

基于数据清洗和知识迁移的变压器故障诊断模型

白 浩¹, 王昱力²

(1. 南方电网科学研究院有限责任公司, 广东 广州 510633;
2. 中国电力科学研究院武汉分院, 湖北 武汉 430072)

摘要: 变压器是保障电网安全运行的重要基础, 本文建立了一种基于数据清洗和知识迁移的变压器故障诊断模型, 用于解决变压器单体故障数据少、总体数据繁杂导致诊断器泛化能力低的问题。首先, 采用 Tanimoto 系数计算待诊断变压器与其他变压器故障的综合相似度, 对辅助故障数据进行一次清洗; 其次, 通过剔除奇异边缘附近故障数据, 对目标和辅助故障数据进行二次清洗; 在两次数据清洗的基础上, 以支持向量机作为迁移学习算法 TrAdaBoost 的基本分类算法, 通过迭代不断调整目标故障数据和辅助故障数据的权重, 将辅助故障数据中的有效知识迁移至故障诊断器, 得到基于迁移学习的变压器故障诊断器模型。测试结果表明数据清洗和知识迁移能够有效提高诊断效率以及故障识别的准确性。

关键词: 变压器; 故障诊断模型; 知识迁移; 数据清洗; TrAdaBoost 算法

DOI: 10. 12067/ATEEE1811049 **文章编号:** 1003-3076(2020)01-0028-08 **中图分类号:** TM76

1 引言

随着数据挖掘技术和机器学习技术的发展, 变压器智能故障诊断技术开始逐步得到研究应用。采用支持向量机、神经网络和极限学习机等智能算法分析绝缘油中氢气(H₂), 甲烷(CH₄)、乙烷(C₂H₆)、乙烯(C₂H₄)、乙炔(C₂H₂)和二氧化碳(CO₂)等气体含量与故障类型的映射关系, 模拟人类专家对变压器的运行状态和故障做出正确判断和决策^[1]。目前的研究主要通过优化算法参数, 尝试新型算法和集成多种算法实现诊断精度的提高。

采用支持向量机的诊断算法性能主要取决于核函数及惩罚因子, 遗传算法有助于选择最优核参数以提高变压器诊断效果^[2]。文献[3, 4]分别采用遗传粒子群算法和混沌粒子群算法优化神经网络参数, 克服常规神经网络在故障诊断中训练收敛速度慢、易陷入局部极小值等缺点。针对传统故障诊断智能方法存在调整参数多及参数确定困难的问题, 文献[5]采用只设置隐含层神经元个数的极限学习机实现快速诊断, 提高诊断效率。文献[6]基于栈

式自编码网络, 利用大量无标签样本进行预训练和少量有标签样本进行微调, 实现了 80% 的诊断正确率。此外自组织抗体网络、云推理模型等新方法在变压器故障诊断中也取得了不错的效果^[7, 8]。组合智能算法能有效弥补单一算法的不足, 提高整体的诊断精确度。文献[9]将模糊 Petri 网的容错能力与 BP 网络的自学习和自适应能力结合使用, 能够完善模糊推理规则, 诊断电力变压器中的多重故障。文献[10]通过集成样本信息提取能力较强的深度信念网络和训练难度小的贝叶斯网络, 提高诊断算法的泛化能力。

为使变压器故障诊断器具有较高的准确率, 文献[3-10]所述诊断器主要以单台变压器故障数据充足为前提, 并且训练集与测试集具有相同的数据分布和特征空间^[11]。实际运行中单台变压器的故障次数有限, 所以能获取的故障数据十分有限并且训练数据与测试数据分布特性存在一定的差异。传统方法多通过整合不同变压器的故障数据以达到提高数据总量的目的, 却忽略了不同变压器类型、运行工况、故障诱因等方面的差异性, 这些都会降低传统机

收稿日期: 2018-11-27
基金项目: 南方电网科技项目(ZBKJXM20180220)
作者简介: 白 浩(1987-), 男, 山西籍, 工程师, 博士, 研究方向为配电自动化、主动配电网和人工智能应用;
王昱力(1989-), 男, 湖北籍, 工程师, 硕士, 研究方向为输电设备智能化运维检修技术。

器学习故障诊断模型的泛化能力和准确率。

迁移学习可以从现有源域中学到知识,然后将这些知识迁移到目标域,以完成目标域的分类,不要求源域和目标域具有相同的数据分布,目前已在电机的故障诊断中得到了应用^[12]。该学习方法适用于目标训练样本不足以训练出一个分类精度较高的分类器的情况^[13],通过辅助数据集的帮助,使得分类器在小样本的学习过程中逐步提高分类精度,可解决变压器单体故障数据少,总体数据繁而杂的问题。

本文建立一种基于数据清洗和知识迁移的变压器故障诊断模型,首先采用 Tanimoto 系数计算待诊断变压器与其他变压器故障的综合相似度,对辅助故障数据进行一次清洗;其次,引入奇异边缘的概念,通过剔除奇异边缘附近故障数据,对目标和辅助故障数据进行二次清洗;在两次数据清洗的基础上,将迁移学习算法应用于变压器的故障诊断中,通过算例分析,验证了本文所提模型的准确性。

