

基于 GAN 模型与随机森林算法的保护系统智能状态评价与预警

张 雷¹, 王光华¹, 曹 磊¹, 戴志辉², 寇博绰²

(1. 国网河北省电力有限公司保定供电公司, 河北 保定 071000; 2. 华北电力大学电力工程系, 河北 保定 071000)

摘 要:为提高继电保护系统状态评价的准确性, 提出一种基于 GAN 模型与随机森林算法的智能状态评价方法。首先, 结合现场情况与专家意见, 建立系统状态指标集, 并针对继保设备状态数据不平衡的问题, 提出基于生成对抗网络的状态数据生成方法; 然后, 建立基于随机森林的继保系统综合评价模型; 最后, 结合设备的历次状态评价结果, 给出设备的健康指数变化曲线及其劣化趋势, 提供相应的状态预警。基于真实数据的实验结果表明, 该方法能较准确评价系统状态, 对合理安排检修周期、制定检修计划具有参考价值。

关 键 词:继电保护; 状态评价; 状态预警; 生成对抗网络; 随机森林算法

DOI:10.19781/j.issn.1673-9140.2021.06.012 中图分类号: TM77 文章编号: 1673-9140(2021)06-0104-09

Smart status evaluation and early warning approach for highly-reliable protection systems based on GAN model and random forest algorithm

ZHANG Lei¹, WANG Guanghua¹, CAO Lei¹, DAI Zhihui², KOU Bochuo²

(1. Baoding Power Supply Company, State Grid Hebei Electric Power Co., Ltd., Baoding 071000, China;

2. Department of Electrical Engineering, North China Electric Power University, Baoding 071000, China)

Abstract: To improve the status evaluation accuracy of protection systems, a smart status evaluation method based on GAN model and random forest algorithm is proposed. Firstly, a system state indicator set is established in combination of the field conditions and expert opinions. To address the problem of the imbalance of relay protection equipment state data, a state data generation method is proposed based on the generation countermeasure network. Then, a comprehensive evaluation model of protection systems based on random forest is established. Finally, combining with the preceding state evaluation results, the health index curve of the equipment and its deterioration trend are given to provide corresponding state early-warning. The real-data experimental results show that this method can more accurately evaluate the system status, and has reference value for rationally arranging the maintenance cycle and formulating the maintenance plan.

Key words: protective device; state evaluation; state early-warning; generating adversarial network; random forest

随着泛在电力物联网的建设和发展,继电保护设备可获取的状态数据及类型在增加。但在数据丰富、评价指标不统一时,仍可能造成评价结果的差异^[1-4]。在泛在电力物联场景下,选取能全面表征继电保护系统状态的指标集、研究继电保护系统综合评价的计算机方法具有重要意义。

近年来,继电保护系统综合评价研究取得了长足发展^[5-10]。文献[5]针对影响智能变电站二次设备状态的因素,利用模糊综合评价法对二次设备进行状态评估;文献[6]将可靠性评估结果用于制定二次设备检修计划中;文献[7]提出一种基于模糊支持向量机的智能算法,根据继保装置的历史运行数据制定评价标准,并结合其当前运行数据进行在线评价,得到继保装置当前的运行状态;文献[8]提出了一种基于变权理论和梯形云模型的二次设备整体性能模糊综合评判方法,并采用层次分析法和变权理论实现评估指标权重的动态修正;文献[9]运用灰色聚类建立二次设备的状态评估模型,采用灰色预测扩充评估样本,结合层次分析法对二次设备状态进行评估;文献[10]采用主、客观权重相结合的方法,利用实时监测信息对二次设备健康状况进行综合评估。整体上,目前对继电保护状态评估方法的研究已较为成熟,但由于可靠性数据不足、缺乏对多源信息的综合利用等问题,导致目前多数基于大样本数据方法的评估结果还不能十分准确地反映保护设备的整体状态。

该文结合现场的实际情况与专家意见,建立继电保护系统状态指标集,并将继保设备的状态划分为“良好”、“正常”、“注意”、“危险”4个等级。由于处于“注意”与“危险”状态的样本数据所占比例极小,故设备的状态数据为严重不平衡数据,直接使用不平衡数据训练模型会影响评价模型的性能。针对这一问题,首先,利用改进的生成对抗网络生成大量的样本数据,经验证,该方法生成的数据与真实数据分布基本一致,可满足继保系统综合评价模型训练的要求;然后,建立基于随机森林的继保系统综合评价模型,将生成的样本数据输入到模型中进行训练,经真实样本验证,该方法得到的评价模型的正确率可以满足要求;最后,结合设备的历次状态评价结果

