集成学习框架与知识蒸馏技术及其变压器故障识别的应用

余盛灿，余涛，冯淼永

（华南理工大学 电力学院，广东 广州 510641）

摘要：准确并快速地识别牵引变压器的故障类型是智能化运维的关键技术。针对目前传统算法中存在单一模型偏差以及复杂模型的迭代速率与部署计算资源之间的约束等问题，提出了一种基于Stacking集成学习框架的牵引变压器故障诊断模型，并融入知识蒸馏技术以压缩模型迭代时间来提高模型的计算性能。首先构造了由变压器油中气体指标组成的评估特征向量，然后基于Stacking集成学习框架将单一的Bagging与Boosting框架算法组合起来，并融入知识蒸馏技术实现对特征向量与故障类型的有效映射。在DGA数据样本中的实际泛化效果表明该方法解决了传统集成模型存在的偏差与方差问题，加快了集成模型的迭代速度，证明了模型的工程应用价值。

关键词：变压器故障诊断；Stacking框架；集成学习；知识蒸馏

中图分类号：TM28 **文献标识码：**A **DOI：**10.19457/j.1001-2095.dqcd25035

Ensemble Learning Framework and Knowledge Distillation Technology and Its Application in Transformer Fault Identification

YU Shengcan，YU Tao，FENG Miaoyong

（*School of Electric Power*，*South China University of Technology*，*Guangzhou* 510641，*Guangdong*，*China*）

Abstract：Accurately and quickly identifying the fault types of traction transformers is a key technology for intelligent operation and maintenance. Aiming at the problems of single model deviation in the current traditional algorithm and the constraints between the iteration rate of complex models and the deployment of computing resources，a traction transformer fault diagnosis model based on the Stacking ensemble learning framework was proposed，and incorporated knowledge distillation technology to compress model iteration time to improve the computational performance of the model. First，an evaluation feature vector composed of gas indicators in transformer oil was constructed，and then the single Bagging and Boosting framework algorithm were combined based on the Stacking integrated learning framework，and knowledge distillation technology was incorporated to realize the effective mapping of feature vectors and fault types. The actual generalization effect in the DGA data sample shows that this method solves the problem of bias and variance in the traditional integrated model，accelerates the iteration speed of the integrated model，and proves the engineering application value of the model.

Key words：transformer fault diagnosis；Stacking framework；ensemble learning；knowledge distillation

牵引电力变压器是高速铁路牵引供电系统中的重要设备之一。对变压器故障的准确判断和快速识别，可以有效地保证铁路供电系统的安全稳定运行。在数字化转型的背景下，在智能化运维牵引变电站中采用大数据与人工智能算法来提高变压器故障诊断的准确性具有重要意义[1]。

目前三比值法溶解气分析（dissolved gas analysis，DGA）仍然是最重要的故障诊断方法[2-3]，但在应用过程中存在漏诊、误诊等问题。针对此类问题，数据驱动方法可以帮助识别潜在的故障模式和提高预测准确性。首先是基于机器学习的方法，这种方法通过对大量的历史数据进行分

基金项目：中国博士后科学基金（2022M721184）

作者简介：余盛灿（1997—），男，硕士研究生，主要研究方向为电力设备智能化运维，Email：1834377961@qq.com

析，识别出电力设备故障的模式和趋势，从而实现准确的故障预测。例如，支持向量机算法（support vector machine，SVM）、随机森林算法（random forest，RF）、神经网络等机器学习算法被广泛应用于电力设备故障预测[4-5]。其次，基于深度学习的方法也被广泛应用，这种方法在机器学习的基础上，采用更深层次的神经网络结构，能够从数据中学习更高级别的特征，并在电力设备故障诊断和预测方面表现出更好的性能。例如，卷积神经网络和循环神经网络被广泛应用于电力设备故障预测。然而上述的单一模型方法仍存在算法收敛速度慢、容易陷入局部最优的缺陷[6-7]。

