

基于 WGAN-GP 的风电机组传动链故障诊断

滕伟¹, 丁显^{2,3}, 史秉帅¹, 徐进^{2,3}, 袁帅¹

(1. 电站能量传递转化与系统教育部重点实验室(华北电力大学), 北京市 102206;

2. 中国绿发投资集团有限公司, 北京市 100020; 3. 都城伟业集团有限公司, 北京市 100020)

摘要: 传动链负责将风电机组叶轮的能量传递至发电机,若传动链中的任一部件,如齿轮、轴承发生异常,风电机组将面临巨大的安全隐患。现有基于深度学习的风电机组故障诊断大多需要人为选择目标变量,所识别故障与所选变量关联性大、通用性不足。梯度惩罚 Wasserstein 生成对抗网络(WGAN-GP)采用 Wasserstein 距离作为量度生成数据与真实数据的代价函数,具有训练结果稳定的优势。文中基于数据采集与监控(SCADA)系统提出两步数据预处理方法进行数据筛选,并基于 WGAN-GP 设计风电机组传动链异常状态分数,进而识别传动链故障。所提方法运用通用 SCADA 参数,无须人为挑选目标变量,可稳定识别风电机组传动链中的非特定故障,具有识别结果准确、泛化能力强等优点。9 台双馈风电机组的状态识别结果验证了所提方法的有效性,可以辅助指导风电场的运行维护。

关键词: 风电机组; 传动链; 梯度惩罚 Wasserstein 生成对抗网络(WGAN-GP); 数据采集与监控(SCADA)系统; 故障诊断

0 引言

风能具有清洁、可再生的特点,在中国新能源产业中占据着重要地位,目前国内安装的风电机组逾 10 万台。风电机组传动链由风轮主轴、增速齿轮箱、发电机轴、各种轴承、联轴器和制动器等部件组成,负责将主轴的机械能传递到发电机^[1]。风电机组常运行于恶劣的环境,工况复杂多变,传动链中的齿轮、轴承等部件很容易出现故障,影响正常的能量传递,造成发电量损失。严重情况下会造成齿轮箱、发电机等关键部件损毁,产生巨大的运维费用^[2],例如:1.5 MW 齿轮箱部件故障若没有及时发现,任由其发展,可能导致整个齿轮箱报废,更换总费用达 150 万元;发电机轴承故障若及时发现并在失效之前更换,产生费用约 2 万元,但任由其发展可能导致发电机扫膛损毁,则更换总费用达 50 万元之多。这一问题在运维困难的海上风电场表现得更为突出。因此,开展风电机组传动链的健康状态识别具有重要的意义。

数据采集与监控(SCADA)系统是风电机组的标准配置,涵盖机组运行的风速、转速、电压、电流、温度、功率等信息,基于 SCADA 系统数据进行风电

机组故障识别与诊断目前有许多研究。文献[3]提出基于深度自动编码器的异常检测方法,提取 SCADA 系统变量之间的关系,用模型输入与输出的重构误差反映部件的健康状况。文献[4]中提出了一种处理原始 SCADA 系统数据的有效方法,在分析 SCADA 系统各参量相关性的基础上,提出了多项式拟合方法实现变工况下风电机组健康状态的定量评估。文献[5]中采用模糊神经网络、快速傅里叶变换和专家系统推理机对风电机组不同类型的故障进行诊断,运用决策融合技术对诊断结果进行优化。文献[6]利用深度自编码对 SCADA 系统数据进行训练,并选用自适应阈值作为风电机组齿轮箱故障检测的决策准则。一种流行的故障检测思路在于:选定目标变量和自变量,使用正常的 SCADA 系统数据训练网络模型,将在线数据输入训练充分的模型(例如深度神经网络(DNN)^[7]、支持向量回归^[8]、长短期记忆(LSTM)网络^[9]、马氏空间^[10]和人工神经网络^[11]等)中,测试残差视为风电机组故障报警的量度。

以上工作能够在一定程度上检测风电机组故障,但大多数方法需要根据故障类别人为确定自变量和目标变量,通用性不强。针对不同类型的故障需要,根据故障机理设计不同的模型及输入输出变量。然而,SCADA 系统监测参数与传动链部件并不直接关联,容易出现所确定的变量与故障匹配不