2 变压器故障特征量

变压器作为一个多元复杂结构体,能够反映故障状态的特征信息很多,通常采用气体含量和气体比值两大类故障特征量选取方法^[14]。气体含量主要是分析变压器油中溶解的 5 种关键气体(H_2 、 CH_4 、 C_2H_6 、 C_2H_4 和 C_2H_2) 的含量与故障类型之间的关联规则,但各特征气体的浓度分散性较大,同一故障类型的气体含量波动范围较广,对变压器故障诊断的准确性有一定影响^[15]。另一类方法,如 IEC 比值法和 DTM 比值法,采用主导气体的相对比值分析变压器故障,但该方法不适用于单一气体的超标情况^[16]。

本文结合关键气体含量和气体比值两类方法,选择 21 种具有较强代表性、能够量化并且准确反映变压器运行状况的特征量作为表征各种故障的故障特征,具体故障特征集如表 1 所示。

由于样本数据波动范围比较大,为了减小输入数据的差值,需要对样本数据进行预处理。在剔除奇异值后,采用式(1)对各特征量进行标准化计算,以提升变压器的故障诊断准确率。

式中, x 表示标准化后的特征值; x_0 表示初始特征值; x_m 表示样本数据中该项特征值的平均值; x_{std} 表

表 1 变压器故障特征集

Tab.1 Transformer fault feature set	
特征量选取方法	故障特征量
关键气体含量	$f_1(H_2)$ 、 $f_2(CH_4)$ 、 $f_3(C_2H_2)$ 、 $f_4(C_2H_4)$ 、 $f_5(C_2H_6)$
与氢气比值	$f_6(CH_4/H_2)$ 、 $f_7(C_2H_2/H_2)$ 、 $f_8(C_2H_4/H_2)$ 、 $f_9(C_2H_6/H_2)$
烃类之间比值	$f_{10}(C_2H_2/C_2H_4)$ 、 $f_{11}(C_2H_2/CH_4)$ 、 $f_{12}(C_2H_6/C_2H_2)$
与总烃含量比值	$f_{13}(CH_4/S)$ 、 $f_{14}(C_2H_2/S)$ 、 $f_{15}(C_2H_4/S)$ 、 $f_{16}(C_2H_6/S)$; $S=CH_4+C_2H_2+C_2H_4+C_2H_6$
与氢气和总烃含量的比值	$f_{17}(H_2/T)$ 、 $f_{18}(CH_4/T)$ 、 $f_{19}(C_2H_2/T)$ 、 $f_{20}(C_2H_4/T)$ 、 $f_{21}(C_2H_6/T)$; $T=S+H_2$

示样本数据中该项特征值的标准差值。

根据 IEC60599 标准规定,变压器故障主要分为热性故障和电性故障,具体可分为中低温过热故障(温度 $\leq 700^{\circ}C$)、高温过热故障(温度 $> 700^{\circ}C$)、低能放电故障、高能放电故障和局部放电故障五种类型,其典型故障实例分析如表 2 所示。

表 2 变压器典型故障

Tab.2 Typical transformer fault		
故障编号	故障类型	实例
1	中低温过热	绕组油流阻塞,绕组中平行相邻导体间绝缘磨损
2	高温过热	铁心和油箱上存在大的环流,铁心硅钢叠片之间发生短路
3	低能放电	套管与箱臂、夹件间、线圈内的高压和地端之间产生电弧或火花放电
4	高能放电	沿面放电、电弧或者闪络
5	局部放电	充气空腔中局部放电

3 变压器故障数据清洗方法

在对某台变压器进行故障诊断时,历史故障诊断指标及所对应的故障类型为目标故障数据,未标注故障类型的测试数据构成数据集 S ,其他变压器的故障信息称为辅助故障数据。若将故障数据中的不良数据引入变压器的故障诊断中,会严重影响变压器的诊断精度,因此需要清洗目标和辅助故障数据,筛选出符合要求的数据对变压器故障诊断器进行训练。

