有10种, 表中只取5种二类不平衡情况进行实验。

由表2~5可知: 训练样本不平衡对KNN分类器影响最大; 训练集二类不平衡对分类器的影响比一类不平衡要大; BPNN和KNN对轴承外圈故障样本不平衡的训练集比较敏感; 在样本极度不平衡时, ELM的辨识精度总体上要高于BPNN和KNN。

表2 2600 r/min时一类不平衡训练样本分类器的分类精度
Tab.2 Classification accuracy of classifiers with a kind of imbalanced training samples at 2600 r/min %

分类器	不平衡类型				
	G	B	N	W	Z
ELM	87.5	84.1	84.1	90.5	80.0
BPNN	83.8	89.5	91.0	72.5	84.0
KNN	59.0	71.0	79.0	79.0	65.0

表3 3000 r/min时一类不平衡训练样本分类器的分类精度
Tab.3 Classification accuracy of classifiers with a kind of imbalanced training samples at 3000 r/min %

分类器	不平衡类型				
	G	B	N	W	Z
ELM	83.7	85.2	90.5	85.0	80.0
BPNN	86.0	86.0	87.2	75.0	77.6
KNN	66.7	83.0	89.0	78.0	62.0

表4 2600 r/min时二类不平衡训练样本分类器的分类精度
Tab.4 Classification accuracy of classifiers with two kinds of imbalanced training samples at 2600 r/min %

分类器	不平衡类型				
	GB	GN	GW	BN	BW
ELM	79.0	87.0	86.0	84.0	83.0
BPNN	88.1	81.0	61.0	82.0	71.3
KNN	35.0	46.0	75.0	48.0	77.0

表 5 3 000 r/min 时二类不平衡训练样本分类器的分类精度
Tab.5 Classification accuracy of classifiers with two kinds of imbalanced training samples at 3000 r/min %

分类器	不平衡类型				
	GB	GN	GW	BN	BW
ELM	80.4	89.0	82.3	82.2	81.5
BPNN	82.0	85.6	69.0	81.8	68.6
KNN	37.0	47.0	80.0	52.0	83.0

3.2 本研究所提样本再生技术的有效性验证

采用本研究的样本再生技术将类别分布不平衡的训练集补充完整后训练分类器,并验证其分类精度。在 3 000 和 3 200 r/min 转速下,训练集经再生技术补充后训练的分类器学习能力分别如表 6,7 所示。

表 6 3 000 r/min 时训练样本经再生技术补充后训练的分类器学习能力

Tab.6 The learning ability of classifiers trained after training samples supplemented by regeneration technology at 3 000 r/min %

分类器	不平衡类型				
	G	B	N	W	Z
ELM	92.4	91.5	93.0	92.5	93.1
BPNN	89.0	87.0	85.0	91.0	88.6
KNN	91.0	98.0	100.0	98.0	100.0

表 7 3 200 r/min 时训练样本经再生技术补充后训练的分类器学习能力

Tab.7 The learning ability of classifiers trained after training samples supplemented by regeneration technology at 3 200 r/min %

分类器	不平衡类型				
	G	B	N	W	Z
ELM	90.0	90.5	92.8	93.6	91.3
BPNN	83.6	87.4	87.0	91.1	92.4
KNN	89.0	95.0	100.0	98.0	99.0

对比表 3 和表 6 发现,经再生样本补充的训练集训练的分类器在分类精度上有较好的提升。对比同转速下不同分类器的分类精度发现,ELM 的分类精度较好,且 ELM 的训练时间(0.021 s)低于 BP 神经网络的训练时间(0.27 s),这说明 ELM 比 BPNN 更适合作为 Bagging 集成学习模型的子分类器。

3.3 本研究的样本再生技术和 E-ELM 结合后对不平衡数据集的学习效果验证

为获得更好的学习效果,不平衡训练集用基于云模型的再生样本补充平衡后训练出了 E-ELM 模型,拟在算法层面提高数据的最终学习效果。结合

方法对不同工况下一类不平衡数据和二类不平衡数据的学习情况分别如表 8,9 所示。

表 8 表明,基于云模型的样本再生技术和 E-ELM 结合后,能大幅度改善一类不平衡数据的学习效果。对比表 9 和表 4、表 5 后发现,本研究方法对于二类不平衡数据同样具有较强的学习能力。