收稿日期: 2021-01-27; 修回日期: 2021-05-31。

上网日期: 2021-09-01。

国家自然科学基金资助项目(51775186)。

准确的问题,造成误诊断或漏诊断。生成对抗网络(generative adversarial network, GAN)是一种依靠生成器努力生成与真实信息真假难辨的虚假信息,而判别器将2种信息区分的新网络,可通过纳什均衡进行模型训练。然而,生成对抗网络在训练过程中需要优化KL(Kullback-Leibler)散度与JS(Jensen-Shannon)散度的差值,容易出现训练不稳定的问题,且由于KL散度的不对称性,生成器生成不真实的数据时,惩罚较高;而生成器生成数据缺乏多样性时,惩罚较低,导致生成信息出现模式混乱。由此,文献[12]提出根据判别器的输入对其梯度范数进行惩罚,即梯度惩罚 Wasserstein 生成对抗网络(Wasserstein generative adversarial network-gradient penalty, WGAN-GP),以获得稳定的训练模型和高质量的生成信息。

本文提出基于WGAN-GP的风电机组传动链健康状态识别方法,将已知健康机组的SCADA系统数据用于WGAN-GP模型的训练,构建基于丰富中间层表征的距离测度表示机组传动链的健康状态,针对训练数据误差自适应设计故障阈值。本文方法不依赖于对特定故障敏感的具体SCADA系统参数,能够稳定识别风电机组齿轮箱、发电机轴承故障下的异常状态,具有较好的泛化能力,该方法在某风电场的9台2 MW双馈式风电机组中得到了验证。

1 SCADA系统数据预处理

受测量噪声、并网政策等影响,SCADA系统数据通常呈现散乱、规律性差等特点。为减小数据挖掘的难度、提高网络训练的准确率,需对原始SCADA系统数据进行预处理。

1.1 限功率数据点的筛除

根据文献[13],风功率曲线可分为3个风速区域,划分如下。

区域1:停机或启动阶段。

区域2:最大风能捕捉阶段。即桨距角很小并固定,风电机组调速以适应风速并尽可能多地捕捉风能。

区域3:恒功率阶段。在该阶段,风速超过了产生额定功率的风速,风电机组必须限制捕获的风能以保持额定功率,从而使风电机组所受载荷不超限。通过控制桨距角可实现恒功率运行。

文献[14-15]建议只保留区域2中的SCADA系统数据,剔除区域1和区域3中的数据来进行模型构建,以使得模型更容易拟合数据。但机组在实际运行中,恒功率阶段在一天中占比较多,若剔除会使预

警结果并不完整。因此,本文保留机组正常运行的全部阶段(区域2和区域3)进行模型构建。

附录A图A1为原始风速-有功功率散点图,根据风电机组的运行机理,采用桨距角与功率之间的关系对机组数据进行初步处理。选取桨距角大于2且有功功率小于最大功率90%的数据进行删除,即删除限功率的散点值,如附录A图A2所示。

1.2 离群值的筛除

去除限功率数据后的SCADA系统数据仍存在部分离群值,为进一步筛除数据,采用“四分法”进行二次处理。将每个风速值对应的功率由小到大排序并划分成4个大小相等的区间。3个分割点对应的值分别记作 Q_1 、 Q_2 和 Q_3 。满足式(1)的数据 x_1 予以删除。

$$\begin{cases} x_1 < Q_1 - kI_Q \text{ 或 } x_1 > Q_3 + kI_Q & k \geq 0 \\ I_Q = Q_3 - Q_1 \end{cases} \quad (1)$$

参数 k 的值由风电场的经验公式确定,通常为1.5。筛除离群值之后的数据如附录A图A3所示。

1.3 输入参数的选择及归一化

根据风力发电原理,风速、转速、发电机功率、电压和电流是风电机组较为关键的参数,能够涵盖传动链的能量传递过程。因此,选择9个通用参数进行传动链健康状态评估模型的训练,如附录A表A1所示。

选择通用参数的目的在于排除特定故障对某种参数的依赖,通过对通用参数的深入挖掘,可以尽可能多地识别传动链的故障类别,实现故障预警。而所选参数存在量纲差异,需要在模型训练前进行归一化,如下所示。

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (2)$$

式中: x 为某个参数在某时刻的值; $x_{\max}$ 和 $x_{\min}$ 分别为风电机组对该参数所允许的最大值和最小值。

2 基于WGAN-GP的传动链健康状态识别

2.1 GAN模型

GAN模型的基本结构如图1所示^[16]。图中,生成器 G 通过迭代训练学习随机信号 z 到真实数据 x 的确定性变换;而判别器 D 在训练过程中区分输入的是真实数据 x 还是生成器的输出。生成器 G 和判别器 D 都是非线性映射函数。判别器 D 输出其输入是真实数据 x 的概率。

在训练过程中,生成器 G 试图生成与真实数据相仿的生成数据,以欺骗判别器 D ;而判别器竭力准确识别其输入是真实数据还是生成数据,两者形成反复的对抗,最终达到纳什均衡^[17]。