本文首先采用 Tanimoto 系数^[17]对目标故障数据和辅助故障数据进行相似度检验,对辅助故障数据进行一次清洗;其次,通过去除位于奇异边缘的故

障数据,对目标故障数据和辅助故障数据进行二次清洗,具体流程如图 1 所示。

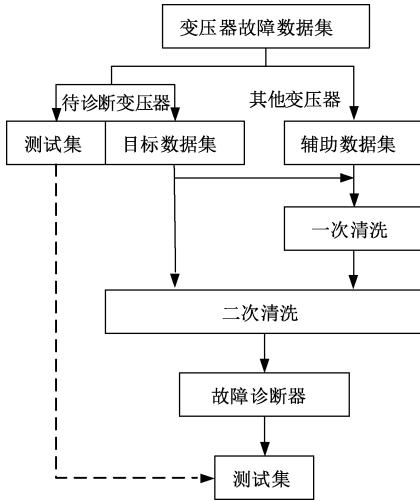


图 1 数据清洗示意图

Fig.1 Schematic of data cleaning

3.1 辅助故障数据的一次清洗

目标故障数据和辅助故障数据具有共性是迁移学习的前提,因此需要筛选出与目标故障数据特征相似度较高的数据作为辅助故障数据。假设变压器故障类型数为 T ,本文中 $T=5$ 。令待诊断变压器及另一其他变压器中故障 t 的故障数据量分别为 $n_{d,t}$ 和 $n_{a,t}$,且在自身故障数据中所占的比例分别为 $p_{d,t}$ 和 $p_{a,t}$ ($t=1,2,\dots,T$),对应的故障诊断指标的标么值分别为 $\mathbf{X}_{d,t}$ 和 $\mathbf{X}_{a,t}$:

$$\begin{cases} \mathbf{X}_{d,t} = \{\mathbf{x}_{d,t,i}\} & i = 1, 2, \dots, n_{d,t} \\ \mathbf{X}_{a,t} = \{\mathbf{x}_{a,t,i}\} & i = 1, 2, \dots, n_{a,t} \end{cases} \quad (2)$$

式中, $\mathbf{x}_{d,t,i}$ 和 $\mathbf{x}_{a,t,i}$ 向量中的元素分别为变压器故障诊断特征集中各变量的标么值。

对 $\mathbf{X}_{d,t}$ 和 $\mathbf{X}_{a,t}$ 中的向量进行平均化处理:

$$\begin{cases} \overline{\mathbf{X}_{d,t}} = \frac{1}{n_{d,t}} \sum_{i=1}^{n_{d,t}} \mathbf{x}_{d,t,i} \\ \overline{\mathbf{X}_{a,t}} = \frac{1}{n_{a,t}} \sum_{i=1}^{n_{a,t}} \mathbf{x}_{a,t,i} \end{cases} \quad (3)$$

引入 Tanimoto 系数,得到待诊断变压器及另一其他变压器故障 t 的相似度 S_t 为:

$$S_t(\overline{\mathbf{X}_{d,t}}, \overline{\mathbf{X}_{a,t}}) = \frac{\overline{\mathbf{X}_{d,t}} \cdot \overline{\mathbf{X}_{a,t}}}{\|\overline{\mathbf{X}_{d,t}}\|^2 + \|\overline{\mathbf{X}_{a,t}}\|^2 - \overline{\mathbf{X}_{d,t}} \cdot \overline{\mathbf{X}_{a,t}}} \quad (4)$$

结合故障分布比例,得到待诊断变压器及另一其他变压器故障的综合相似度 S_T :

$$S_T = \sum_{t=1}^T \frac{(p_{d,t} + p_{a,t}) S_t}{2} \quad (5)$$

式中, $S_T \subseteq [0,1]$, 且越接近 1, 综合相似度越高。定义迁移度阈值 θ , 若 $S_T < \theta$, 则会造成负迁移, 引入辅助数据使得目标分类效果变差, 应剔除该变压器的故障数据; $S_T \geq \theta$, 则会造成正迁移, 引入辅助数据能帮助目标数据学习, 应保留该变压器的故障数据, 记为集合 ω'_a 。计算所有其他变压器故障数据与待诊断变压器故障数据的综合相似度, 筛选出符合要求的数据作为清洗后的辅助故障数据, 组成集合 Ω'_a , 完成辅助故障数据的一次清洗。

3.2 目标和辅助故障数据的二次清洗

目标和辅助故障数据训练样本的选择会直接影响变压器故障诊断器的优劣。差别较小的两组变压器故障数据可能对应着两种不同的故障类型。以图 2 为例, 虚线两边对应于两种故障类型, 本文将虚线定义为奇异边缘。奇异边缘附近(虚线框内)的故障数据太过相近, 将其引入变压器故障诊断器的训练中容易造成过拟合, 降低诊断精度, 因此需要剔除奇异边缘附近的故障数据。

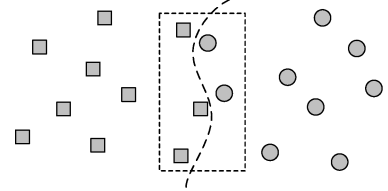


图 2 奇异边缘示意图

Fig.2 Schematic of singular edge

以目标故障数据 $\mathbf{x}_{d,t,i}$ 为例, 说明剔除奇异边缘附近故障数据的方法, 辅助故障数据的筛选方法与此完全相同。

(1) 计算故障 t ($t=1,2,\dots,T$) 中某 $\mathbf{x}_{d,t,i}$ ($i \in 1, 2, \dots, n_{d,t}$) 与同故障中某 $\mathbf{x}_{d,t,j}$ ($j \in 1, 2, \dots, n_{d,t}$ 且 $j \neq i$) 之间的距离 $d_{d,t,ij}$:

$$d_{d,t,ij} = \mathbf{x}_{d,t,i} \cdot \mathbf{x}_{d,t,j} \quad (6)$$

(2) 计算 $\mathbf{x}_{d,t,i}$ 与同故障中所有 $\mathbf{x}_{d,t,j}$ 的距离之和 $d_{d,t,i}$:

$$d_{d,t,i} = \sum_{j=1, j \neq i}^{n_{d,t}} d_{d,t,ij} = \sum_{j=1, j \neq i}^{n_{d,t}} \mathbf{x}_{d,t,i} \cdot \mathbf{x}_{d,t,j} \quad (7)$$