伴随着单一模型的缺陷，组合分类器进入黄金时代。近年来，集成学习作为一种组合多学习器的方法，已经在设备故障诊断领域得到广泛应用。最新研究中，常见的集成学习方法包括Bagging，Boosting，Stacking等[8-10]。Bagging的主要缺点在于样本之间的关联性，由于采样时有重复样本的存在，因此不能完全避免模型的方差问题。Boosting的主要缺点在于容易受到噪声和异常数据的影响，过拟合的风险较高[8]。Stacking集成框架与传统集成学习方法相比在减少模型方差和偏差方面表现出色，也能够减轻模型的过拟合问题。文献[11]表明由于Stacking集成学习算法将多个基学习器进行组合，可以更有效地利用各个基学习器之间的优点，在训练阶段时可以使用K折交叉验证，从而减少过拟合的风险。

现阶段集成学习算法在准确度和测试时间上难以平衡[12]，复杂模型导致部署的计算资源难以满足。虽然集成学习满足了模型上的精确度的提升，但是实际生产中难以满足方法所需算力迭代条件。文献[13]研究知识蒸馏（knowledge distillation，KD）概念，将结构复杂、推理性能优越的大型模型的知识迁移到小型模型，从而解决模型迭代的问题。经过知识蒸馏压缩模型后学习效果和迭代速率达到最优，仅需学生模型的训练时长，便可获得复杂模型迭代速率的提高，可见知识蒸馏在工程应用中有极高的价值[14-15]。

针对以上问题，本文主要贡献如下：考虑到模型的复杂度和迭代速率与部署算力资源之间的约束问题，同时需要保证算法精准度，本文提出了一种基于Stacking集成学习框架的变压器故障诊断算法。基于Stacking集成学习框架将单一的传统集成学习框架算法结合起来，解决了单一模型偏差问题，并融入知识蒸馏技术压缩模型，解决了复杂模型迭代速率问题，实现了设备故障类型的高精度快速识别。最后实际算例结果验证，本文提出的诊断模型在诊断精度和迭代速率的平衡性能上皆优于其他模型。

1　传统集成学习框架方法

当前，人工智能及机器学习技术的快速发展，SVM、神经网络等多种智能算法层出不穷，但是若采用了一种单独方式进行负荷预测，由于状态识别这类分类问题的假设空间很大，可能有多个假设在训练集上达到同等性能，若使用单一模型可能由于随机性而导致泛化性能不佳。因此，寻求使用组合分类器的方式，是进一步提高模型分类精度的必然选择，其中Bagging与Boosting算法是最为经典的集成学习框架。

Bagging是最为典型的一种并行集成学习框架。Bagging可分为自助采样与投票组合两个过程，其算法架构如图1所示。Bagging通过自助采样方式进行采样从而产生随机样本，假设数据集中有m个样本，每次随机有放回地采样1个样本，重复m次得到一个新含有m个样本的新数据集。于是，经过n轮自助采样法，就得到n个样本子数据集，每个数据集都含有m个样本，然后基于这n个数据集训练n个预测模型，最后通过投票结合策略将n个预测模型组合。基于Bagging架构的算法有应用较广的RF算法、提升树算法，贝叶斯集成算法等。

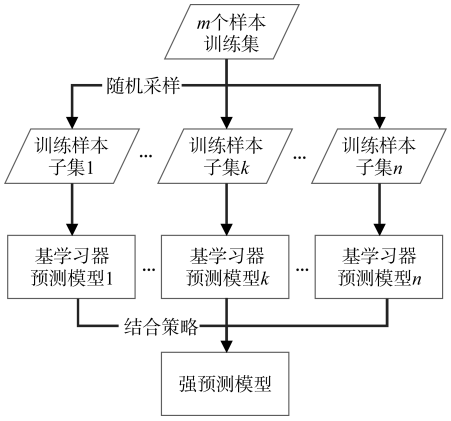

图1　Bagging集成学习框架

Fig.1　Bagging ensemble learning framework

Boosting集成学习框架如图2所示[8]，其基本思想如下：从训练集中进行子抽样组成每个基模型所需要的子训练集，对所有基模型生成的结果进行综合产生最终的分类结果。可见训练过程

为阶梯状,即串联模式,基模型按次序进行训练,实现上可以做到并行,基模型的训练集按照某种策略每次都进行一定的转化。对所有基模型分类的结果进行线性综合产生最终的预测结果。此类算法有GBDT(gradient Boosting decision tree)算法、XG-Boosting算法、Light-GBM(light gradient Boosting machine)等。

