

基于 SE-DenseNet 的变压器故障诊断

郭如雁, 彭敏放, 曹振其

(湖南大学电气与信息工程学院, 湖南 长沙 410082)

摘要: 由于变压器故障数据有限, 采用神经网络诊断容易出现过拟合。通过 Wasserstein 生成对抗网络(Wasserstein Generative Adversarial Networks, WGAN)数据增强, 采用变压器油中溶解气体作为故障诊断参量, 本文提出了 SE-DenseNet 神经网络模型, 该模型使用无编码比值法, 结合深度稠密神经网络(DenseNet)进行数据融合, 特征重复利用以及挤压和激励模块(Squeeze-and-Excitation, SE)的特征图权重值校准的优点, 运用指数衰减学习率、dropout、 L_2 正则化、损失函数 Adam 收敛算法等优化方法, 并且在同样数据集下将深度卷积神经网络、DenseNet 与 SE-DenseNet 的三者性能进行定量对比分析。实验结果表明, 三者之中本文提出的 SE-DenseNet 模型收敛速度快, 准确率最高, 泛化能力最强。

关键词: 数据增强; 无编码比值法; 稠密神经网络; 特征重复利用; 挤压和激励; 权重值校准; 指数衰减学习率

DOI: 10.12067/ATEEE2006023

文章编号: 1003-3076(2021)01-0061-09

中图分类号: TM41; TP183

1 引言

变压器是电力系统中传输电能的重要设备, 关于变压器故障诊断的研究对电力系统的运行、维护和检修具有十分重要的意义。在电力系统中 500/220 kV 和 500/330 kV 油浸式变压器是各变电站常用的设备, 虽然变压器的故障率较低, 但由于附件(如分接开关)质量问题, 线匝、线饼间的绝缘损坏, 内部绝缘距离不够, 绝缘油被污染, 雷击, 出口短路, 长期过载, 过电压运行等内外部原因, 均能造成变压器的重大故障, 给国民经济造成严重损失。为保障电力系统的安全可靠稳定运行, 国内外各学者围绕变压器故障诊断, 做了多个方向的研究。文献[1]通过小波变换对绕组行波的频率响应进行分析, 文献[2]通过测量变压器原边电压电流利用二端口网络知识在线监测短路阻抗, 文献[3]通过提取变压器振动信号的时域特征进行主分量分析, 文献[4]通过迁移学习将故障数据清洗得到的有效知识输入支持向量机进行故障分类, 文献[5]基于变压器油溶解气体(Dissolved Gas Analysis, DGA)的油色谱

数据结合深度神经网络进行分析等。由于前四者分别具有电磁干扰、对故障反映的灵敏度低、感应高压造成的安全隐患、集成多个二分类诊断器效率不高等问题, 目前基于 DGA 的故障诊断^[6,7]依然是变压器故障诊断的主要方法之一。

基于 DGA 诊断有实时在线进行, 安全无干扰、发现变压器早期潜伏性故障^[8]等优点, 结合深度学习神经网络强大的特征提取并进行多分类的能力, 能较大地提高变压器故障诊断的准确率, 并且所建模型具有性能稳定、易收敛的优点, 是一个新颖且发展前景广阔的研究方向。至今国内已知的基于深度学习的变压器故障诊断神经网络有四种^[9-12]。这四种网络均能实现多分类, 但由于变压器故障样本数据较少, 需要使用数据增强进一步提高模型的泛化能力。卷积神经网络擅长提取抽象特征, 能够将高维空间向量非线性变换映射到低维空间从而实现线性可分, 在图像分类、目标检测等问题上取得了很大的成功。本文所提出的基于 SE-DenseNet 变压器故障诊断模型, 就是深度压缩、激励网络和深度稠密网络的结合。模型的输入采用无编码比值法^[13]所提

收稿日期: 2020-06-10

基金项目: 国家自然科学基金项目(61973107&61472128)

作者简介: 郭如雁(1988-), 男, 湖南籍, 硕士研究生, 研究方向为基于深度学习变压器故障诊断;

彭敏放(1964-), 女, 湖南籍, 教授, 博导, 研究方向为电气设备监测诊断与评估、智能信息处理等。

出的 9 种不同组合形式的特征气体比率值,模型的输出为 9 种故障类型的各自概率值,将概率值代入 softmax 交叉熵损失函数中,再使用 Adam 算法最优化该损失函数,最终使 SE-DenseNet 模型的预测准确率收敛。

基于大云物移智技术平台的智能传感器在线监测的变压器油色谱数据不平衡,且故障数据量少,为此提出 WGAN^[14]网络进行数据增强,即样本生成技术,WGAN 可生成同一故障类别同特征的多类样本数据,为样本原始特征非线性化表示提供了充足的数据源。本文在已有故障数据的基础上利用 WGAN 将故障样本数量增加近一倍。实验证明 SE-DenseNet 模型预测效果在同类网络中最好。

2 SE-DenseNet 框架

2.1 SE module 介绍

本文构造一种能学习到特征图各通道的空间相关性和各通道的权重大小的 SE 模块,在 DenseNet 神经网络中加入 SE^[15]操作。SE 模块对输入其中的特征图每个通道的重要性进行学习,将得到的权重与对应的通道相乘,输出各通道权重得到校准后的特征图,即进行特征选择,使有用特征得到加强,无用特征得到削弱,完成特征重标定,以提高神经网络的辨别能力,其功能模块如图 1 所示。

