

文章编号: 2095-2163(2019)05-0326-05

中图分类号: TP391.9

文献标志码: A

基于监测数据的蓄能设备故障预测方法研究

曲晓峰¹, 苗东旭¹, 王树新², 王振羽², 甄宏³, 刘烁⁴

(1 哈尔滨电气集团有限公司, 哈尔滨 150028; 2 丰满大坝重建工程建设局, 吉林 132000;

3 哈尔滨汽轮机厂有限责任公司, 哈尔滨 150040; 4 哈尔滨工业大学 计算机科学与技术学院, 哈尔滨 150001)

摘要: 蓄能设备是发电厂削峰填谷的重要手段之一。针对发电厂蓄能设备故障种类和原因复杂, 提出一种基于运行数据和机器学习方法的蓄能设备故障预测方法。该方法以管理系统监控的历史数据和实时数据为基础, 首先对数据集进行特征提取和充放电周期识别, 然后建立随机森林分类器, 训练分类器模型参数, 以实现蓄能设备运行过程中是否存在故障以及故障类别的预测。在实际运行数据中验证了所提方法能够在故障发生的早期有效识别故障类型。

关键词: 蓄能设备; 故障预测; 机器学习; 支持向量机

Research on fault prediction of energy storage equipment based on monitoring data

QU Xiaofeng¹, MIAO Dongxu¹, WANG Shuxin², WANG Zhenyu², ZHEN Hong³, LIU Shuo⁴

(1 Harbin Electric Group Co., Ltd., Harbin 150028, China; 2 Fengman Dam Reconstruction Engineering Bureau, Jilin 132000, China; 3 Harbin Steam Turbine Factory Co., Ltd., Harbin 150040, China; 4 School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

[Abstract] Energy storage equipment is one of the key devices for peak cutting in a power station. Aiming at the complex types and causes of energy storage device faults in power station, a fault assessment method based on profile data and machine learning method is proposed. In order to predict the faults and their types in the operation of energy storage equipment, based on profile historical data and real-time data, the feature extraction and charge-discharge cycle identification of profile data sets are carried out first. Then a support vector machine classifier is established and trained to obtain the model parameters. The results on the actual operation data of energy storage device verify that the proposed method can effectively identify the fault type in the early stage of fault occurrence.

[Key words] energy storage device; fault prediction; machine learning; support vector machine

0 引言

蓄能设备的管理与诊断是蓄能设备应用技术中的核心之一,及时、准确地诊断蓄能设备故障、保证蓄能设备容量的准确估计,可以延长蓄能设备使用寿命、提高蓄能设备一致性及可靠性^[1]。在产生严重的故障之前,微小缓变故障往往提早发生,并进一步演变为严重故障,因此,及时检测出早期缓变故障对于能源管理与维护有着重要的现实意义。考虑到蓄能设备诊断的故障现象、故障原因及故障机理的复杂性和模糊性,难以借助确定的数学模型来描述,也难以借助确定性的特殊判据来诊断。在安全评估方法中,模糊理论、人工神经网络、基因遗传算法等方法被广泛应用^[2]。但是,由于微小缓变故障相比于严重故障信号特征不明显,浅层的机器学习

算法和一般的统计方法辨识这些故障十分困难^[3]。

故障定义为系统由于出现至少一个或多个参数、特性低于正常指标,导致了系统丧失正常使用功能的行为^[4]。故障诊断是通过各种能够检查和测试的方法,指对发生不正常状态的部位进行诊断,查明其功能失调的原因,对系统出现的异常情况进行判断,确定故障的原因、类型、大小、位置和发生时间^[5]。利用蓄能设备管理系统的监测数据来及时对蓄能设备组的多种故障进行预测,达到目的检测故障,提出设备事故预防措施是控制领域的研究热点^[6]。故障诊断技术的主要任务是能快速准确地对故障做出评价和决策,系统能及时判断故障的大小、位置和发生的时间。一般而言,故障检测容易且耗时少,而故障诊断由于要判断出系统发生故障的位置、时间原因等,而耗时长^[7]。

