

基于极端不平衡学习的泛化低压异常箱表关系识别 研究与应用

管永明^{1,2}, 王 刚³, 骆凯波³, 吕 梁², 吕晓雯², 史玉良^{1,2}

(1. 山东大学软件学院, 山东济南 250101; 2. 山大地纬软件股份有限公司, 山东济南 250100; 3. 国网重庆市电力公司, 重庆 400015)

摘 要: 针对低压配电网箱表关系存在人工核查成本高、异常案例少、难以实现异常规律捕获的问题, 采用极端不平衡分类学习方法实现低压异常箱表关系识别的泛化应用推广. 通过电压原理识别出部分异常箱表关系样本集, 随后构建CNN(卷积神经网络)异常箱表关系识别模型, 通过样本三分类赋权值实现类别均衡处理; 并在模型推广应用过程中, 采用强化学习实现离线模型的在线泛化学习, 并以分组模型交互学习和竞争优化的方式筛选出最优泛化识别模型. 实验证明, 通过人工核查和数据反馈, 该方法可实现模型对异常样本数据分布规律的自拟合学习, 提高模型对不同应用环境的泛化性, 进一步降低人工现场核查工作量, 保障低压台区用户拓扑网络关系的准确性.

关键词: 极端不平衡分类; 电压曲线识别; 卷积神经网络; Adaboost 算法; 分组强化学习

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112(2021)08-1507-08

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20191372

Generalized Identification for Low-Pressure Abnormal Box-Table Relationship Based on Extreme Unbalance Classification Learning

GUAN Yong-ming^{1,2}, WANG Gang³, LUO Kai-bo³, LÜ Liang², LÜ Xiao-wen², SHI Yu-liang^{1,2}

(1. School of Software, Shandong University, Jinan, Shandong 250101, China;

2. Dareway Software Co., Ltd., Jinan, Shandong 250100, China;

3. State Grid Chongqing Electric Power Company, Chongqing 400015, China)

Abstract: Due to high labor cost and few abnormal cases of power box-table relations inspection, which difficulty to obtain the law. The extreme unbalanced classification learning method was used to capture the generalization. Through the principle of voltage, abnormal box-table relationship sample sets were identified. And by three-class weighting balance, the CNN(convolutional neural network) abnormal box-table relationship recognition model was constructed. In addition, the grouped parallel generalization learning of recognition model was realized by reinforcement learning. The experiment proves that, through self-learning the distribution of newly identified abnormal sample data, which improve the generalization to different environments. This reduces the workload of manual on-site verification and ensures the accuracy of the topology network relationship in the low-voltage station area.

Key words: extreme unbalanced classification; voltage curve identification; convolutional neural network; Adaboost algorithm; group reinforcement learning

1 引言

台区箱表关系异常, 是档案录入、现场安装存在问题, 导致用户电能表与计量箱的连接关系的档案记录与现场台区不一致. 随着泛在电力物联网及能源信息化转型发展^[1], 准确的台区箱表关系不仅关系到用户准确缴费、设备快速定位与故障检修, 同时, 底层拓扑数

据是上层应用全方位开展的基础^[2], 是实现能源互联网和泛在电力物联网建设的前提保障.

然而在电力系统建设初期阶段, 低压配电网建设前期规划不够完善, 造成用户拓扑网络无统一标准, 包含部分台区网络未按照规则部署, 仅以用户通电为目的, 缺乏精确箱表信息; 且部分地区用户档案原始记

录多为纸质文本,后经手工录入系统,往往存在错误、遗漏的情况;此外,在配网台区运行过程中,电表存在增删调整,信息变更不及时导致箱表关系错误.针对上述问题,国网公司开展各类低压台区箱表关系核查方案,如通过台区内拉闸观测,以人工普查更正台区内箱表关系^[3],然而这种方式工作量大且存在二次人为误判,难以开展实施.随着用电信息采集系统数据的积累,电力公司欲通过对电能表相关采集数据进行分析,以实现异常低压台区箱表关系检测,但由于异常箱表关系样本不足,且采集数据与识别目标相关特征不直观,单一特征识别导致异常箱表关系识别率低(如当前采用电压曲线识别方法,箱表关系核查正确率仅为整体箱表核查量的48.98%)^[4],且仅为部分异常识别,难以实现对异常箱表关系的全覆盖识别.

针对上述问题,本文采用极端不平衡分类学习构建泛化低压异常箱表关系识别模型,以线上数据处理的方式降低箱表关系人工现场校核工作量.主要贡献如下:

(1)针对正负样本类别不平衡且新识别样本数据量极少^[5],导致模型难以捕获异常数据分布规律问题,采用Adaboost算法^[6]以互补模式获取样本集三分类均衡权重比,并提取复用历史学习框架,以逐步递增的方式增加模型层级,简化模型的训练过程,保障神经网络规模^[7]的正确性和适用性.