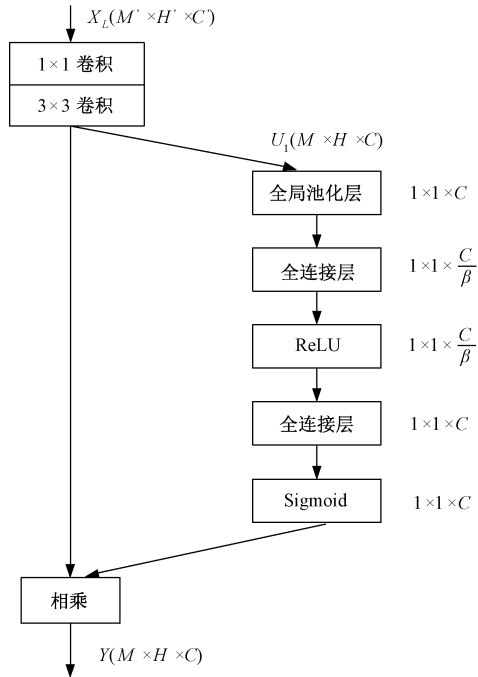


图 1 SE 模块图

Fig.1 SE module

(1)通过卷积操作将第 L 层特征图 X_L 变换为张量 U_1 , 即:

$$U_1 = W^L \otimes f(\text{BN}(W^{L-1} \otimes f(\text{BN}([X_0 X_1 \cdots X_{L-1}])))) \quad (1)$$

式中, BN 表示批量归一化处理; $f(\cdot)$ 为 ReLU 函数; W^{L-1}, W^L 分别为大小为 $1 \times 1, 3 \times 3$ 的卷积核。

(2)挤压操作,此环节是全局池化层,将特征图形状 $[M, H, C]$ 压缩为 $[1, 1, C]$, 得到该层特征图的 c 个通道的数值分布情况,其数学描述如式(2):

$$z_c = F_{\text{sq}}(u_c) = \frac{1}{M \times H} \sum_{i=1}^M \sum_{j=1}^H u_c(i, j) \quad (2)$$

式中, u_c 表示卷积操作后第 c 个通道特征图; z_c 为挤压操作后第 c 个通道特征图。 M, H, C 代表特征矩阵 U_1 的三维信息; c 为第 c 通道。

(3)激励操作,其数学原理是式(3):

$$s_c = F_{\text{ex}}(z, W) = \sigma(g(z, W)) = \sigma(W_2 f(W_1 z_c)) \quad (3)$$

式中, $W_1 \in R^{(C/\beta) \times C}$; $W_2 \in R^{C \times (C/\beta)}$; $f(\cdot)$ 表示 ReLU 激活函数; σ 是 Sigmoid 函数; β 为维度变换率,在本文中取 16,以减少计算量。最后将所得尺度矩阵 s_c 的 c 个元素与特征图 U_1 的 c 个通道一一对应相乘,输出得到 $Y = [y_1 y_2 \cdots y_c]$, 其数学原理是式(4):

$$Y = F_{\text{scale}}(u_c, s_c) = s_c \cdot u_c \quad (4)$$

式中, s_c 表示先挤压后激励所得到的向量,其维数为 c ; u_c 表示卷积操作后的特征图,其通道数为 c 。

由以上分析可知,SE 模块通过先压缩再激励输入,将特征图以通道为单位映射成一个具有全局性的实数,最后将此实数与输入对应相乘,完成特征图各通道的相关性学习。

2.2 SE-DenseNet 网络结构特点

DenseNet^[16]主要特点是将每一层特征图与前面所有层的特征图在通道数的维度上进行累加,网络的每一层对特征值进行少量的学习,降低了冗余性^[17]。Densenet 的优点是通过稠密连接将神经网络的各层特征进行了融合,特征得到了重复利用,可以减少计算量,其跳跃结构使各层输入特征图可以直接与最后的损失函数相连,接受最终损失函数的监督,解决梯度消失问题,使网络的信息流通顺畅。

本文介绍的 SE-DenseNet 将 SE 模块与 Densenet 的优点结合,其网络配置信息如表 1 所示,结构如图 2 所示。由表 1 和图 2 可知,相较于传统 DenseNet 神经网络,SE-DenseNet 的创新之处在于:

(1)将稠密连接结构块上的转换层中包含的平

均池化层去掉,保留 BN 层,ReLU 层和 1×1 的卷积,得到新的转换层,可以保留全局信息,增强网络的稳定性,减少资源占用。

(2)SE 模块再将经稠密连接块得到的特征图与经新的转换层得到的特征图进行各通道权重校准,增强有益特征,抑制无用特征,使网络的性能得到有效提高。

表 1 SE-DenseNet 网络配置信息表

Tab.1 SE-DenseNet network configuration information

网络层	SE-DenseNet		DenseNet	
	矩阵维度	结构配置	矩阵维度	结构配置
结构块	3×3	$\begin{bmatrix} 1 \times 1, 48 \\ 3 \times 3, 12 \\ \text{SE}(24/16, 12) \end{bmatrix} \times 4$	3×3	$\begin{bmatrix} 1 \times 1, 48 \\ 3 \times 3, 12 \end{bmatrix} \times 4$
转换层	3×3	$[1 \times 1, 6]$	3×3 3×3	$[1 \times 1, 6]$ 1×1 平均池化
⋮				
分类层	1×1	全局平均池化,全连接,soft max		