(3) 对所有 $d_{d,t,i}$ ($i \in 1, 2, \dots, n_{d,t}$) 进行降序排列, 得到降序向量 $\mathbf{D}_{d,t}$ 。

(4) 选择剔除系数 $k_{d,t}$ (整数), 将 $\mathbf{D}_{d,t}$ 中最后 $k_{d,t}$ 个元素剔除, 其余元素所对应的 $\mathbf{x}_{d,t,i}$ 即为二次清洗后的目标故障数据, 记为 $\Omega'_{d,t}$ 。

采用上述二次清洗方法处理辅助故障数据,记某台非待诊断变压器二次清洗后的辅助故障数据集为 ω'_a , 而全体非待诊断变压器二次清洗后的辅助故障数据集为 Ω'_a . Ω'_d 和 Ω'_a 中的元素均为变压器故障诊断特征集中的各状态量及与之对应的故障,令经过两次数据清洗后的 Ω'_d 和 Ω'_a 中的元素数量分别为 N_d 和 N_a , 记 Ω'_d 和 Ω'_a 中变压器故障诊断特征集中的各状态量分别为 $\mathbf{x}_{d,i}$ 和 $\mathbf{x}_{a,i}$, 对应的故障类型分别为 $y_{d,i}$ 和 $y_{a,i}$, Ω'_d 和 Ω'_a 的形式如式(8)所示:

$$\begin{cases} \Omega'_d = \{(\mathbf{x}_{d,1}, y_{d,1}), \dots, (\mathbf{x}_{d,i}, y_{d,i}), \dots, (\mathbf{x}_{d,N_d}, y_{d,N_d})\} \\ \Omega'_a = \{(\mathbf{x}_{a,1}, y_{a,1}), \dots, (\mathbf{x}_{a,i}, y_{a,i}), \dots, (\mathbf{x}_{a,N_a}, y_{a,N_a})\} \end{cases} \quad (8)$$

4 基于迁移学习算法的变压器故障诊断

4.1 基于实例的迁移学习算法

迁移学习算法主要包括同构空间下基于实例的迁移学习、同构空间下基于特征的迁移学习以及异构空间下的迁移学习^[18]。其中,基于实例的迁移学习算法 TrAdaBoost 具有更强的知识迁移能力。本文采用 TrAdaBoost 算法构建变压器的故障诊断模型。

基于实例的迁移算法示意图如图3所示。图3(a)中,当单台变压器的训练样本较少时,进行目标数据的故障分类十分困难,无法发现最优边界;图3(b)中,当有大量其他变压器的训练样本作为辅助数据时,则可以根据辅助数据确定目标数据的分类边界;图3(c)中,部分辅助数据会误导分类结果,导致分类边界出现偏差;TrAdaBoost 算法通过降低辅助数据的误分类样本权重,让分类器忽视不相关的标注样本,同时加强目标数据的误分类样本权重,让分类器重视分错的样本。图3(d)中,针对辅助数据和目标数据,采用不同的权重调整机制使得分类边界朝着正确方向移动^[19]。

4.2 变压器故障诊断

基于 TrAdaBoost 算法的变压器故障诊断方法如下:

(1) 设定 TrAdaBoost 算法迭代次数 $Iter$ 及基本分类算法 Learner, 训练数据集 $\Omega = \Omega'_d \cup \Omega'_a$, 测试数据集 S 。

(2) 初始化权重向量 $\mathbf{w}^1 = (w_1^1, w_2^1, \dots, w_{N_d+N_a}^1)$, 其中

$$\begin{cases} w_i^1 = 1/N_d & 1 \leq i \leq N_d \\ w_i^1 = 1/N_a & N_d + 1 \leq i \leq N_d + N_a \end{cases} \quad (9)$$

