

基于 RNN 的低压变压器区域 日线损率基准测试

何伟民, 孙一迪, 姜捷, 金良勇, 毛和云

(国网浙江江山市供电有限公司, 浙江 江山 324100)

摘要: 输变电线路损耗是输变电阶段的固有现象, 其是评价低压变压器区域日线损率的重要指标和基准; 当对大数据样本上进行训练时, 区域的数量通常非常大, 并且线损率数据集包含大量的异常值; 为了准确的计算低压变压器区域日线损率, 提出了一种具有去噪自动编码器 (DAE) 多径网络模型的鲁棒神经网络 (RNN) 方法, 利用丢包层、L2 正则论和 Huber 损失函数的优点获得多种不同的输出, 并利用比较结果计算出基准值和合理区间, 实现了精确评估采样数据集的质量并消除线损率的异常值, 从而提高数据检测的稳定性; 通过与传统的机器学习模型相比, 所提出的 RNN 具有较好的鲁棒性和准确性; 根据所提出的 RNN 的最终结果, 在整个数据点中约有 13% 的异常值, 一个月内线损率无缺失值和异常值的区域仅占 20% 左右, 说明了计电设备可靠性较低。

关键词: 基准计算; 日线损率; 低压电压器区域; 鲁棒神经网络; 去噪自动编码器

Reference Test of Daily Line Loss Rate in Low Voltage Transformer Area based on RNN

He Weimin, Sun Yidi, Jiang Jie, Jin Liangyong, Mao Heyun

(State Grid Jiangshan Power Supply Company, Jiangshan 324100, China)

Abstract: Transmission line loss is an inherent phenomenon in the transmission and transformation stage, and it is an important index and benchmark for evaluating the daily line loss rate in the low-voltage transformer area. When training on large data samples, the number of regions is usually very large, and the line loss rate data set contains a large number of outliers. In order to accurately calculate the daily line loss rate of low-voltage transformers, a robust neural network (RNN) method with a denoising autoencoder (DAE) multipath network model is proposed. It uses the packet loss layer, L2 regularity, and Huber loss. The advantage of the function is to obtain a variety of different outputs, and use the comparison results to calculate the reference value and a reasonable interval, to achieve an accurate evaluation of the quality of the sampled data set and eliminate abnormal values of the line loss rate, thereby improving the stability of data detection. Compared with traditional machine learning models, the proposed RNN has better robustness and accuracy. According to the final result of the proposed RNN, there are about 13% outliers in the entire data point, and the area with no missing values and outliers in the line loss rate within one month only accounts for about 20%, indicating that the reliability of the metering equipment is low.

Keywords: benchmark calculation; daily line loss rate; low voltage regulator area; robust neural network; denoising automatic encoder

0 引言

线损率是电网的一项重要评价指标, 它能反映电网在经济和技术方面的运行和管理水平^[1]。线损一般分为技术性线损和非技术性线损。日线损率值是否在合理的范围内 (即日线损率的合格率) 已成为电网运营商迫切关注的问题, 这就需要从大量采集的样本中直接区分正常线损率值和异常值^[2]。由于日线损率能够作为操作人员更好了解低压变压器区域工作状态的依据, 因此, 日线损率的基准值的准确测量对于提高线损管理水平尤为重要。

在数据挖掘分析领域, 通常有 4 种方法来计算基准值

和检测异常值, 即经验法^[3]、统计法^[4]、无监督法^[5]和监督法^[6]。文献 [3] 利用经验法指出在日线损率的基准测试中, 经验区间通常设置为 -1%~5%。文献 [7] 揭示了由于不可避免的采集误差, 不小于 -1% 的值均可以接受。文献 [4] 指出统计法中的区间界限能够适应不同的检测样本, 但该方法很难利用线损率的影响因素。文献 [5] 利用无监督法中的聚类算法将异常值可以通过数据点与聚类中心的距离来进行识别。监督法利用机器学习模型求解分类问题^[6]和回归问题^[8], 文献 [9] 和文献 [10] 分别设计用于异常值检测和基准计算任务, 分类模型通过对标记样本的学习来区分正常和异常数据。然而, 线损率样本通常没有标记, 因此它无法识别收集到的线损率值是否正常。

本文提出了一种基于鲁棒神经网络 (RNN) 的回归计算方法, 并由去噪自动编码器 (DAE)、多径网络结构、丢

收稿日期: 2020-01-21; 修回日期: 2020-03-18。

作者简介: 何伟民 (1970-), 男, 浙江江山人, 工学学士, 工程师, 主要从事线损管理, 逆变器变换控制技术和电能质量等方向的研究。

包层、Huber 损失函数、L2 正则化和 10 个输出组成。基准是根据 10 个输出的平均值计算得出。经过误差分析, 该方法可以得到合理的区间来检测原始线损率样本的异常值。