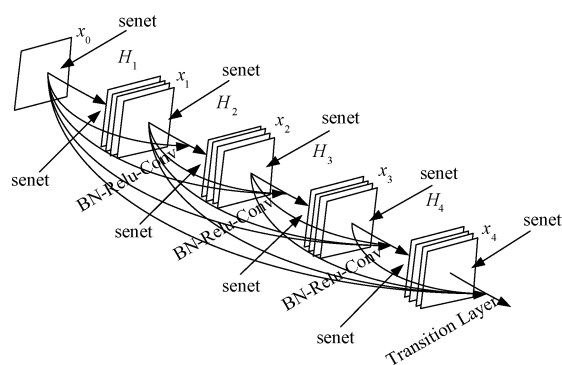


图 2 SE-DenseNet 示意图

Fig.2 SE-DenseNet

3 提高模型性能的方法

本文在 DenseNet 程序的基础上,采用 6 种方法改进 SE-DenseNet 模型的收敛效果和预测准确率:指数衰减学习率确保模型参数较快的更新速度和在全局最优点收敛; L_2 正则化使模型待估参数衰减,以减少计算资源占用;dropout 减弱特征之间的依赖性,以防止过拟合;Adam 算法使用迭代次数和指数衰减率对梯度均值和梯度平方均值进行了校正,使算法对梯度的预测更加精准;使用批量归一化可减少初始化的影响,提高训练速度;运用 ReLU 函数可提高神经网络逼近任意非线性函数的能力。这 6 种方法具体介绍如下。

3.1 指数衰减学习率

学习率是用梯度更新模型待估参数的速率,当

学习率较大时,网络参数更新较快,在模型训练初期局部收敛较快,随着迭代次数增加时,模型会在全局最优点回荡,此时较小的学习率能使模型收敛,因此提出学习率随迭代次数进行指数级衰减,如式 (5):

$$\text{decayed_learning_rate} = \text{learning_rate} \times \text{decay_rate}^{(\text{global_step}/\text{decay_steps})} \quad (5)$$

式中,decayed_learning_rate 为衰减后的学习率;decay_rate 是衰减指数;global_step 是当前的迭代轮数;decay_steps 是衰减速度。

3.2 Dropout 策略

Dropout 是一种正则化方法,它在一次训练中让隐含层神经元以概率 P 抑制,丢弃隐含层一部分神经元,同时保留被丢弃节点的参数值,在误差反向传播时仅更新被激活神经元的参数值,在下一次训练时重复上述过程,每次训练得到一个不同的神经网络,最后集成这些网络。在测试集上运行模型时,将模型学习到的权值矩阵乘以概率 P ,使模型预测准确。dropout 将性质相反的过拟合相互削弱,在不同隐含层子集中均可学到相应特征,以减弱特征之间的依赖性,防止过拟合。

3.3 L_2 正则化

为了将权重和偏置参数收敛为稀疏性更好的矩阵,以减少计算量,通常在目标函数上附加一项参数惩罚项,在用梯度下降更新参数时,权重矩阵会乘以一个小于 1 的缩放因子,将权重正则化至零点,若未添加惩罚项的原损失函数的 Hessian 矩阵为 H ,则 H 的特征值越小,权重收缩至零点的效果越明显, L_2 正则化公式为:

$$\begin{cases} f(\omega, b; X, y) = -\frac{1}{n} \sum_x \sum_i [y_i \cdot \log(y_{\text{predicted}_i}) + (1 - y_i) \cdot \log(1 - y_{\text{predicted}_i})] + \alpha R(\omega, b) \\ R(\omega, b) = \frac{1}{2} \|\omega\|_2^2 + \frac{1}{2} \|b\|_2^2 \end{cases} \quad (6)$$

式中, ω, b 为神经网络模型待寻优参数 θ ; X 为输入向量; $y_{\text{predicted}_i}$ 为预测值; y_i 为真实标签值; α 为正则项系数。 α 能增加输入 X 的方差,因此 L_2 正则化使权重往稀疏性好的方向衰减,在以下要介绍的 Adam 算法中, $f(\theta; X, y)$ 为损失函数。

3.4 Adam 算法

Adam 算法是一种用一阶梯度对损失函数进行最优化的方法,该算法所需存储小,准确率高,能避免模型在最优点大幅振荡,因此适用于大规

模数据集和参数的神经网络。该算法的流程如图 3 所示。

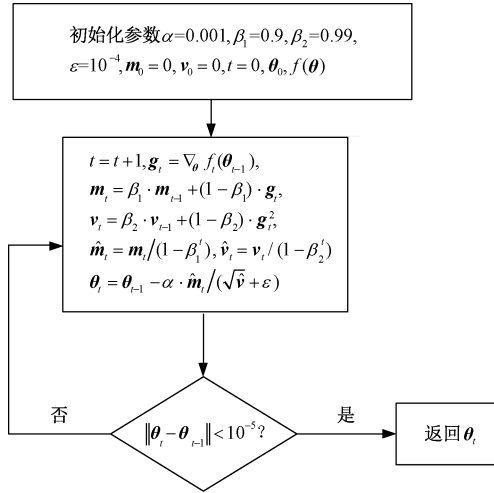


图 3 Adam 算法流程

Fig.3 Adam algorithm flow

(1) 设置超参数学习率 α , 一阶矩估计指数衰减率 β_1 , 二阶原点矩估计指数衰减率 β_2 , α 采用指数衰减学习率, 其值从 0.001 开始随迭代次数 t 按指数规律衰减。初始化参数 θ_0 向量的各元素, 一阶矩向量 m_0 , 二阶矩向量 v_0 。

(2) 求出 $f(\theta)$ 对 θ 的偏导数 $\nabla_{\theta} f(\theta_{t-1})$, 得梯度向量 g_t , 用 g_t 按指数加权平均值公式更新带偏差的一阶矩估计 m_t , 同理, 用梯度向量的逐元素平方更新带偏差的二阶原点矩估计 v_t 。用指数衰减率 β_1, β_2 计算偏差修正的一阶矩估计 \hat{m}_t 和二阶原点矩估计 \hat{v}_t 。将 \hat{m}_t 与 $\sqrt{\hat{v}_t}$ 的比值取代梯度下降法的 g_t , 对参数进行更新, 图 3 中“ \cdot ”表示乘法。