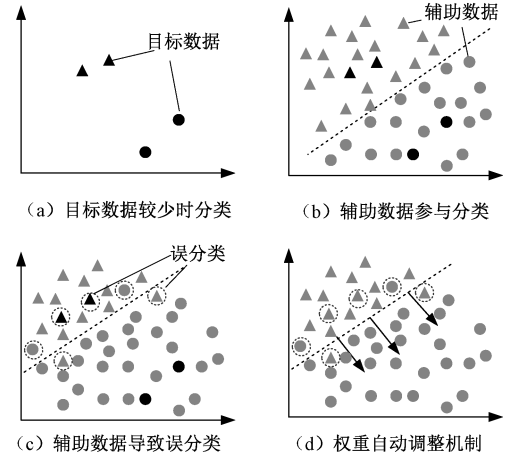


图3 基于实例的迁移算法示意图

Fig.3 Schematic of instance-based transfer learning

(3) 初始化参数:

$$\beta = \frac{1}{1 + \sqrt{2\ln(N_d + N_a)/Iter}} \quad (10)$$

(4) 迭代 $t = 1, 2, \dots, Iter$, 包括以下步骤:

1) 权重归一化:

$$p^t = \frac{w^t}{\sum_{i=1}^{N_d+N_a} w_i^t} \quad (11)$$

2) 调用 Learner, 根据 Ω, p^t 和 S 得到变压器故障弱诊断器 h_t 。

3) 计算弱诊断器 h_t 在 Ω'_d 上面的错误率:

$$e_t = \sum_{i=1}^{N_d} \frac{w_i^t (|h_t(\mathbf{x}_{d,i}) - y_{d,i}|)}{\sum_{i=1}^{N_d} w_i^t} \quad (12)$$

式中, $h_t(\mathbf{x}_i)$ 表示弱诊断器对 \mathbf{x}_i 得到的学习; y_i 表示正确故障类型。

4) 设置弱诊断器权重参数 λ_t 和目标权重调整参数 β_t :

$$\begin{cases} \beta_t = e_t / (1 - e_t) \\ \lambda_t = \ln(1/\beta_t) \end{cases} \quad (13)$$

5) 权重更新, 根据 e_t 值分配下一次迭代的权重, 即增加 Ω'_d 的权重, 减少分类错误的 Ω'_a 权重:

$$w_i^{t+1} = \begin{cases} w_i^t \beta_t^{-\text{sgn}(|h_t(\mathbf{x}_i) - y_i|)} & 1 \leq i \leq N_d \\ w_i^t \beta_t^{\text{sgn}(|h_t(\mathbf{x}_i) - y_i|)} & N_d + 1 \leq i \leq N_d + N_a \end{cases} \quad (14)$$

(5) 根据 $Iter/2 \sim Iter$ 次的分类结果, 输出变压器故障强诊断器:

$$H = \sum_{t=Iter}^{Iter/2} \lambda_t h_t \quad (15)$$

常见的分类算法 Learner 包括决策树法、最小邻居法、贝叶斯法、支持向量机法和神经网络法等。支持向量机对小样本数据的模式识别具有出色的学习泛化能力,考虑到本文的样本规模,选择支持向量机法作为 TrAdaBoost 算法中的 Learner。

5 算例分析

5.1 数据清洗结果

本文选取 35 台变压器的 674 条故障数据作为算例分析对象。选取其中一台变压器作为待诊断变压器,包括数据样本 81 组,其中 54 组作为初始目标故障数据,27 组作为测试集。

数据标准化后,分析 34 台辅助变压器与待诊断变压器故障数据的相似度,计算两组数据之间的 Tanimoto 系数。综合考虑数据样本数量和正迁移效应,迁移度阈值选择为 0.7,筛选出 Tanimoto 系数大于等于 0.7 的 21 台辅助变压器数据,完成一次清洗。辅助数据样本一共 377 组,分布情况如表 3 所示。

表 3 变压器故障样本一次清洗后分布情况

Tab.3 Distribution of transformer fault samples after first data cleaning

故障类型	辅助数据 Ω'_a	目标数据 Ω'_d	测试集 S
1	81	11	6
2	82	10	5
3	60	9	4
4	71	10	5
5	83	14	7
合计	377	54	27

对辅助数据和目标数据进行数据二次清洗,针对辅助数据分别选择 6、6、2、4、6 作为剔除系数,共筛选出 353 组数据。由于目标数据样本量较小,针对局部放电故障类型选择 2 作为剔除系数,其他故障类型不做筛选,共筛选出 52 条故障数据。辅助数据和目标数据的分布情况如表 4 所示。

表 4 变压器故障样本分布情况

Tab.4 Distribution of transformer fault samples

故障类型	辅助数据 Ω'_a	目标数据 Ω'_d	测试集 S
1	75	11	6
2	76	10	5
3	58	9	4
4	67	10	5
5	77	12	7
合计	353	52	27

分别选取 405 组故障数据中包含全部特征值($f_1 \sim f_{21}$)、关键气体容量($f_1 \sim f_5$)和常用比值法中的气体比值($f_{11} \sim f_{16}$)的样本,采用 Scikit-Learn 机器学