(2)针对应用环境改变导致模型对异常样本识别率降低的问题,采用强化学习实现识别模型的分组并行泛化应用推广,通过强化对新识别异常样本数据分布规律的自学习^[8],持续提高模型对不同应用环境的适应性.

(3)为避免强化学习长时间无目标调参训练导致的高试错成本^[9],采用训练样本集分组进行并行训练,从而在保证学习效率的同时,提高识别模型的学习多样性和专向性,且在某个模型陷入局部最优时^[10],仍可通过模型的交互学习与竞争,筛选出最优泛化识别模型,降低模型训练难度的同时,提高模型识别的正确率和召回率^[11].

2 整体过程描述

本文整体模型构建过程如图1所示.

首先对大范围低压用户采集样本电压数据进行聚类识别,获取电压疑似异常样本,并与随机提取的低压电能表样本共同构成模型训练样本集(即20%训练样本集);随后,基于上述20%训练样本集,采用CNN

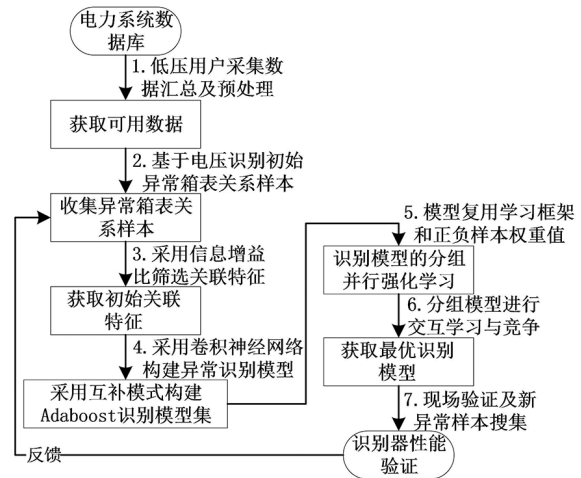


图1 低压用户异常箱表关系识别方法的整体流程图

(Convolutional Neural Network, 卷积神经网络)^[12]构建异常箱表关系识别模型,并以Adaboost集成模型互补模式构建异常箱表关系识别模型集,输出历史复用学习框架和正负样本分配权值,并在历史复用学习框架基础上,采用递增方式构建CNN异常箱表关系识别模型^[13],具体过程如图2所示;接着将70%训练样本集分组,对CNN异常箱表关系识别模型进行分组迭代强化学习训练,并对识别结果进行核查反馈,最终以模型交互学习和轮换竞争的模式生成最优泛化CNN异常箱表关系识别模型,具体过程如图3所示;并以剩余10%的训练样本集对最优泛化模型的准确率和召回率进行验证,并将验证异常数据进行反馈,不断完善和优化模型的适用性与泛化性.