(3) 进行收敛准则校验, 若收敛则停止更新返回参数 θ_t , 否则重新计算步骤(2)、(3)直至收敛。

3.5 批量归一化

对批量归一化 (Batch Normalization, BN) 做出解释: 训练网络时, 对于正向传播, 一个批量的每张特征图输入神经元为 $x = \{x_1, x_2, \dots, x_m\}$, 归一化公式为:

$$y = \gamma \frac{x - u}{\sigma} + \beta, u = \frac{1}{m} \sum_{i=1}^m x_i, \sigma^2 = \frac{1}{m} \sum_{i=1}^m (x_i - u)^2 \quad (7)$$

式中, u 为样本均值; σ^2 为样本方差; 缩放系数 γ 和平移系数 β 为类似于权重的可学习参数。通过变换重构, 使网络学习到所要提取的特征的数据分布, 反

向传播时通过链式求导求得梯度, 从而改变训练权值; 测试网络时, 对于多个批量样本的均值 u 和方差 σ^2 , 计算 u', σ' 时采用滑动平均的技巧:

$$u'^{(t)} = \frac{\alpha u'^{(t-1)} + (1 - \alpha) \theta^{(t)}}{1 - \alpha^t} \quad (8)$$

式中, 变量 u' 在第 t 批次记为 $u'^{(t)}$; $\theta^{(t)}$ 为变量 u' 在第 t 批次取值; $\alpha \in [0, 1)$, 当 $\alpha = 0$ 时, 则不使用滑动平均; 当 $\alpha \neq 0$, 使用滑动平均计算法, 减少内存占用。通过 BN 可以防止梯度弥散和爆炸。

3.6 ReLU 激活函数

使用 ReLU 非线性激活函数可以使神经网络拟合各种函数, ReLU 函数的定义是:

$$f(x) = \max(0, x), x \in (-\infty, +\infty) \quad (9)$$

式中, x 为输入张量。该函数可将输入映射到正数域, ReLU 函数占用计算资源小, 当输入为负数, 该神经元处于抑制状态, 当输入 x 为正值时, 由于其导数为常数 1, 不会导致梯度变小, 模型参数能保持收敛。综上可得基于 SE-DenseNet 变压器故障诊断流程如图 4 所示。

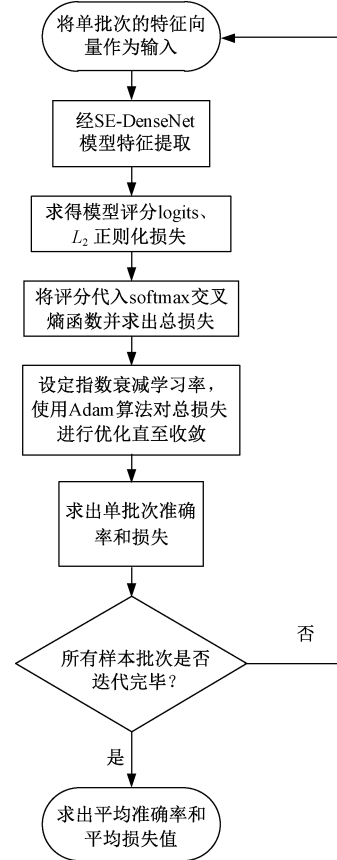


图 4 SE-DenseNet 诊断流程

Fig.4 SE-DenseNet diagnosis process

4 特征向量选取

4.1 无编码比值法及输入特征向量的选取

根据《变压器油中溶解气体分析和判断导则》^[18]所述,变压器发生故障时,变压器油中会产生七种含量不同的气体^[19],将其中的氢气(H_2)、甲烷(CH_4)、乙烷(C_2H_6)、乙烯(C_2H_4)、乙炔(C_2H_2)的体积分数作为观察对象,依据在线监测的气体含量数据,可采用 IEC 三比值法^[20]、四比值法^[21](包括 Dornenburg 法, Rogers 法, 德国四比值法)、大卫三角形法^[22]进行故障诊断,但故障编码不多,由于缺码的原因导致有些故障无法判断。本文采用无编码比值法,即直接由比值确定故障,不用将气体含量的比值进行编码,将变压器的故障区分为 9 种:低能放电兼过热 MF_1 、高能放电兼过热 MF_2 、局部放电 PD、低能放电 D_1 、高能放电 D_2 、低温过热 T_1 ($<300^\circ C$)、中温过热 T_2 ($300\sim 700^\circ C$)、高温过热 T_3 ($>700^\circ C$)、正常 N。将以下特征气体浓度比值作为输入特征量:

$$CH_4/H_2, C_2H_2/C_2H_4, C_2H_4/C_2H_6, C_2H_2/(C_1 + C_2)/\%, H_2/(H_2 + C_1 + C_2)/\%, C_2H_4/(C_1 + C_2)/\%, CH_4/(C_1 + C_2)/\%, C_2H_6/(C_1 + C_2)/\%, (CH_4 + C_2H_4)/(C_1 + C_2)/\%$$

式中, C_1 为一阶碳氢化合物 CH_4 ; C_2 为二阶碳氢化合物 C_2H_6 、 C_2H_4 、 C_2H_2 体积分数之和,输入特征信息含量丰富,适用于 SE-DenseNet 深度神经网络。