1 线损理论计算

本文提出了基于等效电阻法计算技术的理论线损公式, 该方法假定线路的前端存在等效电阻, 其中三相三线和三相四线系统的能量损失可表示为^[11]:

$$\Delta A_b = NK^2 I_{av}^2 R_{eq} T \times 10^{-3} \quad (1)$$

其中: ΔA_b 为三相平衡负载时的理论线损, N 为结构系数, 在三相三线制下等于 3, 在三相四线制下等于 3.5。 K 、 I_{av} 、 R_{eq} 和 T 分别为负荷曲线的形状系数、线首处平均电流 (A)、导体等效电阻 (W) 和工作时间 (h)。此外, R_{eq} 的计算公式为:

$$R_{eq} = \frac{\sum_i N_i A_i^2 R_i}{N (\sum_j A_j)^2} \quad (2)$$

其中: N_i 、 A_i 和 R_i 分别为第 i 个线段的结构系数、计量功率和电阻。 A_j 为从第 j 个电表采集的电量。对于三相平衡负载系统, 理论线损可修正为:

$$\Delta A_{ub} = \Delta A_b \times K_{ub} \quad (3)$$

其中: K_{ub} 为修正系数, 可定义为:

$$K_{ub} = 1 + k\delta_l^2 \quad (4)$$

其中: 当出现单相重负荷和两相轻负荷时, $k = 2$ 。当出现两相重负荷时, $k = 8$ 。 δ_l 为三相负载的不平衡度, 可计算为:

$$\delta_l = \frac{I_{max} - I_{av}}{I_{av}} \quad (4)$$

其中: I_{max} 为来自具有最大负载相的电流。因此, 以上定义的理论线损是不可避免的能量损耗, 即技术线损耗。然而, 电网运营商也担心因窃电引起的非技术性线路损耗。由于非技术性线损情况会导致按日计量的线路损耗率出现异常值, 因此有必要计算合理的时间间隔以进行区分识别。

2 数据集

在实际应用中, 通常国家电网公司每月检查一次低压变压器区域日线损率的合格性。在这种情况下, 本文研究中使用了 2019 年 7 月份的线损率数据集, 该数据集以每日间隔进行采集, 以此检查当月线损率的合格率。合格率指标在 7 月份尤为重要, 这是由于 7 月份通常是夏季的用电高峰期。该数据集选自浙江省江山市共计 19 884 个低压变压器区域, 共有 616 404 个样本, 满足了大数据分析的需要。基于该数据集, 选择约 80% 的样本 (15 907 个低压变压器区域) 作为训练样本, 其余的样本 (3 977 个区域) 作为测试样本。

2.1 数据质量分析

本文的研究对象为日线损率, 一些低压变压器区域日线损率示例, 如图 1 所示。

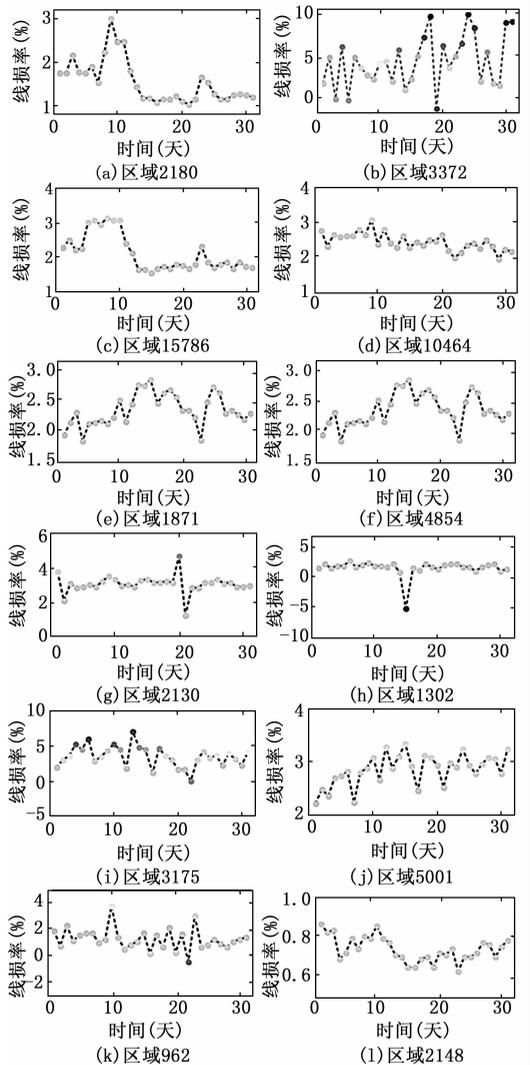


图 1 不同低压变压器区域日线损率示例

本文选取 25% (q_1), 中位数 (q_2), 75% (q_3), 最大值 (max)、最小值 (min)、均值、标准差 (std), 下限值 (l_a) 和上限值 (u_a) 作为研究指标。基于总体线损率数据集的数据质量分析, 如表 1 所示。

表 1 基于总体线损率数据集的数据质量分析

指标	原始数据集	无异常值的原始数据集	插值后原始数据集
均值	-5.16	1.78	-10.96
标准差	2.47×103	1.16	3.19×103
最小值	-1.69×106	-1.57	-1.69×106
最大值	100	5.22	100
l_a	-1.57	-1.10	-1.50
u_a	5.22	4.57	5.26
q_1	1.02	1.02	1.03
q_2	1.74	1.67	1.76
q_3	2.70	2.44	2.72