基金项目: 2017年智能制造新模式应用项目;水力发电设备智能远程运维新模式。

作者简介: 曲晓峰(1976-),男,硕士,高级工程师,主要研究方向:自动控制、工业大数据;苗东旭(1984-),男,硕士,工程师,主要研究方向:自动控制、工业大数据;王树新(1969-),男,高级工程师,主要研究方向:水电厂运行与维护。

通讯作者: 曲晓峰 Email: xiaofengqu@126.com

收稿日期: 2019-07-24

现有技术中,专家系统是一种通过现有知识来推理实际问题的程序,结合专家经验和知识求解那些复杂问题。现有的方法建立在专家库之上,根据输入的故障信息结合相应计算机算法进行推理分类,完成故障诊断和决策,准确度依赖于专家库的完善程度,即需要大量的工程经验知识,存在专家系统中容错率低,对于不确定信息的诊断准确率低、专家知识经验获取困难,以及维护难度大等问题^[8]。

神经网络主要是基于神经元而建立的一种数学模型,输入和输出信息之间通过这些神经元连接。神经网络模型由于其本身特性,对处理不确定性信息的诊断优势较大,但由于神经网络的训练需要大量的样本,大量的故障信息样本数据难以获得,通过结合模糊集理论或区域分区来减少样本数据获取的难度的研究还不够深入,不能很好地适用于蓄能设备故障诊断系统^[8]。

贝叶斯网络将概率论和图形理论相结合,用来推理和解决不确定性知识,在电网故障诊断中处理不确定性信息诊断效果良好^[9]。现有研究中电网故障诊断关于贝叶斯网络诊断方法的研究多结合粗糙集相关理论。贝叶斯网络节点赋值需要统计分析或实际观察确定,如何解决贝叶斯网络在处理复杂情况下的建模、网络训练等问题仍然存在困难^[10]。

借助系统的数学模型,利用一定的技术方法,让系统产生残差信号,再利用生成的残差信号应用相应的评价规则,或者直接让其与给定阈值进行比较,从而达到故障分类的目的。其中包括参数辨识法、状态估计法和等价空间法,主要缺点是依赖于系统的数学模型,数学模型的精确程度将影响故障诊断的效果。而等价空间法的缺点是大部分只能在线性系统中得以运用,因此在非线性系统中就难以发挥作用^[10]。

有向图法和故障树法,实质上就是利用故障种类和故障原因之间的因果逻辑关系对系统故障进行分类。大型复杂系统的故障诊断时,由于系统的故障种类繁多、故障相关性复杂等客观原因,导致其故障诊断准确性难以保证^[11]。

本文针对早期缓变故障特征不明显等特点,以多种机器学习技术为核心,以数据驱动方法为基础,利用数据采集管理系统采集的大量动力蓄能设备运行和EV工况数据,通过逐层特征提取进行安全评估。

1 数据集

本文采用实际运行的数据采集管理系统采集的

4组共3470191条有效数据为数据源,上述数据集具有如下特点:

(1)19个国家标准报警分类:温度差异报警,高温报警,蓄能设备类型过压报警,蓄能设备类型欠压报警,SOC低报警,单体蓄能设备过压报警,单体蓄能设备欠压报警,SOC过高报警,SOC跳变报警,可充电蓄能系统不匹配报警,蓄能设备单体一致性差报警,绝缘报警,DC-DC温度报警,制动系统报警,DC-DC状态报警,驱动电机控制器温度报警,高压互锁报警,驱动电机温度报警,蓄能设备类型过充。

(2)24个特征:设备状态,充电状态,运行模式,车速,累计里程,总电压,总电流,SOC,DC-DC状态,档位,加速踏板行程值,制动踏板状态,燃料蓄能设备电压,最高电压蓄能设备子系统号,最高电压蓄能设备单体代号,蓄能设备单体电压最高值,最低电压蓄能设备子系统号,最低电压蓄能设备单体代号,蓄能设备单体电压最低值,最高温度子系统号,最高温度探针序号,最高温度值,最低温度子系统号,最低温度探针序号。