表 8 结合方法对不同工况下一类不平衡数据的学习情况

Tab.8 The combined method on the learning situation of the one type of unbalanced data under different working conditions %

转速/(r·min ⁻¹)	不平衡类型				
	G	B	N	W	Z
3 200	100.0	97.6	100.0	100.0	99.85
3 000	99.7	99.8	99.8	100.0	99.70
2 800	99.4	99.8	99.9	99.8	99.30
2 600	100.0	100.0	100.0	100.0	100.00

表 9 结合方法对不同工况下二类不平衡数据的学习情况

Tab.9 The combined method on the learning situation of the two types of unbalanced data under different working conditions %

转速/(r·min ⁻¹)	不平衡类型				
	GB	GN	GW	BN	BW
3 200	97.25	100.0	99.90	98.20	97.60
3 000	100.00	99.8	99.85	99.95	100.00
2 800	99.10	99.2	99.25	99.85	99.45
2 600	100.00	100.0	100.00	100.00	100.00

3.4 实验小结

首先,验证了不平衡数据对传统分类器的影响;其次,用所提的样本再生技术平衡了数据,并用传统分类器验证了本方法在不平衡问题上的作用;最后,证明了 E-ELM 作为本方法体系的分类器时,对不平衡学习的促进作用。实验表明,本研究设计的方法能较好地消除训练样本类别不平衡对滚动轴承故障辨识精度的影响,解决了因个别类训练样本较少引起的数据不平衡问题。

4 结 论

1) 基于云模型的样本再生技术本质上是依靠特征自身的分布趋势来构造虚拟样本,可视化效果强,可以直观地了解样本的分布。

2) 采用分层抽样技术从众多再生样本中获取部分用于填充不平衡训练集,在兼顾泛化性的同时尽量使得大部分再生样本靠近期望值。

3) 在分类阶段引入了集成的思想,以消耗一定

时间为代价换取了分类精度的大幅提高。

4) 正态分布并不是所有故障数据的分布形式。有时同一类故障聚类时,类内会聚成若干个小簇。基于云模型的样本再生成技术在处理这种类型的数据时是否有较好的效果,还需进一步探讨。