给出设备的健康指数变化曲线,再根据曲线判断设备状态及其劣化趋势,并给出相应的状态预警决策。

1 继电保护系统状态指标集

根据《继电保护状态检修导则》^[11],结合专家经验以及对现场实际情况的调研结果,选取用于反映继保系统状态的指标集,如表 1 所示,其主要包含装置本体、二次回路、记录型和指标型状态量,可综合反映继电保护设备的状态。其中,前 3 种状态量均可通过运行人员对设备进行运行巡视或专业巡视时采集得到;而指标型状态量可通过基础数据计算得到,用于反映该设备在同类设备中的相对品质优劣。

表 1 继电保护系统状态指标集

Table 1 State index set of relay protection system

状态量	指标
装置本体	运行环境、红外测温、装置采样、装置差流、通讯状况、通道运行情况
二次回路	红外测温、回路绝缘、锈蚀状况、接地网状况
记录型	家族性缺陷、缺陷发生频率、运行时长
指标型	平均失效率、可靠性、平均无故障时间

2 继电保护系统状态数据生成网络

2.1 生成对抗网络模型

以状态评价指标集为基础对保护设备进行综合评价,将评价结果分为“良好”、“正常”、“注意”、“危险”4个等级。由于保护设备本身的可靠性及变电站运维水平的不断提升,大部分保护设备工作于良好或正常状态,只在极少数次检查中出现设备状态为注意或危险的情况。因此,继保设备状态数据为严重不平衡数据。以某变电站为例,从其保护及二次回路专业巡检信息采集表中抽取 100 份,请专家进行状态评价。评价结果显示:7 份记录中的保护设备工作于良好状态,92 份记录中的设备处于正常状态,1 份记录中设备状态达到注意等级。若直接使用这些数据训练保护设备的状态评价模型,会更倾向于将被评价设备的状态分类为正常。因为即使将所有 100 份状态数据均分类为正常,模型的分

正确率也能达到 92%，这样的模型无实际意义。

为解决数据不平衡问题，提高模型的泛化能力，一般可采用过采样与欠采样的方式对训练数据进行调整。过采样通过对小样本重复采样达到扩充小样本容量的目的，同时，小样本中包含的噪声也会成倍放大，影响模型的训练效果；欠采样则是减少大样本的样本容量，达到平衡数据的目的，欠采样会抛弃大量数据，造成数据浪费。无论是过采样还是欠采样，都会增加模型过拟合的风险。

该文采用生成对抗网络产生大量继电保护设备状态数据样本，以解决不平衡数据带来的问题。

2.1.1 经典生成对抗网络

生成对抗网络 (generative adversarial nets, GAN)^[12-13] 是一类以零和博弈为理论基础的生成模型的统称，其生成样本的方式简单，不依赖样本的先验信息，理论上可完全逼近数据的真实分布，在数据生成领域受到广泛关注，其结构如图 1 所示。

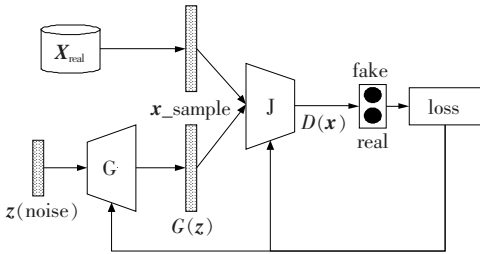


图 1 生成对抗网络结构

Figure 1 Generative adversarial network structure

GAN 的核心思想是构造 2 个神经网络：生成器 G 和判别器 J。生成器接受随机噪声向量 z 作为网络的输入，其训练目标是尽最大可能生成符合真实数据分布的样本；判别器的训练目标是尽最大可能识别输入数据来自真实样本还是生成样本。

训练生成对抗网络的过程就是不断交替训练生成器与判别器的迭代过程，即先固定生成器，训练判别器，再固定判别器，训练生成器，循环往复。在一次迭代内最优化判别器在计算上是十分困难的，且在样本容量有限的情况下易导致过拟合，一般采取每训练 k ($k \geq 1$) 次判别器之后，训练一次生成器的策略。当生成器与判别器的性能均足够好时，二者达到纳什均衡，判别器将无法区分被判别样本来自真实数据还是生成器生成的数据，此时，生成样本的分布与真实样本分布相同。