习框架对待诊断变压器的迁移学习诊断模型进行训练,诊断结果如图 4 所示。

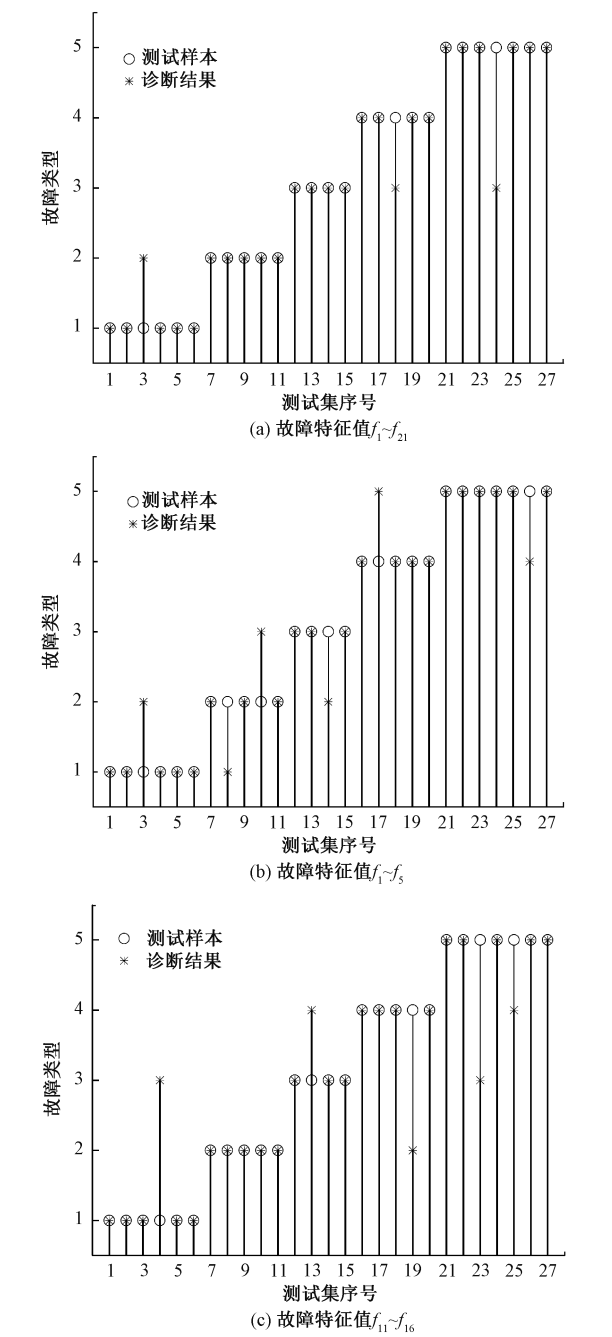


图 4 不同故障特征下测试样本与诊断结果对比
Fig.4 Comparison between test sample and diagnosis result with different fault features

本文所提方法对中低温过热、高温过热、低能放电、高能放电、局部放电五种故障类型的正判数分别为 5、5、4、4 和 6,正判率分别为 83.33%、100%、100%、80%和 85.71%,综合正判率为 88.89%。由图 4(a)可知,除低能放电故障外,其他四种故障类型均出现了一次误判。由图 4(b)可知,只采用关键

气体容量作为特征量时,在高温过热出现较多误判,综合正判率为 77.78%。由图 4(c)可知,只采用气体比值作为特征量时,在局部放电出现较多误判,综合正判率为 81.48%。

5.2 迁移学习对故障诊断的影响分析

为验证本文基于迁移学习算法的变压器故障诊断器的精确性,根据经过两次数据清洗后的目标故障数据和辅助故障数据,采用迁移学习算法对故障诊断器(记为 M1)进行训练;同时,基于待诊断变压器的故障数据,利用支持向量机基本分类器对变压器故障诊断器(记为 M2)进行训练。图 5 为不同迭代次数下的故障诊断精度。

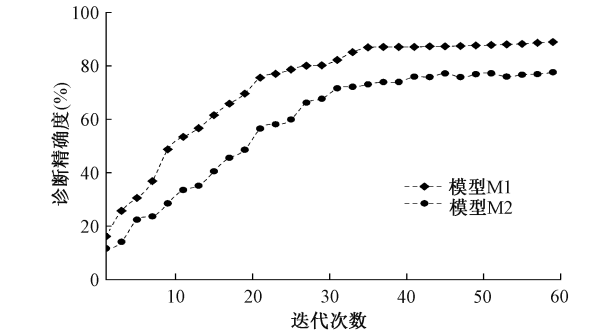


图 5 M1 和 M2 的故障诊断精度对比
Fig.5 Comparison of fault diagnosis accuracy between M1 and M2

从图 5 可以看出,随着迭代次数的增加,M1 的故障诊断精度始终高于 M2,且在迭代次数为 35 次左右时,两种模型的故障诊断精度趋于稳定。这是由于待诊断变压器的故障数据量较少,仅利用这些数据训练的故障诊断器 M2 特征知识图谱构建不完整,泛化能力较弱,对输入得到正确的分类结果相对不稳定。而 M1 借助了其他变压器的故障信息,并通过两次数据清洗使得辅助故障数据与其自身的故障数据相似度较高,用于训练的故障数据不仅在数量上得到了保障,并且在质量上也与待诊断变压器故障数据在同一水平。此外,M1 通过迁移学习不断对故障信息的权重进行迭代,使得对故障诊断精度有利的数据权重提高,而误导故障诊断的数据权重降低,最终提高了变压器故障诊断器的精度。

