

DOI:10.13296/j.1001-1609.hva.2020.09.012

## 基于Spark的高压断路器机械寿命预测评估

段雄英<sup>1</sup>, 赵洋洋<sup>1</sup>, 陈艳霞<sup>2</sup>, 马 畅<sup>1</sup>, 廖敏夫<sup>1</sup>

(1. 大连理工大学电气工程学院, 辽宁 大连 116024; 2. 国网北京市电力公司电力科学研究院, 北京 100075)

**摘要:** 电力系统中高压断路器普遍采用液压操动机构,其机械寿命是考核断路器可靠性和稳定性的基础,文中提出运用数据驱动方法建立退化过程与特征参量之间的数学映射关系,实现高压断路器液压操动机构的机械寿命预测。基于Hadoop+Spark大数据平台及其生态组件,将采集的数据存储在Hadoop分布式文件系统中,利用Spark运算框架处理数据,结合Spark ML机器学习库中的回归算法,训练模型并进行验证评估,从评估结果来看:文中创建的广义线性回归算法模型的均方根误差为52.96,  $R$ -平方系数为0.92,预测值和实际值偏差较小,拟合效果较好,相较于其他回归算法模型性能指标最优。最后得到最优模型及其参数,据此可确定对高压断路器维护检修的最佳时刻,对于切实保护高压断路器设备的运行可靠性、安全性和经济性具有重要的意义。

**关键词:** 断路器液压机构; 机械寿命; Spark; 广义线性回归

## Spark-based Prediction and Evaluation of Mechanical Life for High Voltage Circuit Breakers

DUAN Xiongying<sup>1</sup>, ZHAO Yangyang<sup>1</sup>, CHEN Yanxia<sup>2</sup>, MA Chang<sup>1</sup>, LIAO Minfu<sup>1</sup>

(1. School of Electrical Engineering, Dalian University of Technology, Liaoning Dalian 116024, China; 2. State Grid Beijing Electric Power Research Institute, Beijing 100075, China)

**Abstract:** The high-voltage circuit breakers generally use hydraulic operating mechanisms in power systems. The mechanical life is the basis for evaluating the reliability and stability of circuit breakers. In this paper, the data-driven method is proposed to establish the mathematical mapping relationship between the degradation process and the characteristic parameters. Realize the mechanical life prediction of the hydraulic operating mechanism of the high voltage circuit breaker. Based on Hadoop+Spark platform and its ecological components, the data is collected and stored in the distributed file system HDFS in Hadoop and the Spark computing framework is used to process the data. Combined with the regression algorithm in the Spark ML machine learning library, the model is trained and verified. From the evaluation results: The RMES of the model of generalized linear regression is 52.96, the  $R$ -square coefficient is 0.92, it shows the deviation between the predicted value and the actual value is small, and the fitting effect is better, which is the best compared with other regression algorithm models. Finally, the optimal model and its parameters are obtained. Based on this, the optimal time for maintenance and repair of the high-voltage circuit breaker is determined, which is of great significance for effectively protecting the operational reliability, safety and economy of the high-voltage circuit breaker.

**Key words:** hydraulic mechanisms of circuit breaker; mechanical life; Spark; generalized linear regression

## 0 引言

近几年大数据技术应用在电力系统中已经越

来越普遍,通过构建电力大数据平台,在数据的采集与整合、数据存储与管理、数据分析与挖掘和数据展示与应用等方面进行综合应用,将电力大数据

收稿日期:2020-05-20; 修回日期:2020-07-24

基金项目:国家电网公司总部科技项目。

Project Supported by State Grid Corporation of China Science and Technology Projects.

潜在价值进行深度挖掘和效益转化<sup>[1-2]</sup>。然而,电力大数据平台中设备的健康管理及故障诊断(PHM)系统冗余繁杂,在数据分析与挖掘方面与实际应用融合程度不高,在平台上进行电力设备状态监测、故障诊断和寿命评估等具体应用时仍面临诸多挑战。