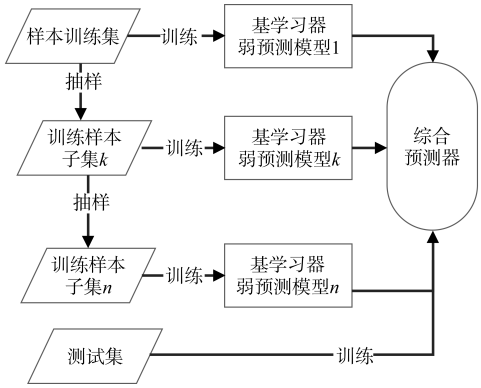

图2　Boosting集成学习框架

Fig.2　Boosting ensemble learning framework

总体来说,集成框架下算法模型的偏差和方差与基模型的偏差和方差息息相关。对于Bagging,整体模型的偏差和基模型近似,随着训练进行,整体模型的方差降低。对于Boosting,整体模型的初始偏差较高,方差较低,随着训练进行,整体模型的偏差降低;当训练过度时,因方差增高,整体模型的准确度反而降低,如表1所示。

表1　Bagging与Boosting框架的对比

Tab.1　Comparison of Bagging and Boosting frameworks

算法名称	训练方式	基学习器	相关性误差类型
Bagging	并行集成	强预测模型	弱相关性减小方差
Boosting	串行集成	弱预测模型	强相关性减小偏差

2　混合集成学习框架

基于上述分析,为了权衡统计学习类算法中的偏差与方差,同时考虑到Bagging集成学习算法与Boosting集成学习算法对于偏差与方差的敏感程度的差异,本文基于Stacking混合集成学习框架,将Bagging与Boosting两种集成学习方式有效结合,充分发挥两个模型优点,平衡单一模型导致的预测偏差,鲁棒性、泛化能力更强[10]。

Stacking集成学习框架首先将原始数据集划分成若干子数据集,输入到第1层预测模型的各个基学习器中,每个基学习器输出各自的预测结果。然后,第1层的输出再作为第2层的输入,对第2层预测模型的元学习器进行训练,再由位于第2层的模型输出最终预测结果。Stacking集成学习的流程伪代码如表2所示,具体训练方式如下:

1)对于数据集$S=\{(y_n,\boldsymbol{x}_n),n=1,\cdots,N\}$,其中$\boldsymbol{x}_n$为第$n$个样本的特征向量,$y_n$为第$n$个样本对应的标签值,$p$为所包含特征数量,即每一个特征向量为$(x_1,x_2,\cdots,x_p)$。随机将数据划分成$K$个大小基本相等的子集$S_1,S_2,\cdots,S_K$。其中$S_{-K}=S-S_K$,分别定义$S_K$和$S_{-K}$为$K$折交叉验证中的第$k$折测试集与训练集。对于第1层预测算法包含$k$个基学习器,对训练集$S_{-K}$用第$k$个算法训练得到基模型$L_k,k=1,2,\cdots,K$。

2)对于K折交叉验证中的第k折测试集S_K中的每个样本$\boldsymbol{x}_n$,基学习器L_k对它的预测表示为z_{kn}。在完成交叉验证过程后,将K个基学习器的输出数据构成新的数据样本,即

$$S_{\text{new}}=\{(y_n,z_{1n},z_{2n},\cdots,z_{kn}),n=1,\cdots,N\} \tag{1}$$

3)新产生的数据集就是Stacking第2层输入数据,使用第2层预测算法对这些数据进行归纳得到的元学习器L_{new}。Stacking的配置方式使得第1层的训练结果能够充分用于第2层算法的归纳过程当中,第2层算法能够发现并且纠正第1层学习算法中的预测误差来提升模型的精度。

表2　Stacking集成学习训练流程伪代码

Tab.2　Stacking ensemble learning training process pseudocode

Input:数据集$S=\{(y_n,x_n),n=1,\cdots,N\}$;
Step1:划分数据集为$S_1,S_2,\cdots,S_K$,并设$S_{-K}=S-S_K$;
Step2:训练第1层的基学习器;
For 1 to K:基于S_{-K}训练第一层的基学习器L_k;End
Step3:构建新的数据集:
$S_{\text{new}}=\{(y_n,z_{1n},z_{2n},\cdots,z_{kn}),n=1,\cdots,N\}$
Step4:基于S_{new}对第2层元学习器模型进行L_{new}的训练

