



引用格式:肖 飞,冷喜武,徐元直,等. 基于数据挖掘的集中监控设备典型缺陷预测方法[J]. 科学技术与工程, 2020, 20(16): 6522-6526
Xiao Fei, Leng Xiwu, Xu Yuanzhi, et al. Research on typical defect prediction method of centralized monitoring equipment based on data mining[J]. Science Technology and Engineering, 2020, 20(16): 6522-6526

基于数据挖掘的集中监控设备典型缺陷预测方法

肖 飞¹, 冷喜武², 徐元直³, 何 忠³

(1. 上海市电力公司, 上海 200122; 2. 国调中心, 北京 100031; 3. 泰豪软件股份有限公司, 南昌 330096)

摘 要 为改善传统方法对监控设备典型缺陷分析结果的不理想,提出了基于数据挖掘技术的集中监控设备缺陷预测方法。预测集中监控设备在未来时间内的相关监控数据,并将其与历史典型缺陷发生期间内的监控数据相似度进行比较,以量化指标的形式对集中监控设备的典型缺陷进行预测。实验结果表明,利用所提方法对实际发生的油温异常典型缺陷进行预测,计算出的相似度指标形成了一个峰值,而缺陷未发生时的相似度指标较低,证明本文方法能够较好地体现缺陷发生的可能性,整体应用性较高。

关键词 数据挖掘;设备缺陷;集中监控;趋势预测

中图法分类号 TP391; **文献标志码** A

Research on Typical Defect Prediction Method of Centralized Monitoring Equipment Based on Data Mining

XIAO Fei¹, LENG Xi-wu², XU Yuan-zhi³, HE Zhong³

(1. State Grid Shanghai Electric Power Company, Shanghai 200122, China;

2. National Electric Power Dispatching and Control Center, State Grid Corporation of China, Beijing 100031, China;

3. Tellhow Software Software co., LTD, Nanchang 330096, China)

[Abstract] In order to improve the traditional method to analyze the typical defect analysis results of monitoring equipment, a method based on data mining technology was proposed to predict the defects of centralized monitoring equipment. Which predicts relevant monitoring data of the centralized monitoring equipment with the results are compared with the similarity of monitoring data during the historical typical defect occurrence period to predict the typical defects of the centralized monitoring equipment in a form of quantitative indicators. The experimental results show that the proposed method can be used to predict the typical oil temperature anomalies with a calculated similarity index forms a peak while the similarity index is lower when the defect does not occur, which proves that the method with a high overall application can be used to reflect the possibility of defects.

[Key words] data mining; equipment defect; centralized monitoring; tendency projection

电力系统集中监控设备的可靠性是保证电网安全运行、提升供电可靠性的基础。在实际运行中,有些电力系统集中监控设备虽然能继续使用,但运行状态可能存在异常或隐患,将影响工作人员人身安全、电网的可靠运行、设备寿命以及电能质量等,这种异常或隐患便称之为缺陷^[1]。对电力系统集中监控设备可能发生的缺陷提前加以控制或消除,能够有效保障电网安全、可靠运行。

当前,电力系统集中监控设备的缺陷分析、预测的技术主要可以分为三大类^[2]:基于物理模型(model-driven)的缺陷预测技术^[3,4]、基于数据驱动(data-driven)的缺陷预测技术^[5]、基于经验预测(experience based)的缺陷预测技术^[6]。传统基于数据驱动的缺陷预测方法一般基于神经网络、回归分

析、时间序列、马尔可夫^[7]等机器学习方法或这些方法的改进、组合(回归神经网络^[8]、模糊神经网络^[9]、小波神经网络^[10]),去尝试对某一个指标进行分析、预测,然后根据该指标的预测值去推断设备缺陷发生的可能性,但该方法具有以下缺点:在电力系统集中监控设备缺陷发生时,往往伴随着不同类型监控数据值的异常变化,但是每一个监控值的异常变化又极其不明显。因此,仅仅对某一类监控数据进行分析根本无法准确预测电力系统集中监控设备存在的缺陷,只有对多类监控数据进行综合分析才能准确预测出可能存在的缺陷。