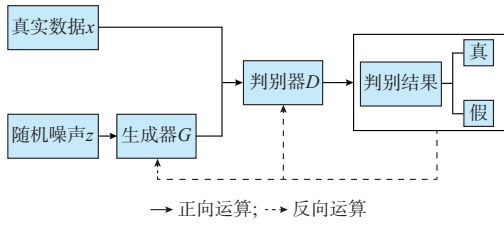


图1 GAN模型的基本结构
Fig. 1 Basic structure of GAN model

2.2 WGAN-GP的损失函数

GAN是一种无监督的神经网络,它基于最小化生成数据和真实数据之间的散度进行模型参数优化,其生成器的输入数据为高斯噪声,损失函数可以表示为:

$$L_G = -E_{z \sim p_z(z)} (D(G(z))) \quad (3)$$

式中: $E(\cdot)$ 表示求期望函数; $G(\cdot)$ 表示生成器函数; $D(\cdot)$ 表示判别器函数; $p_z(z)$ 表示噪声数据 z 的分布。

判别器的输入数据为真实数据或生成器输出的数据,损失函数可以表示为:

$$L_D = -\left(E_{x \sim p_{data}(x)} (\lg D(x)) + E_{z \sim p_z(z)} (\lg (1 - D(G(z)))) \right) \quad (4)$$

式中: $p_{data}(x)$ 表示真实数据 x 的分布。

建立以生成器与判别器对抗的代价函数 $V(D, G)$,可表示为:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} (\lg D(x)) + E_{z \sim p_z(z)} (\lg (1 - D(G(z)))) \quad (5)$$

根据文献[18],当判别器达到最优时,优化的代价函数用JS散度 D_{JS} 表示,即

$$L^* = 2D_{JS}(p_{data} \| p_g) - 2\lg 2 \quad (6)$$

式中: p_g 表示生成数据的分布。

在最优判别器的条件下,可以把GAN定义的生成器损失函数 L_G 等价于真实数据的分布 p_{data} 与生成数据的分布 p_g 之间的JS散度最小化。然而,由于高维分布中 p_{data} 和 p_g 重叠部分较少,难以通过优化JS散度将 p_g 拉向 p_{data} ,即会导致GAN的梯度消失和训练不稳定。

Wasserstein生成对抗网络(WGAN)模型通过最小化Wasserstein距离 $W(p_g, p_{data})$ 代替优化JS散度对真实数据和生成数据进行有效的距离度量^[19]。该距离与JS散度相比的优越性在于,即使2个分布之间没有重叠,依然可以反映2个分布的远近,此距离可视为将生成数据分布 p_g 传递到真实数据分布 p_{data} 的最小成本(该成本为传递量与传递距离的乘积)。理想状态下, $W(p_g, p_{data})$ 总是连续、可微的。

$W(p_g, p_{data})$ 可表示为:

$$W(p_g, p_{data}) = \frac{1}{K} \sup_{\|f\|_L \leq K} E_{x \sim p_{data}(x)} (f(x)) - E_{\tilde{x} \sim p_g(\tilde{x})} (f(\tilde{x})) \quad (7)$$

式中: $\sup$ 表示上确界,即最小上界; K 为Lipschitz常数; $\tilde{x} = G(z)$,表示生成的数据; $\|f\|_L \leq K$ 表示函数 f 满足 K -Lipschitz连续,函数 f 可以用神经网络拟合,通过限制神经网络中所有参数不超过某确信范围使得 K 存在。

根据判别器的输入直接约束判别器输出的梯度范数,实现Lipschitz约束,可弥补WGAN的不足。为简化此约束并提高训练速度,对WGAN判别器输入 $\hat{x} \sim p_{\hat{x}}$ ($p_{\hat{x}}$ 为 $\hat{x}$ 的分布,其中 $\hat{x}$ 为真实数据和生成数据采样点连线上的均匀采样)的梯度范数进行惩罚,得到WGAN-GP的损失函数如下。

$$L = E_{\tilde{x} \sim p_g(\tilde{x})} (D(\tilde{x})) - E_{x \sim p_{data}(x)} (D(x)) + \rho E_{\hat{x} \sim p_{\hat{x}}} \left(\left(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1 \right)^2 \right) \quad (8)$$

式中: ρ 为常数,推荐取10。

WGAN-GP中生成器的损失函数为 $V(G) = 1 - D(G(z))$ 。在网络的训练过程中,随机优化器Adam^[20]通过迭代减小 L 和 $V(G)$ 来分别优化判别器和生成器的参数。