2.1.2 改进的生成对抗网络

为解决经典 GAN 模型存在的训练困难、结果不稳定、模式崩溃等问题，文献[14]提出了 WGAN (wasserstein GAN) 模型，使用 Wasserstein 距离代替 JS 散度来衡量 2 个分部之间的距离；文献[15]在此基础上又提出了 WGAN-GP (WGAN with gradient penalty) 模型，相比 WGAN，WGAN-GP 网络权重分布更均匀，训练过程更稳定，能更快收敛。而条件生成对抗网络 (conditional generative adversarial nets, cGAN)^[16] 解决了 GAN 生成结果不受控制的问题，通过调整标签变量 y ，即可改变生成结果的特征，通过调整输入到生成器中各标签的占比，即可改变生成结果中含有不同特征的样本的比例。故该文融合 WGAN-GP 与 cGAN 模型的结构，生成对抗网络，如图 2 所示。

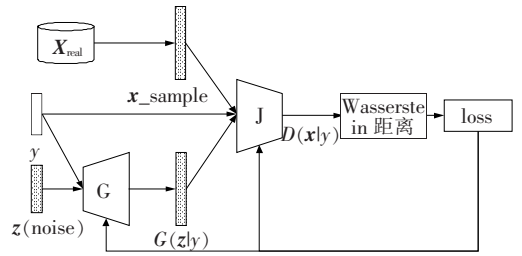


图 2 改进的网络结构

Figure 2 Improved network architecture

在图 2 中， z 是噪声向量，向量维数为 128 维。 X_{real} 为真实样本向量，向量维数为 33 维。 y 是样本标签，该网络中样本标签代表继电保护设备状态评价的结果，分为“良好”、“正常”、“注意”、“危险”4 类，用 4 维 one-hot 向量表示。

G 为生成器，包含 1 个输入层、2 个隐藏层以及 1 个输出层，2 层隐藏层分别含有 64、32 个节点，输出层含有 33 个节点，各层之间采用全连接结构。隐藏层的激活函数为 ReLU，输出层的激活函数为 sigmoid。生成器接受噪声向量 z 和样本标签 y 作为输入，输出结果 $G(z|y)$ 是标签为 y 的归一化的继电保护设备状态向量。

J 为判别器，包含 1 个输入层、2 个隐藏层以及 1 个输出层，2 层隐藏层分别含有 64、32 个节点，输出层含有 1 个节点，各层之间采用全连接结构。隐藏层的激活函数为 ReLU，输出层不使用激活函数。判别器接受样本标签 y 以及真实样本向量 X_{real} 或

生成器的输出 $G(z|y)$ 作为输入,输出结果用于组成训练生成对抗网络时使用的损失函数。训练模型时目标函数为

$$\min_{w \in W} E_{x \sim p_g} [f_w(x|y)] - E_{x \sim p_{data}} [f_w(x|y)] + \lambda E_{\hat{x} \sim p_{\hat{x}}} [(\|\nabla_{\hat{x}} f_w(\hat{x}|y)\|_2 - 1)^2] \quad (1)$$

式中 $x \sim p_g$ 为样本服从生成器拟合分布; $x \sim p_{data}$ 为样本服从真实分布; f_w 为满足 Lipschitz 连续条件的函数,可用含参数 w 的神经网络表示,即网络中的判别器; λ 为惩罚系数; $E_{\hat{x} \sim p_{\hat{x}}} [(\|\nabla_{\hat{x}} f_w(\hat{x}|y)\|_2 - 1)^2]$ 为梯度惩罚项; \hat{x} 为在样本空间中生成样本与真实样本之间连线上的随机采样值,即 $\hat{x} = \epsilon x_{data} + (1 - \epsilon)x_g, 0 < \epsilon < 1$ 。

损失函数为

$$\begin{cases} L_G = -E_{x \sim p_g} [f_w(x|y)] \\ L_J = E_{x \sim p_g} [f_w(x|y)] - E_{x \sim p_{data}} [f_w(x|y)] + \lambda E_{\hat{x} \sim p_{\hat{x}}} [(\|\nabla_{\hat{x}} f_w(\hat{x}|y)\|_2 - 1)^2] \end{cases} \quad (2)$$

其中,惩罚系数 λ 取值过大会使分类错误的惩罚偏重,导致模型过拟合,影响泛化能力; λ 取值过小会导致模型训练不充分,产生欠拟合。该文结合大量有关 WGAN-GP 模型的论文研究经验^[15],并经分析验证,取 λ 为 10。

2.2 生成数据质量评价

在生成对抗网络的经典应用场景中,网络的任务通常是生成图片、音频、文本等类型的数据,这类数据的质量可用“眼观耳闻”的方式评判。而该文继保设备状态数据为高维向量,人们对此类高维数据没有直观感觉,无法直接评判生成数据的质量。因此,采用人工法和非线性降维(t-distributed stochastic neighbor embedding, t-SNE)算法对网络生成的继保设备数据质量进行评价。t-SNE 法常用于高维数据的降维^[17],当数据维度降至 2~3 维时,即可将其分布情况以图像形式表现出来,该算法被广泛用于数据可视化领域。