4.2 故障类型独热编码

针对 9 种故障类型,采用 one-hot 编码。将 MF_1 、 MF_2 、PD、 D_1 、 D_2 、 T_1 、 T_2 、 T_3 、N 依次序编码为: $[0,0,0,0,0,0,0,0,1]$, $[0,0,0,0,0,0,0,1,0]$, $[0,0,0,0,0,0,1,0,0]$, $[0,0,0,0,0,1,0,0,0]$, $[0,0,0,0,1,0,0,0,0]$, $[0,0,0,1,0,0,0,0,0]$, $[0,0,1,0,0,0,0,0,0]$, $[0,1,0,0,0,0,0,0,0]$, $[1,0,0,0,0,0,0,0,0]$ 。针对本文探讨的分类问题,使用 softmax 交叉熵来作为目标函数的一部分。

5 WGAN 的数据增强

WGAN 是有监督的学习,本文使用 WGAN 将数据空间规模扩大,依据原始数据,生成特征一致,但数值不同、故障类型可以被识别为同一种类型的数据,以提高 SE-DenseNet 神经网络的泛化能力。

5.1 GAN 的原理

生成对抗网络(Generative Adversarial Networks, GAN)^[25]由生成器 G 和判别器 D 组成,生成器 G 将

噪音数据生成为与真实数据分布相似的数据,并尽量骗过判别器,使判别器将生成的数据判断为真(1)。判别器的作用是辨别数据的真(1)伪(0)。因此使判别器不断的学习,以提高它的判别能力,让判别器效果更好,在生成器和判别器对抗学习中,最后判别器分辨不出生成数据的真伪,即生成的数据特征和真实数据的特征也就高度一致了,损失函数即为式(10):

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\ln(D(x))] + E_{z \sim p_z(z)} [\ln(1 - D(G(z)))] \quad (10)$$

式中, x 为真实数据; $p_{data}(x)$ 为真实数据分布; z 为噪音数据; $p_z(z)$ 为噪音数据分布; $D(\cdot)$ 为判别器; $G(\cdot)$ 为生成器。损失函数的功能是最大化 D 的判断能力,使最小化 G 和真实数据的分布之间的差异。判别器中使 $D(x)$ 接近 1, $D(G(z))$ 接近 0, 生成器中使 $D(G(z))$ 接近 1。 $E_{x \sim p_{data}(x)}$ 和 $E_{z \sim p_z(z)}$ 表示真实数据和生成数据的概率, E 表示期望。

5.2 WGAN 的说明

WGAN 中使用平滑的 EM(Earth-Mover)距离,用梯度下降法优化参数时,可以提供梯度,实现两分布距离度量,在 EM 距离中加入李普希兹约束,使网络满足稳定性要求。以已知故障类型样本的五种特征气体含量为真实数据,从九种故障中任选一种,样本量 10 左右,在区间 $[-1, 1]$ 中随机选取服从均匀分布的噪音数据,将两者输入 WGAN 模型得生成数据 gen_data ,用关于 gen_data 的一次函数生成同种故障类型的样本数据 $data$ 。总共生成 1500 条样本数据,为验证生成数据与原始数据特征一致性,预先使用 1630 条原始数据训练 SE-DenseNet 模型,再分别将生成的各种故障类型数据作为输入,得到的 SE-DenseNet 模型预测准确率如表 2 所示。

表 2 生成数据作为输入的模型准确率

故障类型	样本量	准确率(%)
MF1	127	98.43
MF2	173	93.06
PD	162	92.88
D1	150	87.33
D2	260	93.75
T1	121	90.16
T2	184	97.83
T3	173	90.17
N	150	86.72
样本总量	1500	90.41

表2的准确率为将生成的某一故障类型数据输入模型,重复预测12次得到的平均值。准确率最低的故障类型为正常,其值为86.72%,最高为低能放电兼过热,其值为98.43%,当将所有生成样本输入模型时得到的准确率值为90.41%,表2数据充分说明生成数据具有与原始数据一致的特征。

6 算例分析

本实验使用的软件框架为 tensorflow 1.12, win10 系统,硬件为 intel core i7, 1.80 GHz, 内存 16 GB, 英伟达显卡 MX-150, 2 GB 显存。本文使用某水电站的 500/330 kV 变压器的 1 630 条数据作为原始数据,使用 WGAN 生成了 1 500 条数据。将生成数据和原始数据合并成 3 130 条数据的样本集,两者随机充分混合后,按 9:1 的比例分为训练集和测试集。运用 t-SNE 技术将样本数据进行可视化,得到初始分类效果如图 5 所示,可知未经 SE-DenseNet 模型特征提取处理的数据分类效果较为分散。

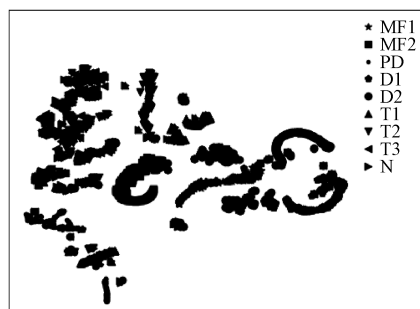


图5 原始分类效果

Fig.5 Original classification effect

在 CNN, Densenet 和 SE-DenseNet 中,将数据输入各自模型,迭代 300 个 epoch (每个 epoch 代表整个数据集完整处理一次),可得训练集和测试集的准确率如图 6 和图 7 所示。