2.3 基于WGAN-GP的健康状态识别方法

风电机组运行工况复杂、故障类型多样,机组的健康状态难以通过某种失效模式下已训练好的模型进行识别。为此,本文在WGAN-GP框架下,融合反卷积/卷积形成生成器和判别器,只采用健康时期的SCADA系统数据进行模型训练,以识别测试数据中的风电机组传动链异常。识别流程如图2所示,可解释如下。

1)将所得的数据进行预处理。

2)选择机组投运初期(健康时期)的SCADA系统数据进行WGAN-GP模型的训练。在训练过程中,生成器的输入数据为高斯噪声,生成器将标准正态分布噪声转换为与真实数据分布相似的数据,称之为伪数据;判别器的输入为生成器输出的伪数据(即与真实数据分布相似的数据)和真实数据,判别器努力区分真实数据和伪数据。

3)按 $1 - D(G(z))$ 和式(8)分别建立生成器和判别器的损失函数,进行生成器和判别器的对抗训练,训练充分后,判别器各层输出能够反映SCADA系统数据的本质特征。受文献[21]启发,在测试阶段,将生成数据和真实数据在判别器中某一层的输出 y_1 和 y_2 之差的绝对值作为机组的健康指标,用以

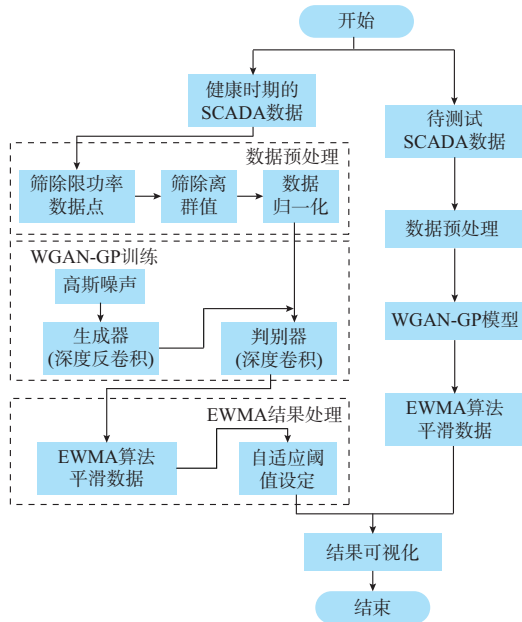


图2 识别方法流程图

Fig. 2 Flow chart of proposed identification method

量化风电机组传动链的故障程度。其中, y_1 和 y_2 的距离测度 θ 可表示为:

$$\theta = |y_1 - y_2| \quad (9)$$

4) 风电机组运行工况多变, 采用距离的绝对值作为异常状态的测度具有较大随机性。在此, 将该距离按指数加权移动平均 (exponentially weighted moving-average, EWMA) 控制图进行距离平滑, 可得 t 时刻故障异常分数 v_t 为:

$$v_t = \lambda v_t + (1 - \lambda) \theta_t \quad (10)$$

式中: θ_t 为 t 时刻 y_1 和 y_2 的距离测度; $\lambda \in (0, 1]$ 为 EWMA 控制图中异常分数的权重, 建议设置为 0.2^[22]。

v_t 的均值 μ_{v_t} 和方差 $\sigma_{v_t}^2$ 可表示为:

$$\begin{cases} \mu_{v_t} = \mu_\theta \\ \sigma_{v_t}^2 = \frac{\sigma_\theta^2}{n_s} \left(\frac{\lambda}{2 - \lambda} \right) [1 - (1 - \lambda)^{2t}] \end{cases} \quad (11)$$

式中: μ_θ 为一定时期内 θ 的平均值; σ_θ 为该时间段内 θ 的标准差; n_s 为 θ 值的数量。

5) 健康状态时, EWMA 控制图的上限和下限可分别表示为:

$$U_{CL} = \mu_\theta + h\sigma_\theta \sqrt{\frac{\lambda[1 - (1 - \lambda)^{2t}]}{2 - \lambda}} \quad (12)$$

$$L_{CL} = \mu_\theta - h\sigma_\theta \sqrt{\frac{\lambda[1 - (1 - \lambda)^{2t}]}{2 - \lambda}} \quad (13)$$

式中: h 为常数, 取值为 1.5。

式(12)中的 U_{CL} 为机组健康状态的上限。待测

试数据同样经过两步数据预处理, 输入由健康数据训练好的 WGAN-GP 模型, 计算得到的 θ_t 经 EWMA 算法平滑之后得到 v_t , 当 v_t 超过 U_{CL} 时, 即可判断机组传动链的健康状态是否出现异常。