3　知识蒸馏应用

为了减少复杂模型带来存储和计算资源需求的指数增长,研究采用知识蒸馏技术应用于Stacking集成学习框架。知识蒸馏是一种模型压缩技术[11],可以将较复杂的模型压缩成一个较简单的模型,同时保留原模型的预测能力。接下来,结合集成学习框架和知识蒸馏技术,以加速模型的迭代速度和提高预测精度。具体步骤如下:

1)基础模型Stacking模型作为教师模型f_t,将待训练的小型模型作为学生模型f_s,本文选择的RF算法是在传统决策树算法基础上应用统计抽样原理构造的集成学习算法,RF算法属于Bag-

ging框架，其计算复杂度低、迭代速度快。

2）通过教师模型将训练数据集进行预测，并将这些预测结果作为训练集的新特征，构成一个增强的训练集，使用增强的训练集对学生模型进行训练。在测试集上评估教师模型和学生模型的性能，并比较两个模型的精度和速度。知识蒸馏技术可以通过下式来实现：

$$L_{KD}(w)=\alpha L_{CE}[f_s(X,w),Y]+(1-\alpha)L_{CE}[f_t(X,w),Y_{soft}] \quad (2)$$

式中：w为模型的参数；X为输入特征；Y为真实标签；Y_{soft}为教师模型的软标签输出；L_{CE}为交叉熵损失函数；α为权衡两个损失函数的超参数。

在式(2)中，第一项是学生模型的交叉熵损失，第二项是教师模型的交叉熵损失，但是教师模型输出的是软标签，是一种概率分布，比硬标签更加平滑。这种平滑有利于学生模型更好地学习教师模型的知识。

在知识迁移的过程应用知识蒸馏于Stacking框架中以得到Stacking-KD模型，该框架如图3所示。

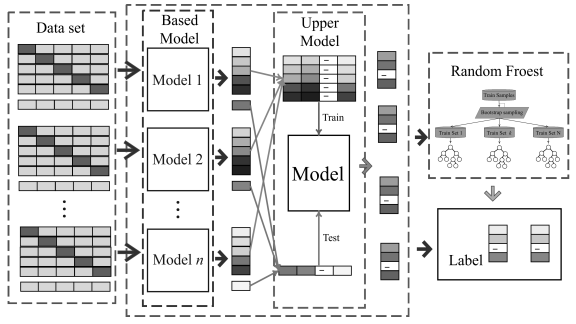

图3　Stacking-KD模型

Fig.3　Stacking-KD model

4　变压器故障识别模型

4.1　故障识别模型

本文提出基于Stacking集成学习框架与知识蒸馏技术的牵引变压器故障模式识别方法，整体的算法流程模型如图4所示，具体如下：

1）特征构建：本文根据牵引变压器油中溶解气体中特征向量的特点，选定One-Hot编码完成故障类型的离散数据编码处理，并设置Max-Min原则作为油中溶解气体的数据归一化方法。

