

# 基于GAN数据增强与改进Bi-LSTM的充电桩故障预测方法

周秋阳<sup>1</sup>, 高辉<sup>1</sup>, 李炜卓<sup>2</sup>, 归耀城<sup>2</sup>

(1. 南京邮电大学自动化学院, 江苏南京210023;

2. 南京邮电大学现代邮政学院、智慧交通学院, 江苏南京210023)

**摘要:**近年来,电动汽车发展迅猛,其充电基础设施建设加快,因此,电动汽车充电设施的可靠性与安全性成为当前研究重点。现有研究采用的数据多为完整且充足的仿真数据,面对实际数据时,往往会因为数据不足或不完整,影响预测精度。为解决上述问题,通过数据驱动,实现充电设备充电过程的故障预警。首先,进行特征选取,选择合适的数据特征。其次,对订单数据进行筛选,构建数据集,并进行归一化处理。再次,将数据集划分为训练组和测试组,训练组用于模型的训练,测试组用于判断模型训练的优劣。然后,利用生成对抗网络(generative adversarial network, GAN)对划分好的训练组进行数据增强,扩充数据规模,形成足量的新数据,并将数据输入双向长短期记忆(bi-directional long-short term memory, Bi-LSTM)网络,采用粒子群优化(particle swarm optimization, PSO)算法对初始参数进行优化,对GAN-PSO-Bi-LSTM进行多次试验,观察模型试验的结果。最后,与其他预测模型进行比较,验证表明GAN-PSO-Bi-LSTM模型的预测性能更高,能够提高充电桩的故障预测准确率。

**关键词:**充电桩;故障预测;数据增强;生成对抗网络(GAN);粒子群优化(PSO);双向长短期记忆(Bi-LSTM)

中图分类号:TM73

文献标志码:A

文章编号:2096-3203(2025)06-0049-13

## 0 引言

随着全球能源危机的不断加剧和环境问题的日益突出,电动汽车越来越受到人们的青睐,电动汽车相比传统汽车在节能减排方面有着巨大的优势<sup>[1]</sup>。据统计,截至2023年底,中国电动汽车保有量达到2 041万辆,并依旧保持高速增长趋势。这也推动了电动汽车充电基础设施的建设,各国政府投入大批资金用于建设电动汽车充电站与充电桩,以满足日益增长的电动汽车充电需求<sup>[2-4]</sup>。我国电动汽车基础设施建设也取得了长足发展,据国家能源局统计,截至2023年底,我国充电基础设施总量达到859.6万台,同比增长65%。随着大量电动汽车充电设备的建设与投运,其充电可靠性和安全性逐渐成为关注的重点<sup>[5-9]</sup>。

目前,主要利用充电设备内部的传感器与通信设备采集充电过程中产生的各项数据,例如温度、电压、电流、频率等,再采用机器学习、神经网络等技术对充电过程加以分析,以判断充电设备是否有故障风险。文献[10]利用电压、电流数据,基于改进支持向量机进行充电桩故障预测,但是算法需要反复试验以获得最佳结果,预测效率不高。文献[11]基于仿真的电流数据,使用随机森林(random

forest, RF)算法进行故障诊断,但当采样频率高于一定阈值时,其预测正确率明显降低。文献[12]采用仿真的电压数据作为输入,利用模型预测控制技术,实现充电桩的故障诊断,但该方法针对不同场景需要设计不同的代价函数,在实际使用过程中泛化性不是很强。文献[13]基于RF算法设计充电桩充电故障自动化检测系统,但其决策树的数量过多,容易出现过拟合的问题。文献[14]基于蛇群算法实现充电桩故障预测,其输入数据选择电子锁驱动信号、急停信号、门禁信号等故障信号,但蛇群算法对参数设置敏感,容易陷入局部最优。文献[15]同样以故障信号作为输入,基于反向传播神经网络进行充电桩故障预测。文献[16]基于广义自回归(autoregressive, AR)模型,综合天气、运维等影响指标,实现充电桩故障预测。国外在这一方面同样进行了大量研究,文献[17]基于小波变换方案,增强对充电桩的故障检测能力,但小波变换对复杂信号的辨析能力不强,当电气信号较为复杂时,其预测精度不高。文献[18]采用仿真的充电桩散热器温度数据作为输入,提出一种基于前馈神经网络(feedforward neural network, FNN)的充电桩异常检测方法,但该方法需要大量数据作为支撑,在实际使用中可能无法满足其数据需求。

绝大多数研究在进行模型验证时都采用模拟仿真数据,实际数据一方面获取途径困难,另一方面由于充电桩内部传感器或通信设备故障等原因,

收稿日期:2025-05-26;修回日期:2025-07-30

基金项目:国家自然科学基金资助项目(52077107);国家电网有限公司总部科技项目(5400-202416211A-1-1-ZN)

往往会在数据缺失、数据量不足等问题<sup>[19]</sup>。文献[20]在传统的故障因素以外,将环境因素纳入考虑,构建综合预测模型,但模型测试样本较少。为了解决实际充电数据在预测过程出现的问题,往往会在模型验证前,对实际数据进行数据增强<sup>[21]</sup>。文献[22]在解决实际充电数据样本量少的问题时,采用变分自编码器(variational auto encoder, VAE)进行数据增强,但如果原始数据质量不佳,VAE生成的样本会与真实数据存在明显差异。文献[23-24]构建充电设施故障树,充分挖掘原始数据,尽可能多地挖掘数据特征,但过度依赖人工检测与分析模型,可能出现故障原因考虑不周的问题,导致判断准确率不高。

上述研究大多基于充电桩充电过程中产生的大量电压、电流数据进行故障预测,这类数据由充电桩内部安装的各类传感器收集,并通过通信单元模块传输给服务器进行存储。但在充电桩使用过程中,各类传感器或通信单元可能发生故障,且在存储电压、电流数据过程中,由于数据量庞大,一般储存一段时间后,新的电压、电流数据会覆盖旧的电压、电流数据,导致采集的电压、电流数据不完全或不足。现有研究大多基于完整的仿真数据,在实际运用中,数据不足往往会导致模型预测精度降低。文中从充电账单数据与维护工单数据的角度研究充电桩故障预测,聚焦于充电订单中的充电时间、充电结束原因等。充电账单与维护工单数据存储成本低,往往会保存较长时间,这解决了传统通过电压、电流数据进行故障诊断的数据缺失问题。此外,文中所用数据为某市充电站充电账单数据与维护工单数据,由于选址、电动汽车用户数量少等原因,该充电站订单数据相对较少。其中,充电账单数据中的充电桩数量为30,维护工单的维修记录共323条,这个数据规模对于模型训练是不够的,会导致模型训练不充分,出现过拟合现象,影响实际使用中故障预测的精确性。

文中针对这一类数据量较少的充电站,提出一种基于生成对抗网络(generative adversarial network, GAN)-粒子群优化(particle swarm optimization, PSO)-双向长短期记忆(bi-directional long-short term memory, Bi-LSTM)的充电桩故障预测方法。在构建数据集时,为避免传感器故障导致电压、电流数据缺失,从充电账单数据与维护工单数据的角度,构建时序数据集;为解决数据量较少影响故障预测精度的问题,采用GAN对原始数据集进行数据增强,扩充数据规模;采用Bi-LSTM进行预测,相较于其

他预测模型,能更好地挖掘时间数据序列的依赖和影响因素特征依赖;使用PSO对初始参数进行优化,能提高模型的预测精度。

## 1 充电桩特征集与数据增强

### 1.1 充电桩故障分析

充电桩的故障状态主要可以划分为两大类:充电时发生的故障和非充电时发生的故障。对于非充电时的故障,其主要成因为外部环境的影响以及运维的不足,这类故障存在较大的不可预测性,主要解决方案是定期检查和维护。而根据统计,大部分故障发生在充电过程中,因此文中研究充电时发生的故障。根据某市电动汽车用户充电订单数据,可知充电接口电子锁故障占比43.79%,功率控制模块通信故障占比15.53%,电池管理系统(battery management system, BMS)通信故障占比14.60%,其他故障占比26.08%。

充电接口电子锁故障出现的主要原因是公共充电桩使用频繁且环境恶劣,充电枪头可能会变形,使电子锁解锁装置损坏或卡滞,导致无法正常充电。功率控制模块通信故障、BMS通信故障都属于通信故障,其中BMS负责监控和管理电池的状态和性能,其故障会导致无法准确获取电池的状态信息。因此要在故障严重到无法充电之前,及时发现故障,避免对充电桩以及电动汽车电池造成不良影响。文中主要研究上述故障,针对原始的电动汽车充电订单数据集,构建特征集,为模型的训练与测试做准备。

### 1.2 充电桩特征集构建

文中数据为某市充电站充电账单数据与维护工单数据。充电账单数据类型包括充电桩编号、所属充电站、起始时间、终止时间、充电车型、是否完成充电以及充电结束原因。维护工单数据则包括充电桩编号、所属充电站、故障原因、故障等级、报修时间以及完成时间。