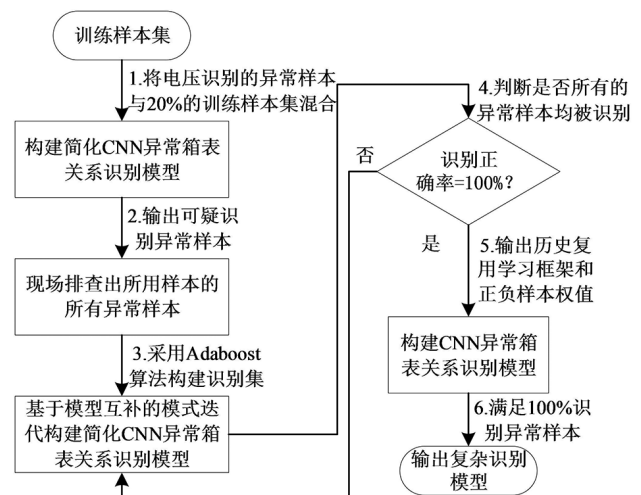


图2 CNN简化异常箱表关系识别模型集流程图

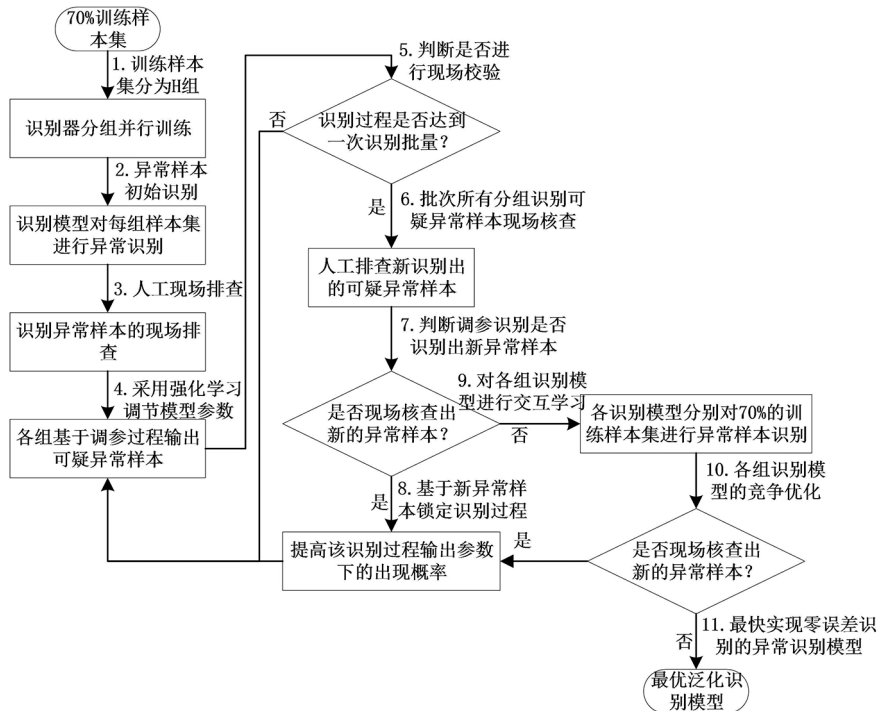


图3 最优泛化CNN异常箱表关系识别模型

3 初始异常样本获取过程

根据电路基本原理,同一导体上电压处处相等,因此同一表箱内同一相序下电能表电压也应时刻相等^[14]。

根据上述原理,采用电压曲线识别方法实现对初始异常箱表关系样本集的搜集,首先,定义同一台区为同一计量点,在同一计量点进行电压曲线归一化;随后采用K-Means聚类算法^[15],将同一箱表内根据三相电压连接方式聚类分为3类,疑似异常类判定规则如下。

电压聚类数量若满足:

$$N_i < \min(N_c, N_{\text{总}}/M) \quad (1)$$

则此类中所有数据即为异常,其中 $i \in [0, 2]$, 表示同一计量箱下电能表电压曲线可能存在至多3种模式; N_c 为固定阈值,表示同一箱内某类电压曲线聚类数量少于 N_c , 则可能存在户变异常, 设定为2; N_i 为某三相电压连接方式聚类数量; $N_{\text{总}}$ 为该电能表箱内总的电能表数

量; M 为自定义阈值, 设为5效果最好。

然而上述方法在实际应用中存在缺陷,如固定阈值设定导致的误判情形、计量故障导致电压判别户变关系失效的情形。上述问题说明基于单一特征的固化判定模式导致户变关系判定精准度缺失,而实际中的户变关系异常往往导致多类采集数据存在异常。

综上所述,采用电压曲线识别方法判断户变关系,虽精准度不足,但针对低压智能电能表数量亿级而异常箱表关系数量占比不足1%,导致缺乏异常样本的问题,仍可线上识别出大范围区域的可用的部分异常箱表关系样本集。故本文基于此数据集基础对低压异常箱表关系识别精准度进行提升。初始异常样本集获取过程如下。

基于某地区的用电信息采集系统,获取2019年8月可用约12万余例低压电能表相关采集数据为样本进行电压曲线识别聚类,共计获取疑似异常样本465例,现场核查结果如表1所示。

表1 基于电压曲线一致性标准识别的现场核查结果

核查电能表	465块	涉及计量箱	176个
异常分类	涉及电能表数	涉及计量箱数	占有核查计量箱比例
箱表关系异常	243	86	48.98%
计量故障	72	65	36.73%
疑似采集异常	135	32	18.37%
接线错误	6	5	3.06%
异常总数	456	167(累计188次)	94.90%
误判异常样本	9	9	5.10%

通过人工对 465 块疑似异常样本进行核查,获取箱表关系异常样本 243 例,作为初始异常样本集。

4 构建异常箱表关系识别模型

由于训练样本集正负样本数据分布不平衡,本文采用 Adaboost 算法辅助完成 CNN 异常箱表关系识别模型的构建,具体过程如下。

4.1 CNN 简化异常箱表关系识别模型

将电压曲线识别出的疑似异常箱表关系样本 243 例与 20% 的训练样本集混合后形成初始训练样本集,首先基于初始训练样本集构建 CNN 简化异常箱表关系识别模型,过程如下。

训练样本的日采集量为 24 点,共计涉及 126 类相关特征数据,非 24 点采集数据以累计均值或恒值计入,故输入数据为 126×24,卷积层共享同一卷积核,采用了 2-GPU 并行结构,基于 keras 深度学习框架,最终模型架构如下所示。