2）样本划分：将数据集按照交叉验证准则划分为训练集和测试集。训练集用于训练识别模型，而测试集用于验证模型的准确性和泛化性。

3）训练分类器：将故障数据训练集输入到Stacking-KD模型中，设置好学习器类型和固定的模型参数，利用输出故障类型预测结果。

4）测试过程：训练好的Stacking-KD模型具有分类能力，利用训练完成的模型可进行变压器故障类型识别，从而测试算法准确性和迭代速度。

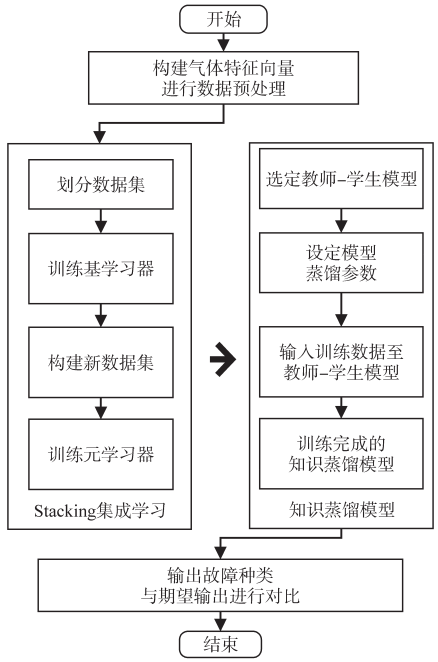

图4　变压器故障识别算法

Fig.4　Transformer fault identification algorithm

4.2　模型参数设置

4.2.1　设定集成学习参数

在选择基学习器时，需要考虑多个方面，例如模型的预测性能、模型的复杂度、模型的计算效率等。本文所研究的Stacking混合集成学习框架，选择RF，GBDT，XG-Boosting，Light-GBM这4种基于Bagging或Boosting的集成学习模型作为基学习器，选定决策树模型为元学习器，设定决策树的数量为60，每个决策树的最大深度为5，每个内部节点至少包含的样本数为2，最大训练数*Epoch*为100。RF和GBDT作为传统的集成学习模型，其基础模型基于决策树为主，已经被广泛应用于数据挖掘和机器学习领域，并且在各种比赛中都取得了不错的成绩，计算效率比较高，因为它们的基模型是决策树，而决策树可以通过并行计算来加速。XG-Boosting和Light-GBM是近年来非常流行的集成学习模型，具有出色的预测性能和速度。虽然采用了复杂的模型，如GBDT和神经网络，但是它们通过使用一些技巧，如正则化、特征采样等来控制模型的复杂度。这4种基学习器在实际应用中都有着较好的预测性能，

并且可以避免过拟合问题[10]。

4.2.2　知识蒸馏模型参数

在知识蒸馏中教师模型的输出经过一个softmax函数得到概率分布,而温度参数T可以控制概率分布的"平缓程度",即温度越高,概率分布越平缓。通过网格搜索法可得T=2时效果最优,较好地平衡了软标签和真实标签的信息分布。知识蒸馏中的损失函数由两部分组成,即对于硬标签的交叉熵损失和对于软标签的散度损失,而损失权重控制两个损失函数的相对重要性,因此设置α为0.5,最大训练数$Epoch$为50。

5　实例分析

5.1　特征选择

在变压器故障诊断领域,通过分析变压器油中产生的气体类型及其体积分数,可以确定是否发生了异常和特定的故障类型,因而在工程实践中具有广泛应用。为验证本文所提混合集成算法的有效性,从智能化运维示范工程牵引变压器的实际运行记录中选取1 000个故障DGA数据样本进行仿真测试,其中包括变压器的5种状态:高能量放电故障(HD)、低能放电故障(LD)、高温过热故障(HT)、低温过热故障(LT)和正常状态(NS)。部分测量样本及分布如表3所示,按照交叉验证4:1的比例,随机选取数据组成训练集和测试集。

表3　变压器DGA数据样本及其分布

Tab.3　DGA data samples and their distribution

故障类型	溶解气体浓度/(μL·L^{-1})					样本数量
	H_2	CH_4	C_2H_6	C_2H_4	C_2H_2	
HD	217.5	40	4.9	51.8	67.5	125
	⋮	⋮	⋮	⋮	⋮	
LD	345	112.25	27.5	51.5	58.75	304
	⋮	⋮	⋮	⋮	⋮	
HT	172.9	334.1	172.9	812.5	37.7	328
	⋮	⋮	⋮	⋮	⋮	
LT	181	262	210	528	0	199
	⋮	⋮	⋮	⋮	⋮	
NS	7.5	5.7	3.4	2.6	3.2	44
	⋮	⋮	⋮	⋮	⋮	

5.2　算例结果分析

本文除提出的Stacking-KD算法外,选取了经典的变压器故障诊断方法进行仿真,为了体现和参考引言内容相关的数据驱动发展研究历程,对照组设置了包括从早期的IEC故障诊断标准的三比值法、机器学习中的单一分类器SVM算法到组合分类器典型集成学习的RF算法等代表变压器故障诊断模型发展历程各个阶段的重要算法,进一步对比证明了本文方法的优越性和可行性,上述方法的参数设置如表4所示。