在选择合适的特征值前,从维护工单数据出发,根据充电桩编号,在充电账单数据中寻找到对应的充电桩。因为维护工单中的报修时间这一项可以表示充电桩损坏的大致时间,因此,根据维护工单数据中充电桩的报修时间,选取报修时间前7天的充电账单数据作为负例,观察这7天的是否完成充电和充电结束原因,揭示充电桩在损坏前后一段时间的数据特征。随后,根据维护工单数据中充电桩的完成时间,选取完成时间后7天的充电账单数据作为正例,观察这7天的是否完成充电和充

电结束原因, 揭示充电桩正常充电的数据特征。按照上述思想, 统计维护工单中出现的每一个充电桩对应的正例与负例, 形成用于构建特征工程的受维护充电桩的充电数据集。

基于受维护充电桩的充电数据集, 对充电订单记录进行分类, 分为正常充电订单、无效充电订单以及故障充电订单。正常充电订单是指充电结束原因为达到设置充电金额停止、用户远程(手机、网站)终止、电动汽车充满等正常结束, 且充电时长超过 1 min 的充电订单记录; 无效充电订单是指充电结束原因为达到设置充电金额停止、用户远程(手机、网站)终止、电动汽车充满等正常结束, 但充电时长不超过 1 min 的充电订单记录; 故障充电订单则是指充电结束原因为检测点电压检测故障、电表通信故障、充电接口电子锁故障等异常结束充电的充电订单记录。

进行特征工程构建时, 考虑充电订单类型、充电订单数量等, 选取如下特征:

(1) 充电订单汇总数据。订单数量变化往往能反映充电桩状况, 当充电桩出现问题时, 异常充电订单数量会变多, 正常充电订单数量会变少, 总充电订单数量也会变少。该类特征包括总充电订单数、故障充电订单数占比、无效充电订单数占比、最后一天充电订单数占比, 反映各类型充电订单在总充电订单数中所占比重。

(2) 订单数量每日变化。这一类特征也是反映充电订单数量的变化, 但是较充电订单汇总数据更加细致, 主要反映 7 天内每一天的充电订单数量占比、故障充电订单数量占比、无效充电订单数量占比变化, 关注离故障时刻远近对充电订单数量占比的影响。

(3) 充电结束原因。充电桩异常结束往往预示着充电桩可能出现问题, 因此, 将每一种充电结束原因都列为一个特征因素。根据充电账单数据统计, 共 97 种充电结束原因, 其中正常结束充电原因 12 种, 异常结束充电原因 85 种。

综上所述, 共得到 122 种特征构建特征集, 如图 1 所示。基于充电账单与维护工单数据, 根据每个充电桩的订单时间, 构建时序数据集, 得到 600 条数据的原始数据集, 为后续数据增强与 Bi-LSTM 模型的训练及测试做准备。

### 1.3 基于 GAN 的数据增强

文中采用 GAN 实现数据增强<sup>[25-26]</sup>。GAN 由生成器和判别器组成, 在 GAN 中, 生成器通过给定一个随机的噪声信号作为输入, 经过多层隐含层变

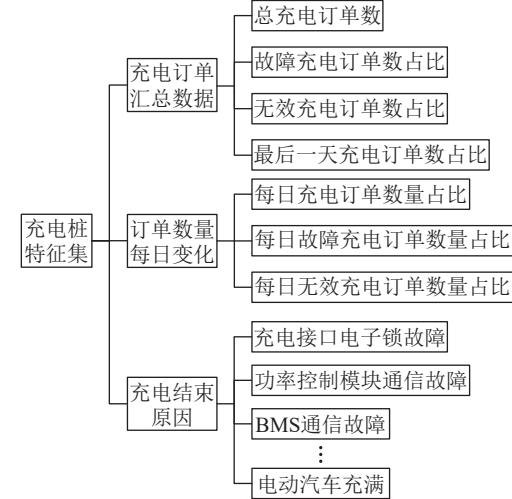


图 1 充电桩特征集

Fig.1 Charging station feature set

换, 将输入空间映射到样本空间, 生成器的输出会尽量模仿训练集中的真实样本。生成器网络模型如下:

$$x = G(z; \theta_{(G)}) \quad (1)$$

式中:  $x$  为生成样本数据;  $z$  为噪声信号;  $G$  为生成器神经网络模型, 负责生成样本数据;  $\theta_{(G)}$  为生成器网络参数。

判别器则是将真实样本和生成样本随机抽取作为输入, 经过训练后, 判别器会尽量分辨出输入的是真实样本还是生成样本。判别器网络模型如下:

$$s = D(x; \theta_{(D)}) \quad (2)$$

式中:  $s$  为鉴别样本数据真假的标签输出;  $D$  为判别器神经网络模型, 负责鉴别样本数据真假;  $\theta_{(D)}$  为判别器网络参数。

训练过程分为 2 步, 判别器和生成器轮流升级, 相互竞争。第一步是优化判别器, 判别器的输入包括 2 种数据, 一种是真实数据, 另一种是还未训练过的生成器生成的假数据, 此时生成的假数据与真实数据相差较大, 因此判别器可以准确判别数据的真假。第二步是优化生成器, 将具有一定判别能力的判别器固定, 生成器根据判别器回传的数据和梯度信息优化自身, 提升生成数据的能力。通过这样的相互对抗, 生成器在对抗过程中生成的数据会越来越接近真实数据, 判别器对真实数据和生成数据的辨别能力也会不断加强, 并最终达到纳什均衡<sup>[27]</sup>, 此时, 可以用生成器生成足够真实的数据, 进行数据增强。整个对抗过程为博弈问题, 其目标函数如下:

$$\min_G \max_D V(D, G) = E_{x \sim P_r(x)}(\lg(D(x))) + E_{z \sim P_z(z)}(\lg(1 - D(z))) \quad (3)$$

式中:  $E$  为期望;  $P_r(x)$  为真实数据分布;  $P_z(z)$  为采样噪声先验分布;  $\lg(D(x))$  为判别器正确识别真实数据的对数概率;  $\lg((1-D(z)))$  为判别器错误识别生成数据的对数概率。训练时, 判别器力求最大化  $\lg(D(x))$ , 以准确区分真实数据与生成数据, 生成器则力求最小化  $\lg(1-D(z))$ , 以让生成数据欺骗判别器。训练过程中固定一方, 更新另一个网络的参数, 交替迭代, 使得对方的错误最大化, 最终,  $G$  能估测出样本数据的分布, 也就可以生成足够真实的数据。

目前, GAN 在数据增强方面已实际应用。例如通过 GAN 能够生成与真实图像难以区分的新图像, 从而扩展数据集的规模和多样性, 通过增加数据的多样性可以提高机器学习的性能和泛化能力<sup>[28]</sup>。GAN 常常被用于人脸识别系统, 包括人脸生成与人脸修复, 改善图像质量, 扩充训练数据规模, 能显著提高人脸识别准确率。GAN 也被应用于轴承寿命的预测, 基于原始监测数据, 训练近似于原始数据分布的数据, 再通过深度学习实现对轴承寿命的预测。

文中使用 GAN 进行数据增强, 将构建好的数据集输入到 GAN 中, 训练得到近似于原始数据的时序分布数据, 其原理如图 2 所示。

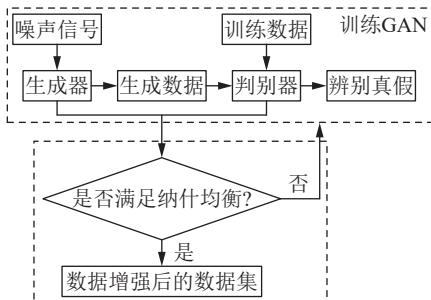


图 2 GAN 基本架构  
Fig.2 Basic architecture of the GAN

## 2 基于改进的 Bi-LSTM 的充电桩故障预测模型

### 2.1 Bi-LSTM 模型

LSTM 用门控机制控制信息的更新或丢弃, 引入输入门、遗忘门、输出门, 以此去除对当下情况不重要的内容, 从而延长信息保存时间, 解决循环神经网络中梯度爆炸和消失的问题。LSTM 中的输入  $\mathbf{X}_t$  均为当前时间步输入与上一时间步隐藏状态  $\mathbf{H}_{t-1}$  的组合, 输出由激活函数 Sigmoid 函数的全连接层计算得到。其总体框架如式(4)—式(8)所示, 其中式(4)为输入门, 式(5)为遗忘门, 式(6)、式(7)为门控单元, 式(8)为输出门。

$$\mathbf{I}_t = \sigma(\mathbf{X}_t \mathbf{W}_{xi} + \mathbf{H}_{t-1} \mathbf{W}_{hi} + \mathbf{b}_i) \quad (4)$$

$$\mathbf{F}_t = \sigma(\mathbf{X}_t \mathbf{W}_{xf} + \mathbf{H}_{t-1} \mathbf{W}_{hf} + \mathbf{b}_f) \quad (5)$$

$$\tilde{\mathbf{C}}_t = \tanh(\mathbf{X}_t \mathbf{W}_{xc} + \mathbf{H}_{t-1} \mathbf{W}_{hc} + \mathbf{b}_c) \quad (6)$$

$$\mathbf{C}_t = \mathbf{F}_t \odot \mathbf{C}_{t-1} + \mathbf{I}_t \odot \tilde{\mathbf{C}}_t \quad (7)$$

$$\mathbf{O}_t = \sigma(\mathbf{X}_t \mathbf{W}_{xo} + \mathbf{H}_{t-1} \mathbf{W}_{ho} + \mathbf{b}_o) \quad (8)$$