为了解决传统的基于数据驱动的缺陷预测方法存在的问题,现提出一种基于数据挖掘的集中监控设备典型缺陷预测方法。该方法主要通过引入

收稿日期: 2019-04-11; 修订日期: 2019-07-31

第一作者:肖 飞(1976—),男,汉族,江西吉安人,博士,高级工程师。E-mail: xiaofei@sh.sgcc.com.cn

投稿网址:www.stae.com.cn

大数据相关技术及支撑平台来克服传统缺陷预测方法的不足。根据薛禹胜等^[11]中对大数据进行的分析、总结,所提方法在以下方面体现了大数据的特点:①该系统构建在大数据领域主流技术之上,基于 Hbase 存储平台,实现海量数据的高效存取,并采用 Spark Streaming 及 MapReduce 等计算引擎实现海量信息的分析及处理;②在进行电网设备缺陷预测时,本文所提方法利用概率转移矩阵确定每一类监控数据的权重,以此为基础实现基于多监控数据的设备缺陷预测。

1 设备缺陷预测方法设计

在数据挖掘过程中,可以使用许多不同的模型,如分类模型、回归模型、时间序列模型、聚类模型和关联规则模型等。针对同一模型,可以使用不同的算法进行数据挖掘。算法的目的就是找到适合于当前场景的模型。在所提的集中监控设备典型缺陷预测方法中采用了多种算法结合的方式,体现了大数据的特点,以克服传统预测技术的主要缺陷,实现了对异构数据的综合分析、处理,达到了提高预测准确度的目标。算法主要分为三个关键环节:关键监控数据搜索与权重分析、监控数据变化趋势分析及设备缺陷预测分析。

(1) 关键监控数据搜索与权重分析

对典型缺陷的历史监控数据进行分析,找出缺陷发生前及发生时数据特征跟设备状态正常时相比发生异常的监控数据,并根据异常发生的时间计算缺陷开始的实际周期。将异常监控数据与设备正常状态下的理论监控数据进行对比得出监控数据异常变化的幅度,并将异常变化的幅度大小作为后续相似度指标计算的权重。

(2) 监控数据变化趋势分析

根据历史监控数据的特征,选择合适的时间序列模型,并根据历史数据求出模型的参数,最终得到监控数据随时间和其他相关历史监控数据变化的数学模型作为监控数据未来时间取值的预测公式。

(3) 典型设备缺陷预测

根据典型缺陷周期和预测时间点得到需预测的时间段,将需预测的时间段一一代入所有该典型缺陷发生时异常监控数据的预测公式中,得到每个监控数据未来时间的变化趋势。将监控数据未来时间的变化趋势与该典型缺陷历史发生时的实测变化趋势进行相似度对比,并将对比结果按照上述算法求出的权重进行加权平均,最终得到的相似度系数即可评估该设备未来发生该典型缺陷的可能性。

1.1 关键监控数据搜索与权重分析

1.1.1 关键监控数据搜索

关键监控数据是在统计的基础上对某一监控数据未来缺陷发生产生重要影响的其他监控数据的集合。本节将计算关键监控数据集合和权重模型。

由于电力设备的部分潜伏性缺陷发展缓慢,当设备刚开始处于异常状态时往往巡视人员难以察觉,缺陷实际发生的时间往往早于缺陷历史记录中记录的时间。因此需要对缺陷发生的周期进行分析,以找出缺陷发生的时间,获得该缺陷发生周期内相关监控数据的统计特征量。

概率转移矩阵模型相较文献[5]方法有如下两点优势:①在整合多个监控数据上,提供了统一的框架和计算方法;②可以从观测数据中,不断学习参数、优化结构。

概率转移模型为

$$P(x_1, x_2, \cdots, x_N) = \prod_{i=1}^N P[x_i/P_{AR}(x_i)] \tag{1}$$

式(1)中: $P(x)$ 表示监控数据在时间点 x 上取值的概率函数; $P_{AR}(x)$ 表示监控数据在时间点 x 之前时间点上取值的概率函数; $P[x_i/P_{AR}(x_i)]$ 表示监控数据在时间点之间取值的概率函数。