参 考 文 献

- [1] 雷亚国,贾峰,孔德同,等. 大数据下机械智能故障诊断的机遇与挑战[J]. 机械工程学报, 2018, 54(5): 94-104.
LEI Yaguo, JIA Feng, KONG Detong, et al. Opportunities and challenges of machinery intelligent fault diagnosis in big data era[J]. Journal of Mechanical Engineering, 2018, 54(5): 94-104. (in Chinese)
- [2] 汤健,柴天佑,丛秋梅,等. 基于EMD和选择性集成学习算法的磨机负荷参数软测量[J]. 自动化学报, 2014, 40(9): 1853-1866.
TANG Jian, CHAI Tianyou, CONG Qiumei, et al. Soft sensor approach for modeling mill load parameters based on EMD and selective ensemble learning algorithm[J]. Acta Automatica Sinica, 2014, 40(9): 1853-1866. (in Chinese)
- [3] 赵荣珍,赵孝礼,何敬举,等. 相关流形距离在转子故障数据集分类中的应用方法[J]. 振动与冲击, 2017, 36(18): 125-130.
ZHAO Rongzhen, ZHAO Xiaoli, HE Jingju, et al. Application of correlation manifold distance in the classification of rotor fault data set[J]. Journal of Vibration and Shock, 2017, 36(18): 125-130. (in Chinese)
- [4] LIM P, GOH C K, TAN K C. Evolutionary cluster-based synthetic oversampling ensemble (ECO-learning) for imbalance learning[J]. IEEE Transactions on Cybernetics, 2017, 47(9): 2850-2861.
- [5] 费洪磊,袁琦,郑玉叶. 基于深度学习的癫痫脑电不平衡分类方法[J]. 仪器仪表学报, 2021, 42(3): 231-240.
FEI Honglei, YUAN Qi, ZHENG Yuye. Imbalanced classification for epileptic EEG signals based on deep learning[J]. Chinese Journal of Scientific Instrument, 2021, 42(3): 231-240. (in Chinese)
- [6] 李敏波,董伟伟. 面向不平衡数据集的汽车零部件质量预测[J]. 中国机械工程, 2022, 33(1): 88-96.
LI Minbo, DONG Weiwei. Research on quality prediction of automotive parts for imbalanced datasets[J]. China Mechanical Engineering, 2022, 33(1): 88-96. (in Chinese)
- [7] 张建军. 静态与动态的不平衡分类问题研究[D]. 广州:华南理工大学, 2020.
- [8] 董勋,郭亮,高宏力,等. 代价敏感卷积神经网络:一种机械故障数据不平衡分类方法[J]. 仪器仪表学报, 2019, 40(12): 205-213.
DONG Xun, GUO Liang, GAO Hongli, et al. Cost sensitive convolutional neural network: a classification method for imbalanced data of mechanical fault [J]. Chinese Journal of Scientific Instrument, 2019, 40(12): 205-213. (in Chinese)
- [9] 刘鹏,杜佳芝,吕伟刚,等. 面向不平衡数据集的一种改进的k-近邻分类器[J]. 东北大学学报(自然科学版), 2019, 40(7): 932-936.
LIU Peng, DU Jiazhi, LÜ Weigang, et al. A modified KNN classifier for unbalanced dataset [J]. Journal of Northeastern University (Natural Science Edition), 2019, 40(7): 932-936. (in Chinese)
- [10] 范雨强,崔晓钰,韩华,等. 不平衡数据技术在冷水机组故障诊断中的应用[J]. 工程热物理学报, 2019, 40(6): 1219-1228.
FAN Yuqiang, CUI Xiaoyu, HAN Hua, et al. Chiller fault diagnosis with the technology of imbalance data [J]. Journal of Engineering Thermophysics, 2019, 40(6): 1219-1228. (in Chinese)
- [11] 李艳霞,柴毅,胡友强,等. 不平衡数据分类方法综述[J]. 控制与决策, 2019, 34(4): 673-688.
LI Yanxia, CHAI Yi, HU Youqiang, et al. Review of imbalanced data classification methods [J]. Control and Decision, 2019, 34(4): 673-688. (in Chinese)
- [12] 谢鹏,吕鹏飞. 基于改进云模型-IAHP的涉爆粉尘企业安全风险评估[J]. 中国安全生产科学技术, 2018, 14(8): 187-192.
XIE Peng, LÜ Pengfei. Assessment on safety risk of enterprises involved in explosive dust based on improved cloud model and IAHP [J]. Journal of Safety Science and Technology, 2018, 14(8): 187-192. (in Chinese)
- [13] 唐耀华,高静怀,包乾宗. 一种新的选择性支持向量机集成学习算法[J]. 西安交通大学学报, 2008, 42(10): 1221-1225.
TANG Yaohua, GAO Jinghuai, BAO Qianzong. Novel selective support vector machine ensemble learning algorithm[J]. Journal of Xi'an Jiaotong University, 2008, 42(10): 1221-1225. (in Chinese)
- [14] HUANG G B, ZHOU H, DING X, et al. Extreme learning machine for regression and multiclass classification[J]. IEEE Transactions on Systems, Man and Cybernetics, Part B :Cybernetics, 2012, 42(2):513-529.



第一作者简介:马森财,男,1994年7月生,硕士。主要研究方向为旋转机械故障诊断技术和故障数据挖掘。
E-mail:masencaibangong@163.com

通信作者简介:赵荣珍,女,1960年12月生,博士、教授、博士生导师。主要研究方向为旋转机械故障诊断和动态测试技术。
E-mail:zhaorongzhen@lut.edu.cn