基于电力大数据技术应用在高压断路器方面研究尚在起步阶段,邱剑<sup>[3]</sup>等人通过数据挖掘技术研究断路器全寿命状态评估模型,展示断路器全寿命状态信息流,为实现断路器全寿命周期管理提供了基础。断路器的机械操动机构可靠动作是断路器全寿命中的关键,机械寿命是其稳定性和可靠性的重要体现,另外,一些研究单位<sup>[4]</sup>开展高压断路器的机械特性试验,针对不同寿命阶段的机械状态进行分析,研究各机械状态特征量的变化规律。杨秋玉<sup>[5]</sup>等考虑触头超行程和平均分闸速度两种机械特征量,研究随机模糊双重不确定性影响下的高压断路器剩余机械寿命,但是模型容易陷于局部最优。

文中依托于电力大数据平台中 PHM 系统,针对高压断路器液压操动机构提出一种新的机械寿命预测评估方法,搭建 Hadoop+Spark 的生态体系架构<sup>[6-8]</sup>,利用大数据技术综合多种特征参量进行数据挖掘与分析,研究高压断路器机械寿命的影响因素规律,结合 Spark ML 机器学习回归算法,建立广义线性回归算法模型,并对模型验证评估得到最优模型,从而实现预测评估高压断路器机械寿命,为高压断路器的检修和替换提供合理性建议。

## 1 高压断路器机械寿命

### 1.1 高压断路器液压操动机构

液压操动机构在高压断路器领域应用较为广泛,主要由储能元件(储压器、滤油器)、控制元件(电磁铁、阀系统)、执行元件(工作缸、活塞杆)及辅助元件(油压开关、安全阀、压力表)四部分组成<sup>[9]</sup>。通常断路器根据运行年数和运行状态判断其是否需要替换,对于液压操动机构,在空载或额定长期负荷轻载下的分合循环操作次数即是机械寿命,一般可达 5 000 次以上<sup>[10]</sup>。但是在实际运行中由于开断次数较少,而其内部机构元件逐渐老化,再加上在运行期间受到的各种不确定因素的影响,实际的机械寿命或已达到需要维修或替换的状态。因此,文中研究机械寿命试验中特征参量与剩余使用寿命的关联性,建立退化过程与特征参量之间的数学映射关系,基于断路器机械寿命试验数据训练初始机器学习模型,另外,实际运行中的检测数据与试验中的特征参量相互对应,基于此模型利用工况下检测

数据进行预测机械寿命,所得预测结果可为高压断路器维修策略提供合理性建议。

### 1.2 影响因素分析

机械寿命的影响因素主要与机械机构组成部分有关,主要包括机构中各元件的老化磨损,操作电压和系统压力的变化,结构工艺参数等,比如当工作缸中主活塞杆的缓冲行程参数有所变化时,其分合闸速度和时间必受影响,对机械强度、耐机械冲击性能和耐磨损能力等带来一定的考验,从而影响操动机构的机械寿命<sup>[11]</sup>。但是,对于断路器本体机械机构性能和结构工艺参数的综合因素,很难做到直接监测并综合评估其带来的影响。

而在《高压交流断路器标准》<sup>[12]</sup>中可知,由机械操作试验可得到机械特性曲线及相关数据,因此文中利用相关试验数据作为等效参量,主要是初始设置参量、机械特性参量和其他参量数据,以分闸操作试验为例:

- 1)初始设置参量。操作电压和系统压力(油压);
- 2)机械特性参量。主触头行程、分闸时间和速度、分闸同期性等;
- 3)其他参量。电流持续时间、振动噪声和环境温度等。

这些高压断路器操动机构特征参量与机械寿命之间存在某种相关性,文中基于 Spark 建立数学模型,研究分析特征参量与机械寿命之间的关系。

## 2 机械寿命预测模型

### 2.1 高压断路器液压操动机构

文中基于 Hadoop+Spark 的生态体系架构创建预测模型,将传感器采集的数据进行流式处理,数据存储利用 Hadoop 平台中的 HDFS<sup>[13]</sup>(分布式文件系统)完成,数据预处理采用归一化和主成分分析方法,另外调用 Spark ML<sup>[14]</sup>机器学习库构建算法模型,并对其进行训练、验证及评估等,具体实现流程见图 1。

Spark 强劲的并行计算能力及内核处理机制可以高效完成算法模型流程。一方面,通过创建弹性分布式数据集<sup>[15]</sup>(RDD)处理数据,极大的提高数据查询速度,实现数据的快速转换迭代,并缓存后续的频繁访问需求。将创建的 RDD 转换为 DataFrame 再进行数据的预处理操作,使得在已有的分布式数据集的基础上附加 Schema 信息,即每一列都带有名称和类型,从而获得更多的结构信息;另一方面,通过创建训练模型的机器学习工作流(Pipeline)处理数据<sup>[16]</sup>,Pipeline 处理流程见图 2。