基于贝叶斯理论,通过计算监控数据在缺陷发生时每个时间点其最可能发生的取值及其发生概率,即

$$\hat{\theta}_{MLE}(x) = \operatorname{argmax}_{\theta} f(x|\theta) \tag{2}$$

式(2)中: θ 表示概率函数的参数; $f(x|\theta)$ 表示监控数据在 n 时间点取得的观测值; $\operatorname{argmax}_{\theta} f(x|\theta)$ 表示使得 $f(x|\theta)$ 最大化的参数 θ 。

缺陷周期计算的主要过程为:将要进行异常检测的缺陷记录开始时间及开始时间前一天的所有该监控数据实测值代入上述转移概率矩阵,获得该监控数据的概率序列。将不在置信区间内的点设置为异常点,该点对应的的时间即认为是该监控数据发生异常的时间。根据其关联的各个监控项异变的时间,取异变最早发生的时间作为缺陷开始时间,缺陷记录中的消缺时间为缺陷结束时间,从而得到每个缺陷发生的时间周期。并根据时间周期得出缺陷发生时各监控数据随时间变化的缺陷特征曲线。

1.1.2 关键监控数据权重分析

本文通过计算在缺陷发生时其监控数据发生异常变化的幅度,作为相似度分析时的权重。计算关键监控数据权重的思想为:计算监控数据在正常状态下缺陷发生时间段内的概率转移矩阵的参数,

并将该概率转移矩阵与缺陷发生时的异常数据概率转移矩阵做对比。根据两个转移概率矩阵计算每一个监控变量对应状态取值的差异绝对值及该差异的置信区间。将置信区间中的每一个差异乘以其对应发生的概率(最大后验概率),得到最大后验计算值,即:

$$\hat{\theta}_{\text{MAP}}(x) = \arg \max_{\theta} \frac{f(x|\theta)g(\theta)}{\int_{\theta} f(x|\theta')g(\theta')d\theta'} \quad (3)$$

式(3)中: $g(\theta)$ 表示监控数据 $n-1$ 时间点,对 n 时间取值的先验概率函数。

最后计算其期望值,该期望值体现了监控数据异常变化的幅度,过程如式(4)所示:

$$E(x) = \int_{\hat{\theta}_{\text{MAP}}(x)} xP(x)dx \quad (4)$$

1.2 遥测量变化趋势分析

确定了缺陷发生周期内的监控数据异常变化后,利用 ARMA(auto-regressive and moving average model) 时间序列模型分析其变化趋势得出其监控数据随时间变化的函数,从而对监控数据在未来时间段的变化趋势进行预测。计算过程如式(5)所示:

$$\begin{aligned} Th = & c_1 \ln \left\{ \sum_{i=0}^m \text{lag}[Ih(t)] \right\} + \\ & c_2 \ln \left\{ \sum_{i=0}^n \text{lag}[Im(t)] \right\} + \\ & c_3 \ln \left(\sum_{i=0}^v \text{abs} \{ \text{lag}[Ph(t)] \} \right) + \\ & c_4 \sum_{i=0}^r \ln \{ f(t) \text{lag}[Th(t)] \} - \alpha \end{aligned} \quad (5)$$

式(5)中: α 表示预测模型的截距; t 表示目标监控数据的变化周期; $\text{lag}[Ih(t)]$ 表示以预测时间点为基准、目标监控数据在 t d 前的实际值; $\text{lag}[Ih(5)]$ 表示目标监控值在预测时间点 5 d 前的实测值; $\text{lag}[Im(t)]$ 表示其相关的监控数据 Im 在 t d 前的实际值; $\text{lag}[Ph(i)]$ 表示其相关的监控数据 Ph 在 t d 前的实际值; $c_1、c_2、c_3、c_4$ 表示预测模型的系数。预测值所选择的相似时间数据需根据预测时间段与历史时间段的相关性选取,最终得到所有与主变油温异常缺陷发生有关系的监控数据的变化趋势预测公式。相关目标监控数据通过皮尔逊系数检验获得。