原始数据集和插值后数据集的方框图, 如图 2 所示。下限值 (l_a) 和上限值 (u_a) 是基于 25% 的 (q_1)

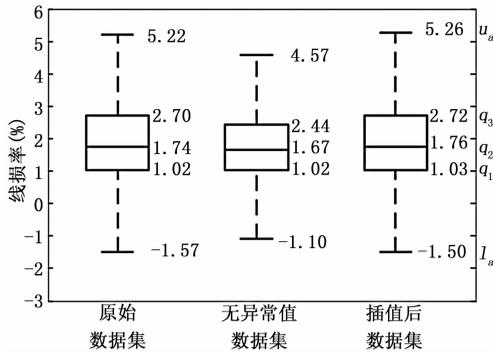


图 2 原始数据集和插值后数据集的方框图

和 75% 的 (q_3) 计算得出, 其中超出界限范围的值可以视为异常值:

$$\begin{cases} l_a = q_1 - 1.5 \times (q_3 - q_1) \\ u_a = q_3 + 1.5 \times (q_3 - q_1) \end{cases} \quad (6)$$

根据曲线和数据质量分析, 日线损率的数据特征总结如下:

(1) 线损率数据日变化规律性很小, 但波动性很大。从图 1 可以看出, 不同低压变压器区域的线损率曲线随着时间的推移变化很大, 历史线损率很难用来估计进一步的数值。因此, 选取线损率的影响因素是本文研究的重点。

(2) 数据集中异常值的偏差有时偏离正常值较大, 这表明计电装置和通信设备的可靠性较低。根据表 1 和图 2, 对比图中原始数据集的上下限值分别为 -1.57% 和 5.22% , 与项目标准 (-1% 和 5%) 相当接近。然而, 所收集的线损率的最大值和最小值分别为 100% 和 $-1.69 \times 10^6\%$, 与界限有很大的不同。在这种情况下, 基准线损率在实际应用中仍然重要。

(3) 数据集的质量较差, 无法直接使用。数据质量分析的组成结果, 如图 3 所示。其中, 正常值 84.61% , 异常值 8.67% , 缺失值 6.72% , 因此, 存在大量的异常值和缺失值, 并且分别占整个数据集的 8.67% 和 6.72% 。本研究利用样条插值法来填补缺失值。从表 1 和图 2 可以看出, 插值后的数据集与原始数据集的分布相似。相反, 虽然可以根据 l_a 和 u_a 可以直接消除异常值, 但分布会发生变化, 并且很难计算出准确的合理区间。

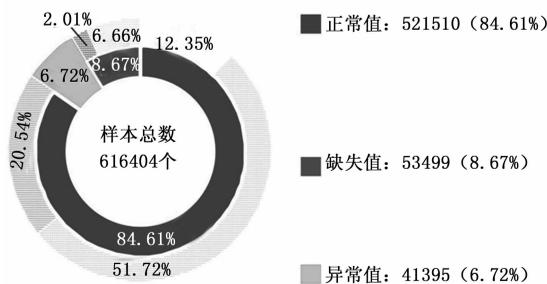


图 3 数据质量分析的组成结果

2.2 线损率的影响因素

考虑到可能的影响因素和记录的信息, 本文共选择 12

个因素作为回归模型的输入, 如表 2 所示。其中, 第三因素和第四个因素是 1 bit 字符, 其他都为数值。

表 2 线损率的影响因素

因素	参数
数据	—
变压器容量	—
变压器类型	公共变压器(=0);专用变压器(=1)
区域所属网络类型	城市电网(=0);农村电网(=1)
月线损率	前 3 个月的负荷
日负荷率	日负荷率=日功率
日最大负荷率	—
日平均功率因数	—
用户数量	—
每个用户的平均变压器容量	平均容量=变压器容量/用户数量
居住容积率	容积率=住宅容量/变压器容量
供电时间	—

3 基准值和合理区间的计算

根据数据质量分析, 原始数据集中含有大量的异常值, 这些异常值超出合理范围较远, 很难得到准确的结果。因此, 本文的任务是利用具有鲁棒的学习模型来获得异常值稳定的回归结果, 如图 4 所示。

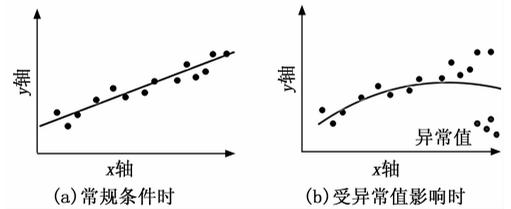


图 4 传统的学习模型容易受到异常值的影响

通常, 学习模型需要手动设置阈值, 并根据这些阈值从数据集中删除异常值, 数据集的其余部分可用于训练机器学习模型, 然而确定准确的阈值成为研究的难点。此外, 学习模型合理区间的计算范围可能接近人工阈值, 从而影响原始数据集的分布, 并使得训练概率学习模型实效。在这种情况下, 本文提出了基于 RNN 的计算方法, 如图 5 所示。