3 案例分析

3.1 数据来源与机组介绍

应用上述方法对中国山东某风电场 9 台 2 MW 双馈风电机组传动链进行异常检测。机组的编号分别为 2、3、6、21、22、23、46、47、48。采集 9 台机组的 SCADA 系统数据, 采样频率为 10 min, SCADA 系统监测参数如附录 A 表 A1 所示, 数据长度为 3 年。21 号、22 号、23 号机组的采样时间为 2015-05-01 至 2018-05-15, 其余 6 台机组的采样时间为 2015-09-01 至 2018-09-01。其中, 2 号机组齿轮箱出现行星轴承磨损, 在 2018 年 4 月 14 日进行了更换。2 号机组发电机非驱动端轴承出现故障, 在 2018 年 3 月 17 日进行了更换。46 号机组的故障类型为齿轮箱高速轴的后轴承故障, 于 2018 年 5 月 1 日进行了更换。3 号、6 号、21 号、22 号、47 号、48 号是正常机组, 分别是 2 号、23 号和 46 号故障机组的对照。

3.2 SCADA 系统数据预处理结果

2 号机组的预处理结果见附录 A 图 A1 至图 A3。其余 8 台机组按本文提出的两级预处理方法处理前后的对比如附录 A 图 A4 至图 A11 所示。可以看出, 本文所提预处理方法能够有效去除由限功率等引起的离群点, 为后续传动链的故障识别提供合格数据。

3.3 基于 WGAN-GP 的故障识别结果

用于传动链健康状态识别的 WGAN-GP 模型如图 3 所示, 生成器的输入为 1×100 的张量, 来自标准正态分布的随机采样。生成器包含 4 个隐藏层, 神经元个数分别为 256、128、64 和 1。前 2 个隐藏层是全连接层, 后 2 个隐藏层是一维反卷积, 卷积核的大小为 5, 并且在前 3 层中采用 LeakyReLU 激活函数, 最后一层为 Tanh 激活函数。生成器输出与实际数据维度 1×9 相同的伪数据。判别器的输入是生成数据或真实数据。判别器包含 4 个隐藏层, 神经元个数分别为 64、128、256 和 1。前 2 个隐藏层是一维卷积层, 卷积核的大小为 5, 后 2 个隐藏层是全连接层, 其中前 3 层的激活函数为 LeakyReLU 函数, 由于 WGAN-GP 模型的结构特点, 最后一层没有激活函数。

每台机组训练一个健康状态识别模型, 经过预处理后的 SCADA 系统数据分成训练集和测试集。认为机组投运的前两年数据来自健康机组, 作为

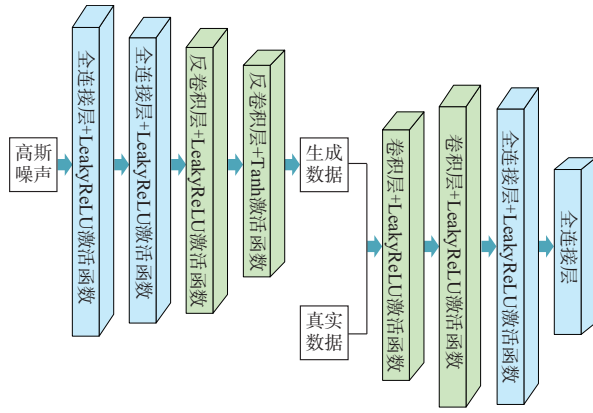


图3 用于传动链健康状态识别的WGAN-GP模型
Fig. 3 WGAN-GP model for health status identification of wind turbine drivetrain

WGAN-GP模型的训练集,后续的数据作为测试集。图4为随着训练次数的增加2号机组对应的生成器与判别器损失的变化过程。可以看出,随着迭代过程的进行,损失函数收敛于稳定值,指引线所指数字为在该损失下生成数据和真实数据的误差值(误差的计算方式为:在数据归一化的条件下,生成数据与真实数据的均方误差)。可以看出,随着损失趋于平稳,误差值稳定在0.07~0.08附近。

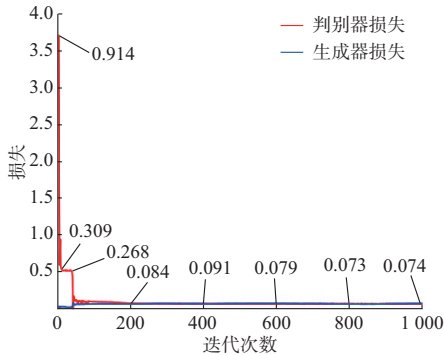


图4 生成器损失和判别器损失的变化
Fig. 4 Changes in generator loss and discriminator loss

图5(a)和(b)为网络训练10次的风速-功率和风速-主轴转速的真实数据与生成数据的对比图,图5(c)和(d)为网络训练1000次的风速-功率和风速-主轴转速的真实数据与生成数据的对比图。可以看到,与真实数据相比,生成器可以生成视觉上逼真的伪数据。当损失收敛的时候,生成数据已经逼近真实数据。根据图4中随着训练次数增加模型损失及误差的变化情况,本文采用1000次训练后的模型作为最终模型。