首先,在网络生成的数据中随机抽取少量样本,请专家根据这些数据,结合《继电保护状态检修导则》中的相关规定评估设备的状态,将结果与样本的标签进行对比,验证生成数据的正确率。其次,利用

t-SNE 法对数据进行降维,并将降维结果以图形化的方式呈现,比较生成数据与真实数据的分布情况是否一致,以验证生成样本的合理性。

需要说明的是这里的降维主要是将高维数据的特征映射至二维,目的在于直观地可视化地了解数据的结构,证明:①在此种映射下,真实样本和生成样本的分布情况基本一致;②在高维空间中存在超平面,可以将 4 种状态的数据进行较好地分割。低维空间中存在的这种特征是高维空间中数据特征的反映,因此不会影响结论的正确性。

3 基于随机森林模型的保护系统状态预警

3.1 随机森林模型

随机森林模型结构如图 3 所示。样本集 D 中包含 n 个样本,记样本向量为 x ,样本标签为 y 。随机森林模型在样本集中进行 bootstrap 抽样,有放回地抽取 n 个样本,并将抽样结果作为新的样本集,记作 D_i ,没有被抽到的样本组成验证集,称为包外样本(out-of-bag),共抽取 m 次。之后,模型生成 m 棵决策树,每棵决策树选择一组样本集进行训练,从样本的全部属性 A 中随机抽取 a 个属性,每当决策树生成新的结点时,从 a 个属性中选择最优划分属性。在训练完成后,对于任意样本 x' ,每一棵决策树都对其有一个分类结果,记为 $h_i(x')$,随机森林模型输出 $H(x')$ 由所有决策树投票决定。

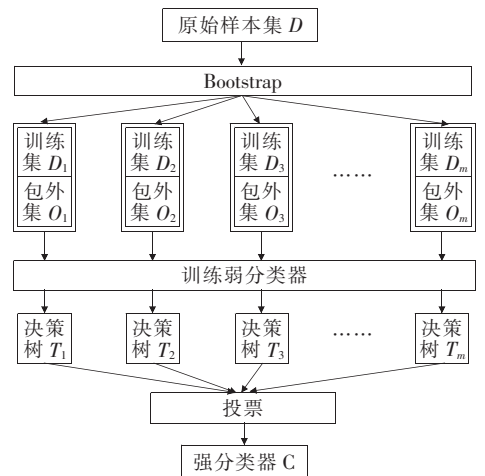


图 3 随机森林结构

Figure 3 Structure of random forest

3.2 继电保护系统状态预警

由 t-SNE 法对状态数据的评价结果可知,设备的状态差别越大,则设备的状态向量在空间中的距离越远。状态向量 x 距离 4 种分类中心的距离可用随机森林的分类概率值 p_i 表示, p_i 为向量 x 属于第 i 种分类的概率。 p_i 越大,表示 x 距离第 i 种分类中心的距离越近。

在应用大数据技术后,可以便捷地获取继保设备历次巡视时取得的状态数据记录及其评价结果,根据这些数据可得设备的状态变化趋势,进而对设备状态的异常状态及状态的异常变化给出预警。

结合随机森林模型的评估结果,定义继电保护设备的健康指数(health index)为

$$H_1 = \sum h_{ik} p_k \quad (3)$$

式中 p_k 为随机森林模型给出的设备处于第 k 个状态的概率, p_k 越大,表示设备的状态向量与第 k 个状态空间中心的距离越近; h_{ik} 为设备处于第 k 个状态空间中心时的健康指数,即 $p_k = 1$ 时设备的健康指数。

结合专家意见,该文将设备状态分为“良好”、“正常”、“注意”、“危险”4 种类型, h_i 取值如表 2 所示。也可以针对不同的应用场景灵活调整,对于重要程度较高的设备,应将告警定值适当提高,对于重要程度较低的设备,可适当放宽告警定值。