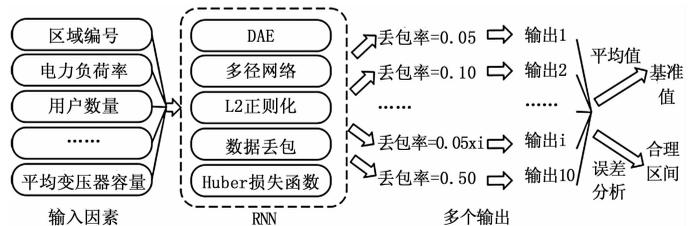


图 5 基于 RNN 的基准值和合理区间计算方法的流程图

具体计算步骤如下。

步骤 1: 建立 RNN。为了充分扩展其鲁棒性, 本文利用 DAE、多径结构、L2 正则化、丢包层和 Huber 损失函数等方式进行分析。由于 RNN 具有 10 个输出节点, 其中每

个节点以不同的丢包率 (从 0.05 到 0.50) 连接到一个层。

步骤 2: 根据 10 种不同的输出计算平均值, 即线损率的最终基准值:

$$\tilde{y}_i = \frac{1}{10} \sum_{n=1}^{10} y_i^n \quad (7)$$

其中: \tilde{y}_i 为第 i 个基准值, y_i^n 为第 i 个线损率的第 n 个输出。

步骤 3: 根据误差分析获得合理的区间。本文不仅计算了基准值与实际线损率的绝对误差, 还计算了不同输出的方差。根据区间结果, 不在区间范围内的数据点认为是异常值, 具体的计算公式如下:

$$e_1 = \sqrt{\frac{1}{n_s} \sum_{i=1}^n (\tilde{y}_i - y_i^*)^2} \quad (8)$$

$$e_2 = \sqrt{\frac{1}{10-1} \sum_{i=1}^{10} (\tilde{y}_i - y_i^n)^2} \quad (9)$$

$$\begin{cases} l_i = \tilde{y}_i - e_1 - e_2 \\ u_i = \tilde{y}_i + e_1 + e_2 \end{cases} \quad (10)$$

其中: e_1 和 e_2 是误差分析的结果, e_1 为常数, e_2 根据 i 而变化, n_s 为训练样本数量, y_i^* 为第 i 个实际线损率, l_i 和 u_i 分别为合理间隔的下限和上限。此外, 由于实际线损率值中存在异常值, 这可能会影响 e_1 的结果。因此, 本文利用双尾检验来消除小于 0.7% 和大于 99.3% 的可能异常 y_i^* 值, 如图 6 所示。

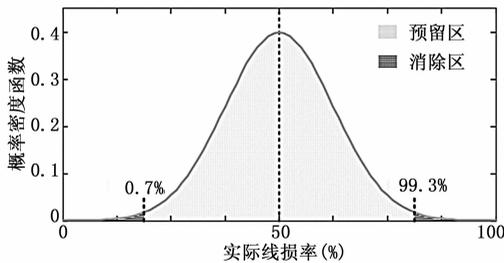


图 6 消除可能异常线损率值的双尾检验

4 鲁棒神经网络 (RNN)

本文使用 RNN 算法^[12]进行鲁棒学习, 其结构如图 7 所示。其由三条主要路径组成, 这些路径通过串联组合在一起, 并且每条主路径上都有一个 DAE。为了进一步提高系统的鲁棒性, 将串联后的输出节点放在同一层中, 这些层表示从原始输入中提取的高阶特征, 并在层中采用 L2 正则化来限制这些节点的输出值。然后, 在高阶特征层之后叠加 10 个具有不同丢包率的丢包层, 并得到 10 个输出。本文对 10 个输出进行分析, 并计算基准值和合理区间。

4.1 去噪自动编码器 (DAE)

本文所提出的 DAE 的结构, 如图 8 所示。它是自动编码器的鲁棒变体, 在编码器之前具有一个噪声层^[13], 例如正常 (高斯) 噪声层:

$$x_{i,n} = x_i + N(0, \sigma^2) \quad (11)$$

其中: x_i 和 $x_{i,n}$ 分别为噪声层的第 i 个输入和第 i 个输

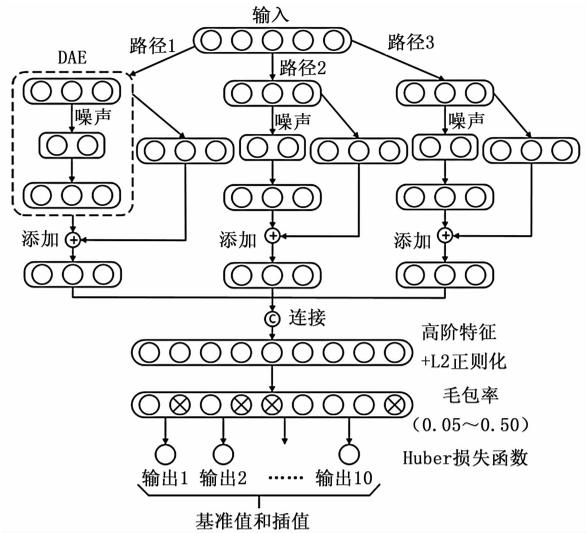


图 7 鲁棒神经网络 (RNN) 的结构

出。 $N(0, \sigma^2)$ 为正态分布, 其平均值为 0, 方差值为 σ^2 。在本文研究中, 当输入标准化为 $[0, 1]$ 时, σ 设置为 0.05。