1.3 典型设备缺陷预测

在得到典型缺陷发生期间发生异常的监控数据及其对应的变化趋势函数后,即可将根据函数得到未来时间段的相关监控数据变化曲线。将变化曲线跟该典型缺陷的多个历史缺陷记录进

行对比求出相似度后,取相似度最高的为该时间该典型缺陷的相似度指标。其具体步骤如图 1 所示。

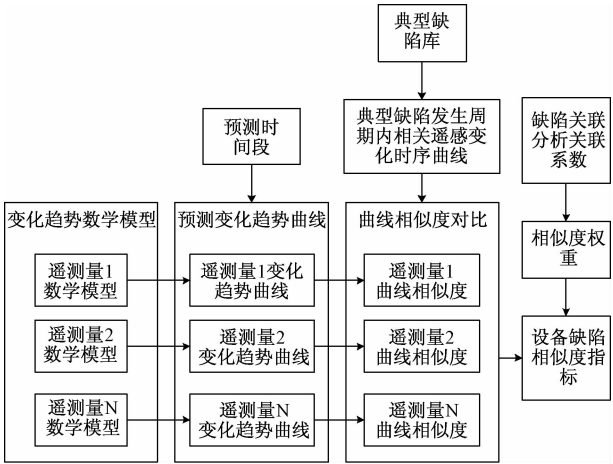


图1 典型设备缺陷预测过程

Fig. 1 The process of predicting of typical equipment defect

(1) 根据需预测的时间点及该典型缺陷的周期得出需预测的时间段。

(2) 将需预测的时间段——代入该典型缺陷相关的监控数据预测函数中,得出每个相关监控数据在该时间段的变化趋势曲线。

(3) 将得到的变化趋势曲线与该典型缺陷的历史发生时间内的监控数据实测曲线进行相似度对比,得出每个监控数据的相似度系数。

(4) 将每个监控数据的相似系数进行加权平均计算,其中权重为各监控数据在该缺陷发生时数据异常变化的程度,最终结果即为该时间点的设备缺陷相似度指标。

(5) 依次计算每个需要预测的时间点,得出预测时间段的设备缺陷相似度曲线,即可根据曲线的取值高低判断每个时间点发生该典型缺陷的可能性。

下面主要针对典型设备缺陷预测中涉及的曲线相似度计算和缺陷相似率指标计算进行详细讨论。

1.3.1 曲线相似度对比

曲线相似度对比采用皮尔逊系数来衡量两条曲线的相似性,可以避免两条曲线形状相似而取值差距很大被认为不相关的情况。

计算过程如下:

$$Y_{XY} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}} \quad (6)$$

式(6)中: $X_i、Y_i$ 表示变量 $X、Y$ 的观测值; $\bar{X}、\bar{Y}$ 表示变量 $X、Y$ 的样本均值。

1.3.2 缺陷相似率指标计算

由于影响缺陷可能性评价的因素往往是众多而复杂的,如果仅从每一项单一的维度曲线相似度指标上对缺陷可能性进行综合评价不尽合理,因此需要将反映缺陷可能性的多项指标的信息加以汇集,得到一个综合指标,以此来从整体上反映缺陷可能性的整体情况。综合指标处理的方法有多种,由于每个典型缺陷其相关维度都经过了筛选,因此影响缺陷的各个指标之间没有很强的相关性。在计算综合指标时,不需要考虑指标间的相关关系,只需要采用以加权平均为基础的指标评分法即可。其实质是将多个指标用加权平均的方法综合成一个评价价值。

2 案例分析

未检验所提方法的有效性,设计如下实验进行验证。

测试数据来自于某电力公司 500 kV 主变历史油温异常缺陷记录,实验在 Hadoop 平台上进行。实验所需的设备监控信息存储在 Hbase 中,利用离线计算组件完成所需模型的训练,并利用实时计算组件实现电力设备缺陷的在线预测。