式中:  $\mathbf{I}_t$  为输入门在给定时间步长  $t$  的输出;  $\mathbf{X}_t$  为给定时间步长  $t$  的小批量输入向量;  $\mathbf{W}_{xi}$ 、 $\mathbf{W}_{hi}$  为输入门的权重矩阵;  $\mathbf{b}_i$  为输入门的偏置项;  $\sigma(\cdot)$  为 Sigmoid 函数;  $\mathbf{F}_t$  为遗忘门在给定时间步长  $t$  的输出;  $\mathbf{W}_{xf}$ 、 $\mathbf{W}_{hf}$  为遗忘门的权重矩阵;  $\mathbf{b}_f$  为遗忘门的偏置项;  $\tilde{\mathbf{C}}_t$  为在给定时间步长  $t$  下短期记忆需要计算的候选记忆细胞;  $\tanh(\cdot)$  为双曲正切函数;  $\mathbf{W}_{xc}$ 、 $\mathbf{W}_{hc}$  为门控单元的权重矩阵;  $\mathbf{b}_c$  为门控单元的偏置项;  $\mathbf{C}_t$  为当前时间步长门控单元的状态;  $\mathbf{C}_{t-1}$  为上一时间步长门控单元的状态;  $\mathbf{O}_t$  为输出门在给定时间步长  $t$  的输出;  $\mathbf{W}_{xo}$ 、 $\mathbf{W}_{ho}$  为输出门的权重矩阵;  $\mathbf{b}_o$  为输出门的偏置项。使用值域在  $[0,1]$  的 Sigmoid 函数作为激活函数, 通过元素乘法  $\odot$  控制隐藏状态中的信息流动。

输出门  $\mathbf{O}_t$  控制记忆细胞到隐藏状态的信息流动, 得到最终输出  $\mathbf{H}_t$  为:

$$\mathbf{H}_t = \mathbf{O}_t \odot \tanh(\mathbf{C}_t) \quad (9)$$

Bi-LSTM 由前向 LSTM 和后向 LSTM 组合而成<sup>[29-31]</sup>, 在提取数据特征时会考虑到隐藏在数据中的整体信息, 通过正向和反向 2 个角度进行特征提取, 然后将双向提取的结果按照特定方式进行结合。其最终输出如下:

$$\mathbf{H}_{t1} = f_{\text{LSTM}}(\mathbf{X}_t, \mathbf{H}_{t-1}) \quad (10)$$

$$\mathbf{H}_{t2} = f_{\text{LSTM}}(\mathbf{X}_t, \mathbf{H}_{t+1}) \quad (11)$$

$$\mathbf{Y}_t = \mathbf{w}_{hi} \mathbf{H}_{t1} + \mathbf{w}_{h2} \mathbf{H}_{t2} + \mathbf{b}_y \quad (12)$$

式中:  $\mathbf{H}_{t1}$ 、 $\mathbf{H}_{t2}$  分别为  $t_1$ 、 $t_2$  时刻的隐藏状态输出;  $f_{\text{LSTM}}(\cdot)$  为 LSTM 模型;  $\mathbf{w}_{hi}$  为前向 LSTM 层的权重矩阵;  $\mathbf{w}_{h2}$  为后向 LSTM 层的权重矩阵;  $\mathbf{b}_y$  为输出层的偏执矩阵;  $\mathbf{Y}_t$  为  $t$  时刻 Bi-LSTM 的最终输出。Bi-LSTM 摆脱了只能依据之前时刻的时序信息来预测下一时刻的输出的限制, 能更好地结合上下文进行输出。这种特性使得 Bi-LSTM 在处理时序数据时具有更高的灵活性和准确性。

### 2.2 PSO-Bi-LSTM 模型

若采取随机初始化 Bi-LSTM 网络参数, 很难准确得到较为精确的结果。为了更好地初始化 Bi-LSTM 网络参数, 利用 PSO 算法对 Bi-LSTM 模型的学习率、隐含层神经元个数进行训练, 从而优化 Bi-LSTM 模型<sup>[32]</sup>。PSO 算法是一种基于群体智能的优化算法, 其基本原理是模拟粒子在搜索空间中的移动和信息交流, 考虑自己搜索到的历史最优位

置和种群内其他粒子的历史最优位置, 并据此更新自己的速度和位置, 通过迭代过程, 逐渐收敛到问题的最优解<sup>[33]</sup>。利用 PSO 算法对 Bi-LSTM 模型进行优化的具体步骤如下。

步骤 1: 初始化粒子群参数设置。设置粒子群的数量  $N_p=100$ , 迭代次数  $t_p=50$ , 搜索维度  $D_p=2$ 。

步骤 2: 对 PSO 算法的位置和速度进行初始化设置。粒子的初始位置随机, 随机生成的第  $i$  个粒子的位置为  $X_i = (l, h)$ , 其中,  $l$  为 Bi-LSTM 模型的学习率, 学习率取值过大会导致迭代过程中训练不稳定且跳过最优解, 因此, 文中学习率取值范围为  $[0.0001, 0.001]$ ,  $h$  为隐藏层神经元数量, 其取值范围为  $[10, 300]$ 。

步骤 3: 确定 PSO 算法的适应度函数。利用 Bi-LSTM 模型对训练集进行训练, 并将实际结果与预测值进行比较, 适应度函数如下:

$$f(y) = \frac{y'}{y} \quad (13)$$

式中:  $f(y)$  为适应度函数;  $y'$  为预测值与实际结果相同的数量;  $y$  为测试集的总个数。

步骤 4: 迭代计算。对每一次迭代后粒子的新位置进行计算, 并通过调整粒子的位置和速度, 使得粒子不断逼近最优位置, 从而最终获得最优解。速度和位置的更新公式如下:

$$V_{i,t+1} = wV_{i,t} + c_1r_1(P_{i,t} - X_{i,t}) + c_2r_2(P_{g,t} - X_{i,t}) \quad (14)$$

$$X_{i,t+1} = X_{i,t} + V_{i,t+1}T \quad (15)$$

$$n = 1 + \ln t_{\max} \quad (16)$$

$$T = \frac{1}{\sqrt{t}} \quad (17)$$

式中:  $V_{i,t}$ 、 $V_{i,t+1}$  分别为第  $t$  次迭代和第  $t+1$  次迭代时粒子  $i$  的速度;  $X_{i,t}$ 、 $X_{i,t+1}$  分别为第  $t$  次迭代和第  $t+1$  次迭代时粒子  $i$  的位置;  $w$  为惯性权重;  $c_1$ 、 $c_2$  为学习因子;  $r_1$ 、 $r_2$  为  $[0, 1]$  上的随机数;  $P_{i,t}$ 、 $P_{g,t}$  分别为第  $t$  次迭代时粒子  $i$  最优位置和全局最优位置;  $t_{\max}$  为最大迭代次数;  $T$  为时间因子;  $n$  为时间因子  $T$  随当前迭代次数变化的变化速率。

通过迭代得到最优解后, 将最优学习率、隐藏层神经元数量输入 Bi-LSTM 模型。

步骤 5: 充电桩故障预测。在确定最优参数后, 对模型进行训练, 并将测试集输入到训练好的模型中, 输出充电桩故障预测结果。

结合 GAN 数据增强, 最终 GAN-PSO-Bi-LSTM 优化模型的故障预测流程如图 3 所示。

### 3 算例分析

文中算例数据为某市充电站充电账单数据与

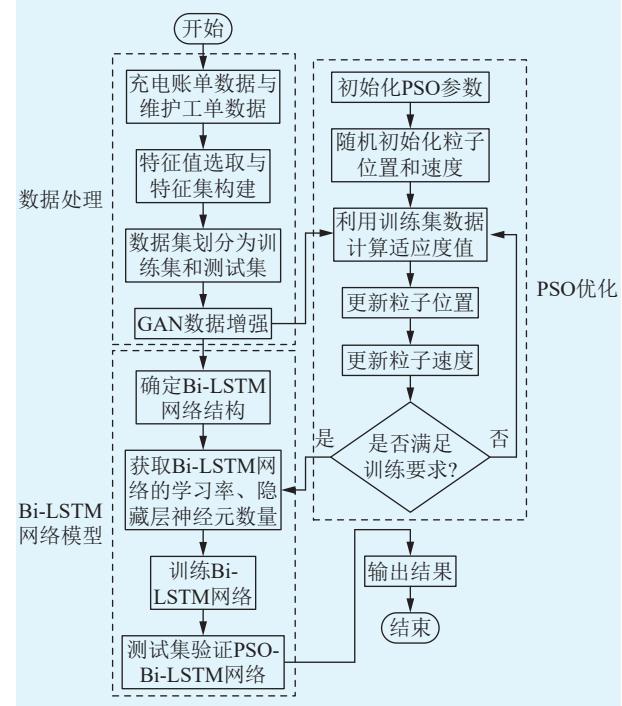


图 3 GAN-PSO-Bi-LSTM 故障预测流程

Fig.3 GAN-PSO-Bi-LSTM fault prediction process

维护工单数据, 在构建原始数据特征集后, 基于 GAN 对原始数据特征集进行数据增强。

#### 3.1 GAN 数据增强结果分析

在数据增强过程中, 对生成器和判别器进行训练。生成器网络包含一个线性层和激活函数 ReLU; 判别器包含一个线性层和激活函数 ReLU, 并通过最后一层的 Sigmoid 函数对输入数据进行分类, 输出一个 0~1 的概率值, 表示输入数据为真实数据的概率。将梯度惩罚项设定为 10, 模型采用 Adam 优化器训练, 生成器和判别器的学习率为 0.0001。GAN 生成数据的数据规模对故障预测结果存在影响, 因此, 采用改进的 Bi-LSTM 模型对不同数据规模的数据集进行测试, 以故障预测准确率为评估指标, 测试结果如图 4 所示。