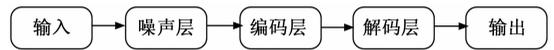


图 8 去噪自动编码器 (DAE) 的结构

此外, DAE 中的编码器层和解码器层均由传统的全连接 (FC) 层组成, 其方程可以表示为:

$$y_i^l = w_{ij} x_j^{l-1} + b_j \quad (12)$$

其中: y_i^l 和 x_j^{l-1} 分别为第 l 层的第 i 个输出和第 j 个输入, w_{ij} 和 b_j 分别为连接第 j 个输入和第 i 个输出的 FC 层的权重和偏差。对于编码器层, 输出的数量小于输入的数量, 即 $i < j$; 并且解码器层中的输出的数量等于原始输入的数量。因此, 在计算 DAE 之后, 输入的维数和大小保持不变, 而特征的鲁棒性增加, 这是由于 DAE 处理后能够抵抗一定程度的噪声干扰。

4.2 通过加法和串联组合的多路径

在 RNN 中共有三条主要路径, 它们具有相似的层, 其输出在串联操作下可以组合起来:

$$\begin{cases} y_c = C[f_k(x, \{w_k^n, b_k^n\})] \\ y_{k,mp} = f_k(x, \{w_k^n, b_k^n\}) \end{cases}, \quad k = 1, 2, 3 \quad (13)$$

其中: $C[\cdot]$ 为串联运算, 它将不同层的输出节点合并为一个完整层, y_c 和 $y_{k,mp}$ 分别为串联的输出矢量和第 k 条主路径的输出矢量, $f_k(x, \{w_k^n, b_k^n\})$ 为第 k 条主路径的计算结果, x 为 RNN 的输入矢量, w_k^n 和 b_k^n 分别为第 k 条主路径的权重矩阵和偏差向量, n_w 和 n_b 分别为权重和偏差的数量。

此外, 主路径由两个子路径形成, 即 DAE 子路径和 FC 层子路径。两个子路径的输出作为主路径的输出相加, 如下所示:

$$y_{k,mp} = g_k(x, \{w_k^{n-1}, b_k^{n-1}\}) + (w_k^p x + b_k^p) \quad (14)$$

其中: $g_k(\cdot)$ 为第 k 条主路径上 DAE 的计算, w_k^p 和 b_k^p 分别为第 k 条主路径上 FC 层子路径的权重矩阵和偏移

向量。

4.3 丢包层

丢包层作为一种特殊的层，其可以有效地防止过度拟合^[14]。丢包过程可概括为两个阶段，即训练阶段和应用阶段。对于公式 (12) 所示的传统 FC 层，存在 j 个输入节点。在训练阶段中，输入节点将以概率 p ($0 < p < 1$) 被丢弃，其中被丢弃的节点不能连接到输出节点^[15]，如图 9 所示。在训练之后，所有的输入节点都在应用阶段中完成，且权重值乘以概率 p ，可以得到：

$$y_j^i = p w_{ij} x_j^{i-1} + b_j \quad (15)$$

其中： p 为丢包率，其设定在 0.05 到 0.50 之间（步长为 0.05），以便在研究中获得 10 种不同的输出。

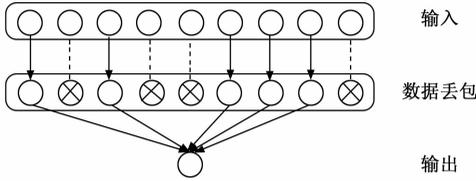


图 9 在训练阶段的丢包原则

4.4 Huber 损失函数

神经网络的训练过程是设置损失函数，利用 BP 梯度下降算法逐层更新参数。均方误差 (MSE)：

$$MSE = \frac{1}{n_s} \sum_{i=1}^n (y_i - y_i^*)^2 \quad (16)$$

其中： n_s 为训练样本的数量， y_i 和 y_i^* 分别为第 i 个预测的输出和第 i 个实际的输出。平均绝对误差 (MAE)，其定义如下：

$$MAE = \frac{1}{n_s} \sum_{i=1}^n |y_i - y_i^*| \quad (17)$$

其中： MSE 和 MAE 也可作为 L1 损失和 L2 损失，这是由于 MSE 和 MAE 分别使用了线性项和二次项。

MSE 与 MAE 相比， MSE 具有更光滑的导数函数，这有利于梯度下降算法的计算，而 MAE 的微小差异可能导致参数更新的巨大变化。相反，在对抗异常值时， MAE 表现出比 MSE 更好的性能^[16]。在这种情况下，Huber 损失函数的原理，如图 10 所示。

本文采用 Huber 损失函数^[17]，该函数结合了 MSE 和 MAE 的优点：

$$Huber = \begin{cases} \frac{1}{n_s} \sum_{i=1}^n \left[\frac{1}{2} (y_i - y_i^*)^2 \right], & |y_i - y_i^*| \leq \delta \\ \frac{1}{n_s} \sum_{i=1}^n \left[\delta |y_i - y_i^*| - \frac{1}{2} \delta^2 \right], & |y_i - y_i^*| > \delta \end{cases} \quad (18)$$