表4　对比算法模型参数设定

Tab.4　Comparison algorithm model parameter setting

算法	参数设定
Three-Ratio	具体参考IEC三比值法的编码规则
SVM	正则化参数C:1.0;核函数:RBF函数;$Gamma$:1/8
RF	决策树数量60;最大深度5

在具体的训练计算资源和训练平台中,本文提出的模型使用Python 3.6进行开发,采用的编译平台为Pycharm专业版,所提的基于Stacking架构和知识蒸馏模型主要采用目前较为流行的Pytorch深度学习框架实现。以上计算过程通过一台配备Intel(R)Xeon(R)Platinum 8160 CPU @ 2.10 GHz, GPU: 4×GeForce RTX 2080 Ti, RAM 125 G的服务器上完成模型的训练和测试。

图5显示了不同方法诊断结果的准确性。具体数据如表5所示。参与对比的方法有传统经典识别方法以及基于Stacking框架的混合集成学习模型,还有知识蒸馏模型压缩后的Stacking-KD模型。从结果上来看,Stacking模型在训练样本集和测试样本集上的诊断准确率要高于传统的识别方法。同时,Stacking方法在训练集和测试集上的性能都是最好的,对比Bagging集成学习框架下的RF模型而言,Stacking方法平均准确率要高出10.78%,可见采用元学习器来融合不同基学习器的预测结果,可以更充分地利用基学习器之间的差异性,避免单一集成模型的偏差影响,提高模型的性能。

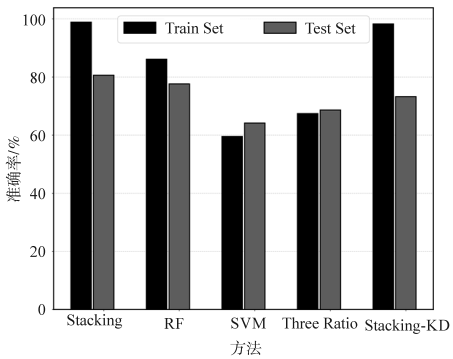

图5　训练集和测试集的准确率对比

Fig.5　Accuracy comparison between training set and test set

表5 训练集和测试集的准确率对比

Tab.5 Accuracy comparison between training set and test set

算法类型	准确率/%		
	训练集	测试集	整体
Stacking	98.87	80.59	95.21
RF	86.14	77.61	84.43
SVM	59.55	64.17	60.47
Three Ratio	67.41	68.65	67.66
Stacking-KD	98.28	73.26	93.26

此外，本文提出的Stacking-KD模型平均准确率与压缩前的Stacking模型较为接近，总体仍比传统方法的精度要高，在训练集上的准确率与压缩前的效果几乎相同，仅降低了0.59%，说明教师-学生模型在最优知识传递上达到了预期效果，但在测试集上降低较为明显，比压缩前降低了7.33%，表明模型在泛化性能上存在调整空间。

为了进一步评价不同方法的性能，本文计算了上述方法识别不同类型故障在训练集和测试集上的准确率，比较结果如图6和图7所示。由图6可以看出，训练集中的Stacking模型和Stacking-KD模型对任何状态样本的识别准确率都在96%以上，是所有的方法中准确率最高的。由图7可见，Stacking模型对HT，LT，LD的识别精度不是最优的，但从图5的结果来看，Stacking模型在测试集的平均表现的整体性能仍然是最好的，并且所有方法对LT和LD的识别都存在较大的误差，说明LT和LD之间的数据差异不明显，可以加入其他监测数据辅助分类。