输入:126×24,步长为 1。

第一层卷积:12×3 大小的卷积核 24 个,每个 GPU 上 12 个。

第一层 max-pooling:2×2 的核。

第二层卷积:12×3 卷积核 48 个,每个 GPU 上 24 个。

第二层 max-pooling:2×2 的核。

第一层全连接:120 维,即将第四层卷积的输出连接成为一个一维向量,作为该层的输入。

输出层:采用 Sigmoid 函数,输出为 2 分类,即箱表关系正常类 $C = 0$ 和箱表关系异常类 $C = 1$ 。

基于上述架构^[16],以反向梯度调参,采用二值自定义权值交叉熵作为损失函数,并设定二分类初始权重 w ,从而降低初始正负样本不平衡分类的影响,损失函数如下

$$L(y, \hat{y}) = -w[y \log \hat{y} + (1 - y) \log(1 - \hat{y})] \quad (2)$$

$$w = \begin{cases} 4, & C = 1 \\ 1, & C = 0 \end{cases} \quad (3)$$

$$\hat{y} = \text{Sigmoid}(\tilde{y}) = \frac{1}{1 + e^{-\tilde{y}}} \quad (4)$$

其中, y 为样本的期望输出, \hat{y} 为样本的实际输出, \tilde{y} 为 CNN 的原始输出值。

以单数据点验证模型后,采用迭代反向梯度训练,目标为 CNN 简化异常箱表关系识别模型 100% 将 243 例疑似异常样本分类为负样本,此时对初始训练样本集的所有样本进行人工现场核查,输出 CNN 简化异常箱表关系识别模型的识别误差率。

4.2 Adaboost 集成算法辅助优化模型

根据 CNN 简化异常箱表关系识别模型的识别误差率,调整样本权重值,即提高识别错误样本的权重,降

低识别正常样本的权重,构建 CNN 简化异常箱表关系识别模型集,首先,以 4.1 节构建的识别模型作为首个识别模型,识别误差率为

$$\varepsilon_t = \frac{\sum w_t^n \delta(y_t \neq \hat{y}_t)}{\sum w_t^n} \quad (5)$$

其中, n 表示第 n 个样本; y_t 为样本在第 t 个模型中的期望输出; $\delta(y_t \neq \hat{y}_t)$ 表示预测值与实际值不一致,则为 1,一致则为 0; ε_t 为第 t 个识别模型的识别误差; w_t^n 为第 t 个模型第 n 个样本的权重;初始各样本权重 $w^n = 1$ 。

基于识别误差 ε_t ,更新初始训练样本集的第 n 个样本构建第 $t+1$ 个模型权重,即当 $\varepsilon_t < 0.5$ 时,如果第 n 个样本在第 t 个模型识别中分类正确,则其在第 $t+1$ 个模型中的样本权重为 $w_{t+1}^n = w_t^n \times e^{-\alpha_t}$;如果第 n 个样本在第 t 个模型识别中分类不正确,则其在第 $t+1$ 个模型中的样本权重为 $w_{t+1}^n = w_t^n \times e^{\alpha_t}$,其中, $\alpha_t = \ln \sqrt{(1 - \varepsilon_t) / \varepsilon_t}$,表示基于第 t 个模型的识别错误率,调整构建第 $t+1$ 个模型时训练样本的权重值,即提高第 t 个模型判别错误的样本权重,降低其判断正确的样本权重,从而在构建第 $t+1$ 个模型时,模型更关注第 t 个模型判别错误的样本。

将各个样本权重更新后,根据更新的训练样本集构建第 $t+1$ 个 CNN 简化异常箱表关系识别模型,模型构建方法与 4.1 节相同,并将 $t+1$ 个生成的 CNN 简化异常箱表关系识别模型形成识别集,判断模型集是否将初始训练样本集内的异常箱表关系样本无误差识别出,即判断最终各个样本分类结果 H 与样本实际正负类别是否一致:

$$H = \text{sign} \left(\sum_{t=1}^{t+1} \alpha_t y_t \right) \quad (6)$$

其中, α_t 是基于第 t 个模型的识别误差率计算出的,故可基于第 t 个模型的性能优劣给出判别占比权重。例如:模型识别误差率为 $\varepsilon_t = 0.1$,则 $\alpha_t = 1.1$,模型判别比重提高;模型识别误差率为 $\varepsilon_t = 0.4$,则 $\alpha_t = 0.2$,模型判别比重降低。

若不一致,则继续以互补模式构建新的 CNN 简化异常箱表关系识别模型;若一致,则计算模型集中,所涉及各样本权重值累计值,标准化比例正样本权重值为 1,则获取正负样本比重值为 $1:\alpha$,从而获取为平衡样本不均衡分布的负样本权重比例值,并提取最末次构建的 CNN 简化异常箱表关系识别模型的样本最大权重值 β 作为后期新识别异常样本的设定权值;同时,基于最末次构建的 CNN 简化异常箱表关系识别模型作为历史复用学习框架,采用跨层共享参数、逐层增加隐藏层的方式完成 CNN 异常箱表关系识别模型的构建。

5 模型泛化学习推广过程

本文通过对 CNN 异常箱表关系识别模型进行分组强化学习^[17]获取最优泛化识别模型。

将随机提取的 70% 训练样本集分为 H 组,将上一小节训练获取的 CNN 异常箱表关系识别模型对每组样本集进行异常识别,识别出初始可疑异常样本后进行人工现场排查,从而识别出各组内实际存在的异常样本;随后,采用强化学习对 CNN 异常箱表关系识别模型的首层卷积神经网络参数进行调整,过程如下。