表 2 健康指数

Table 2 Health index value

状态	h_i	状态	h_i
良好	100	注意	60
正常	80	危险	0

为了反映继电保护设备状态变化的速度,定义设备劣化速度:

$$v_d = - \frac{dH_1}{dt} \quad (4)$$

式中 dH_1 为相邻 2 次状态评价的健康指数变化量; dt 为相邻 2 次状态评价的时间间隔。

当继电保护设备的健康指数低于告警定值,或继电保护设备的劣化速度超过告警定值时,将发出

预警,提醒工作人员注意设备的状态。设备状态预警流程如图 4 所示。

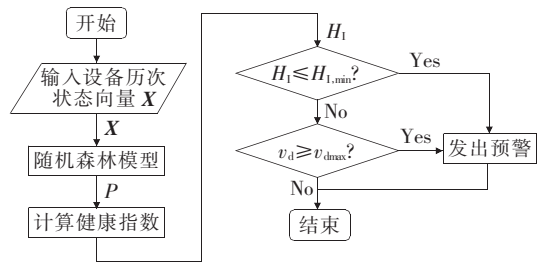


图 4 状态预警流程

Figure 4 Flow chart of status warning

4 算例分析

为验证算法的有效性并避免不平衡样本对评价结果的影响,在某供电公司的保护设备状态评价数据中,随机抽取处于“良好”、“正常”、“注意”、“异常”状态的样本数据各 64 份组成训练样本,总数据量 256 份。训练之前数据均经过归一化处理。若有充足的历史数据和产品家族数据可作为模型训练集,则尽可能采用,训练的模型性能会更优。

该文采用的归一化方案:①对于数据类型为连续型数据的指标,如红外测温、平均无故障时间等指标,将其转化为[0, 1]区间内的浮点数;②对于数据类型为离散型数据的指标,将其转化为[0, 1]区间内的离散值。如家族性缺陷指标,此项数据是根据设备家族性缺陷的严重程度给出的得分值,若设备不存在家族性缺陷,分值记为 1,若设备存在性质为一般、严重、危急的家族性缺陷,得分分别记为 0.5、0.3、0。

算例采用融合 WGAN-GP 和 cGAN 模型结构的生成对抗网络,网络结构如文 2.2 中所述。基于 Windows10 平台训练模型,所用软件包为 TensorFlow 1.13.1,Python 版本为 3.7。模型训练过程中交替训练生成器与判别器,每训练 5 次判别器,训练 1 次生成器。使用的 Adam 优化器参数设置如表 3 所示,模型训练过程如图 5 所示。

在 GAN 模型训练完成后,利用其生成大量的继保设备数据样本。首先,采用人工法对生成的继

保设备状态数据进行质量评价,该文从生成的数据中随机抽取少量样本,其中 4 类状态数据各 50 份,请专家对样本的状态进行评价,将专家评价结果与样本标签进行对比,结果如表 4 所示,可知采用改进生成对抗网络生成数据的平均正确率达到 91.5%,且没有发生设备分类误差超过一个等级的严重错误,因此生成数据的准确率满足要求。

表 3 优化器参数

Table 3 Optimizer parameters table

α	β_1	β_2	ϵ
1×10^{-5}	0.5	0.999	1×10^{-8}

令 $\lambda = 10, n_{critic} = 5$, 每次输入模型的训练样本容量 $m = 32$

开始

初始化判别器参数 ω_0 , 初始化生成器参数 θ_0 .

while θ 不收敛 do

for $t = 1, \dots, n_{critic}$ do

for $i = 1, \dots, m$ do

抽样真实样本 $\mathbf{x} \sim P_r$, 随机噪声向量 $\mathbf{z} \sim p(\mathbf{z})$.

生成随机数 $\epsilon \sim U[0, 1]$

$\hat{\mathbf{x}} \leftarrow G_\theta(\mathbf{z} | y)$, $\hat{\mathbf{x}} \leftarrow \epsilon x + (1 - \epsilon)\tilde{\mathbf{x}}$

$L^{(t)} \leftarrow D_w(\hat{\mathbf{x}} | y) - D_w(x | y) + \lambda(\|\nabla_{\hat{\mathbf{x}}} D_w(\hat{\mathbf{x}} | y)\|_2 - 1)^2$

end for

$\omega \leftarrow \text{Adam}(\nabla_\omega \frac{1}{m} \sum_{i=1}^m L^{(i)}, \omega, \alpha, \beta_1, \beta_2)$

end for

抽取一组随机噪声向量 $\{\mathbf{z}^{(i)}\}_{i=1}^m \sim p(\mathbf{z})$.