由图 6 和图 7 可知, CNN, Densenet, SE-DenseNet 训练集、测试集准确率最终收敛值如表 3 所示。

表3 各神经网络的准确率

Tab.3 The accuracy of each neural network

模型 样本	CNN	Densenet	SE-DenseNet
训练集	84.30	90.23	94.08
测试集	81.47	85.58	88.46

此处训练集样本量为 2 816, 测试集样本量为

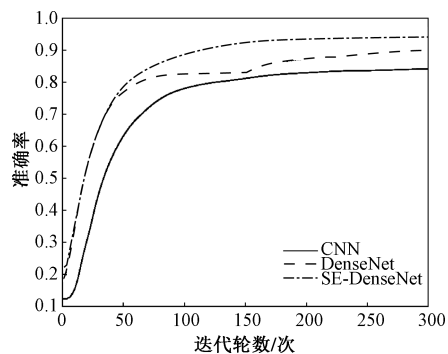


图6 训练集准确率图

Fig.6 Accuracy graph of training set

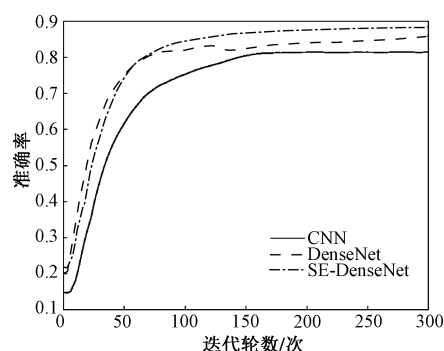


图7 测试集准确率图

Fig.7 Test set accuracy chart

312。训练集准确率的定义是将单批次 64 个样本作为模型输入,共 44 个批次,将单批次的准确率累加,再除以 44 求得平均准确率。单批次准确率是模型评分 logits 向量中的最大值的索引号与真实标签(已转为独热编码向量)中的最大值索引号相等的次数之和除以单批次样本量。每 44 个批次为一回迭代轮数,共 300 回迭代轮数,模型采用最终收敛的平均准确率。测试集是单批次 26 个样本,共 12 个批次,其平均准确率同理可得。

由图 6 可知,当数据为训练集时, CNN、Densenet、SE-DenseNet 分别在迭代 220、250、200 次后趋于稳定。通过观察可得 DenseNet 准确率上升的速度最快,SE-DenseNet 由于挤压和激励环节操作,准确率上升速度位居其次,在迭代 48 次后,SE-DenseNet 的训练集的准确率高于 Densenet,最终 SE-DenseNet 网络在训练集上的准确率为三者最高: 94.08%, 分别比 DenseNet 和 CNN 高出了 3.85% 和 9.78%。由图 7 可知,当数据为测试集时, CNN、Densenet、SE-DenseNet 分别在迭代 160、135、140 次后收敛,在迭代初期, DenseNet 准确率上升的速度最快,迭代 65 次后, SE-DenseNet 的准确率高于

Densenet,最终 SE-DenseNet 在测试集上的准确率为 88.46%,在三者之中最高,分别比 DenseNet 和 CNN 高出了 2.88%和 6.99%。

CNN 由于数据样本量不够,提取特征的能力相较于数据特征信息流通更为顺畅的 DenseNet 和 SE-DenseNet 较低,容易出现过拟合,所以准确率最低,SE-DenseNet 由于在 DenseNet 基础上嵌入 SE 操作,使得网络能更快的学习到重要特征,降低了 DenseNet 网络融合数据特征时产生的冗余性,所以准确率最高。

从图 8 和图 9 可以看出 SE-DenseNet 模型在训练集和测试集上的最终损失值分别为 0.270 1 和 0.605 8,由此得出 SE-DenseNet 损失值很小,说明 SE-DenseNet 模型预测很准确。

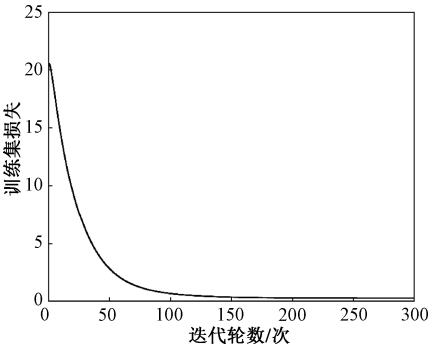


图 8 SE-DenseNet 的训练集损失函数下降图
Fig.8 SE-DenseNet training set loss function decline graph

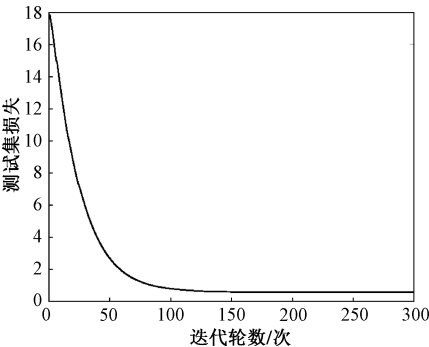


图 9 SE-DenseNet 的测试集损失函数下降图
Fig.9 SE-DenseNet test set loss function decline graph

为了验证 SE-DenseNet 模型的优良的预测性能,将其与 DenseNet 在与本节上文同样的数据集的条件下进行训练对比,在测试集的结果如表 4 和表 5 所示。

此处准确率定义同表 3,由表 4 可得模型改进后,在相同的网络层数下,SE-DenseNet 比 DenseNet 测试集准确率都要高,平均高出 1.504%,为验证模

表 4 模型改进前后测试集分类准确率对比

Tab.4 Comparison of classification accuracy (%)		
网络层数	DenseNet	SE-DenseNet
18	85.26	86.54
26	85.58	86.86
35	85.90	87.18
48	86.22	87.82
56	86.38	88.46