由于训练集样本正负样本数据分布极端不平衡,对正负样本的权重分配为 $1:\alpha$,对历次新识别异常箱表关系样本权重值设为 β ,定义卷积神经网络的参数调节过程 τ :

$$\tau = \{s_1, a_1, r_1, s_2, a_2, r_2, \dots, s_t, a_t, r_t, \dots, s_T, a_T, r_T\} \quad (7)$$

其中, τ 为从模型某一参数调节开始到识别出某一可疑异常箱表关系样本的一次识别过程, s 表示模型当前所有参数值和识别结果的状态, a 表示参数调节动作,即对模型参数 θ 值进行增值或减值调整, θ 表示模型的参数集合, r 为调参识别成功异常样本获取的识别奖励值, T 为一次识别总调节次数。

$$R(\tau) = \sum_{t=1}^T r_t \quad (8)$$

其中, $R(\tau)$ 表示一次识别模型识别异常样本获取的奖赏。

$$\bar{R}_\theta = \sum_{\tau} R(\tau) P(\tau/\theta) \approx \frac{1}{N} R(\tau^n) \quad (9)$$

其中, \bar{R}_θ 为模型参数为 θ 时, N 次识别过程的平均奖赏期望, N 为历次 T 的累计值, $P(\tau/\theta)$ 为模型参数为 θ 时 τ 这一调参识别过程出现的概率。

\bar{R}_θ 为最大化训练目标, θ 为模型强化学习训练完成后的输出值,采用梯度上升法最大化 \bar{R}_θ :

$$\theta^{n+1} \leftarrow \theta^n + \eta \nabla R_{\theta^n} \quad (10)$$

其中, θ^n 为第 n 次识别过程的模型参数值, $n \in [1, N]$, η 为自定义的学习率,将其设置为一个阈值上限, ∇R_{θ^n} 为第 n 次的参数调节值后,根据识别结果反馈的梯度值。

$$\nabla R_{\theta^n} \approx \frac{1}{N} \sum_{n=1}^N \sum_{t=1}^T R(\tau^n - b) \frac{\nabla P(a_t^n/s_t^n, \theta)}{P(a_t^n/s_t^n, \theta)} \quad (11)$$

其中, τ^n 为第 n 次识别过程, s_t^n 为第 n 次识别过程中第 $t-1$ 次模型调参后的当前模型所有参数值和识别结果的状态, a_t^n 为第 n 次识别过程中第 $t-1$ 次模型调参后模型的参数调节动作, b 为自定义阈值,用以避免采样不均衡, a_t^n/s_t^n 表示在 s_t^n 状态下采取 a_t^n 动作,若 $R(\tau^n - b)$ 为正值,即异常样本识别正确,则 $P(a_t^n/s_t^n, \theta)$ 变大,若

$R(\tau^n - b)$ 为负值,即异常样本识别错误,则 $P(a_t^n/s_t^n, \theta)$ 变小。

基于上述过程迭代调参进行模型强化学习训练,直至各组识别模型均完全识别出当前已核实异常样本,并输出各组历次迭代构建识别出的新的未核实的可疑异常样本。设定 N 次识别过程为一次识别批次,若现场核查验证出新异常样本,对新异常样本赋权值 β ,并提高该识别过程输出参数下的出现概率后,继续进行迭代调参强化学习训练。若否,则基于分组训练获取的各组识别模型,对总体 70% 训练样本集进行异常样本识别,输出各组识别的可疑异常样本;若存在新核实的异常样本,则对新异常样本赋权值 β 并进行模型强化学习训练,直至无新核实异常样本出现,筛选最优识别模型。

6 实验与分析

本文以某市用电信息采集系统为实验背景平台, CNN 异常箱表关系识别模型训练采用 Transformer 编码器,实验结果为累计 5 次实验数据的均值,具体实验过程如下所示。

6.1 实验数据获取及处理

本文以用电信息采集系统历史累计数据及计量设备实时采集数据为主要数据源,结合用电信息采集系统与营销信息化系统之间接口集成获取相关用户数据。共计获取 2019 年 8 月可用约 6 万余例低压电能表相关数据构建训练样本集,集合用电行为、区域数据、运行数据等多维度数据共计 237 例进行筛选^[18],样本集采集数据进行数据清洗、数据补全、归一化处理,采用信息增益比进行初始特征筛选,筛选阈值为 0.23,提取低压电能表异常箱表关系识别特征 126 个。

6.2 CNN 异常箱表关系识别模型识别结果与分析

随机提取 20180 例样本集为训练样本集,从训练样本集随机抽取 4136 例样本(内含基于电压曲线一致性标准识别的 16 例异常箱表关系样本),将电压曲线识别出的异常箱表关系样本 243 例与 4136 例训练样本集混合后形成初始训练样本集,构建 CNN 简化异常箱表关系识别模型,并对 4136 例样本进行逐一排查,获取新异常箱表关系样本 28 例,具体现场核查结果如表 2 所示。