$\theta \leftarrow \text{Adam}(\nabla_\theta \frac{1}{m} \sum_{i=1}^m -D_w(G_\theta(\mathbf{z} | y)), \theta, \alpha, \beta_1, \beta_2)$

end while

结束

图 5 模型训练过程

Figure 5 Model training process

表 4 生成样本评价结果

Table 4 Generated sample evaluation results

样本标签	评价结果正确率/%			
	良好	正常	注意	危险
良好	98	2	0	0
正常	6	86	8	0
注意	0	8	84	8
危险	0	0	2	98

然后,采用 t-SNE 法对生成数据进行质量评价,从生成对抗网络生成的状态样本及真实样本中抽取一定量的数据并降至 2 维,所用软件包为 scikit-learn 0.20.2。t-SNE 参数设置如表 5 所示。

表 5 t-SNE 参数

Table 5 Parameter table of t-SNE

n_components	perplexity	n_iter	init
2	104	5 000	pca

降维后的数据分布如图 6 所示。观察样本分布可得如下结论:

1) 真实样本数据分布基本成线性排列,状态接近的继电保护设备在样本空间中的距离也更近,比如“良好”与“正常”状态的设备之间的距离小于“良好”与“注意”状态的设备之间的距离,这为继电保护设备的状态预警提供依据;

2) 生成样本与真实样本紧密混合,表示 2 种样本的分布情况基本一致,说明生成器已经学习到真实数据的分布特点,生成样本满足要求。

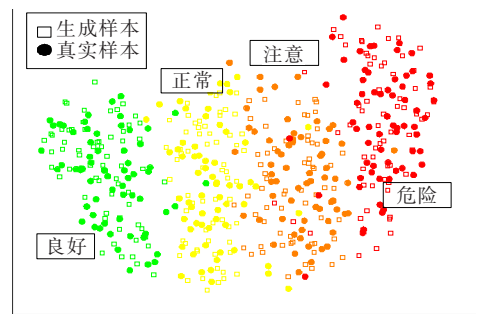


图 6 状态数据分布情况

Figure 6 State data distribution

以生成对抗网络生成的 4 096 份继保设备状态样本作为随机森林模型的训练集,其中“良好”、“正常”、“注意”、“危险”状态的样本各 1 024 份,所用样本数据均采用极值处理法进行归一化。

随机森林模型中包含的决策树的数量是一个超参数,为选择合适的值,该文采用包外误差作为衡量模型性能的指标,定义:

$$\epsilon^{oob} = \frac{1}{|D_o|} \sum_{(\mathbf{x}, y) \in D_o} I(H^{oob}(\mathbf{x}) \neq y) \quad (5)$$

式中 \mathbf{x} 为样本向量; y 为样本标签; D_o 为包外样

本集; $H^{oob}(x)$ 为模型对属于包外样本集的 x 的分类结果; $I(\cdot)$ 为指示函数。

包外误差随决策树数量变化曲线如图 7 所示, 可以看出, 当决策树的数量超过 300 时, 包外误差逐渐稳定, 该文取 450 作为决策树的棵数。

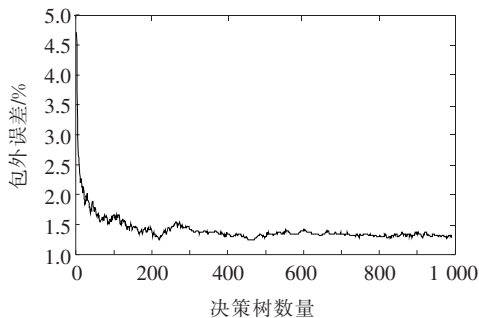


图 7 包外误差

Figure 7 Out-of-bag error

在训练完成后, 以供电公司提供的 256 份样本作为测试集来测试模型性能, 测试结果如表 6 所示, 可知设备状态评价模型的准确率达到 96.09%, 且没有发生严重分类错误, 可以满足继电保护设备状态评价的要求。

表 6 测试集评价结果

Table 6 Evaluation results of test set

样本标签	评价结果正确率/%			
	良好	正常	注意	危险
良好	98.44	1.56	0.00	0.00
正常	3.12	92.19	4.69	0.00
注意	0.00	3.12	93.76	3.12
危险	0.00	0.00	0.00	100.00

对于 220 kV 及以上变电站, 运行巡视要求每周至少 1 次, 专业巡视要求每半年至少一次, 每次巡视之后即可利用最新的数据更新评估结果, 每半年可进行一次全面的状态评估。若保护装置寿命为 10 年, 则在保护装置寿命周期内至少可进行 20 次全面的状态评估。对于 110 kV 及以下变电站, 运行巡视要求每月至少 1 次, 专业巡视要求每年至少一次, 则每年可进行一次全面的状态评估。