针对油温异常缺陷计算其发生前一天及缺陷发生时有异常变化的监控数据。以主变 A 相油温 1 为例。

(1)针对主变 A 相油温 1 的历史数据集合,划分成等距长度区间,再对区间里的数据集进行建模,获得区间里数据集的统计特征。

(2)将每一个分布所代表的结点用全连通的无向图连接并存储。

(3)对每一条结点之间的边进行独立性统计假设检验,利用深度优先算法,对每一条边,使用最大拟然估计对边的转移概率分布的参数进行学习。

如表 1 所示为主变 A 相油温 1 在学习完成后的转移概率矩阵计算结果。

将缺陷发生时的主变 A 相油温 1 历史数据代入到上述的概率转移矩阵中,得到如表 2 所示的转移概率序列。

表 1 主变 A 相油温 1 转移概率矩阵

Table 1 The transition probability matrix about a phase of main transformer oil temperature 1

状态	区间 1		区间 2		区间 3		区间 4	
	油温/ ℃	概率	油温/ ℃	概率	油温/ ℃	概率	油温/ ℃	概率
1	24.12	0.145	29.02	0.226	31.77	0.255	32.21	0.373
2	24.32	0.252	29.05	0.147	31.78	0.185	32.24	0.415
3	24.37	0.126	28.93	0.379	31.78	0.192	32.24	0.300
4	24.45	0.115	29.00	0.204	31.79	0.206	32.25	0.474

表 2 主变 A 相油温 1 概率序列

Table 2 The probabilistic sequences of a phase of main transformer oil temperature 1

时间状态	主变 A 相油温 1 计算值	转移概率
3:40	27.744 995	0.312 5
3:41	27.812 141	0
3:42	27.830 452	0
⋮	⋮	⋮
3:20	27.854 874	0.232 8

根据转移概率序列,在 3:41 时出现大批量 0 值。因此主变 A 相油温 1 的异常时间点为第一天 3:41,依次类推求出其他该缺陷相关监控数据的异常时间点。其中主变 A 相油温 1 的异常时间点最早。根据历史记录,该缺陷结束时间为第二天3:20,由此得出该缺陷记录的实际发生周期为 23 h 39 min。

根据式(4)计算各监控数据随时间变化的时间序列模型,预测各监控数据未来时间的变化曲线。采用式(5)计算未来时间变化曲线与历史缺陷实测曲线的相似性。对于油温异常缺陷,将同一主变第二次发生油温异常缺陷的时间点代入,得到各监控数据在缺陷周期内的相似性计算结果如表 3 所示。

表 3 各监控数据相似性计算结果

Table 3 The results of similarity calculation about remote measurements

监控项	相似度	权重	结果
A 相油温 1	0.91	0.26	0.236 6
A 相油温 2	0.88	0.18	0.158 4
B 相油温 1	0.85	0.18	0.153
高压侧 A 相电压	0.95	0.02	0.019
高压侧 B 相电压	0.98	0.02	0.019 6
高压侧电流值	0.52	0.03	0.015 6
低压侧无功	0.85	0.17	0.144 5
低压侧电流值	0.82	0.11	0.090 2
高压侧功率因数	0.89	0.03	0.026 7
$\bar{x}$			0.863 6

经过加权平均后,最终典型缺陷相似性指标为 0.863 6,同时计算出缺陷发生时间前后的相似度部分指标计算值如图 2 所示。

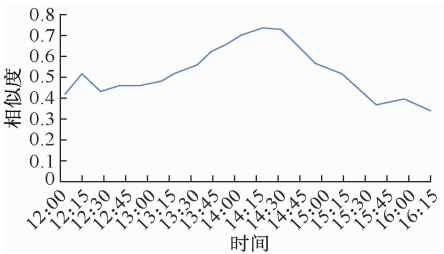


图 2 各时间点相似度指标计算结果

Fig. 2 The results of similarity calculation in different time points

综合上述实验结果可知,在缺陷实际发生时,根据所提方法计算出的相似度指标形成了一个峰值,而缺陷未发生时的相似度指标较低。这证明所提方法能够较好地体现缺陷发生的可能性,对设备发生缺陷的情况具有较好的预测效果。