运用WGAN-GP模型计算每台风电机组每天所有 θ 值的平均值,然后基于平均值绘制EWMA算法控制图,得到每台风电机组传动链的健康状态识

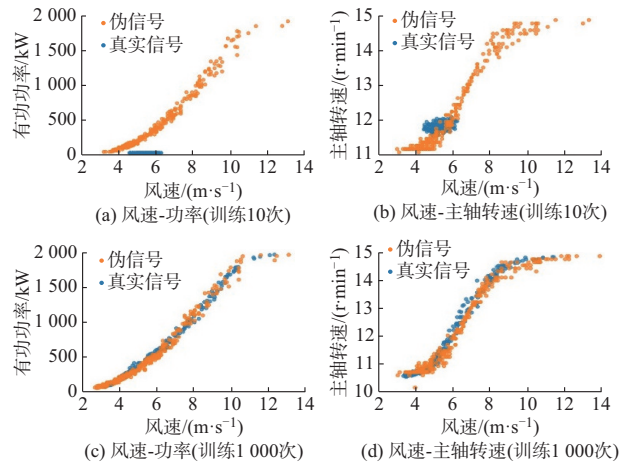


图5 真实信号和伪信号的散点对比
Fig. 5 Comparison of scatter points for real and fake signals

别结果。各机组传动链的检测结果如附录B图B1至图B9所示。

在附录B图B1中,蓝色实线在2018-04-14之前超过了上限,实现了2号风电机组传动链的故障预警。与2号机组同组的3号和6号机组未发生故障,在它们各自的预警图中,测试数据都没有超过阈值,如附录B图B2和图B3所示。

在附录B图B6中,蓝色实线在2018-03-14之前超过了上限,实现了23号机组传动链的故障预警。与23号机组同组的21号和22号机组未发生故障,在各自的预警图中,测试数据都没有超过阈值,如附录B图B4和图B5所示。

在附录B图B7中,蓝色实线在2018-05-01之前超过了上限,实现了46号机组传动链的故障预警。与46号机组同组的47号和48号机组未发生故障,在各自的预警图中,测试数据都没有超过阈值,如附录B图B8和图B9所示。

附录B图B1至图B9的结果表明,采用本文提出基于WGAN-GP的方法可以成功在线识别传动链故障机组,在维修之后,表征健康状态的异常分数回归阈值之内。6台健康机组的异常分数始终处于阈值范围内,没有出现误报警的现象。

附录C图C1为故障机组的拆解图片,其中2号机组齿轮箱中行星轮出现轴向窜动,行星轮与行星架碰磨,该故障由行星轴承打滑所致;23号机组发电机非驱动端轴承内圈滚道剥落;46号机组齿轮箱高速轴后轴承内圈滚道出现剥落。故障图片分别对应了附录B图B1、图B6、图B7的健康状态异常识别结果,验证了本文方法的有效性。

3.4 与其他方法的结果对比

本节基于DNN和生成对抗网络对9台机组进

行分析,以便对本文方法的有效性形成对比。其中,DNN采用附录A表A1中的转速作为目标变量,其余8个变量作为输入变量,共4层,每层神经元个数为8、100、100、1,使用Adam优化器,学习率为0.01。GAN的参数与本章案例中WGAN-GP参数相同。分析结果如附录D图D1至图D9所示,由于每台机组异常分数的幅值并不相同,为将每台机组的2种方法展示在同一张图中,对异常分数和阈值进行了等比例缩放,使之具有同一阈值。

运用DNN方法进行分析,存在故障的2号机组并没有发现阈值超标,而健康的6号机组出现超阈值的故障预警现象,与实际的故障情况不符。在分析健康的21号和22号机组时,DNN没有出现分数超过阈值的情况,在存在故障的23号机组中,DNN的异常分数超过阈值,但与工单记录的故障时刻不符,属于误报警。运用DNN分析的46号、47号和48号机组,其异常分数值都超过训练阈值,无法有效给出针对故障机组的预警。运用GAN方法,几乎所有机组在测试阶段都出现超过阈值的情况,与机组实际的故障状态不吻合。可以看出,DNN和GAN方法的故障识别效果不理想,预警结果与实际状态差异较大,无法给出合理的运维指导意见。相比之下,WGAN-GP方法拥有稳定的检测效果,完全匹配机组的实际健康状态。