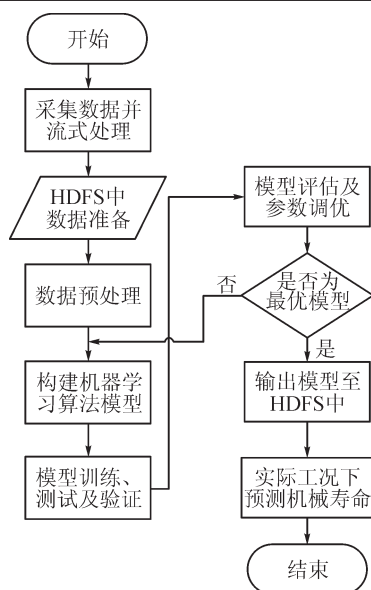


图1 Spark实现算法模型流程图

Fig. 1 The algorithm model flow chart on Spark

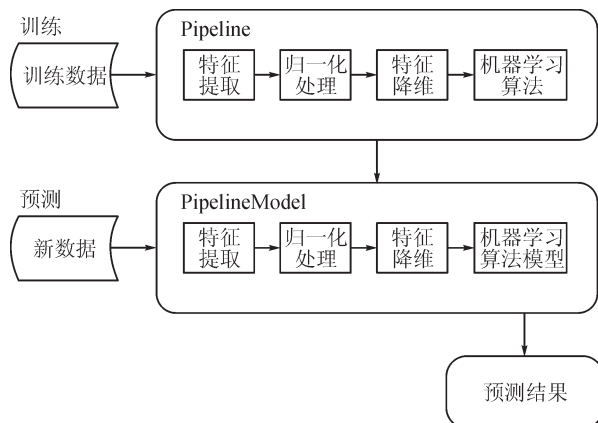


图2 模型在Pipeline上的流程

Fig. 2 Training model flow on the Pipeline

在Pipeline中利用机器学习算法训练数据,执行完毕后产生一个PipelineModel, Pipeline和PipelineModel有助于确保训练数据和测试数据的特征处理流程保持一致。一个Pipeline在结构上包含一个或多个Stage模块进行数据处理,每个Stage都会完成一个数据处理任务,包括特征提取、归一化处理和特征降维等,有了这些处理特定问题的各个Stage并按照具体的逻辑顺序组织创建一个Pipeline,使得流水线式机器学习比单个步骤独立建模更加高效、快速。

## 2.2 机器学习算法

文中考虑断路器机械寿命与特征参量之间的关联性,利用机器学习中的回归算法进行数据挖掘与分析。以高压断路器操动机构的特征参量作为输入变量,包括操作电压、碟簧压缩量(油压)、分合闸主触头行程、合闸超程和动作时间等机械特性数据,另外以高压断路器机械寿命的剩余使用寿命

(RUL)作为输出响应,基于Spark ML机器学习库建立回归算法模型,拟合输入变量与输出响应之间的函数关系。首先给定输入变量数据集:  $D=(X_1, y_1), (X_2, y_2), \dots, (X_n, y_n)$ , 共有  $n$  组数据, 其中特征集合  $X_i$  有  $m$  个特征属性, 表示为  $X_i=(x_{i1}, x_{i2}, \dots, x_{im})$ , 定义  $y_i$  为输出响应, 将数据集按比例可划分为训练集  $D_t$  和测试集  $D_e$ 。

线性回归是通过  $m$  个特征属性的线性组合对输出响应变量  $y_i$  进行拟合, 考虑到文中的输入变量和输出变量的函数关系并不一定是线性关系, 因此选择广义线性回归算法<sup>[17]</sup> (generalized linear regression, GL) 放宽这个假设, 具体来讲, 一般线性回归模型中要求因变量是连续的且服从正态分布, 而在广义线性模型中, 因变量的分布可扩展到非连续的形式, 如伽玛分布、Poisson分布、负二项分布等; 另外, 在一般线性回归模型中, 自变量的线性预测值就是因变量的估计值, 而在广义线性模型中, 自变量的线性预测值是因变量的函数估计值。文中假设特征向量  $X_i$  作为输入变量与目标响应  $y_i$  作为输出响应之间存在某种数学映射关系, 通过引入一个链接函数( $G$ ), 将非线性输出转换为一个线性的响应, 从而拟合一个更为广义的线性模型, 定义广义线性回归模型的基本结构如下:

$$G(\mu_i) = \sum_{i=1}^n X_i \beta_i^T + \varepsilon_i \quad (1)$$

$$\mu_i = E(y_i | X_i; \beta_i^T) \quad (2)$$

式(1)-(2)中:  $\beta_i^T$  是特征向量  $X_i$  的系数向量;  $\varepsilon_i$  是白噪声项;  $\mu_i$  是目标响应变量  $y_i$  的期望, 表示给定参数  $\beta_i^T$  时, 在  $X_i$  发生了的情况下  $y_i$  发生的概率。

随机变量  $y$  的概率分布形式符合指数分布族的范围之内, 指数分布族的概率密度函数可表示为

$$f(y; \theta) = h(y, \varphi) \exp\left[\frac{\theta \cdot y - b(\theta)}{a(\varphi)}\right] \quad (3)$$

式(3)中:  $a(\cdot)$ 、 $b(\cdot)$ 、 $h(\cdot)$  是已知函数;  $\theta$  是自然参数;  $\varphi$  是离散参数。由于断路器开断次数是计数值属性, 而机械寿命的剩余使用寿命作为输出响应, 因此定义链接函数  $G(\mu) = \ln(\mu)$ , 将计数值转换成线性输出, 由此可得概率密度函数为

$$f(y_i; \theta_i) = \exp\left[\sum_{i=1}^n \frac{\theta_i \cdot y_i - b(\theta_i)}{a(\varphi)} + c(y_i, \varphi)\right] \quad (4)$$

该分布的期望(3.10)即是文中要优化的目标函数, 建立目标函数的似然函数, 并通过最大化似然函数进行参数估计。其对数似然函数为

$$l(\theta_i; y_i) = \frac{\theta_i \cdot y_i - b(\theta_i)}{a(\varphi)} + c(y_i, \varphi) \quad (5)$$



两边对  $\theta_i$  微分可得得分统计量  $U$

$$U(\theta_i; y_i) = \frac{dl(\theta_i; y_i)}{d\theta_i} = \frac{y_i - b'(\theta_i)}{a(\varphi)} \quad (6)$$

两边对  $\theta_i$  二阶微分可得 Hessian 矩阵

$$H_k^{-1} = U'(\theta_i; y_i) = -\frac{1}{a(\varphi)} b''(\theta_i) \quad (7)$$

在机器学习的优化问题中,使用迭代思想更新方程,而牛顿法仅适用于求一元函数,针对高压断路器寿命影响因素特征集合的多元性,文中引入拟牛顿法中的 L-BFGS 优化算法,通过迭代计算来逼近  $H_k^{-1}$ ,求解得分统计量  $U(\theta_i)$  的零点,求出似然估计  $\hat{\theta}_i$  后,进而求得广义线性回归模型中参数  $\beta_i^T$ 。

### 2.3 模型评估性能指标

算法模型构建以后,需要对模型的性能、预测目标的拟合度等进行一些评估,因此,计算预测值和实际值的误差,并用一些度量方式进行评估,选取的评估方法<sup>[18]</sup>有均方误差(mean squared error, MSE)、均方根误差(root mean squared error, RMSE)、R-平方系数(R-squared coefficient)。公式如下:

$$MSE = \frac{\sum_{i=1}^n (y - \hat{y})^2}{n} \quad (8)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y - \hat{y})^2}{n}} \quad (9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y - \hat{y})^2}{\sum_{i=1}^n (y - \bar{y})^2} \quad (10)$$

式(8)-(10)中:  $\hat{y}$  是断路器机械寿命的预测值;  $y$  是试验时的观测值;  $\bar{y}$  是  $y$  的平均值,  $MSE$  计算所有样本机械寿命预测值和实际值平方差之和;  $RMSE$  是  $MSE$  的平方根,能更好的衡量机械寿命预测值和真值之间的偏差,  $RMSE$  值越小说明偏差越小,但是过小容易导致过拟合;  $R$ -平方系数也称之为判定系数,具体测量目标变量的变异度,用来评估算法模型拟合数据的好坏,反映预测值误差的实际情况,最终结果在  $[0, 1]$  区间,数值越接近于 1 表明拟合数据效果越好。