3 结论

(1)设计了一种基于大数据挖掘的集中监控设备典型缺陷预测方法,分析缺陷发生时每种不同监控数据的异常变化幅度,并根据这种异常变化幅度对每一种监控数据赋予不同的权重,实现基于多种不同类型的监控数据对设备缺陷的预测,以避免传统的基于数据驱动的缺陷预测方法由于仅考虑某一种监控数据所带来的预测准确度不高的缺点。

(2)所提方法为了降低监控数据的预测误差,引入了时间序列预测技术,对大量统计数据进行数学处理,确定因变量与某些自变量的相关关系,以建立回归方程(函数表达式),用于预测今后的因变量的变化,能够提高预测准确性。

(3)为了提高设备缺陷预测的准确度,所提方法综合考虑多种不同的监控数据变化曲线与典型缺陷发生时的监控数据变化曲线的相似度,并在此基础上获得一个综合指标,以实现从整体上客观评价缺陷发生的可能性,使得所提方法更具应用优势。

参 考 文 献

- 1 荣 辉,李 勋,吕启深,等. 基于数据挖掘的电力设备缺陷分析研究[J]. 电气应用, 2015, 21(2): 46-50.
Rong Hui, Li Xun, Lü Qishen, et al. Research on defect analysis of power equipment based on data mining[J]. Electrotechnical Application, 2015, 21(2): 46-50.
- 2 彭 宇,刘大同,彭喜元. 故障预测与健康管理技术综述[J]. 电子测量与仪器学报, 2010, 24(1): 1-9.

Peng Yu, Liu Datong, Peng Xiyuan. Summary of fault prediction and health management technology[J]. Journal of Electronic Measurement and Instrumentation, 2010, 24(1): 1-9.

- 3 Vachtsevanos G, Lewis F, Roemer M, et al. Fault diagnosis and prognosis performance metrics[C]// Intelligent Fault Diagnosis and Prognosis for Engineering Systems. New York: John Wiley & Sons, Inc. 2007.
- 4 Adams D E. Nonlinear damage models for diagnosis and prognosis in structural dynamic systems[J]. Proceedings of SPIE: The International Society for Optical Engineering, 2002, 47(33): 180-191.
- 5 严英杰,盛戈皞,江秀臣. 基于大数据分析的输变电设备状态数据异常检测方法[J]. 中国电机工程学报, 2015, 32(1): 52-59.
Yan Yingjie, Sheng Gehao, Jiang Xiuchen. Anomaly detection method for state data of power transmission and transformation equipment based on big data analysis[J]. Proceedings of the CSEE, 2015, 32(1): 52-59.
- 6 王林川,宋超翼,吴铁山,等. 管理信息专家系统在电厂设备缺陷管理中的应用[J]. 东北电力学院学报, 2001, 25(4): 71-73.
Wang Linchuan, Song Chaoyi, Wu Tieshan, et al. Application of management information expert system in power plant equipment defect management[J]. Journal of Northeast Electric Power University, 2001, 25(4): 71-73.
- 7 Wang W Q, Golnaraghi M F, Ismail F. Prognosis of machine health condition using neuro-fuzzy systems[J]. Mechanical Systems & Signal Processing, 2004, 18(4): 813-831.
- 8 Zhang S, Ganesan R. Diagnostics of slow rotating bearings using a novel DAI Based on acoustic emission[C]// Advances in Condition Monitoring of Machinery in Non-Stationary Operations. Berlin: Springer, 2016.
- 9 Wang P, Vachtsevanos G. Fault prognostics using dynamic wavelet neural networks[J]. Artificial Intelligence for Engineering Design Analysis & Manufacturing, 2002, 15(4): 349-365.
- 10 Yildirim M, Gebrael N Z, Xu A S. Integrated predictive analytics and optimization for opportunistic maintenance and operations in wind farms[J]. IEEE Transactions on Power Systems, 2017, 32(6): 4319-4328.
- 11 薛禹胜,赖业宁. 大能源思维与大数据思维的融合(一)大数据与电力大数据[J]. 电力系统自动化, 2016, 40(1): 1-8.
Xue Yusheng, Lai Yening. Integration of macro energy thinking and big data thinking[J]. Automation of Electric Power Systems, 2016, 40(1): 1-8.