表 5 深层 SE-DenseNet 测试集分类准确率

Tab.5 Deep SE-DenseNet accuracy (%)	
网络层数	SE-DenseNet
83	88.46
120	88.14

型的稳定性,继续对 SE-DenseNet 进行深层训练,由表 4 和 5 可得 SE-DenseNet 模型在 56 层时达到稳定状态,准确率维持在 88.46%。经表 4 和表 5 的对比可知,由于 SE 模块的特征重标定,结合 DenseNet 的特征重复利用的优点,本文所提出的 SE-DenseNet 模型有着优良的性能和很高预测准确率。

在每一单批次迭代过程中,经 SE-DenseNet 模型的特征提取后,将模型得到的评分数据 logits 输入 softmax 交叉熵函数之前,将评分数据 logits 保存至文档中,然后再进行分类训练。当评分数据迭代更新至第 300 轮,即分类训练结束时,将最终的评分数据提取出来并进行 t-SNE 的可视化,可得最终分类效果如图 10 所示。由图 10 可得经本文提出的 SE-DenseNet 模型特征提取后样本数据分类效果更集中。结合图 6 和图 7、表 3 和表 4 中的数据对比以及对图 8~图 10 的分析,可知本文提出的 SE-DenseNet 诊断模型收敛性能最好,泛化能力最强。



图 10 最终分类效果
Fig.10 Final classification effect

7 结论

针对变压器在已有故障数据条件下,故障类型难以判断、诊断准确率较低的情况,本文结合 WGAN 数据增强提出了 SE-DenseNet 模型,能细致地检测出故障数据与故障类型的内在联系,结论如下:

(1)基于无编码比值法原理的 CNN, DenseNet, SE-DenseNet 变压器故障诊断准确率普遍较高, SE-DenseNet 加入了特征图校准权重的 SE 模块,抑制无用特征,增强有用特征,同时充分利用 DenseNet 的梯度流通顺畅,计算冗余性较低的优点,结合运用指数衰减学习率、dropout、 L_2 正则化,运用 Adam 收敛算法,批量归一化及 ReLU 函数,经表 3 和表 4 的对比可知,相较于传统的 CNN 和 DenseNet,本文提出的 SE-DenseNet 性能更好,预测准确率最高。

(2)无编码比值法在变压器的故障判断中,能将复合故障和单一故障区分开来,使模型判断能力得到很大的提高,泛化能力增强。在 WGAN 数据增强时,虽然生成了特征一致的同故障类型数据,但还可进一步研究关于生成数据 gen_data 的经验公式,以提高特征一致性程度和模型的预测准确率。

(3)SE-DenseNet 神经网络模型的加深与样本空间规模的扩大有着密切联系,通过实验已找到本文数据集下 SE-DenseNet 具有最佳性能的网络层数,今后的工作将研究模型性能的提高与故障样本规模扩大之间的关联,以及使用更有效的迭代收敛算法,对原模型进行优化,使模型收敛速度更快,泛化能力更优秀。

参考文献 (References):

- [1] 李成祥, 夏麒, 朱天宇, 等 (Li Chengxiang, Xia Qi, Zhu Tianyu, et al.). 基于脉冲频率响应法的变压器绕组变形带电检测仪 (Live transformer winding deformation detector based on impulse frequency response method) [J]. 电力自动化设备 (Electric Power Automation Equipment), 2018, 38 (1): 110-115, 120.
- [2] 刘勇, 杨帆, 张凡, 等 (Liu Yong, Yang Fan, Zhang Fan, et al.). 检测电力变压器绕组变形的扫频阻抗法研究 (Study on sweep frequency impedance to detect winding deformation within power transformer) [J]. 中国电机工程学报 (Proceedings of the CSEE), 2015, 35 (17): 4505-4516.
- [3] 张国栋, 苗青, 滕伟, 等 (Zhang Guodong, Miao Qing, Teng Wei, et al.). 基于振动特征主分量分析的电力变压器故障诊断 (Power transformer fault diagnosis based on vibration component principal component analysis) [J]. 变压器 (Transformer), 2019, 56 (5): 62-65.
- [4] 白浩, 王昱立 (Bai Hao, Wang Yuli). 基于数据清洗和知识迁移的变压器故障诊断模型 (Transformer fault diagnosis model based on data cleaning and transfer learning) [J]. 电工电能新技术 (Advanced Technology of Electrical Engineering and Energy), 2020, 39 (1): 28-35.
- [5] 张朝龙, 何怡刚, 杜博伦, 等 (Zhang Chaolong, He Yigang, Du Bolun, et al.). 基于深度学习的电力变压器智能故障诊断方法 (Intelligent fault diagnosis method of power transformer using deep learning) [J]. 电子测量与仪器学报 (Journal of Electronic Measurement and Instrumentation), 2020, 34 (1): 81-89.
- [6] 李春茂, 周妹末, 袁海满, 等 (Li Chunmao, Zhou Meimo, Yuan Haiman, et al.). 基于 DGA 粗糙集与决策信息融合变压器故障诊断 (Fault diagnosis of transformer based on rough set theory and decision information fusion) [J]. 电工电能新技术 (Advanced Technology of Electrical Engineering and Energy), 2018, 37 (1): 84-90.
- [7] 赵峰, 李硕 (Zhao Feng, Li Shuo). 基于熵权和灰关联度的变压器故障诊断 (Fault diagnosis in transformer based on entropy weighted and grey incidence) [J]. 电工电能新技术 (Advanced Technology of Electrical Engineering and Energy), 2015, 34 (1): 57-61.
- [8] 刘占元, 牛晓晨, 苏红, 等 (Liu Zhanyuan, Niu Xiaochen, Su Hong, et al.). 一种全光纤电流互感器早期故障诊断方法研究 (Research on early fault diagnosis method for all optical fiber current transformers) [J]. 电工电能新技术 (Advanced Technology of Electrical Engineering and Energy), 2019, 38 (11): 10-16.
- [9] 王德文, 雷倩 (Wang Dewen, Lei Qian). 基于贝叶斯正则化深度信念网络的电力变压器故障诊断方法 (Fault diagnosis of power transformer based on BR-DBN) [J]. 电力自动化设备 (Electric Power Automation Equipment), 2018, 38 (5): 129-135.
- [10] 石鑫, 朱永利, 宁晓光, 等 (Shi Xin, Zhu Yongli, Ning Xiaoguang, et al.). 基于深度自编码网络的电力变压器故障诊断 (Transformer fault diagnosis based on deep auto-encoder network) [J]. 电力自动化设备 (Electric Power Automation Equipment), 2016, 36 (5): 122-126.
- [11] 代杰杰, 宋辉, 盛戈皞, 等 (Dai Jiejie, Song Hui, Sheng Gehao, et al.). 采用 LSTM 网络的电力变压器运行状态预测方法研究 (Prediction method for power