以华北地区某变电站的保护设备 A、B 为例, 通过向随机森林模型中输入该设备的状态量, 可得该设备的状态评价结果, 便可计算出设备的健康指数。

调取设备 A、B 历次状态评价结果, 如表 7 所示; 设备健康指数随时间变化的曲线如图 8 所示。

表 7 设备状态评价结果

Table 7 Equipment condition evaluation results

设备	日期	评价结果	健康指数
A	20150614	[0.016, 0.976, 0.009, 0]	80.22
	20160520	[0, 0.978, 0.022, 0]	79.56
	20170710	[0.013, 0.827, 0.160, 0]	77.06
B	20150613	[0.058, 0.750, 0.191, 0]	77.26
	20160519	[0.065, 0.700, 0.218, 0.016]	75.58
	20170709	[0.086, 0.723, 0.138, 0.052]	74.72

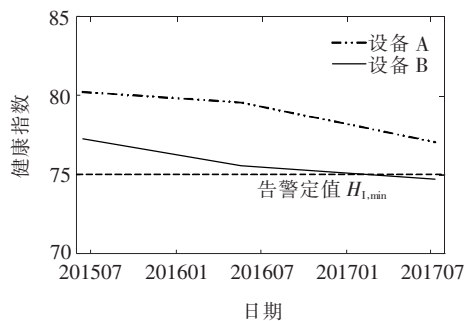


图 8 健康指数变化曲线

Figure 8 Health index change curve

由图 8 可知, 设备 A 健康指数高于最小告警定值且基本保持平稳, 没有发生较大变化, 因此, 不需要增加额外的关注。设备 B 的健康指数自 2017 年 2 月下旬开始低于最小告警定值, 应适当缩小巡视周期, 并合理安排检修计划, 评价结果与现场实际情况相符。

受从业年限、业务素质、自身经验等多种因素影响, 不同的巡视人员针对同一设备做出的评价结果往往存在差异, 应用该文模型可消除这种随意性。随着智能变电站的不断发展, 将利用先进巡检手段实时采集的设备状态输入到模型中, 即可实现状态的在线自动判别, 模型相对于人工评价的优势也将愈发明显。

5 结语

以电力物联为背景, 建立了综合反映继电保护

设备状态的评价指标集。构建了继电保护系统状态数据的生成网络,以此扩充样本容量,解决了以往继电保护设备状态信息样本数量不足导致评估模型性能较差的问题。其次,提供了一种以若干棵随机选择属性的决策树为基学习器的集成学习算法。相较于单棵决策树,随机森林算法可降低模型选择错误假设的风险,提高模型的抗干扰能力与泛化能力。较之其他算法,该算法应用在大数据情形下时计算开销更小,拥有更强的分类与泛化能力。

基于GAN模型、随机森林算法的继电保护系统状态评价和预警方法与常规的继电保护设备状态评价方法相比,不受状态样本数量较少的限制,能以较高的准确率综合评价继电保护设备的状态,并描绘出设备状态在一段时间内的变化趋势。对状态评价结果较差以及劣化趋势明显偏快的设备,模型将发出告警,提醒工作人员注意。与例行试验检测等方式相比,使用该方法可减轻工作人员的负担、提高工作效率,对合理安排检修周期、制定检修计划有现实意义。该文暂未考虑数据漏报的问题,但随机森林模型具有一定的抗野值能力,因此不受少量误报信息的影响。

参考文献:

- [1] 徐长宝,赵立进,高吉普,等. 基于灰色马尔科夫链的继电保护装置寿命研究[J]. 电力科学与技术学报,2019,34(3):114-119.
XU Changbao, ZHAO Lijin, GAO Jipu, et al. Research of relay protection device life cycle based on gray Markov mode[J]. Journal of Electric Power Science and Technology, 2019, 34(3): 114-119.
- [2] 周虎兵,王玉磊,张曼. 熵值法赋权的智能变电站二次设备效能模糊综合评价[J]. 电测与仪表,2018,55(3):73-79.
ZHOU Hubing, WANG Yulei, ZHANG Man. Comprehensive fuzzy evaluation of effectiveness of secondary equipment in smart substation based on entropy method [J]. Electrical Measurement & Instrumentation, 2018, 55(3): 73-79.
- [3] 黄炳翔,梁欣瑜,李伟,等. 基于全程状态监测量的继电保护故障率模型[J]. 电力科学与技术学报,2019,34(4):148-157.
HUANG Bingxiang, LIANG Xinyu, LI Wei, et al. Research on fault rate model of relay protection based on the whole process condition monitoring[J]. Journal of Electric Power Science and Technology, 2019, 34(4): 148-157.
- [4] 杨国生,戴飞扬,王文焕,等. 基于灰度关联法和TOPSIS法的继电保护状态评估综合算法研究与应用[J]. 中国电力,2019,52(2):94-103.
YANG Guosheng, DAI Feiyang, WANG Wenhuan, et al. Research and application of comprehensive algorithm of relay protection status assessment based on gray correlation analysis and TOPSIS method[J]. Electric Power, 2019, 52(2): 94-103.
- [5] 吴俊杰,陈程,程林,等. 基于模糊综合评价法的智能变电站二次设备状态评价研究[J]. 电测与仪表,2018,55(8):72-76.
WU Junjie, CHEN Cheng, CHENG Lin, et al. Study on state evaluation of secondary equipment in intelligent substation based on fuzzy comprehensive evaluation method[J]. Electrical Measurement & Instrumentation, 2018, 55(8): 72-76.
- [6] 郭升,童晓阳,王洪彬. 计及设备老化与计划检修的智能变电站保护系统可靠性评估方法[J]. 电力系统保护与控制,2019,47(9):97-104.
GUO Sheng, TONG Xiaoyang, WANG Hongbin. Intelligent substation protection system reliability assessment method considering the aging of equipment and planned maintenance[J]. Power System Protection and Control, 2019, 47(9): 97-104.
- [7] 曹海欧,张沛超,高翔. 基于模糊支持向量机的继电保护状态在线评价[J]. 电力系统保护与控制,2016,44(20):70-74.
CAO Haiou, ZHANG Peichao, GAO Xiang. Online condition evaluation of relay protection based on fuzzy support vector machine[J]. Power System Protection and Control, 2016, 44(20): 70-74.
- [8] 张友强,王洪彬,刁兴华,等. 计及保护失效的智能变

- 站二次系统综合风险评估研究[J]. 电力系统保护与控制, 2018, 46(22): 155-163.
- ZHANG Youqiang, WANG Hongbin, DIAO Xinghua, et al. Integrated risk assessment of intelligent substation secondary system considering the protection failure[J]. Power System Protection and Control, 2018, 46(22): 155-163.
- [9] 钱海, 贾松江, 杨飞, 等. 基于移动互联的继电保护设备智能运维技术研究[J]. 智慧电力, 2019, 47(11): 60-66.
- QIAN Hai, JIA Songjiang, YANG Fei, et al. Intelligent operation and maintenance technology for relay protection equipment based on mobile internet[J]. Smart Power, 2019, 47(11): 60-66.
- [10] Xu Changbao, Zhao Lijin, Gao Shuping, et al. State comprehensive evaluation of secondary equipment in an intelligent substation based on real-time information[C]//IEEE Transportation Electrification Conference and Expo, Guangzhou, China, 2018.
- [11] Q/GDW 1806—2013 继电保护状态检修导则[S].
- [12] 黄文琦, 许爱东, 明哲. 基于生成对抗网络的变电站工作人员行为预测的方法[J]. 南方电网技术, 2019, 13(2): 45-50.
- HUANG Wenqi, XU Aidong, MING Zhe. Prediction method for the behavior of substation staff based on generative adversarial network[J]. Southern Power System Technology, 2019, 13(2): 45-50.
- [13] 王守相, 陈海文, 潘志新. 采用改进生成式对抗网络的电力系统量测缺失数据重建方法[J]. 中国电机工程学报, 2019, 39(1): 56-64.
- WANG Shouxiang, CHEN Haiwen, PAN Zhixin. A reconstruction method for missing data in power system measurement using an improved generative adversarial network[J]. Proceedings of the CSEE, 2019, 39(1): 56-64.
- [14] Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks[C]//34th International Conference on Machine Learning, Sydney, NSW, Australia, 2017.
- [15] Gulrajani I, Ahmed F, Arjovsky M. Improved training of wasserstein gans[C]//Advances in Neural Information Processing Systems, Long Beach, California, USA, 2017.
- [16] 何子庆, 聂红玉, 刘月. 基于条件梯度 wasserstein 生成对抗网络的图像识别[J]. 计算机测量与控制, 2019, 27(6): 157-162.
- HE Ziqing, NIE Hongyu, LIU Yue. Image recognition with conditional wasserstein generative adversarial networks with gradient penalty[J]. Computer Measurement & Control, 2019, 27(6): 157-162.
- [17] 王振浩, 杜虹锦, 李国庆. 基于 t-分布邻域嵌入的同调机群无监督识别[J]. 电力系统保护与控制, 2018, 46(22): 64-71.
- WANG Zhenhao, DU Hongjin, LI Guoqing. Unsupervised identification of coherent generators based on t-distributed stochastic neighbor embedding[J]. Power System Protection and Control, 2018, 46(22): 64-71.