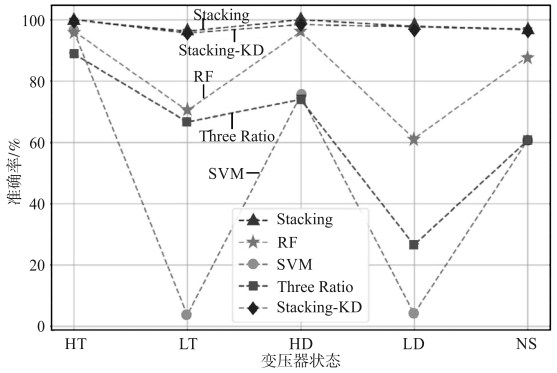

图6 训练集中故障类型准确率比较

Fig.6 Comparison of the accuracy of fault types in the training set

为了进一步从计算性能和迭代速率的指标上体现Stacking-KD的优越性，从训练测试时间和计算量上来对算法进行能效分析。由表6对比可知，Stacking框架的学习精度优于Bagging框架下的RF算法，并且可以看出经过知识蒸馏压缩模型后改进的Stacking-KD框架的学习效果和迭代速率达到最优，其训练和测试的平均准确率可以与Stacking模型非常接近，由此可见仅需单一学生模型的训练时长，便可获得复杂模型迭代速率的提高，因此，Stacking-KD模型在工程应用中有极高的价值，知识蒸馏可以在应用中大大减少模型空间和时间的消耗，可以在不增加计算资源的情况下加快算法迭代时间，以节省工程的计算资源。

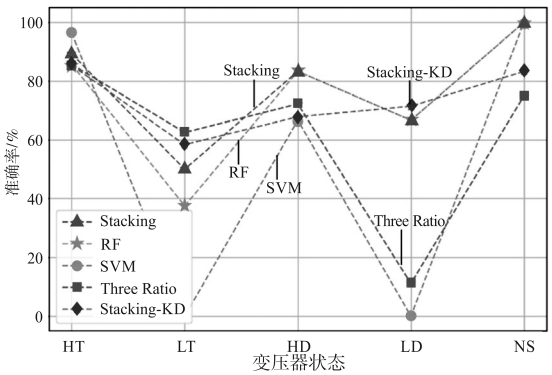

图7 测试集中故障类型准确率比较

Fig.7 Comparison of the accuracy of fault types in the test set

表6 不同算法计算性能对比

Tab.6 Computational performance comparison of different algorithms

算法类型	训练时间/（s·epoch^{-1}）	测试时间/（s·epoch^{-1}）	计算量/FlOPs	准确率/%
Stacking	0.48	0.16	45 308	95.21
RF	0.032	0.021	2 265	84.43
SVM	0.032	0.021	-	60.47
Three Ratio	0.022	0.005	-	67.66
Stacking-KD	0.22	0.010	2 265	93.26

对比组中将Staking方法与单一Bagging框架的典型集成学习算法RF算法进行了对比，说明Stacking框架解决了偏差与方差的问题。更为重要的是，知识蒸馏技术在复杂集成模型不增加计算资源的前提下，将教师模型的测试速率从0.16 s/epoch压缩到了0.01 s/epoch，可见Stacking-KD模型的迭代速率提升了16倍左右。因此，在保证了集成学习模型精确度不相上下的前提下，衡量优劣的指标就变成了迭代每次训练需要的时间，显然，立足于Stacking的基本框架，融入知识蒸馏技术的Stacking-KD模型在计算速度的横向对比中获得了更好的成绩。

6 结论

为了解决变压器故障诊断中传统统计学习类算法中存在的偏差与方差问题，同时解决传统算法处理大规模模型过大、参数冗余、测试耗时等问题，本文提出了一种基于Stacking集成学习

框架的变压器故障诊断模型,并融入知识蒸馏压缩模型技术应用,从而提高模型在实际应用中的效率和速度。对比该方法在DGA样本中的泛化效果,得出以下结论与展望:

1)本文提出的基于Stacking集成学习框架,将单一的Bagging与Boosting框架算法组合起来,并根据实际算例结果,证明此方法能够实现变压器故障类型的高精度识别,充分发挥两个模型优点,获得更稳定、泛化能力更强的模型。可见,多模型融合Stacking集成学习的技术路线值得探索,可成为各领域亟待研究的可行方向。

2)针对工程应用部署计算资源不足的问题,研究融入知识蒸馏将Stacking模型压缩为轻量级模型,减少了模型的存储和计算资源需求。由于教师模型已经学习训练集中的复杂特征,学生模型只需要关注更简单的特征,同时简单模型通常比复杂模型更容易进行调整和优化,可以更好地适应新的数据集和环境,从而提高模型在未知数据上的泛化能力,因此Stacking-KD模型具备较高的工程应用价值。

参考文献

[1] ENDRENYI J, ANDERS G J, LEITE da SILVA A M. Probabilistic evaluation of the effect of maintenance on reliability: an application to power systems[J]. IEEE Transactions on Power Systems, 1998, 13(2): 576-583.