4.3 数据清洗对故障诊断的影响分析

为验证数据清洗对变压器故障诊断精确性的影响,根据迁移学习所利用的故障数据定义几种变压器故障诊断器,如表 5 所示。

表 5 变压器故障诊断器分类

Tab.5 Transformer fault diagnostics classification	
诊断器名称	诊断器描述
M1	进行数据一次和二次清洗
M1_1	只进行数据一次清洗
M1_2	只进行数据二次清洗
M1_3	不进行数据清洗

分别利用表 5 中的各种诊断器对变压器进行故障诊断,图 6 为不同迭代次数下的故障诊断精度。模型 M1 在不同的迭代次数下诊断精确度均最高;模型 M1_1 在迭代初期诊断精度比 M1_2 的诊断精度高,而 M1_2 最终稳定的诊断精度高于 M1_1,而 M1_3 的故障诊断精度始终最低。这说明数据的一次清洗和二次清洗均是必要的,两个步骤互为补充,共同起到剔除不良数据、保证故障诊断精确度的作用。

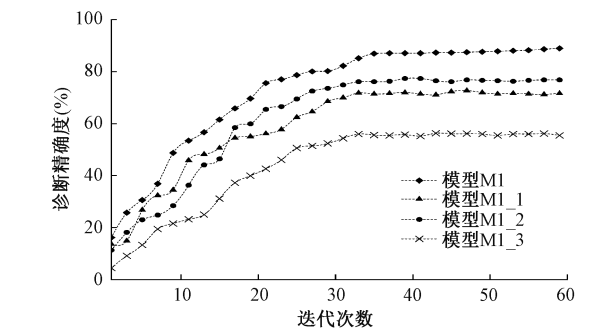


图 6 不同诊断器的故障诊断精度对比
Fig.6 Comparison of fault diagnosis accuracy with different diagnostics

6 结论

本文建立了一种基于数据清洗和知识迁移的变压器故障诊断模型,用于解决变压器单体故障数据少、总体数据繁而杂、诊断器泛化能力低的问题。通过理论推导和算例分析,可以得出以下结论:

- (1) 本文综合考虑气体含量和气体比值,提出的 21 种特征量能充分反映变压器故障状态的特征信息,提高诊断模型的故障辨识能力。
- (2) 数据的一次和二次清洗在变压器故障诊断中互为补充,均起到剔除不良数据、保证故障诊断精确度的作用,有必要在变压器故障诊断之前对故障数据进行清洗。
- (3) 迁移学习可有效提取辅助故障数据的有效信息,进而帮助变压器故障诊断器进行训练,达到提高故障诊断精确度的目的。

(4)本文辅助数据样本中故障类型均衡,下一步需研究故障类型不均衡时的迁移学习方法。

参考文献 (References):