其中： δ 为需要手动设置的超参数，在本文研究中设定为 10%。

4.5 L2 正则化

在本文研究中，L2 正则化旨在为具有较大激活输出的

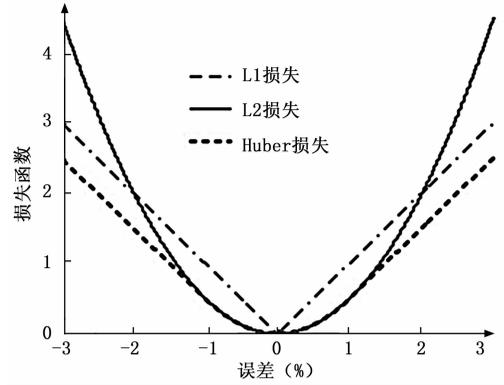


图 10 Huber 损失函数的原理

节点设定惩罚项，以此防止过度拟合，并提高神经网络的鲁棒性。正则化在训练阶段起到作用，在训练损失函数中加入两个范数的惩罚项，其表达式为：

$$L = \lambda \|y_c\|_2 + Huber \quad (19)$$

其中： L 为模型训练的最终损失函数， λ 为惩罚项的超参数，在本文研究中设置为 0.001。

5 实验分析

本文所提出的 RNN 结构和超参数如表 3 所示。

表 3 RNN 的结构和超参数

路径	层	超参数
DAE 子路径(1~3)	FC 层 0	节点数:64 激活方式:ReLU(整流线性单元)
DAE 子路径(1~3)	噪声层	标准差:0.05
DAE 子路径(1~3)	FC 层 1(编码器)	节点数:8 激活方式:ReLU
DAE 子路径(1~3)	FC 层 2(解码器)	节点数:64
FC 子路径(1~3)	FC 层 3	节点数:64
主路径(1~3)	添加层	输入:FC 层 2 和 FC 层 3
—	连接层	输入:添加层 节点数:192 激活方式:ReLU 正则化:L2=1×10 ⁻³
—	丢包层(1~10)	丢包率:0.05~0.50
—	FC 层 4 (输出 1~10)	节点数:1 激活方式:ReLU

考虑到训练样本数量较多的特点，本文建立了 k 近邻 (KNN)、决策树回归 (DTR) 和单隐层人工神经网络 (ANN) 进行比较，在大数据集上具有较高的训练效率。在 NVIDIA GTX 1080 GPU 的计算机上，采用 Python 3.5 和 Tensorflow 1.4 对深度 RNN 模型进行训练。RNN 的所有超参数和训练配置以及超参数（即 σ 、 δ 和 λ ）通过基于整体训练数据集的三重交叉验证的网络搜索进行选择。参数的搜索空间和最终结果，如表 4 所示。

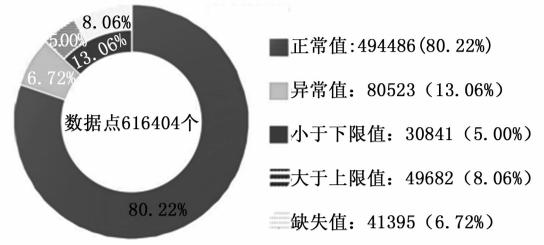
5.1 计算结果

在本文中，从测试样本中随机选择 6 个低压变压器区域作为展示示例，如图 11 所示。

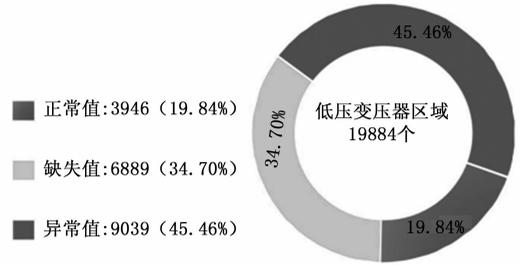
表 4 RNN 中选定超参数的搜索空间

超参数	搜索空间	结果
σ	[0.00;0.05;0.50]	0.05
δ	[0.00;0.10;0.50]	0.10
λ	$1 \times 10^{[-4;1;0]}$	1×10^{-3}
子路径中的节点数	{16,32,64,128}	64
子路径数	[2:1:5]	3
训练优化器	{Adadelata, Adam}	Adam
遍历次数	[50:50:200]	100
单次训练样本数	{64,128,256,512,1024,2048}	1024
学习率	$1 \times 10^{[-5;1;0]}$	1×10^{-4}