由表 2 可得,基于人工逐一识别的方法对异常箱表关系识别效率低,难以获取异常样本集,故本文基于当前电压识别方法实现对前期异常样本集的获取。

将 CNN 简化异常箱表关系识别模型作为首个识别模型,对 4379 例样本的次级训练样本集进行识别,初始各样本权重 $w^n = 1$,识别出实际异常样本 240 例,识别误差 $\varepsilon_{t=1} = 11.76\%$,尽管初始的训练样本存在误分类

表2 基于人工逐一识别的现场核查结果

核查电能表总数	4120块	涉及计量箱总数	623个
异常分类	涉及电能表数量	涉及计量箱数量	占有核查计量箱比例
箱表关系错误	28	21	3.37%
电压曲线识别	16	—	—
异常总数	44	—	—

的情况, CNN简化异常箱表关系识别模型仍可基于大类分类趋势捕获潜在的异常箱表关系识别规律. 继续以互补模式构建新的简化CNN异常箱表关系识别模型, 迭代共计生成6个CNN简化异常箱表关系识别模型100%识别出异常箱表关系样本, 将第6个简化CNN异常箱表关系识别模型作为历史复用学习框架, 采用跨层参数共享、逐层增加的方式增加模型的复杂度.

即基于历史复用学习框架共计5层隐藏层的神经网络及其各层参数数据, 对其第一卷积层和第二卷积层均进行逐次增加, 并在参数微调阈值百分比为2%的前提下进行跨层参数共享, 全连接层数量保持不变, 采用历史复用学习框架的均值识别准确率为79.22%, 随着第一第二卷积层数量的增加, 模型为第一卷积层6层和第二卷积层4层时, 5次实验结果准确率均在98.46%~99.83%, 实验结果相对稳定说明模型鲁棒性较好, 准确率均值为99.16%, 随后随着卷积层数量的增加, 准确率及鲁棒性均降低. 故最终获取隐藏层为13层的CNN异常箱表关系识别模型.

6.3 分组强化学习优化模型实验结果与分析

随机提取的13908例训练样本集分为7组, 根据实际低压台区箱表关系的拓扑布局, 7组训练样本集的数量分别为2132例、1875例、1928例、2075例、2048例、1957例和1893例, 基于分组数据开展并行训练, 提升强化学习效率的同时, 避免CNN异常箱表关系识别模型陷入局部最优, 以分组1训练样本集的训练过程为例进行泛化性能提升说明.

如图4所示, 本文所提强化学习优化模型方法, 通过迭代函数逼近的方式提高异常箱表关系识别模型在实际推广应用对环境的适应度, 设定参数 θ 内所有参数值的学习均方差阈值调整范围为 $\pm 0.45%$, 采用强化学习微调参数的方式实现模型泛化性应用, 并以分组方式提高强化学习过程的稳定性、定向引导和目标低延迟. 如图4所示, 分组1的实验数据说明, 该方法在大约440次强化学习迭代实验后, 模型参数收敛稳定, 说明本文采用分组强化学习直接逼近最优调参的泛化策略的可行性.

通过7组样本集的分组强化学习优化后, 以样本集交互学习与模型竞争的方式, 筛选获取最优泛化CNN异常箱表关系识别模型, 并基于泛化应用学习过程共计现场验证样本数量696例可疑样本, 自学习泛化识别

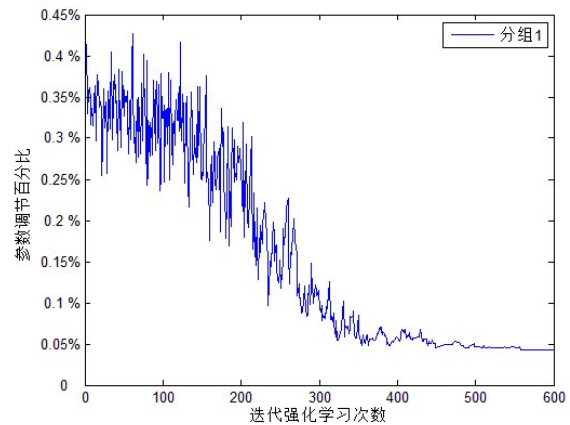


图4 迭代强化学习训练模型参数寻优过程图示

出新异常样本43例.

基于剩余2136例的训练样本集对最优泛化CNN异常箱表关系识别模型的准确率和召回率进行验证, 人工进行现场逐一排查, 识别结果如表3所示.

表3 异常箱表关系识别准确率验证

识别模型	强化学习前			强化学习后		
	异常	正常	识别总数	异常	正常	识别总数
实际异常	29	7	36	35	1	36
实际正常	58	2042	2100	34	2066	2100
实际总数	87	2049	2136	69	2067	2136

表3为迭代强化学习训练前后的异常样本识别数据对比.