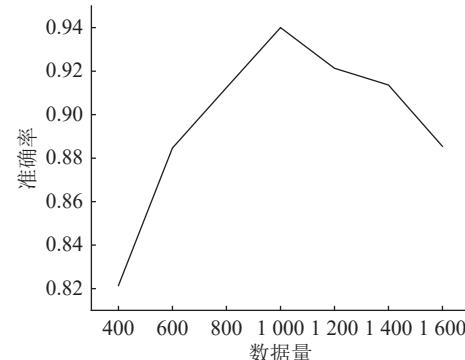


图 4 不同数据规模的故障预测结果

Fig.4 Fault prediction results for different data scales

由图 4 可知, 生成数据过多反而会导致数据质

量下降,影响故障预测结果,且数据过多也会降低模型的运行效率。数据较少则会导致故障预测模型训练不充分,故障预测结果不够精确。因此,综合考虑数据质量和故障预测需求,文中通过GAN生成1 000条新数据。训练过程中,生成器与判别器的损失函数如图5所示。

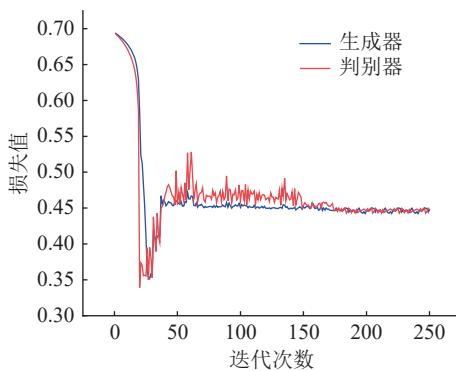


图5 GAN训练损失值  
Fig.5 GAN training loss value

由图5可知,生成器与判别器的损失值随着迭代次数的增加,先减少再上升,并快速收敛,近乎稳定于一个值,这表明GAN训练完成,生成器与判别器的优化达到纳什均衡,生成器生成的数据分布已经和真实数据分布几乎一致。

在评估2个数据集之间的相似度时,常采用推土机距离(earth mover's distance, EMD)和Kullback-Leibler散度(Kullback-Leibler divergence, KL-D)作为评估指标<sup>[34-35]</sup>。其中,EMD用来度量2个分布的相似程度,该值越小,表明生成样本越接近原始样本。KL-D为2个概率分布之间差异的非对称的度量,主要用于度量2个随机变量的差异度,该值越小,两者的概率分布相似度越高。EMD与KL-D的计算方式如下:

$$W_{\text{EMD}}(p, q) = \inf_{\gamma \sim \Pi(p, q)} E_{a, b \sim \gamma}(\|a - b\|) \quad (18)$$

$$D_{\text{KL}}(p \parallel q) = \sum_{i=1}^N p(a_i) \cdot \log \frac{p(a_i)}{q(a_i)} \quad (19)$$

式中:  $W_{\text{EMD}}(p, q)$ 、 $D_{\text{KL}}(p \parallel q)$ 分别为真实样本分布  $p$  与生成样本分布  $q$  的 EMD 值与 KL-D 散度;  $\inf$  为取下界;  $\Pi(p, q)$  为  $p$  和  $q$  组合起来的所有可能的联合分布的集合,对于每一个可能的联合分布  $\gamma$  可以从中采样得到一个样本  $a$  和  $b$ ,并计算出其样本距离;  $E_{a, b \sim \gamma}(\|a - b\|)$  为样本对的期望距离;  $N$  为样本离散取值的数量;  $p(a_i)$ 、 $q(a_i)$  分别为第  $i$  个真实样本分布和第  $i$  个生成样本分布的概率。

为了更好地验证生成样本的合理性和可用性,分别采用GAN、VAE、流模型和AR模型生成1 000

条样本数据,并对其进行评估,评估结果如表1所示。

表1 生成样本评估结果

Table 1 Generate sample evaluation results

评估指标	VAE	流模型	AR模型	GAN
EMD	0.132	0.117	0.095	0.084
KL-D	0.169	0.141	0.124	0.108

由表1可知,GAN的EMD与KL-D值均最小,这表明GAN生成样本与真实样本之间的相似度最高,生成样本足够真实,可以用于后续的故障预测。与GAN相比,VAE由于数据分布恢复和损失函数计算方式的原因,生成数据往往较为模糊,与原始数据相差最大;流模型基于可逆变换,在处理多峰分布数据时表现不佳,而文中数据是一个多峰分布数据,因此流模型生成的数据真实性不如GAN;AR模型生成数据时难以捕捉数据中的长期依赖关系,对于时间跨度较长的数据集,其生成数据的真实性不高。而GAN通过生成器与判别器的相互对抗与优化,能够更好地学习时序数据的内在分布,在保证数据多样性的基础上,模拟数据的内在变化,生成与真实数据相似的样本,因此其评估结果较其他数据增强方法更好。

在数据增强的研究中,向原始数据集中添加噪声,是一种提高模型泛化能力、增加数据多样性的方法,但引入噪声也可能导致数据质量降低,影响生成模型的性能。因此,为研究数据集中的噪声对GAN生成样本的真实性和多样性的影响,在原始数据集中引入噪声,比较原始数据集与引入噪声后的数据集经过GAN数据增强后的新数据集质量。

向数据集中增加噪声,主要分为随机噪声、结构化噪声和特定噪声3种类型。随机噪声一般通过添加高斯噪声、泊松噪声等实现。结构化噪声一般通过添加特定的模式,如波形、方格等实现。特定噪声通常通过添加任务所需的特定信息实现。由于文中原始数据为时序数据集,且没有需要突出的特征类型,所以在增加噪声时选择随机噪声中的高斯噪声,高斯噪声能够模拟真实世界中连续变化过程的随机性。向原始数据集中添加噪声分为4步,具体如下。

步骤1:读取原始数据集  $L$ 。

步骤2:生成高斯噪声信号,高斯噪声是指概率密度函数服从高斯分布的噪声。其噪声位置固定,每一点都有噪声,但噪声幅值随机。高斯分布与高斯噪声生成公式如下。

$$p(\varphi) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(\varphi-\mu)^2/(2\sigma^2)} \quad (20)$$

$$n_{\text{Gauss}} = N(0, \sigma^2) \quad (21)$$

式中:  $p(\varphi)$  为  $\varphi$  的概率密度函数;  $\varphi$  为灰度值;  $\mu$  为  $\varphi$  的期望值;  $\sigma$  为  $\varphi$  的标准差;  $\sigma^2$  为  $\varphi$  的方差;  $n_{\text{Gauss}}$  为高斯噪声信号。文中高斯噪声的均值取 0, 方差分别取 0.01、0.05、0.10。

步骤 3: 将噪声数据添加至原始数据集中, 添加噪声后的数据集如下。

$$L' = L + n_{\text{Gauss}} \quad (22)$$

式中:  $L'$  为添加噪声后的数据集。

步骤 4: 保存添加噪声后的数据集, 并经过 GAN 进行数据增强, 得到增强后的新数据集。

为验证引入噪声对 GAN 生成样本的影响, 分别对原始数据集与添加噪声后的数据集进行数据增强, 生成 1 000 条样本数据, 并对其进行评估, 评估结果如表 2 所示。

表 2 添加噪声后的生成样本评估结果

Table 2 Generate sample evaluation results after introducing noise

评估指标	方差 0.01	方差 0.05	方差 0.10	无噪声
EMD	0.116	0.185	0.378	0.084
KL-D	0.127	0.214	0.442	0.108

由表 2 可知, 无噪声时数据集的 EMD 与 KL-D 值最小, 且随着噪声方差增加, 噪声数据的分散程度越高, 噪声的波动越大, EMD 与 KL-D 值随之增加。这说明原始数据集经 GAN 得到的生成样本与真实样本之间的相似度高, 而引入噪声并没有提高 GAN 的性能, 反而使得生成样本的真实性降低。这是因为文中数据为某市充电站充电账单与维护工单的真实数据, 由于现实环境等原因, 数据中存在一定的噪声, 再引入新的噪声, 反而使得数据质量下降, 进而影响 GAN 生成样本的真实性, 所以采用原始数据集已经足够支持文中模型的训练和预测。