### 3 算例结果与分析

文中基于实验室搭建的 Hadoop+Spark 大数据平台,由 4 台 8 核处理器电脑搭建,内存为 16 G,其中 1 台的作为主节点,其余 3 台为从节点,操作系统为 Ubuntu16.04,部署 Hadoop 和 Spark 集群环境,组成完全分布式的大数据平台。

以某台 1 100 kV 高压断路器 M2 级机械寿命的型式试验为例,按照试验规程对分闸、合闸及重合闸进行循环操作共 10 000 次左右,配置传感器采集试验中机械特性相关数据共 10 000 组,每组数据包括操作电压、碟簧压缩量(油压)、主触头行程、合闸超程、动作时间、动作速度、分合闸同期性等。在试验过程中,额定操作电压为 220 V,额定碟簧压缩量(系统压力)为 120 mm,通过设置改变操作电压(143、220、240 V)和碟簧压缩量(51、120、122 mm)更加全面的考核断路器的稳定性,另外按照试验规程对分闸、合闸及重合闸进行循环操作,随着断路器动作次数的增加,机械特性参量和其他参量数据随之波动变化,提取部分试验分合闸数据见图 3。

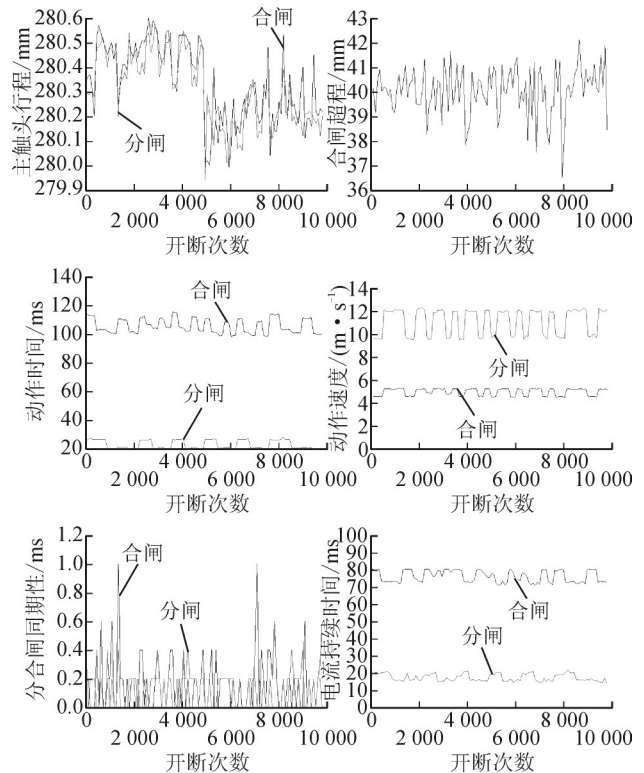


图3 部分特征参量与机械动作次数关系

Fig. 3 Relationship between some characteristic parameters and mechanical action times

由于设定参量操作电压和系统压力的不同,曲线变化呈规律性波动,而整体趋势随着开断次数的增加而变化,可以明显看出主触头行程在 5 000 次左右之后位移量减小,合闸超程波动越来越大,分闸动作时间呈缓慢减小的趋势,分合闸同期性波动呈浴盆曲线规律,而分闸速度和电流持续时间变化较小,这也表明此断路器机械稳定性较高。

为了解决采集单台断路器数据建立模型所带来的通用性问题,文中引入 SparkStream 组件进行实时数据流处理<sup>[19]</sup>,可将传感器采集的数据实时上传到 HDFS 中,为后续批量查询、快速训练模型及更新

模型提供基础。

对源数据预处理,以操作电压、碟簧压缩量和主触头行程等特征参量作为特征向量组(features),机械寿命为目标变量(label),在 Spark 平台上创建 RDD 并进行数据预处理,通过 Normalizer 函数对 features 进行归一化处理,归一化后的 features 维数较多且各自变量之间相互影响,从而造成多重共线性,容易降低预测模型的准确性并增加算法的计算复杂度。因此,采用主成分分析方法<sup>[20]</sup>(PCA)优化选择 features 并剔除冗余变量,PCA 是通过计算数据集的协方差矩阵的特征值和向量,从而确定 features 中重要的主成分,  $k$  个分量的累计方差曲线见图4。其中纵轴代表降维后的各主成分的方差值占总方差值的比例,这个比例越大,则说明越是重要的主成分,横轴是降维后的特征维度数目。