[2] WELTE T M. Using state diagrams for modeling maintenance of deteriorating systems[J]. IEEE Transactions on Power Systems, 2009, 24(1): 58-66.

[3] ABEYGUNAWARDANE S K, JIRUTITIJAROEN P. New state diagrams for probabilistic maintenance models[J]. IEEE Transactions on Power Systems, 2011, 26(4): 2207-2213.

[4] 周爱国,王嘉立,杨思静,等. 基于K-means和K近邻的DPF设备故障分类算法[J]. 内燃机与配件,2019(12):57-59.
ZHOU Aiguo, WANG Jiali, YANG Sijing, et al. DPF equipment fault classification algorithm based on K-means and K nearest neighbors[J]. Internal Combustion Engines and Parts, 2019(12): 57-59.

[5] 田有文,于琳琳,杨镔. 基于支持向量机的电气设备运行状态图像识别方法研究[J]. 沈阳农业大学学报,2008(5):638-640.
TIAN Youwen, YU Linlin, YANG Bin. Recognition method of operation condition image of electricity equipments based on SVM[J]. Journal of Shenyang Agricultural University, 2008(5): 638-640.

[6] 吴海洋,陈鹏,郭波,等. 基于注意力机制和LSTM的电力通信设备状态预测[J]. 计算机与现代化,2020(10):12-16.
WU Haiyang, CHEN Peng, GUO Bo, et al. Attention and LSTM based state prediction of equipment on electric power communication networks[J]. Computer and Modernization, 2020(10): 12-16.

[7] 赵振兵,齐鸿雨,聂礼强. 基于深度学习的输电线路视觉检测研究综述[J]. 广东电力,2019,32(9):11-23.
ZHAO Zhenbing, QI Hongyu, NIE Liqiang. Research overview on visual detection of transmission lines based on deep learning [J]. Guangdong Electric Power, 2019, 32(9): 11-23.

[8] 张文生,于廷照. Boosting算法理论与应用研究[J]. 中国科学技术大学学报,2016(3):222-230.
ZHANG Wensheng, YU Tingzhao. Research on Boosting theory and its application[J]. Journal of University of Science and Technology of China, 2016(3): 222-230.

[9] 史佳琪,张建华. 基于多模型融合Stacking集成学习方式的负荷预测方法[J]. 中国电机工程学报,2019,39(14):4032-4042.
SHI Jiaqi, ZHANG Jianhua. Load forecasting based on multi-model by Stacking ensemble learning[J]. Proceedings of the CSEE, 2019, 39(14): 4032-4042.

[10] ZHOU Zhihua. Ensemble methods: foundations and algorithms [M]. New York: Taylor & Francis, 2012.

[11] 邓威,郭钇秀,李勇,等. 基于特征选择和Stacking集成学习的配电网网损预测[J]. 电力系统保护与控制,2020,48(15):108-115.
DENG Wei, GUO Yixiu, LI Yong, et al. Power losses prediction based on feature selection and Stacking integrated learning[J]. Power System Protection and Control, 2020, 48(15): 108-115.

[12] 侯慧,陈希,李敏,等. 一种Stacking集成结构的台风灾害下停电空间预测方法[J]. 电力系统保护与控制,2022,50(3):76-84.
HOU Hui, CHEN Xi, LI Min, et al. A space prediction method for power outage in a typhoon disaster based on a Stacking integrated structure[J]. Power System Protection and Control, 2022, 50(3): 76-84.

[13] PSRK W, KIM D, LU Y, et al. Relational knowledge distillation [C]//Computer Vision and Pattern Recognition, Los Angeles, United States: IEEE, 2019: 3967-3976.

[14] YANG C, XIE L, SU C, et al. Snapshot distillation: teacher-student optimization in one generation[C]//Computer Vision and Pattern Rccognition, Los Angeles, United States: IEEE, 2019: 2859-2868.

[15] YE H J, LU S, ZHAN D C. Distilling cross-task knowledge via relationship matching[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle WA USA: IEEE/CVF, 2020: 12396-12405.

收稿日期:2023-03-17
修改稿日期:2023-03-30