因此, 经过特征集构建与 GAN 对故障数据进行增强, 共得到 1 600 条故障数据, 其中原始数据 600 条, GAN 生成数据 1 000 条。将其划分为训练集与测试集, 其中训练集 1 120 条, 测试集 480 条, 为确保测试结果可靠性, 测试集的 480 条数据均为原始数据, 不使用 GAN 生成数据作为测试集。利用文中所提 GAN-PSO-Bi-LSTM 模型对该数据进行分析, 具体的故障预测任务和相对应的条数如表 3 所示。

表 3 故障预测任务和相对应的条数

Table 3 Fault prediction task and corresponding number of items

故障预测任务	训练集		测试集		
	二分类 故障预测	故障 (672条)	正常 (448条)	故障 (297条)	正常 (183条)
充电接口电子锁 故障(268条)、 功率控制模块				充电接口电子 锁故障(176条)、 功率控制模块	
多分类 故障预测	通信故障 (120条)、 BMS 通信故障 (102条)、 其他故障(182条)	正常 (448条)	通信故障 (58条)、 BMS 通信故障 (36条)、 其他故障(27条)	正常 (183条)	

### 3.2 充电桩故障预测二分类结果分析

首先利用 PSO 算法通过训练集对 Bi-LSTM 参数隐藏层神经元数量和学习率进行寻优, 设置学习因子  $c_1$ 、 $c_2$  值为 2, 惯性权重  $w$  值为 0.9, 粒子数量为 100, 最大迭代次数为 50, 学习率取值范围为 [0.000 1, 0.001], 隐藏层神经元数量取值范围为 [10, 300], 种群初始分布状态与迭代完成后的最优位置如图 6 所示, 适应度变化曲线如图 7 所示。

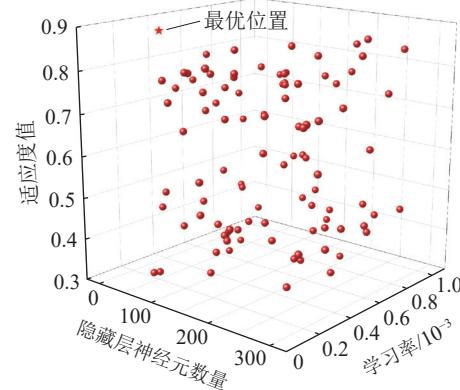


图 6 二分类模型的种群初始分布状态与最优位置

Fig.6 Initial distribution state and optimal position of the population for a binary classification model

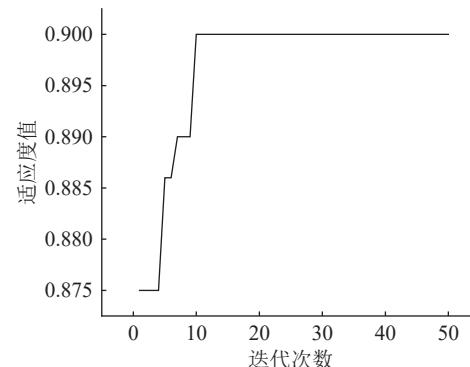


图 7 二分类模型的适应度变化曲线

Fig.7 Fitness change curve for a binary classification model

由图 6 和图 7 可知, 初始时种群较为平均地分布在空间中, 此时初始适应度值为 0.875, 迭代

10 次后,适应度值达到最大,为 0.90,此后不再变化。因此,种群的全局最优位置为(75,  $1.56 \times 10^{-4}$ , 0.90),该位置的数据即为对应的最佳参数。设置 Bi-LSTM 的参数为隐藏层神经元数量  $h=75$ ,学习率  $l=1.56 \times 10^{-4}$ ,考虑模型为二分类问题,激活函数为 Sigmoid,损失函数为 binary\_crossentropy。模型训练过程和测试过程得到的准确率及损失值如图 8、图 9 所示。

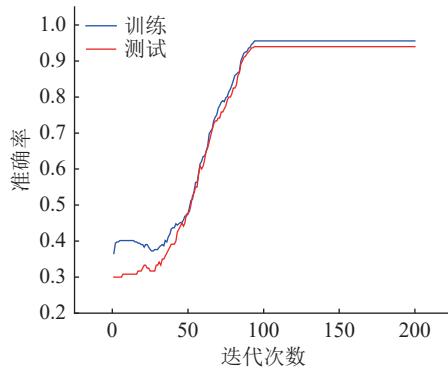


图 8 GAN-PSO-Bi-LSTM 二分类模型的测试和训练预测准确率

Fig.8 Test and training prediction accuracy of the GAN-PSO-Bi-LSTM binary classification model

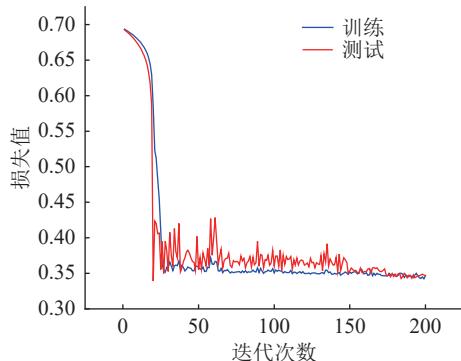


图 9 GAN-PSO-Bi-LSTM 二分类模型的测试和训练预测损失值

Fig.9 Testing and training prediction loss values of the GAN-PSO-Bi-LSTM binary classification model

由图 8 和图 9 可知,迭代 100 次后,故障预测结果趋于稳定,在训练集上达到 95.57% 的正确率,在测试集上达到 94.00% 的正确率,损失值也在迭代过程中减少并最终趋于稳定,说明模型训练完成。

为验证文中模型的改进效果,分别将文中模型与 Bi-LSTM、GAN-Bi-LSTM、PSO-Bi-LSTM 模型进行消融实验。其中 Bi-LSTM 和 PSO-Bi-LSTM 模型选用数据增强前的数据样本,以验证 GAN 数据增强的效果;GAN-Bi-LSTM 选用数据增强后的数据样本,以验证 PSO 优化初始参数的效果。选用准确率  $F_{\text{Acc}}$ 、精确率  $F_{\text{Pre}}$ 、召回率  $F_{\text{Rec}}$ 、F1 值  $F_1$  作为评估指标,计算如下:

$$F_{\text{Acc}} = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \times 100\% \quad (23)$$

$$F_{\text{Pre}} = \frac{T_p}{T_p + F_p} \times 100\% \quad (24)$$

$$F_{\text{Rec}} = \frac{T_p}{T_p + F_n} \times 100\% \quad (25)$$

$$F_1 = \frac{2F_{\text{Pre}}F_{\text{Rec}}}{F_{\text{Pre}} + F_{\text{Rec}}} \times 100\% \quad (26)$$

式中:  $T_p$  表示正确预测为正常充电的故障数据集数量;  $T_n$  表示正确预测为故障的故障数据集数量;  $F_p$  表示错误预测为正常充电的故障数据集数量;  $F_n$  表示错误预测为故障的故障数据集数量。

各模型的评估结果如表 4 所示。可知,GAN-PSO-Bi-LSTM 的 4 项评估指标值均最高,各项指标相较 PSO-Bi-LSTM 模型分别提高了 4.00、5.64、4.68、5.16 个百分点。Bi-LSTM、PSO-Bi-LSTM 模型所用数据集没有经过数据增强,数据量较少,对模型的训练不充分,因此最终的故障预测结果不佳,这验证了 GAN 在数据增强方面的有效性与合理性。而 GAN-Bi-LSTM 模型正确率也远低于文中模型,这是因为没有进行参数寻优,其初始参数并没有达到最佳,模型在训练过程中极易陷入局部最优解,这验证了 PSO 寻优对预测结果的提升效果。

表 4 二分类模型消融实验的评估指标

Table 4 Evaluation indexes for binary model ablation experiment 单位: %

对比模型	准确率	精确率	召回率	F1值
Bi-LSTM	84.17	80.80	82.14	81.46
GAN-Bi-LSTM	87.08	85.81	84.60	85.20
PSO-Bi-LSTM	90.00	87.56	87.95	87.75
GAN-PSO-Bi-LSTM	94.00	93.20	92.63	92.91

为进一步验证文中模型的可靠性,选取 FNN、卷积神经网络(convolutional neural network, CNN)、RF、梯度提升决策树(gradient boosting decision tree, GBDT)、残差网络(residual network, ResNet),对相同的数据增强后的数据集进行训练处理,评估指标对比如表 5 所示。

表 5 不同模型的二分类故障预测指标

Table 5 Binary fault prediction indicators of different models 单位: %

对比模型	准确率	精确率	召回率	F1值
FNN	88.75	84.71	85.71	85.21
CNN	90.42	90.23	88.39	89.30
RF	84.82	86.87	81.70	84.21
GBDT	88.13	86.36	85.94	86.15
ResNet	88.93	88.02	86.38	87.19
GAN-PSO-Bi-LSTM	94.00	93.20	92.63	92.91