2 分类器模型

2.1 支持向量机分类器

采用支持向量机实现单个分类器的故障分类功能,实现过程可表述如下。

SVM支持向量机的原理就是在不同分类的样本点中找到一个正确分割数据的超平面,并使得各分类支持向量距离此平面的距离最大。超平面方程如式(1)所示:

$$w^T x + \gamma = 0, \quad (1)$$

分类间隔 d 如式(2)所示:

$$d = \frac{|w^T \cdot x + \gamma|}{\|w\|}, \quad (2)$$

算法标记样本点为1和-1,如果超平面可以完美分割样本点,那么变量就会满足式(3)所示方程:

$$\begin{cases} w^T x_i + \gamma > 0, & y_i = 1; \\ w^T x_i + \gamma < 0; & y_i = -1. \end{cases} \quad (3)$$

由此可得,变量的约束条件如式(4)所示:

$$y_i(w^T x_i + \gamma) \geq 1 \quad \forall x_i, \quad (4)$$

最小化分类间隔即最大化其分母 $\|w\|$,所以可以得到如式(5)所示的目标函数:

$$\min_{\alpha} \frac{1}{2} \|w\|^2 \text{ s. t. } y_i(w^T x_i + \gamma) \geq 1 \quad i = 1, 2, \dots, n, \quad (5)$$

求解SVM的目标函数属于最优化问题,而这个

问题既属于凸优化问题、又满足 KKT 约束条件,通过拉格朗日对偶可以将优化问题转化为等价的对偶问题求解,将目标函数转换为求最小值问题,变形为式(6):

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^m \alpha_i s. t. \quad (6)$$

$$\sum_{i=1}^m \alpha_i y_i = 0; \alpha_i \geq 0 \quad i = 1, 2, \dots, n,$$

用 SMO 算法求解得到满足式(6)最小对应的 α 向量的值 α^* , 计算式(7)得到超平面权重系数 w^* :

$$w^* = \sum_{i=1}^m \alpha_i^* y_i x_i, \quad (7)$$

找出所有的 S 个支持向量,即满足 $\alpha_s > 0$ 对应的样本 (x_s, y_s) , 通过式(8)计算出每个支持向量 (x_s, y_s) 对应的 b_s^* , 即:

$$y_s \left(\sum_{i=1}^m \alpha_i y_i x_i^T x_s + b \right) = 1. \quad (8)$$

所有的 b_s^* 的平均值即为最终的 b^* 。将结果代入分类决策函数以及超平面公式得到模型训练结果。

2.2 逻辑回归分类器

逻辑回归分类器的样本输入是多维特征构成的特征向量 x , 输出为 0 或 1 表示样本属于负类或正类。逻辑回归从输入到输出需要用 sigmoid 函数完成映射, 函数的自变量就是多维特征加权向量, 代入后得到如式(9)的预测函数:

$$h_{\theta}(x) = g(\theta^T x) = \frac{1}{1 + \exp(-\theta^T x)}, \quad (9)$$

LR 分类器判断输入属于正类的概率可以用式(10)表示:

$$P(y = 1 | x; \theta) = h_{\theta}(x) = \frac{1}{1 + \exp(-\theta^T x)}, \quad (10)$$

向分类器输入训练集样本, 由于样本彼此独立, 那么使得所有样本得到相对应正确分类结果的概率就如式(11)所示, 也就是似然函数:

$$L(\theta) = \prod P(y_i = 1 | x_i)^{y_i} (1 - P(y_i = 1 | x_i))^{1-y_i}, \quad (11)$$

使得似然函数取值最大的参数即为最优参数。

在 LR 中定义损失函数为式(12):

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^i) - y^i)^2, \quad (12)$$

根据极大似然估计可将该损失函数改写为式(13), 使用梯度上升/下降法求解最优解:

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m \text{Cost}(h_{\theta}(x^i), y^i), \quad (13)$$

其中,

$$\text{Cost}(h_{\theta}(x), y) \begin{cases} -\log(h_{\theta}(x)) & \text{if } y = 1 \\ -\log(1 - h_{\theta}(x)) & \text{if } y = 0 \end{cases}$$

逻辑回归分类器主要用于二分类问题, 用逻辑回归实现多分类就是搭建多个二分类器, 求多分类中的最大概率分类即为最后输出的分类结果。

2.3 朴素贝叶斯分类器

朴素贝叶斯分类器假设所有属性相互独立。分类器对任一测试样本 x , 利用贝叶斯定理求出后验概率最大的输出类。假设 y 一共可以取 N 个标签, y_c 代表第 c 类, 贝叶斯定理如式(14)所示:

$$P(y_c | x) = \frac{P(y_c)P(x | y_c)}{P(x)}, \quad (14)$$

其中, $P(y_c)$ 作为先验概率, 来自于客观统计或主观给出。

对于分类问题, 分母 $P(x)$ 的取值对不同的类是一样的, 用来归一化, 对分类效果无影响, 所以主要需计算 $P(x | y_c)$ 。因为假设所有属性都相互独立, 各自独立地对分类结果产生影响, 那么:

$$P(x | y_c) = \prod P(x_i | y_c), \quad (15)$$

直接计算在第 c 类样本上第 i 个属性上取值为 x_i 的样本占第 c 类样本总数的比例。如果属性值连续, 那么采用概率密度函数 $p(x_i | y_c)$ 代替 $P(x_i | y_c)$, 假设 $p(x_i | y_c) \sim N(\mu_{y_c, i}, \sigma_{y_c, i}^2)$, 研究可得:

$$p(x_i | y_c) = \frac{1}{\sqrt{2\pi}\sigma_{y_c, i}} \exp\left(-\frac{(x_i - \mu_{y_c, i})^2}{2\sigma_{y_c, i}^2}\right). \quad (16)$$

3 基于机器学习分类器的故障预测方法

结合图 1 所示, 本项目给出一种蓄能设备故障的诊断方法, 包括: 以单分类器为基础, 构造集成分类器, 建立蓄能设备故障分类预测模型; 从蓄能设备数据采集管理系统中提取特征, 并作为蓄能设备故障分类预测模型的输入。

4 实验结果与分析

目前蓄能设备开放性项目包括 NASA 的开源数据集, 马里兰大学先进寿命周期工程中心的锂电池数据集, 上海市 30 万辆新能源设备大数据平台及服务能力建设等项目等。本实验环境采用 Python 3.6 版本, 在 NVIDIA GeForce GTX 1050 Ti 上运行程序。

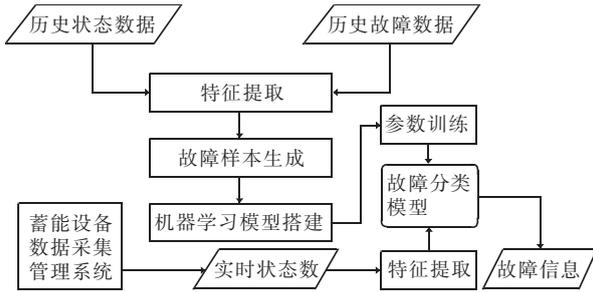


图1 基于机器学习的故障预测模型

Fig. 1 Fault evaluation based on machine learning method

由于充电过程中,总电压、单体电压均处于上升状态,无法准确辨别蓄能设备本身真实性能状态,所以实验截取出蓄能设备运行的每一个放电周期作为训练集的一条样本。根据模式识别分析,实验选择每个放电周期的最低总电压、最低单体电压最低值、最高单体最高最低电压压差值作为特征输入,报警信息作为标签。