- [1] 辜超, 杨祎, 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.
- [2] 吐松江, 高文胜, 张紫薇, 等 (Tu Songjiang, Gao Wensheng, Zhang Ziwei, et al.). 基于支持向量机和遗传算法的变压器故障诊断 (Power transformer fault diagnosis model based on support vector machine and a genetic algorithm) [J]. 清华大学学报 (自然科学版) (Journal of Tsinghua University (Science and Technology)), 2018, 58 (7): 623-629.
- [3] 程声烽, 程小华 (Cheng Shengfeng, Cheng Xiaohua). 基于改进粒子群算法的小波神经网络在变压器故障诊断中的应用 (Application of wavelet neural network with improved particle swarm optimization algorithm in power transformer fault diagnosis) [J]. 电力系统保护与控制 (Power System Protection and Control), 2014, 42 (19): 37-42.
- [4] 苗长新, 申坤 (Miao Changxin, Shen Kun). 基于优化神经网络和 DGA 的变压器故障诊断 (Power transformer fault diagnosis based on improved neural network and DGA) [J]. 高压电器 (High Voltage Apparatus), 2016, 52 (11): 163-168.
- [5] 苑津莎, 张利伟 (Yuan Jinsha, Zhang Liwei). 基于极限学习机的变压器故障诊断方法研究 (Study of transformers fault diagnosis based on extreme learning machine) [J]. 电测与仪表 (Electrical Measurement & Instrumentation), 2013, 50 (12): 21-26.
- [6] 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.
- [7] 苑津莎, 张利伟 (Yuan Jinsha, Zhang Liwei). 基于互补免疫算法的变压器故障诊断 (Fault diagnosis of transformers based on complementary immune algorithm) [J]. 电工技术学报 (Transactions of China Electrotechnical Society), 2015, 30 (24): 67-75.
- [8] 周淦, 孙超, 廖瑞金, 等 (Zhou Quan, Sun Chao, Liao Ruijin, et al.). 基于云理论的变压器多重故障诊断及短期预测方法 (Multiple fault diagnosis and short-term forecast of transformer based on cloud theory) [J]. 高电压技术 (High Voltage Engineering), 2014, 40 (5): 1453-1460.
- [9] 公茂法, 张言攀 (Gong Maofa, Zhang Yanpan). 基于 BP 网络算法优化模糊 Petri 网的电力变压器故障诊断 (Fault diagnosis of power transformers based on back propagation algorithm evolving fuzzy Petri nets) [J]. 电力系统保护与控制 (Power System Protection and Control), 2015, 43 (3): 113-117.
- [10] 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 BRDBN) [J]. 电力自动化设备 (Electric Power Automation Equipment), 2018, 38 (5): 129-135.
- [11] 兰华, 郑广瑞, 廖志民 (Lan Hua, Zheng Guangrui, Liao Zhimin). 基于模糊 TOPSIS 和 BP 网络的电力变压器故障诊断方法 (Fuzzy TOPSIS and BPANN based fault diagnosis method for power transformer) [J]. 南方电网技术 (Southern Power System Technology), 2011, 5 (1): 70-73.
- [12] 沈飞, 陈超, 严如强 (Shen Fei, Chen Chao, Yan Ruqiang). 奇异值分解与迁移学习在电机故障诊断中的应用 (Application of SVD and transfer learning strategy on motor fault diagnosis) [J]. 振动工程学报 (Journal of Vibration Engineering), 2017, 30 (1): 118-126.
- [13] 陈超, 沈飞, 严如强 (Chen Chao, Shen Fei, Yan Ruqiang). 改进 LSSVM 迁移学习方法的轴承故障诊断 (Enhanced least squares support vector machine-based transfer learning strategy for bearing fault diagnosis) [J]. 仪器仪表学报 (Chinese Journal of Scientific Instrument), 2017, 38 (1): 33-40.
- [14] 李春茂, 周妹末, 袁海满, 等 (Li Chunmao, Zhou Momo, Yuan Haiman, et al.). 基于 DGA 的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.
- [15] 代杰杰, 宋辉 (Dai Jiejie, Song Hui). 基于油中气体分析的变压器故障诊断 ReLU-DBN 方法 (Dissolved gas analysis of insulating oil for power transformer fault diagnosis with ReLU-DBN) [J]. 电网技术 (Power System Technology), 2018, 42 (2): 176-181.
- [16] 赵文清, 李庆良, 王德文 (Zhao Wenqing, Li Qingliang, Wang Dewen). 基于多模型的变压器故障组合诊断研究 (Combinational diagnosis of transformer faults

based on multi-models) [J]. 高电压技术 (High Voltage Engineering), 2013, 39 (2): 302-309.

[17] 胡军, 尹立群, 李振, 等 (Hu Jun, Yin Liqun, Li Zhen, et al.). 基于大数据挖掘技术的输变电设备故障诊断方法 (Fault diagnosis method of transmission and transformation equipment based on big data mining technology) [J]. 高电压技术 (High Voltage Engineering), 2017, 43 (11): 3690-3697.

[18] Weiss K, Khoshgoftaar T M, Wang D D. A survey of transfer learning [J]. Journal of Big Data, 2016, 3 (1): 9-49.

[19] 戴文渊 (Dai Wen Yuan). 基于实例和特征的迁移学习算法研究 (Instance-based and feature-based transfer learning) [D]. 上海: 上海交通大学 (Shanghai: Shanghai Jiao Tong University), 2009.

Transformer fault diagnosis model based on data cleaning and transfer learning

BAI Hao¹, WANG Yu-li²

- (1. Electric Power Research Institute, China Southern Power Grid, Guangzhou 510633, China;
2. Wuhan Branch, China Electric Power Research Institute, Wuhan 430072, China)

Abstract: Transformer safety is an important basis for ensuring the safe operation of distribution power system. This paper establishes a transformer fault diagnosis model based on data cleaning and transfer learning, which is used to solve the problem that single transformer fault data is small, the overall data is complicated and the generalization ability of the diagnosis model is low. Firstly, the Tanimoto coefficient is introduced to calculate the comprehensive similarity between the transformer to be diagnosed and other transformers, and the auxiliary fault data is cleaned once. Secondly, the target and auxiliary fault data are cleaned again by eliminating the fault data near the singular edge. Based on the data cleaning twice, the support vector machine is used as the basic classification learner of the transfer learning algorithm (TrAdaBoost). Through adjusting the weight of the target and the auxiliary fault data to transfer the effective information of the auxiliary fault data, the fault diagnosis model can be obtained. The results demonstrate that data cleaning and transfer learning can improve the efficiency of fault diagnosis and the accuracy of fault identification.

Key words: transformer; fault diagnosis model; transfer learning; data cleaning; TrAdaBoost algorithm