由表 5 可知, GAN-PSO-Bi-LSTM 模型 4 项评估指标值均最高, 与效果较优的 CNN 模型相比, 文中模型各项指标分别提高了 3.58、2.97、4.24、3.61 个百分点。这是因为文中模型采用 Bi-LSTM 对时序数据集进行处理, 相较于其他算法, 该模型可以有效地处理时间序列数据中的长时依赖关系, 能够充分考虑时间序列数据前后文的数据信息, 抓住数据变化特征。因此, 从对 Bi-LSTM 改进的角度, 以及横向对比其他算法, 算例结果都验证了文中模型在充电桩故障二分类预测中的可行性和可靠性。

### 3.3 充电桩故障预测多分类结果分析

为验证文中模型在多分类问题中的性能, 选取数据中出现较多的故障问题, 即充电接口电子锁故障、功率控制模块通信故障和 BMS 通信故障, 剩余出现数量较少的故障分类为其他故障。利用训练集得到粒子群优化的 Bi-LSTM 参数, 粒子群参数设置同上, 种群初始分布状态与迭代完成的最优位置如图 10 所示, 适应度变化曲线如图 11 所示。

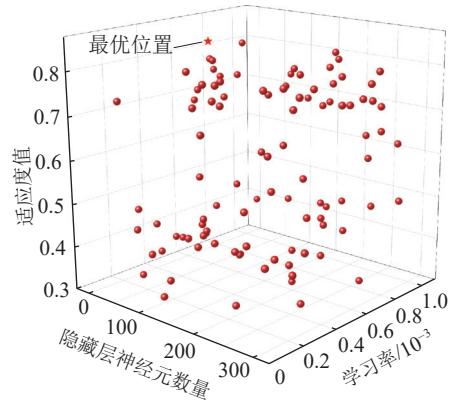


图 10 多分类模型的种群初始分布状态与最优位置  
Fig.10 Initial distribution state and optimal position of the population for a multi classification model

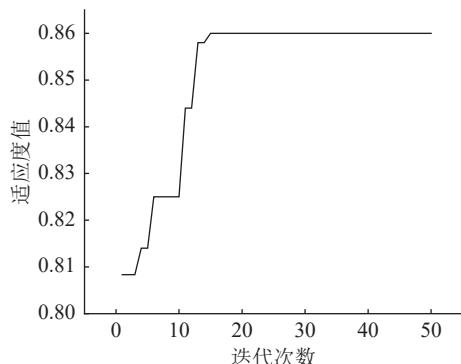


图 11 多分类模型的适应度变化曲线  
Fig.11 Fitness change curve for a multi classification model

由图 10 和图 11 可知, 初始时种群较为平均地分布在空间中, 此时的初始适应度值约为 0.81, 迭

代 16 次后, 适应度值达到最大, 为 0.86, 此后不再变化。因此种群的全局最优位置为  $(105, 4.16 \times 10^{-4}, 0.86)$ , 该位置的数据即为对应的最佳参数。设置隐藏层神经元数量  $h=105$ , 学习率  $l=4.16 \times 10^{-4}$ , 考虑模型为多分类问题, 激活函数为  $relu$ , 损失函数为  $categorical\_crossentropy$ 。GAN-PSO-Bi-LSTM 模型的训练过程和测试过程得到的准确率及损失值如图 12、图 13 所示。

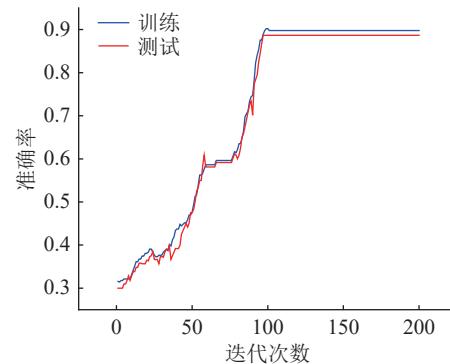


图 12 GAN-PSO-Bi-LSTM 多分类模型的测试和训练预测准确率

Fig.12 Test and training prediction accuracy of the GAN-PSO-Bi-LSTM multi classification model

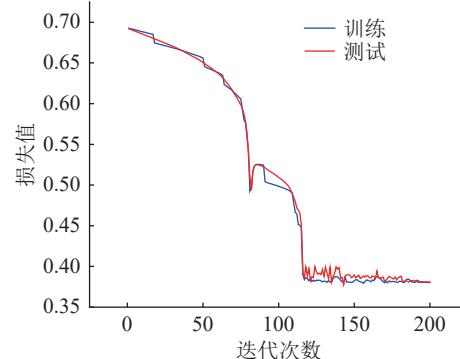


图 13 GAN-PSO-Bi-LSTM 多分类模型的测试和训练预测损失值

Fig.13 Testing and training prediction loss values of the GAN-PSO-Bi-LSTM multi classification model

由图 12 和图 13 可知, 迭代 100 次后, 故障预测结果趋于稳定, 在训练集上达到 89.78% 的正确率, 在测试集上达到 88.66% 的正确率, 损失值也在迭代过程中减少并最终趋于稳定, 与二分类相比, 多分类预测精度相对下降。

与 3.2 节类似, 比较 Bi-LSTM 改进效果, 并与其他算法进行比较, 比较结果如表 6 和表 7 所示。

由表 6 与表 7 可知, 文中模型评估指标值依旧最高。在消融实验中, 与表现较优的 PSO-BiLSTM 模型相比, 文中模型各项指标分别提高了 5.53、5.95、5.56、5.75 个百分点。在与其他模型相比时, 文中模型与表现较优的 CNN 相比, 各项指标分别

表 6 多分类模型消融实验的评估指标

Table 6 Evaluation indexes for multi model ablation experiment

单位: %

对比模型	准确率	精确率	召回率	F1值
Bi-LSTM	75.00	70.25	73.24	71.71
GAN-Bi-LSTM	79.17	74.69	77.97	76.29
PSO-Bi-LSTM	83.13	83.84	82.39	83.11
GAN-PSO-Bi-LSTM	88.66	89.79	87.95	88.86

表 7 不同模型的多分类故障预测指标

Table 7 Multi classification fault prediction indicators of different models

单位: %

对比模型	准确率	精确率	召回率	F1值
FNN	81.25	82.35	80.42	81.37
CNN	84.58	81.29	80.83	81.06
RF	76.79	75.28	78.96	77.08
GBDT	82.86	80.21	79.17	79.69
ResNet	83.13	80.45	80.42	80.43
GAN-PSO-Bi-LSTM	88.66	89.79	87.95	88.86

提高了 4.08、8.5、7.12、7.8 个百分点。与二分类故障预测相似, Bi-LSTM 模型在处理时序数据集上具有优势, 文中模型依旧故障预测性能最优。

以测试集中某编号的充电桩为例进行分析, 观察不同模型在多分类故障预测方面的表现, 该充电桩全年的维护工单记录为 3 条, 如表 8 所示。

表 8 某充电桩维护工单记录

Table 8 A certain charging pile maintenance work order record

真实故障原因	报修时间	完成时间
其他故障	3月7日	3月7日
充电接口电子锁故障	5月29日	5月29日
充电接口电子锁故障	8月18日	8月18日

根据上文所述特征集构建方法, 统计每次报修时间前 7 天与完成维修后 7 天的订单类型分布与充电结束原因, 如表 9 所示。

表 9 报修前后订单类型分布与充电结束原因

Table 9 Distribution of order types before and after repair and reasons for end of charging

阶段	时间	正常充电订单占比/%	无效充电订单占比/%	故障充电订单占比/%	正常结束占比/%	充电接口电子锁故障数量	其他故障数量
报修时间前7天	3月7日	7.68	30.76	61.56	2	0	15
	5月29日	0	40.00	60.00	0	7	5
	8月18日	12.50	25.00	62.50	2	9	1
完成维修后7天	3月7日	96.08	1.92	2.00	50	0	1
	5月29日	94.44	5.56	0	17	0	0
	8月18日	91.12	5.88	3.00	31	0	1

由表 9 可以看出, 报修时间前 7 天与完成维修后 7 天的订单总数不同, 这显示了订单数目变化与充电桩状态的关联。在报修时间前 7 天, 充电桩发生故障, 无效充电订单与故障充电订单占比高; 在完成维修后 7 天, 充电桩恢复正常, 订单总数激增, 正常充电订单占比占据绝对数量。不同模型对该充电桩的多分类预测结果如表 10 所示。

表 10 不同模型对该充电桩的预测结果

Table 10 Prediction results of different models on this charging pile

对比模型	3月7日	5月29日	8月18日
FNN	其他故障	其他故障	充电接口电子锁故障
CNN	其他故障	其他故障	充电接口电子锁故障
RF	其他故障	其他故障	充电接口电子锁故障
GBDT	其他故障	其他故障	充电接口电子锁故障
ResNet	其他故障	其他故障	充电接口电子锁故障
GAN-PSO-Bi-LSTM	其他故障	充电接口电子锁故障	充电接口电子锁故障