迭代强化学习前:

$$\text{准确率} = \frac{29 + 2042}{2136} = 96.96\%$$

$$\text{召回率} = \frac{29}{29 + 7} = 80.56\%$$

迭代强化学习后:

$$\text{准确率} = \frac{35 + 2066}{2136} = 98.36\%$$

$$\text{召回率} = \frac{35}{35 + 1} = 97.22\%$$

迭代强化学习前后的识别准确率和召回率都得到提升, 验证了本文所述迭代强化学习方法的可用性.

取其训练过程中历次根据现场排查迭代进行模型强化学习后的模型, 其识别结果与识别误差如图5所示.

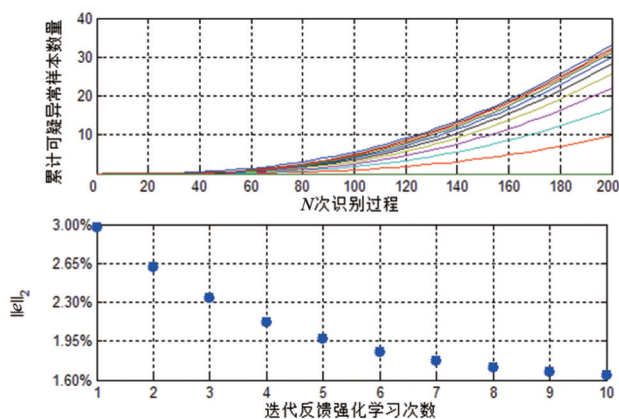


图5 迭代强化学习后模型识别结果与识别误差

图5上图为 $N=200$ 的一次批量识别+现场验证过程,图中曲线从上到下为顺位第一次到第十次迭代强化学习过程,纵轴为历次迭代识别出新可疑异常样本数量;虽然每次识别过程均会识别出可疑疑似异常样本,但样本存在重复性,且随着迭代强化学习过程的深入,识别新可疑异常样本的数量逐渐降低.图5下图为对应上图的10次迭代强化学习过程的识别误差,由3%降至1.64%,识别误差呈下降趋势;验证了本文所提方法的泛化性,显著降低现场人工核查工作量.

7 总结

本文借鉴“迁移学习”理念^[19]提高异常识别方法的有效性和泛化性.首先,通过电压原理识别出大范围内可用的部分异常箱表关系样本集,避免异常样本集的缺失问题;随后,采用Adaboost辅助构建简化CNN异常箱表关系识别模型集,以互补模式获取三分类均衡权重比^[20],并提取最末次简化识别模型作为复用历史学习框架;此外,在模型推广应用过程中,采用强化学习实现识别模型的分组并行泛化学习,一方面仅对识别出的异常样本进行人工核查,降低人工现场核查工作量,另一方面通过强化对新识别异常样本数据分布规律进行自学习,持续提高模型对不同应用环境的泛化性.实验证明本文所述方法的有效性和优越性,在现场检修和维护的过程中,大幅降低现场工作人员的工作量,提高快速排除故障能力,提升配网的服务效率和质量.

参考文献

[1] Su Y, Liang X R, Wang H, et al. A maintenance and troubleshooting method based on integrated information and system principles[J]. IEEE Access, 2019, 7: 70513 – 70524.
 [2] 吴志鹏, 杨革, 姜新宇, 等. 在线监测信息系统结构层次设计与实践[J]. 电子学报, 2018, 46(11): 2803 – 2808.

Wu Z P, Yang P, Jiang X Y, et al. Hierarchical structure design and application for online monitoring information system[J]. Acta Electronica Sinica, 2018, 46(11): 2803 – 2808. (in Chinese)

- [3] 高泽璞, 赵云, 余伊兰, 等. 基于知识图谱的低压配电网拓扑结构辨识方法[J]. 电力系统保护与控制, 2020, 48(2): 34 – 43.
 Gao Z P, Zhao Y, Yu Y L, et al. Low-voltage distribution network topology identification method based on knowledge graph[J]. Power System Protection and Control, 2020, 48(2): 34 – 43. (in Chinese)
 [4] 肖勇, 赵云, 涂治东, 等. 基于改进的皮尔逊相关系数的低压配电网拓扑结构校验方法[J]. 电力系统保护与控制, 2019, 47(11): 37 – 43.
 Xiao Y, Zhao Y, Tu Z D, et al. Topology checking method for low voltage distribution network based on improved pearson correlation coefficient[J]. Power System Protection and Control, 2019, 47(11): 37 – 43. (in Chinese)
 [5] Hang Q, Yang J H, Xing L N. Diagnosis of rolling bearing based on classification for high dimensional unbalanced data [J]. IEEE Access, 2019, 7(1): 79159 – 79172.
 [6] Zhang P B, Yang Z X. A novel AdaBoost framework with robust threshold and structural optimization[J]. IEEE Transactions on Cybernetics, 2018, 48(1): 64 – 76.
 [7] Kim C, Shin D, Kim B, et al. Mosaic-CNN: A combined two-step zero prediction approach to trade off accuracy and computation energy in convolutional neural networks[J]. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 2018, 8(4): 770 – 781.
 [8] Dilokthanakul N, Kaplanis C, Pawlowski N, et al. Feature control as intrinsic motivation for hierarchical reinforcement learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(11): 3409 – 3418.
 [9] Barriga N A, Stanescu M, Besoain F, et al. Improving RTS game AI by supervised policy learning, tactical search, and deep reinforcement learning[J]. IEEE Computational Intelligence Magazine, 2019, 14(3): 8 – 18.
 [10] Bruin T D, Kober J, Tuyls K, et al. Integrating state representation learning into deep reinforcement learning [J]. IEEE Robotics and Automation Letters, 2018, 3(3): 1394 – 1401.
 [11] Sevakula R K, Verma N K. Assessing generalization ability of majority vote point classifiers[J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 28(12): 2985 – 2997.
 [12] 于妍妍, 沈艳霞, 陈杰, 等. 考虑概率区间的微电网短期负