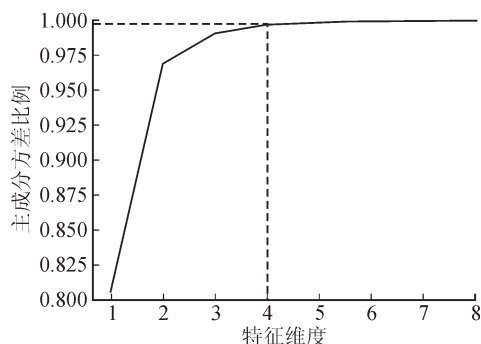


图4  $k$  个分量的累计方差曲线

Fig. 4 Cumulative variance curve of  $k$  components

结果显示  $k=4$  时,主成分比例已接近 100%,即当主触头行程、合闸超程、动作时间、分合闸同期性 4 组特征向量时性能最优。将预处理后的数据集按照 [0.7, 0.3] 随机切分为训练集和测试集,训练集转换成 DataFrame 格式后在 Pipeline 中训练机器学习算法模型,在这一过程中基于流处理方式可持续训练并优化模型。

另外将 GL 算法模型与其他回归算法模型对比,选取加速时间失效算法模型(accelerated failure time, AFT)和决策树算法模型(decision tree, DT)作为比较对象,AFT 是将线性回归模型的建模方法引入到生存分析的领域,研究多协变量(features)与对数生存时间(label)之间的回归关系;DT 是通过一系列规则对数据进行分类的算法,可利用回归树处理连续变量,研究树结构中节点(features)与叶子结点(label)之间的分支关系。

分别求出每个模型的  $MSE$ 、 $RMSE$  和  $R_2$ ,见表1。由表1可以发现 GL 模型的评估指标  $MSE$  和  $RMSE$  都是最小;而且  $R_2$  较接近于 1,说明模型拟合效果较

好;另外统计训练模型时间,结果显示 AFT 模型训练速度最快,但是  $R_2$  结果为负值,说明数据源不适合此模型,综合来讲,GL 模型性能指标最优。

表1 算法模型评估性能指标

Table 1 The evaluate performance indicators of algorithm model

参数	AFT	DT	GL
$MSE$	221 593.00	9 588.50	5 242.12
$RMSE$	470.737 0	97.920 9	72.402 5
$R_2$	-5.376	0.676	0.875
训练用时/ms	4.32	5.45	5.16

由于源数据目标变量机械寿命不为负数且值域较大,因此文中通过变换目标变量对 GL 模型做进一步的优化,对目标变量值取平方根,使数据更加均匀,相较于原始目标变量更接近正态分布,使得 GL 模型训练效果更好。

将平方根变换的目标变量转换为原始的值,并通过式(6)计算回归系数向量组及截距,另外结合交叉验证(cross validator, CV)方法,将数据集切分为  $K$  折数据集,并计算  $K$  折数据集拟合出的模型上的平均评估指标,得出优化后 GL 模型评估性能结果见表2。

表2 优化后 GL 模型评估结果汇总

Table 2 The summary of evaluation results of optimized GL model

特征参量	回归系数	系数标准差	$P$ 值
主触头行程	-0.810 469	0.417 579	0.027 427 1
合闸超程	0.768 676	0.245 018	0.007 862 6
动作时间	-0.164 143	0.120 380	0.001 958 6
分合闸同期性	-2.683 404	1.327 404	0.016 430 7
截距	224.110 000	$MSE$	2 805.110 000 0
AIC	8.096 000	$RMSE$	52.963 200 0
—	—	$R_2$	0.920 439 0

可以看出,系数标准差较小, $P$  值小于 0.05,表明所求回归系数变异程度较小,差异不显著,计算结果可信度高。另外将优化后的模型评估参数和原训练模型比较, $RMSE$  为 52.96 比原 GL 模型更小, $R_2$  为 0.92 更接近于 1,表明测试集中的目标变量预测值与真实值拟合效果更好。