由表 10 可知, 不同算法在对该充电桩进行故障预测时, 都可以预测出其发生了故障, 但是进行多分类故障预测时, 对于 5 月 29 日的故障原因, 除了 GAN-PSO-Bi-LSTM 外, 其他模型都无法正确预测出故障原因为“充电接口电子锁故障”, 都错误地预测为其他故障。观察报修前后订单类型分布与充电结束原因, 可以看出 5 月 29 日前 7 天订单的结束原因中, “充电接口电子锁故障”和“其他故障”数量相似, 这一定程度上影响了算法的预测结果。实际上, 最开始几日, 故障不明显, 此时对故障的标记可能不准确, 所以越靠近报修时间, 对于充电结束原因的标记越可靠, 根据维护工单, “充电接口电子锁故障”才是 5 月 29 日充电桩的故障原因。其他算法受到“其他故障”数量的影响, 无法准确预测, 而 GAN-PSO-Bi-LSTM 模型采用改进的 Bi-LSTM, 相较于其他预测模型, 能更好地挖掘时间数据序列的依赖和影响因素特征依赖, 能抓住更靠近报修时间的“充电接口电子锁故障”数量这一特征, 并使用 PSO 对初始参数优化, 提高模型的预测精确度。因此, 从对 Bi-LSTM 改进的角度, 以及对比其他算法, 算例结果都验证了文中模型在充电桩故障多分类预测中的可行性和可靠性。

## 4 结论

文中所提 GAN-PSO-Bi-LSTM 模型能够实现对充电桩的故障预测, 提高充电设施的充电可靠性与安全性。针对充电订单数据较少的充电桩, 采取

GAN 进行数据扩充, 扩大数据集数量, 为后续模型的训练做准备。并且在构建数据集时, 区别于传统的电压、电流数据, 从充电账单数据与维护工单数据的角度, 聚焦于充电订单中的充电时间、充电结束原因等数据, 构建时序数据集, 避免了实际数据中常见的电压、电流数据不完整问题。在算例分析中, 验证了对 Bi-LSTM 模型的改进效果, 并与其他算法进行横向比较, 结果表明文中所提 GAN-PSO-Bi-LSTM 模型能有效提高故障预测精度, 可为其他领域的故障预测提供借鉴。

#### 参考文献:

- [1] 王斐, 李正砾, 叶萌, 等. 电动汽车充电对电网的影响及其优化调度研究述评[J]. 南方电网技术, 2016, 10(6): 70-80.  
WANG Fei, LI Zhengshuo, YE Meng, et al. Review on research of impact of electric vehicles charging on power grids and its optimal dispatch[J]. Southern Power System Technology, 2016, 10(6): 70-80.
- [2] 张洪程, 颜冰, 伞晨峻, 等. 电动汽车充电站投资策略分析[J]. 电力科学与技术学报, 2024, 39(5): 262-269.  
ZHANG Hongcheng, YAN Bing, SAN Chenjun, et al. Analysis of investment strategies for electric vehicle charging stations[J]. Journal of Electric Power Science and Technology, 2024, 39(5): 262-269.
- [3] 庞松岭, 赵海龙, 张晨佳. 考虑及充电需求差异的电动汽车充电设施协同优化配置[J]. 电测与仪表, 2024, 61(12): 171-177.  
PANG Songling, ZHAO Hailong, ZHANG Chenjia. Collaborative optimization configuration of electric vehicle charging facilities considering differences in charging demand[J]. Electrical Measurement & Instrumentation, 2024, 61(12): 171-177.
- [4] 张智禹, 王致杰, 杨皖昊, 等. 基于充电需求预测的电动汽车充电站选址规划研究[J]. 电测与仪表, 2024, 61(10): 39-49.  
ZHANG Zhiyu, WANG Zhijie, YANG Wanhai, et al. Research on location planning of electric vehicle charging station based on prediction of charging demand[J]. Electrical Measurement & Instrumentation, 2024, 61(10): 39-49.
- [5] 潘鸣宇, 孙绪坤, 李香龙, 等. 基于成本控制的集合覆盖模型的充电桩布局优化[J]. 现代电力, 2021, 38(2): 230-236.  
PAN Mingyu, SUN Xukun, LI Xianglong, et al. Optimization of charging pile layout based on cost control and collective coverage model[J]. Modern Electric Power, 2021, 38(2): 230-236.
- [6] 张震, 郭恩伯, 郭恩仲, 等. GBDT 与感知机融合的充电桩故障诊断方案[J]. 自动化与仪表, 2024, 39(4): 153-157.  
ZHANG Zhen, GUO Enbo, GUO Enzhong, et al. Fault diagnosis of charging piles by fusing gradient boosting decision tree and perceptron[J]. Automation & Instrumentation, 2024, 39(4): 153-157.
- [7] 臧斌斌, 高辉, 刘建. 充电设施安全运维服务模式及评价体系[J]. 广东电力, 2022, 35(10): 47-55.  
ZANG Binbin, GAO Hui, LIU Jian. Safe operation and maintenance service mode and evaluation system of charging facilities[J]. Guangdong Electric Power, 2022, 35(10): 47-55.
- [8] 吴豫, 董智, 赵阳, 等. 基于 LSTM 算法的配电网分布式电源和电动汽车充电站联合优化规划[J]. 供用电, 2023, 40(6): 64-74.  
WU Yu, DONG Zhi, ZHAO Yang, et al. Joint optimization planning of distribution network DG and EV charging station based on LSTM algorithm[J]. Distribution & Utilization, 2023, 40(6): 64-74.
- [9] 朱群波, 郝思鹏. 储能式充电站运行状态综合评价[J]. 电测与仪表, 2023, 60(8): 125-132.  
ZHU Qunbo, HAO Sipeng. Comprehensive evaluation of operation state of energy storage charging station[J]. Electrical Measurement & Instrumentation, 2023, 60(8): 125-132.
- [10] 张梅, 高犁, 陈万利. 基于 FASSA-SVM 的充电桩故障预测算法研究[J]. 电子测量技术, 2022, 45(12): 48-53.  
ZHANG Mei, GAO Li, CHEN Wanli. Research on fault prediction algorithm of charging pile based on FASSA-SVM[J]. Electronic Measurement Technology, 2022, 45(12): 48-53.
- [11] 王群飞, 尹忠东, 鄂涛. 基于随机森林算法的 V2G 充电桩故障诊断研究[J]. 电测与仪表, 2024, 61(8): 111-118.  
WANG Qunfei, YIN Zhongdong, E Tao. Research on fault diagnosis of V2G charging pile based on random forest algorithm[J]. Electrical Measurement & Instrumentation, 2024, 61(8): 111-118.
- [12] 董丽燕, 李雪刚, 傅博, 等. 基于模型预测控制的电动汽车直流充电桩功率器件开路故障诊断方法[J]. 电工技术, 2023(15): 41-46.  
DONG Liyan, LI Xuegang, FU Bo, et al. An open-circuit fault diagnosis method for EV DC charging pile power device based on model predictive control[J]. Electric Engineering, 2023(15): 41-46.
- [13] 郭静, 郭雅娟, 姜海涛, 等. 基于随机森林算法的电动汽车群体充电桩充电故障自动化检测系统[J]. 自动化与仪器仪表, 2023(12): 252-256.  
GUO Jing, GUO Yajuan, JIANG Haitao, et al. Automatic detection system of charging fault of electric vehicle group charging pile based on random forest algorithm[J]. Automation & Instrumentation, 2023(12): 252-256.
- [14] 刘裕舸. 基于 SO 优化神经网络的新能源汽车充电桩故障预测系统研究[J]. 广西电力, 2023, 46(4): 30-37.  
LIU Yuge. Research of fault prediction system for new energy vehicle charging pile based on SO optimized neural network[J]. Guangxi Electric Power, 2023, 46(4): 30-37.
- [15] 茅敏, 窦真兰, 陈良亮, 等. 基于 BOA-SSA-BP 神经网络的充电桩故障诊断方法[J]. 吉林大学学报(信息科学版), 2024, 42(2): 269-276.  
MAO Min, DOU Zhenlan, CHEN Liangliang, et al. Fault diagnosis method of charging pile based on BOA-SSA-BP neural network[J]. Journal of Jilin University (Information Science