基于本文所提出的 RNN, 可以分析线损率的通过百分比结果, 如图 12 所示。



(a) 数据点分析



(b) 低压变压器区域分析

图 12 基于鲁棒神经网络的线损率合格率分析

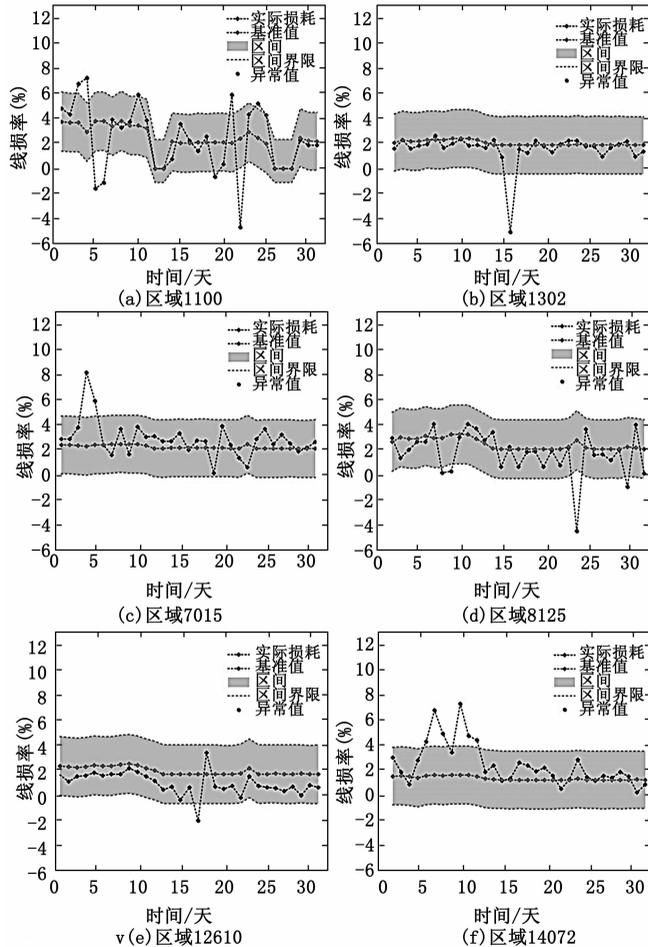


图 11 6 个实验区域的基准值和合理区间的结果

低压变压器区域的编号分别为 1 100、1 302、7 015、8 125、12 610 和 14 072。结果表明, 合理区间的界限可以根据多个输入因素进行自适应调整, 例如, 在区域 1 100 和区域 8 125 中。距离基准值较远的异常值可以有效地剔除, 虽然这些异常值可以在 $-1\% \sim 5\%$ 之间, 但是合理区间的结果要优于 $1\% \sim -5\%$ 之间的固定区间。此外, 基准值与实际线损率相比波动较小, 表明日线损率的估计具有较高的可靠性。基准值能够根据相关因素的变化自适应地反映低压变压器区域的日常运行状况, 而不是根据原始数据集计算出的平均值或中值。

对于线损率的数据点分析, 由于所提出的方法能够准确地识别出与基准值相差较大的异常值, 因此异常值的数目比图 3 中的异常值要多。此外, 虽然所有数据点的缺失值和异常值的百分比都不算大, 分别为 6.72% 和 13.06% , 但一个月内没有缺失值和异常值的区域仅占整个数据集的 19.84% , 这说明当前计电设备的可靠性较低。

5.2 性能分析

为了评估本文所提出方法的鲁棒性和准确性, 首先建立 KNN、DTR 和 ANN 的超参数, 如表 5 所示。

表 5 KNN、DTR 和 ANN 的超参数

超参数
邻居数量:10;加权方法:基于距离
拆分方法:选择最佳拆分
隐藏层中的节点数:64;激活方式:逻辑结构

(1) 鲁棒性分析: 为了评估所提方法的鲁棒性, 本文分析了基于不同测试模型的计算基准值的分布, 如图 13 所示。

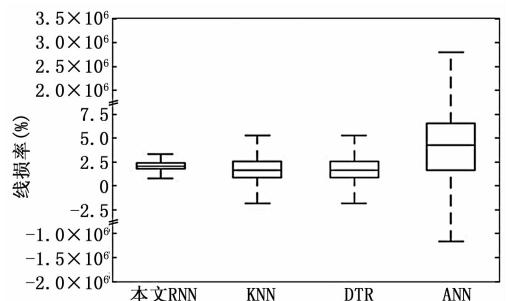


图 13 基于不同测试模型的计算基准值分布

分布指标的详细数值, 如表 6 所示。

表 6 不同测试模型的鲁棒性分析结果

指标	本文 RNN	KNN	DTR	ANN
均值	2.03	1.03	-4.81	-5.83×10^6
标准差	0.80	185.69	194.78	1.87×10^7
最小值	0.00	-8.13×10^4	-8.13×10^4	-8.13×10^7
最大值	33.61	100.00	100.00	4.49×10^6
l_a	0.81	-1.80	-1.80	-1.17×10^6
u_a	3.36	5.25	5.24	2.80×10^6
q1	1.76	0.84	0.84	3.23×10^5
q2	2.09	1.65	1.65	8.47×10^5
q3	2.40	2.60	2.60	1.32×10^6