4 结语

针对风电机组故障率高、故障预警需求迫切的现状,本文基于WGAN-GP模型具有训练稳定的优势,提出针对风电机组传动链的健康状态识别方法。首先,提出两步SCADA系统预处理方法,筛选无规律的运行数据。然后,构建基于丰富中间层表征的距离测度表示机组传动链的健康状态,针对训练数据误差自适应设计故障阈值。最后,运用EWMA算法控制图将传动链健康状态的异常程度予以平滑表征。9台实际风电机组的识别结果表明:本文所提方法不依赖于具体的监测参数,能够从含通用参数的SCADA系统数据中识别出传动链的异常状态并给出故障预警,进而指导风电场的运行维护工作。

本文提出的研究方法适用于发现风电机组传动链的异常健康状态,进一步需要研究基于SCADA系统数据的故障定位方法以使得现场运维更具有针对性。

附录见本刊网络版(<http://www.aeps-info.com/aeps/ch/index.aspx>),扫英文摘要后二维码可以阅读网络全文。

参考文献

- [1] 申屠东华,何先照,卢江跃,等.大型风电机组传动系统支撑方式特性分析[J].装备制造技术,2019(1):129-133.
SHENTU Donghua, HE Xianzhao, LU Jiangyue, et al. Analysis of characteristics of transmission system support of large wind turbine [J]. Equipment Manufacturing Technology, 2019 (1): 129-133.
- [2] 李焱,朱才朝,陶友传,等.风电机组可靠性研究现状与发展趋势[J].中国机械工程,2017,28(9):1125-1133.
LI Yao, ZHU Caichao, TAO Youchuan, et al. Research status and development tendency of wind turbine reliability [J]. China Mechanical Engineering, 2017, 28(9): 1125-1133.
- [3] ZHAO H S, LIU H H, HU W J, et al. Anomaly detection and fault analysis of wind turbine components based on deep learning network[J]. Renewable Energy, 2018, 127: 825-834.
- [4] YANG W X, COURT R, JIANG J S. Wind turbine condition monitoring by the approach of SCADA data analysis [J]. Renewable Energy, 2013, 53: 365-376.
- [5] 任岩,张楷.风电机组传动链综合状态监测与故障诊断技术[J].排灌机械工程学报,2018,36(7):613-616.
REN Yan, ZHANG Kai. Integrated condition monitoring and fault diagnosis technology for wind turbine drive-train[J]. Journal of Drainage and Irrigation Machinery Engineering, 2018, 36(7): 613-616.
- [6] 刘辉海,赵星宇,赵洪山,等.基于深度自编码网络模型的风电机组齿轮箱故障检测[J].电工技术学报,2017,32(17):156-163.
LIU Huihai, ZHAO Xingyu, ZHAO Hongshan, et al. Fault detection of wind turbine gearbox based on deep autoencoder network [J]. Transactions of China Electrotechnical Society, 2017, 32(17): 156-163.
- [7] TENG W, CHENG H, DING X, et al. DNN-based approach for fault detection in a direct drive wind turbine [J]. IET Renewable Power Generation, 2018, 12(10): 1164-1171.
- [8] 梁颖,方瑞明.基于SCADA和支持向量回归的风电机组状态在线评估方法[J].电力系统自动化,2013,37(14):7-12.
LIANG Ying, FANG Ruiming. An online wind turbine condition assessment method based on SCADA and support vector regression[J]. Automation of Electric Power Systems, 2013, 37 (14): 7-12.
- [9] 黄荣舟,汤宝平,杨燕妮,等.基于长短期记忆网络融合SCADA数据的风电齿轮箱状态监测[J].太阳能学报,2021,42(1): 235-239.
HUANG Rongzhou, TANG Baoping, YANG Yanni, et al. Condition monitoring of wind turbine gearbox based on LSTM neural network fusing SCADA data [J]. Acta Energiæ Solaris Sinica, 2021, 42(1): 235-239.
- [10] JIN X H, XU Z W, QIAO W. Condition monitoring of wind turbine generators using SCADA data analysis [J]. IEEE Transactions on Sustainable Energy, 2021, 12(1): 202-210.
- [11] DAMESHGHI A, REFAN M H. Wind turbine gearbox condition monitoring and fault diagnosis based on multi-sensor information fusion of SCADA and DSER-PSO-WRVM method [J]. International Journal of Modelling and Simulation, 2019, 39(1): 48-72.
- [12] GULRAJANI I, AHMED F, ARJOVSKY M, et al.