- Edition), 2024, 42(2): 269-276.
- [16] 傅军, 孙志杰, 孟尚雄, 等. 充电桩运行异常率预测 [J]. 数理统计与管理, 2019, 38(1): 87-93.
- FU Jun, SUN Zhijie, MENG Shangxiong, et al. Forecasting of the abnormal rate of EV charging infrastructure [J]. Journal of Applied Statistics and Management, 2019, 38(1): 87-93.
- [17] GUPTA A, SARANGI S, SINGH A K. Wavelet based enhanced fault detection scheme for a distribution system embedded with electric vehicle charging station [C]//2023 5th International Conference on Power, Control & Embedded Systems (ICPCES). Allahabad, India. IEEE, 2023: 1-6.
- [18] GÓMEZ P I, MORENO A B, LIN J, et al. Data-driven thermal modelling for anomaly detection in electric vehicle charging stations [C]//2022 IEEE/AIAA Transportation Electrification Conference and Electric Aircraft Technologies Symposium (ITEC+EATS). Anaheim, CA, USA. IEEE, 2022: 1005-1010.
- [19] 袁单, 刘鸿鹏, 陈良亮, 等. 基于改进决策树的充电桩故障预测方法 [J]. 电气自动化, 2023, 45(6): 92-94, 103.
- YUAN Dan, LIU Hongpeng, CHEN Liangliang, et al. Fault prediction method of charging pile based on improved decision tree [J]. Electrical Automation, 2023, 45(6): 92-94, 103.
- [20] 吴丹, 王俊, 许燕. 基于深度学习的公共充电桩故障预测模型 [J]. 电子设计工程, 2022, 30(5): 127-130, 135.
- WU Dan, WANG Jun, XU Yan. Fault prediction model of common charging point based on deep learning [J]. Electronic Design Engineering, 2022, 30(5): 127-130, 135.
- [21] 李翔, 高辉, 陈良亮. 基于 GAIN 的数据插补及 Bi-GRU 在充电桩预警中的应用 [J]. 广东电力, 2022, 35(12): 22-31.
- LI Xiang, GAO Hui, CHEN Liangliang. Data imputation based on GAIN and application of Bi-GRU in charging pile early warning [J]. Guangdong Electric Power, 2022, 35(12): 22-31.
- [22] 臧斌斌, 高辉, 刘建, 等. 基于改进 GRU 模型的直流充电桩故障预测 [J]. 计算机系统应用, 2023, 32(2): 356-363.
- ZANG Binbin, GAO Hui, LIU Jian, et al. Fault prediction of DC charging pile based on improved GRU model [J]. Computer Systems and Applications, 2023, 32(2): 356-363.
- [23] 吴玮怡. 数据挖掘在电动汽车充电桩充电故障检测中的应用 [J]. 汽车与新动力, 2023, 6(2): 81-84.
- WU Weiyi. Application of data mining in charging fault detection of electric vehicle charging pile [J]. Automobile and New Powertrain, 2023, 6(2): 81-84.
- [24] 许笑, 高翔, 李光熹, 等. 基于数据挖掘的电动汽车充电设施充电安全故障特性分析 [J]. 电器与能效管理技术, 2022(3): 15-22.
- XU Xiao, GAO Xiang, LI Guangxi, et al. Analysis on charging safety fault characteristics of electric vehicle charging facilities based on data mining [J]. Electrical & Energy Management Technology, 2022(3): 15-22.
- [25] 蔡榕, 杨雪, 田江, 等. 基于相关性分析和生成对抗网络的电网缺失数据填补方法 [J]. 电力工程技术, 2024, 43(1): 229-237.
- CAI Rong, YANG Xue, TIAN Jiang, et al. A power system missing data filling method based on correlation analysis and generative adversarial network [J]. Electric Power Engineering Technology, 2024, 43(1): 229-237.
- [26] 李华瑞, 李文博, 李铮, 等. 基于生成对抗网络与度量学习的数据驱动频率安全评估 [J]. 电力系统保护与控制, 2024, 52(18): 101-111.
- LI Huarui, LI Wenbo, LI Zheng, et al. Data-driven frequency security assessment based on generative adversarial networks and metric learning [J]. Power System Protection and Control, 2024, 52(18): 101-111.
- [27] 宋辉, 范龙祥, 郭双权. 基于数据增强和特征注意力机制的灰狼优化算法-优化残差神经网络变压器故障诊断方法 [J]. 现代电力, 2024, 41(2): 392-400.
- SONG Hui, YUAN Longxiang, GUO Shuangquan. GWO-ResNet power transformer fault diagnosis method based on data augmentation and feature attention mechanism [J]. Modern Electric Power, 2024, 41(2): 392-400.
- [28] 芦肇基, 沈艳霞, 谭永强. 基于多结构融合 WGAN 的模糊绝缘子图像复原方法研究 [J]. 电力系统保护与控制, 2024, 52(22): 166-175.
- LU Zhaoji, SHEN Yanxia, TAN Yongqiang. A WGAN blur insulator image restoration method based on multi-structure fusion [J]. Power System Protection and Control, 2024, 52(22): 166-175.
- [29] 杨茜. 基于 Bi-LSTM 和图注意力网络的多标签文本分类算法 [J]. 计算机应用与软件, 2023, 40(9): 145-150, 183.
- YANG Xi. Multi label text classification algorithm based on Bi-LSTM and graph attention network [J]. Computer Applications and Software, 2023, 40(9): 145-150, 183.
- [30] 朱轶伦, 俞一峰, 虞明智, 等. 基于多数据源融合的电网故障判别与告警技术研究 [J]. 电气自动化, 2024, 46(2): 32-35, 39.
- ZHU Yilun, YU Yifeng, YU Mingzhi, et al. Research on power grid fault identification and warning technology based on multiple data source fusion [J]. Electrical Automation, 2024, 46(2): 32-35, 39.
- [31] 黄君泽, 吴文渊, 李轶, 等. 面向动态公交的离散分层记忆粒子群优化算法 [J]. 计算机工程, 2024, 50(4): 20-30.
- HUANG Junze, WU Wenyuan, LI Yi, et al. Discrete hierarchical memory particle swarm optimization algorithm for dynamic public transport [J]. Computer Engineering, 2024, 50(4): 20-30.
- [32] 刘斌, 吉春霖, 曹丽君, 等. 基于自适应噪声完全集合经验模态分解与 Bi LSTM-Transformer 的锂离子电池剩余使用寿命预测 [J]. 电力系统保护与控制, 2024, 52(15): 167-177.
- LIU Bin, JI Chunlin, CAO Lijun, et al. Prediction of remaining service life of lithium-ion batteries based on complete ensemble empirical mode decomposition with adaptive noise and BiLSTM-Transformer [J]. Power System Protection and Control, 2024, 52(15): 167-177.

- [33] 李强, 车文龙. 基于改进粒子群优化神经网络的电机故障诊断[J]. 电气传动, 2020, 50(1): 103-108.  
LI Qiang, CHE Wenlong. Motor fault diagnosis based on improved particle swarm optimization neural network[J]. Electrical Drive, 2020, 50(1): 103-108.
- [34] 曹洁, 尹浩楠, 王进花. MACDCGAN 的发电机轴承故障诊断方法[J]. 振动与冲击, 2024, 43(11): 227-235.  
CAO Jie, YIN Haonan, WANG Jinhua. Fault diagnosis of generator bearings based on MACDCGAN[J]. Journal of Vibration and Shock, 2024, 43(11): 227-235.
- [35] 李晖照, 王雪, 郭莹. 基于 KL 变换和 KL 散度的电网数据特征提取与分类[J]. 电测与仪表, 2019, 56(6): 87-92.  
LI Huizhao, WANG Xue, GUO Ying. Feature extraction and

classification in smart grid data based on KL-divergence and KL transform[J]. Electrical Measurement & Instrumentation, 2019, 56(6): 87-92.

作者简介:



周秋阳

周秋阳(2000), 男, 硕士在读, 研究方向为充电桩故障预警与诊断(E-mail: [1222056606@njupt.edu.cn](mailto:1222056606@njupt.edu.cn));

高辉(1981), 男, 博士, 教授, 通信作者, 研究方向为电动汽车充电安全大数据预警及智能保护控制;

李炜卓(1989), 男, 博士, 讲师, 研究方向为知识表示、自然语言处理。

## A fault prediction method for charging pile based on GAN data enhancement and improved Bi-LSTM

ZHOU Qiuyang<sup>1</sup>, GAO Hui<sup>1</sup>, LI Weizhuo<sup>2</sup>, GUI Yaocheng<sup>2</sup>

(1. College of Automation, Nanjing University of Posts and Telecommunications, Nanjing 210023, China; 2. School of Modern Posts, School of Intelligent Transportation, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

**Abstract:** In recent years, the rapid development of electric vehicles (EVs) has likewise led to the construction of EV charging infrastructure, so the research on charging reliability and safety of EV charging facilities has become a focus of attention. However, most of the data used in existing research are complete and sufficient simulation data, when faced with actual data, the prediction accuracy is often affected by insufficient or incomplete data. To solve these problems, a data-driven approach is used to achieve early warning of faults during the charging process of charging equipment. Firstly, feature selection is performed to select appropriate data features. Secondly, the order data is filtered, the dataset is constructed and normalized. Secondly, the dataset is divided into a training group and a test group, the training group is used for model training, and the test group is used to judge the advantages and disadvantages of model training. Then, the divided training group is augmented with generative adversarial networks (GAN) to expand the data size and form a sufficient amount of new data. Subsequently, the data are inputted into bi-directional long-short term memory (Bi-LSTM) and the initial parameters are optimized using particle swarm optimization (PSO). A number of trials are conducted to observe the results of the modeling tests. Finally, in comparison with other prediction models, it is verified that the GAN-PSO-Bi-LSTM model has higher prediction performance, which improves the fault prediction accuracy of charging piles.

**Keywords:** charging pile; fault prediction; data enhancement; generative adversarial network (GAN); particle swarm optimization (PSO); bi-directional long short-term memory (Bi-LSTM)

(编辑 方晶)