由于实验需要测试模型精度,需要预先对数据进行标注。实验将标签分为4类:0表示正常无任何报警;1表示单体欠压与蓄能设备类型欠压复合报警;2表示单体欠压与单体一致性差复合报警;3表示其它报警。为排除数据采集管理系统传输数据丢失或误报警情况对标签的影响,实验基于蓄能设备标准安全阈值对单体一致性差报警、蓄能设备欠压复合报警进行了重新标注,保证训练集与测试集的标签都基于故障模式准确标注。

数据经处理后得到完整样本集,共6865个样本。其中,正常无报警5198条;单体欠压与蓄能设备类型欠压复合报警200条;单体欠压与单体一致性差复合报警47条;其它报警1420条。

考虑到其它报警情况对输入特征的影响较为复杂,除掉其它报警得到样本集1,共5445个样本。此时正负例比达到21:1,严重样本失衡。

删除正例样本,直至与负例样本平衡,得到样本集2,共447条样本。其中,正常无报警231条;单体欠压与蓄能设备类型欠压复合报警200条;单体欠压与单体一致性差复合报警47条。

各机器学习分类器精度表现见表1。

表1 机器学习分类器精度

Tab. 1 Precision results of machine learning methods

分类器模型	完整样本集	样本集1	样本集2
Logistic Regression	0.757 611	0.955 189	0.574 324
MultinomialNB	0.757 174	0.954 638	0.484 276
SVM	0.758 191	0.958 124	0.733 456

分类器最主要的目的在于准确识别出故障蓄能

设备,而不是识别正常蓄能设备模式。各机器学习分类器在样本集1上表现出的高精度是基于大量正例样本的条件,而对于负例样本诊断的准确率很差。

在研究中的3个样本集上,表现最好的为支持向量机,在完整样本集上,SVM分类器预测1373个样本,其中44个负例样本,预测错误29个;在样本集1上,SVM分类器预测1089个样本,其中51个负例样本,预测错误39个;而在样本集2上,SVM分类器预测96个样本,其中42个负例样本,预测错误1个。其它2个分类器虽然性能弱于支持向量机,但在经过平衡化处理的样本集1和样本集2上,其对负例样本的分类性能均有提高。

由此可见,在平衡样本集上,分类器准确率虽然没有达到0.9,但对于故障的评估是十分精确的。

5 结束语

本文探讨了基于集成学习进行特征重要度排序的原理,并利用集成学习boosting算法XGBoost以及bagging算法随机森林对蓄能设备的故障特征进行了重要度排序,得到对蓄能设备报警影响程度最高的几个特征:蓄能设备单体电压最高/低值、总电压(总电流)、SOC、累计里程、设备状态、充电状态、DC-DC状态,以及对蓄能设备类型欠压报警影响程度最高的几个特征:蓄能设备单体电压最高/低值、总电压、SOC、累计里程。未来工作中将进一步讨论其它故障与数据采集管理系统数据的相关性。进而通过提高发电厂蓄能设备健康管理水平来保障发电厂削峰填谷能力。

参考文献

- [1] 苏伟,钟国彬,沈佳妮,等. 锂电子电池故障诊断技术进展[J]. 储能科学与技术,2019,8(2):225-236.
- [2] GAO Zuchang, CHENG S C, WOO W L, et al. Genetic algorithm based back-propagation neural network approach for fault diagnosis in lithium-ion battery system [C]//6th IEEE International Conference on Power Electronics Systems and Applications (PESA). Hong Kong: IEEE,2016:1-6.
- [3] CHEN Wen, CHEN Weitian, SAIF M, et al. Simultaneous fault isolation and estimation of lithium-ion batteries via synthesized design of Luenberger and learning observes [J]. IEEE Transactions on Control Systems Technology,2014,22(1):290-298.
- [4] 徐佳宁,梁栋滨,魏国,等. 串联电池组接触电阻故障诊断分析[J]. 电工技术学报,2017,32(18):106-112.
- [5] XU Jianing, LIANG Dongbin, WEI Guo, et al. Series battery pack's contact resistance fault diagnosis analysis[J]. Transactions of China Electrotechnical Society,2017,32(18):106-112.

(下转第333页)