- Improved training of Wasserstein GANs[EB/OL]. [2021-01-05]. <https://arxiv.org/pdf/1704.00028.pdf>.
- [13] KIM K, PARTHASARATHY G, ULUYOL O, et al. Use of SCADA data for failure detection in wind turbines[C]// Proceedings of ASME 2011 5th International Conference on Energy Sustainability, August 7-10, 2011, Washington, USA.
- [14] IGBA J, ALEMZADEH K, DURUGBO C, et al. Analysing RMS and peak values of vibration signals for condition monitoring of wind turbine gearboxes[J]. Renewable Energy, 2016, 91: 90-106.
- [15] 张帆,刘德顺,戴巨川,等.一种基于SCADA参数关系的风电机组运行状态识别方法[J].机械工程学报,2019,55(4):1-9.
- ZHANG Fan, LIU Deshun, DAI Juchuan, et al. An operating condition recognition method of wind turbine based on SCADA parameter relations [J]. Journal of Mechanical Engineering, 2019, 55(4): 1-9.
- [16] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.
- [17] CHAKERI A, SHEIKHOLESLAM F. Fuzzy Nash equilibriums in crisp and fuzzy games[J]. IEEE Transactions on Fuzzy Systems, 2013, 21(1): 171-176.
- [18] ARJOVSKY M, BOTTOU L. Towards principled methods for training generative adversarial networks[EB/OL]. [2021-01-05]. <https://arxiv.org/abs/1701.04862>.
- [19] BHATIA R, JAIN T, LIM Y. On the Bures-Wasserstein distance between positive definite matrices [J]. Expositiones Mathematicae, 2019, 37(2): 165-191.
- [20] HAN Z D. Dyna: a method of momentum for stochastic optimization[EB/OL]. [2021-01-05]. <https://arxiv.org/abs/1805.04933>.
- [21] SCHLEGL T, SEEBÖCK P, WALDSTEIN S M, et al. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery[EB/OL]. [2021-01-05]. https://www.researchgate.net/publication/318017139_Unsupervised_Anomaly_Detection_with_Generative_Adversarial_Networks_to_Guide_Marker_Discovery.
- [22] WANG L, ZHANG Z, XU J, et al. Wind turbine blade breakage monitoring with deep autoencoders [J]. IEEE Transactions on Smart Grid, 2016, 9(4): 2824-2833.

滕 伟(1981—),男,通信作者,博士,教授,主要研究方向:风电机组智能诊断及预测。E-mail:tengw@ncepu.edu.cn

丁 显(1983—),男,博士,主要研究方向:新能源场站的智能监控。E-mail:fd_dingxian@163.com

史秉帅(1994—),男,硕士研究生,主要研究方向:风电机组的数据挖掘与故障诊断。E-mail:bingshuai134@126.com

(编辑 鲁尔姣)

Fault Diagnosis of Wind Turbine Drivetrain Based on Wasserstein Generative Adversarial Network-Gradient Penalty

TENG Wei¹, DING Xian^{2,3}, SHI Bingshuai¹, XU Jin^{2,3}, YUAN Shuai¹

(1. Key Laboratory of Power Station Energy Transfer Conversion and System, Ministry of Education

(North China Electric Power University), Beijing 102206, China;

2. China Green Development Investment Group Co., Ltd., Beijing 100020, China;

3. Duchengweiye Group Co., Ltd., Beijing 100020, China)

Abstract: The drivetrain is responsible for the energy transfer from rotor hub to generator in wind turbines. If any part of the drivetrain, such as gears and bearings, is abnormal, the wind turbine will face a huge safety hazard. Now most of the current wind turbine fault diagnosis based on the deep learning need to select target parameters artificially, and the identified fault has a close correlation with the selected variables, resulting in insufficient versatility. Wasserstein generative adversarial network-gradient penalty (WGAN-GP) uses Wasserstein distance between the generated data and the real data as a measurement for the cost function, which has the advantage of stable training results. This paper proposes a two-step data preprocessing method for data screening based on the supervisory control and data acquisition (SCADA) system, and designs anomaly state score of the wind turbine drivetrain based on the WGAN-GP model to identify the drivetrain faults. The proposed method uses common SCADA parameters, does not need to manually select target variables, and can stably identify non-specific faults in the wind turbine drivetrain, which has the advantages of accurate identification results and strong generalization ability. The status identification results of nine doubly-fed wind turbines verify the effectiveness of the proposed method, which can assist in guiding the operation and maintenance of wind farms.

This work is supported by National Natural Science Foundation of China (No. 51775186).

Key words: wind turbine; drivetrain; Wasserstein generative adversarial network-gradient penalty (WGAN-GP); supervisory control and data acquisition (SCADA) system; fault diagnosis