在模型中对每一组测试集特征变量进行预测,选取测试集中前 75 组预测结果排序后见图5。

其中纵轴表示断路器剩余使用机械寿命的百分比,由于测试集与训练集的切分,使得曲线有阶段性的陷落,可以看出测试集中预测结果偏差较小,浴盆曲线中稳定阶段预测精确度较高。

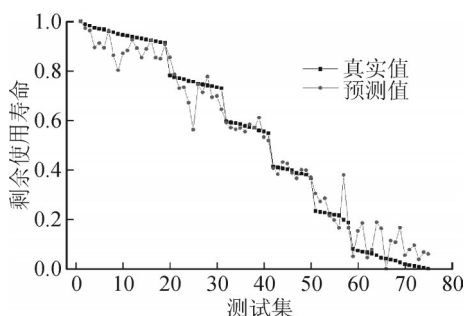


图5 测试集中部分预测结果

Fig. 5 The part of the predicted results in the test set

## 4 结论

1)文中运用数据驱动方法实现断路器机械寿命预测评估,基于 Hadoop+Spark 大数据平台结合 HDFS 存储技术和 Spark 运算框架,建立机器学习算法模型,综合多种影响因素对机械寿命进行预测。通过实例分析验证,对比不同算法模型的评估性能指标参数,广义线性回归算法模型的预测误差最小,拟合效果更好。

2)流处理方式提高模型的通用性,可以针对不同断路器创建独立的机器学习模型,另外文中未探究特征参量影响机械寿命的内在机理,而是基于数据更多的关注模型是否有准确的预测结果。

3)针对广义线性回归模型做进一步优化,一方面,通过处理数据使其更符合算法模型架构,从而减小预测误差。另一方面,利用正则化及交叉验证方法控制过拟合现象,确定一个泛化能力最强、评估指标更优的算法模型。

文中研究成果可作为第三方应用程序嵌入电力大数据平台完善 PHM 系统,通过对海量现场数据的不断训练优化模型,解决现有“离线建模——在线预测”的弊端,助力区域配电网运维模式从故障维修、计划检修向预测性维修转变。

## 参考文献:

- [1] 饶 玮,蒋 静,周爱华,等. 面向全球能源互联网的电力大数据基础体系架构和标准体系研究[J]. 电力信息与通信技术,2016, 14(4):1-8.  
RAO Wei, JIANG Jing, ZHOU Aihua, et al. Research on power big data basic architecture and standard system for global energy interconnection[J]. Electric Power Information and Communication Technology, 2016, 14(4): 1-8.
- [2] 陈敬德,盛戈皞,吴继健,等. 大数据技术在智能电网中的应用现状及展望[J]. 高压电器,2018,54(1):35-43.  
CHEN Jingde, SHENG Gehao, WU Jijian, et al. Application status and prospect of big data technology in smart grid

- [J]. High Voltage Apparatus, 2018, 54(1): 35-43.
- [3] 邱 剑,王慧芳,应高亮,等. 文本信息挖掘技术及其在断路器全寿命状态评价中的应用[J]. 电力系统自动化, 2016, 40(6):107-112.  
QIU Jian, WANG Huifang, YING Gaoliang, et al. Text mining technique and application of lifecycle condition assessment for circuit breaker[J]. Automation of Electric Power Systems, 2016, 40(6): 107-112.
- [4] 王建华,张国钢,宋政湘,等. 物联网+大数据+智能电器——电力设备发展的未来[J]. 高压电器,2018,54(7):1-9.  
WANG Jianhua, ZHANG Guogang, SONG Zhengxiang, et al. Internet of things, big data, intelligent electrical apparatus—the future development of power equipment[J]. High Voltage Apparatus, 2018, 54(7): 1-9.
- [5] 杨秋玉,彭彦卿,庄志坚,等. 基于随机模糊理论的高压断路器剩余机械寿命评估[J]. 高压电器,2016, 52(8): 161-165.  
YANG Qiuyu, PENG Yanqing, ZHUANG Zhijian, et al. Residual mechanical life evaluation for high voltage circuit breaker based on random fuzzy theory[J]. High Voltage Apparatus, 2016, 52(8): 161-165.
- [6] SVEN V P, ZHANG Zhongheng, SCHMITZ M, et al. Scalable predictive analysis in critically ill patients using a visual open data analysis platform[J]. PLOS One, 2016, 11 (1): e0145791.
- [7] GUPTA A, THAKUR H K, SHRIVASTAVA R, et al. A big data analysis framework using apache spark and deep learning[C]//IEEE International Conference on Data Mining Workshops (ICDMW). New Orleans, LA, USA: IEEE, 2017: 9-16.
- [8] 王保义,王冬阳,张少敏. 基于 Spark 和 IPPSO\_LSSVM 的短期分布式电力负荷预测算法[J]. 电力自动化设备, 2016, 36(1):117-122.  
WANG Baoyi, WANG Dongyang, ZHANG Shaomin. Distributed short-term load forecasting algorithm based on Spark and IPPSO\_LSSVM[J]. Electric Power Automation Equipment, 2016, 36(1): 117-122.
- [9] 孙艳玲,王永良,雷 琴,等. 特高压断路器液压操动机构仿真研究[J]. 高压电器,2014, 50(10): 15-19.  
SUN Yanling, WANG Yongliang, LEI Qin, et al. Research on hydraulic operating mechanism of UHV circuit breaker [J]. High Voltage Apparatus, 2014, 50(10): 15-19.
- [10] 王阳英夫. 高压断路器的机械可靠性分析及寿命评估 [D]. 南京:东南大学,2017.  
WANG Yangyingfu. Mechanical reliability analysis and Life evaluation of high voltage circuit breakers[D]. Nanjing: Southeast University, 2017.
- [11] 黄 凯. 高压断路器剩余寿命评估研究[D]. 北京:华北电力大学,2013.  
HUANG Kai. Research on residual life assessment of high