- 荷多目标预测方法[J]. 电子学报, 2017, 45(4): 930 – 936.
- Yu X Y, Shen Y X, Chen J, et al. A multi-objective prediction method for short-term microgrid load considering interval probability[J]. Acta Electronica Sinica, 2017, 45(4): 930 – 936. (in Chinese)
- [13] Xiang W M, Tran H D, Johnson T T, et al. Output reachable set estimation and verification for multilayer neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(11): 5777 – 5783.
- [14] Lu Y J, Wu H F, Tu B, et al. Ultra-wide output voltage range power supply based on modular switched-converter principle[J]. IEEE Transactions on Power Electronics, 2020, 35(1): 94 – 106.
- [15] Kim C, Shin D, Kim B, et al. Mosaic-CNN: A combined two-step zero prediction approach to trade off accuracy and computation energy in convolutional neural networks [J]. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 2018, 8(4): 770 – 781.
- [16] Passalis N, Tefas A. Training lightweight deep convolutional neural networks using gag-of-features pooling[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(6): 1705 – 1715.
- [17] Chen G, Colin I J D, Zhang M J. Accuracy-based learning classifier systems for multistep reinforcement learning: A fuzzy logic approach to handling continuous inputs and learning continuous actions[J]. IEEE Transactions on Evolutionary Computation, 2016, 20(6): 953 – 971.
- [18] Wang Y J, Sun F M, Li B. Multiscale neighborhood normalization-based multiple dynamic PCA monitoring method for batch processes with frequent operations[J]. IEEE Transactions on Automation Science and Engineering, 2018, 15(3): 1053 – 1064.
- [19] Tang Z Q, Zhou X J, Zhang F M, et al. Migration modeling and learning algorithms for containers in fog computing[J]. IEEE Transactions on Services Computing, 2019, 12(5): 712 – 725.
- [20] 胡峰, 王蕾, 周耀. 基于三支决策的不平衡数据过采样方法[J]. 电子学报, 2018, 46(1): 135 – 144.
- Hu F, Wang L, Zhou Y. An oversampling method for imbalance data based on three-way decision model[J]. Acta Electronica Sinica, 2018, 46(1): 135 – 144. (in Chinese)

作者简介



管永明 男, 1984年1月出生, 山东济南人. 2009年于山东大学计算机软件与理论专业获得硕士学位, 其后在山大经纬软件股份有限公司从事智能电网应用、数据挖掘工作.
E-mail: guanyongming@dareway.com.cn



王 刚 男, 1988年2月出生, 重庆人. 2010年于重庆大学电气工程与自动化专业获得学士学位, 其后在国网重庆市电力公司从事电力营销工作.
E-mail: wangg@eq.sgcc.com.cn

骆凯波 男, 1982年6月出生, 河北保定人. 2004年于重庆大学计算机科学与技术专业获得学士学位, 高级工程师, 其后在国网重庆市电力公司从事用电信息采集建设管理工作.
E-mail: luokaibo777@126.com

吕 梁 男, 1985年7月出生, 山东济南人. 2010年于中国海洋大学信号与信息处理专业获得硕士学位, 其后在山大经纬软件股份有限公司从事电力系统设计与分析工作.
E-mail: ll@dareway.com.cn

吕晓雯 女, 1988年11月出生, 山东青岛人. 2015年于山东大学控制科学与工程专业获得硕士学位, 其后在山大经纬软件股份有限公司从事电力系统自动化工作.
E-mail: lxw@dareway.com.cn

史玉良(通信作者) 男, 1978年10月出生, 山东威海人. 教授、博士生导师、山东省社会保障大数据工程实验室主任、兼任中国计算机学会协同计算专委会委员、服务计算专委会委员. 2003年、2006年分别于山东大学、复旦大学获得硕士学位、博士学位, 2006年后在山东大学从事大数据、服务计算、人工智能等方面的研究工作.
E-mail: shiyuliang@sdu.edu.cn