结果表明, 测试的 ANN 模型性能最差, 完全无法计算出有效的基准值。ANN 的最大值和最小值分别为 $4.49 \times 10\%$ 和 $-8.26 \times 10\%$, 因此, 其很难作为基准值。根据分布, KNN 和 DTR 得到了相似的结果。它们都利用接近未知测试区域的大量训练样本来确定新的基准值。因此, 在本文中, KNN 和 DTR 比 ANN 具有更好的鲁棒性, 并且在大多数低压变压器测试区域都具有可行性。然而, 这两个模型的最小基准为 $-8.13 \times 10^4\%$, 仍然不是合理的基准值, 而且 RNN 在 4 个测试模型中取得了最好的结果, 其中计算的基准值在合理的范围内。利用 RNN 计算得出的基准值标准差仅为 0.80% , 这表明该方法得到的结果稳定且可靠。

(2) 精度分析: 本文利用 MAE、MSE 和 Huber 损耗 3 个损耗指标来比较 4 个测试模型。在使用双尾检验进行损失计算之前, 测试样本中的异常值被消除, 如图 6 所示。不同测试模型的精度分析结果, 如表 7 所示。

表 7 不同测试模型的精度分析结果

指标	本文 RNN	KNN	DTR	ANN
MAE(%)	1.46	0.70	5.01	7.26×10^6
MSE(% ²)	13.70	574.58	3.90×10^5	3.54×10^{14}
Huber 损耗	4.44	5.90	48.19	7.26×10^7

结果表明, 由于 ANN 的 3 个损失指标远高于其他模型, 因此其性能最差。当直接对具有极端异常值的样本进行训练时, ANN 并不适用。虽然 KNN 和 DTR 具有相似的鲁棒性, 但它们的精度指标却有很大的不同。由于 KNN 计算出的异常值较少, 因此 KNN 得到的 MAE 指标最好, 而 KNN 的 MSE 值大于所提出的 RNN 的 MSE 值。综合比较这些指标, 本文提出的 RNN 具有最高的性能, 在 MAE 值较小的情况下获得了最佳的 MSE 和 Huber 损耗指标。

6 结束语

日线损率作为考核低压变压器区域性能的重要指标, 其对供电企业的利润有很大的影响。为了更好地管理线损水平, 为低压变压器区域的建设和运行提供指导, 本文研

究开发的日线损率基准值计算方法, 有助于发现异常线损率值, 也有助于运行人员对异常运行情况进行检查和确认。从实例分析和比较结果来看, 传统的 ANN 模型不能处理异常值, 无法计算出基准的结果。在案例分析中证明了 KNN、DTR 和所提出的 RNN 的适用性, 其中所提出的 RNN 优于其他两个模型。在所有的测试模型中, 该方法具有较高的精度和鲁棒性。此外, 根据所提出的 RNN 的最终结果, 在整个数据点中约有 13% 的异常值。一个月内线损率无缺失值和异常值的区域仅占 20% 左右, 说明了计电设备可靠性较低。因此, 目前电网中仍需要一套可靠的线损数据监测与管理系统。

参考文献:

- [1] 梁丽莉. 小水电并网引起线损偏高原因及降损措施分析 [J]. 通讯世界, 2019, 26 (10): 199-200.
- [2] 伍慧君, 吴丽贤. 公变台区线损率异常原因检测系统设计 [J]. 自动化与仪器仪表, 2018 (12): 123-125.
- [3] Schnell J, Nentwich C, Endres F, et al. Data mining in lithium-ion battery cell production [J]. Journal of Power Sources, 2019, 413: 360-366.
- [4] Pourghasemi H R, Yousefi S, Kornejady A, et al. Performance assessment of individual and ensemble data-mining techniques for gully erosion modeling [J]. Science of the Total Environment, 2017, 609: 764-775.
- [5] 张海林, 李琳, 夏传良. 改进 K-means 算法的馈线线损计算 [J]. 软件导刊, 2019, 18 (12): 22-25.
- [6] Yang D. Application of Data Mining Technology in the Subject Tactical Teaching of Badminton [J]. International Journal of Emerging Technologies in Learning (iJET), 2018, 13 (7): 30-42.
- [7] 韦杏秋, 陈俊, 龙东, 等. 基于时序差分算法的线损异常判别优化研究 [J]. 科技通报, 2017, 33 (3): 100-103.
- [8] 王海燕. 基于核偏最小二乘回归分析的线损率预测 [J]. 计算机仿真, 2012, 29 (11): 323-326.
- [9] 王守相, 周凯, 苏运. 基于随机森林算法的台区合理线损率估计方法 [J]. 电力自动化设备, 2017, 37 (11): 39-45.
- [10] Quansah D A, Adaramola M S. Assessment of early degradation and performance loss in five co-located solar photovoltaic module technologies installed in Ghana using performance ratio time-series regression [J]. Renewable Energy, 2019, 131 (3): 900-910.
- [11] Ajithkumar K, Amrutha P C, Vinitha U C, et al. Analysis of mortality and loss to follow-up during the rollout of the second-line antiretroviral therapy: An observational study from South India [J]. Journal of Skin and Sexually Transmitted Diseases, 2019, 1 (1): 19-23.
- [12] 孙晓丽, 郭艳, 李宁, 等. 基于改进 RNN 的多变量时间序列缺失数据填充算法 [J]. 信息技术与网络安全, 2019, 38 (11): 47-53.