- transformer running state based on LSTM network) [J]. 高电压技术 (High Voltage Engineering), 2018, 44 (4): 1099-1106.
- [12] 贾京龙, 余涛, 吴子杰, 等 (Jia Jinglong, Yu Tao, Wu Zijie, et al.). 基于卷积神经网络的变压器故障诊断方法 (Fault diagnosis method of transformer based on convolutional neural network) [J]. 电测与仪表 (Electrical Measurement & Instrumentation), 2017, 54 (13): 62-67.
- [13] 杜洋 (Du Yang). 用“无编码比值法”分析和判断变压器故障性质 (Analysis and judgement for transformer fault by using “noncode ratio method”) [J]. 变压器 (Transformer), 1999, 36 (3): 32-36.
- [14] 刘云鹏, 许自强, 和家慧, 等 (Liu Yunpeng, Xu Ziqiang, He Jiahui, et al.). 基于条件式 Wasserstein 生成对抗网络的电力变压器故障样本增强技术 (Data augmentation method for power transformer fault diagnosis based on conditional Wasserstein generative adversarial network) [J]. 电网技术 (Power System Technology), 2020, 44 (4): 1505-1513.
- [15] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks [A]. 2018 IEEE Conference on CVPR [C]. Salt Lake City, USA, 2018. 7132-7141.
- [16] Huang G, Liu Z, Weinberger K Q, et al. Densely connected convolutional networks [A]. 2017 IEEE Conference on CVPR [C]. Honolulu, USA, 2017. 2261-2269.
- [17] 辜超, 杨伟, 张晓星, 等 (Gu Chao, Yang Yi, Zhang Xiaoxing, et al.). 基于最大相关最小冗余准则的变压器故障诊断特征选择 (Feature selection for transformer fault diagnosis based on maximal relevance and minimal redundancy criterion) [J]. 电工电能新技术 (Advanced Technology of Electrical Engineering and Energy), 2018, 37 (7): 84-89.
- [18] DL/T 722—2014 变压器油中溶解气体分析和判断导则 (Guide to the analysis and the diagnosis of gases dissolved in transformer oil) [S].
- [19] 刘一, 倪远平 (Liu Yi, Ni Yuanping). 三比值灰关联度分析的变压器故障诊断法 (A transformer faults diagnosis method based on the grey correlation theory) [J]. 高电压技术 (High Voltage Engineering), 2002, 28 (10): 16-17, 27.
- [20] 杨廷方, 刘沛, 李景禄, 等 (Yang Tingfang, Liu Pei, Li Jinglu, et al.). FCM 结合 IEC 三比值法诊断变压器故障 (New fault diagnosis method of power transformer by combination of FCM and IEC three ratio method) [J]. 高电压技术 (High Voltage Engineering), 2007, 33 (8): 66-70.
- [21] 邹剑, 陆瑾, 周晓帆 (Zou Jian, Lu Jin, Zhou Xiaofan). 四比值法在变压器过热性故障判断中的应用 (Application of four ratio method in transformer overheating fault diagnosis) [J]. 变压器 (Transformer), 2011, 48 (10): 66-67.
- [22] Lakehal A, Tachi F. Bayesian Duval triangle method for fault prediction and assessment of oil immersed transformers [J]. Measurement and Control, 2017, 50 (4): 103-109.

Fault diagnosis of power transformer based on SE-DenseNet

GUO Ru-yan, PENG Min-fang, CAO Zhen-qi

(College of Electrical and Information Engineering, Hunan University, Changsha 410082, China)

Abstract: Due to the limited failure data of transformer, using neural network diagnosis is prone to overfitting. In this paper, by data enhancement of WGAN (Wasserstein Generative Adversarial Networks), the dissolved gas in the transformer oil is used as the fault diagnosis parameter and the SE-DenseNet neural network model was proposed. The model uses non-encoding ratio method combining with data fusion of DenseNet, feature reuse and feature map weight value calibration advantages of SE (Squeeze-and-Excitation) module, and using exponential decay learning rate, dropout, L_2 regularization, loss function Adam convergence algorithm and other optimization methods. This paper analyzed the performance of CNN (Convolutional Neural Network), DenseNet and SE-DenseNet under the same data set. The experimental results show that among the three kinds of neural networks, the SE-DenseNet model proposed in this paper has faster convergence speed, highest accuracy and strongest generalization ability.

Key words: data enhancement; non-encoding ratio method; DenseNet; feature reuse; squeeze-and-excitation; weight value calibration; exponential decay learning rate