- voltage circuit breaker[D]. Beijing: North China Electric Power University, 2013.
- [12] 高压交流断路器:GB 1984—2014[S]. 2014.  
High voltage AC circuit breaker:GB 1984—2014[S]. 2014.
- [13] SHVACHKO K, KUANG Hairong, RADIA S, et al. The hadoop distributed file system[C]//IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST). Incline Village, NV, USA: IEEE, 2010: 1-10.
- [14] MENG X R, BRADLEY J, YAVUZ B, et al. MLlib: Machine learning in apache spark[J]. Journal of Machine Learning Research, 2016, 17(1):1235-1241.
- [15] 张少敏, 毛冬, 王保义. 大数据处理技术在风电机组齿轮箱故障诊断与预警中的应用[J]. 电力系统自动化, 2016, 40(14):129-134.  
ZHANG Shaomin, MAO Dong, WANG Baoyi. Application of big data processing technology in fault diagnosis and early warning of wind turbine gearbox[J]. Automation of Electric Power Systems, 2016, 40(14): 129-134.
- [16] SVYATKOVSKIY A, IMAI K, KROEGER M, et al. Large-scale text processing pipeline with apache spark[C]//IEEE International Conference on Big Data (Big Data). Washington, DC, USA: IEEE, 2016: 3928-3935.
- [17] MADSEN H, THYREGOD P. Introduction to general and generalized linear models[M]. Boca Raton: CRC Press, 2011.
- [18] 刘嘉蔚, 李奇, 陈维荣, 等. 基于核超限学习机和局部加权回归散点平滑法的PEMFC剩余使用寿命预测方法[J/OL]. 中国电机工程学报, 2019, 39(24):7272-7279.  
LIU Jiawei, LI Qi, CHEN Weirong, et al. PEMFC remaining service life prediction method based on nuclear overrun learning machine and local weighted regression scattering smoothing method[J]. Proceedings of the CSEE, 2019, 39(24):7272-7279.
- [19] ZAHARIA M, DAS T, LI H, et al. Discretized streams: Fault-tolerant streaming computation at scale[C]//Proceedings of the Twenty-fourth ACM Symposium on Operating Systems Principles(SOSP). New York, NY, US: [s.n.], 2013:423-438.
- [20] TIPPING M E, BISHOP C M. Probabilistic principal component analysis[J]. Journal of the Royal Statistical Society, 2010, 61(3): 611-622.
- 段雄英(1974—), 女, 博士, 教授, 研究方向为智能开关电器理论及应用、电力设备在线检测与绝缘诊断等方面的研究工作(通讯作者)(E-mail: xyduan@dlut.edu.cn)。
- 赵洋洋(1994—), 男, 硕士, 研究方向为电力设备在线检测与寿命诊断(E-mail: z1398655303@163.com)。