

大数据挖掘分析在电力设备状态评估中的应用

宫宇, 吕金壮

(中国南方电网超高压输电公司, 广州 510663)

摘要: 为提高电力设备状态评估的准确性和效率, 提出了将大数据挖掘分析应用于电力设备状态评估的思路和方法。介绍了大数据挖掘分析的架构, 将电力设备状态的多维度数据解析为静态、动态、准动态和外部参数四大类, 分析了数据关联规则、关联度和权重, 最后给出了大数据挖掘分析技术的应用前景。

关键词: 电力设备状态评估; 多维度数据; 大数据挖掘; 关联分析

Application of Big Data Mining Analysis in Power Equipment State Assessment

GONG Yu, LÜ Jinzhuang

(EHV Power Transmission Company, CSG, Guangzhou 510663, China)

Abstract: To improve the accuracy and efficiency of the power equipment condition assessment, a big data mining analysis method is proposed to evaluate the power equipment condition. The structure and phases of big data mining analysis are introduced. Multi-dimensional data in power equipment conditions are categorized into four types as static, dynamic, quasi-dynamic and exterior parameters. The basic rules in data association rule, association ratio and weight are analyzed. The trend of big data mining analysis towards power equipment condition assessment is put forward.

Key words: power equipment condition assessment; multi-dimensional data; big data mining; Correlation Analysis

电力设备状态检修与设备运行的可靠性、经济性紧密相关。设备状态评价结果可用于电网设备规划设计、物资采购、退役报废等资产生命周期环节的管理工作, 以及故障预测、故障类型判断及故障处理等环节。关注电力设备状态检修是交直流互联电网发展、国民经济命脉产业提升的保障性工作^[1]。

目前, 电力设备状态评估的方案包括基于标准导则和打分制的评估方法、基于有限数据的设备故障诊断专家系统、基于远程专家介入的自学习诊断系统, 以及基于多维度数据和传统机器学习的设备状态评价专家系统等^[2]。用于设备故障诊断^[2]的传统机器学习方法通常基于有限的样本训练, 构建输入量对输出量的可预测关系。其中输入量、输出量在构建的算法模型中都是事前确定无法变动的。基于远程专家介入的诊断系统将专家意见作为新样本

训练系统, 提高后续诊断结果的准确可靠性。

大数据挖掘技术为电力设备状态评价开辟了一条全新的路线, 并对现有设备状态监测参数提出了更高的要求^[3-5]。该方法引入了数理统计、模式识别等方面的理论和工具, 在大规模数据分析的基础上, 重点挖掘不确定模型条件下被分析因素之间的关联性。本文以大电网设备状态评估为目的, 在传统评估方法基础之上, 提出状态评估的大数据概念, 旨在将传统的大电网行业注入新兴概念与算法。

1 设备状态评估大数据挖掘关键问题

通常认为, 大数据具有“5V”特点: 数据量 (volume)、多样性 (variety)、速度 (velocity)、价值 (value) 和真实性 (veracity)^[3]。数据库在分析前需要“缓慢”导入数据, 数据库的统计与建模并不与大量数据有效平行^[3]。而大数据变革的关键是样本即为整体, 所有量都是状态量, 无输入输出之分、无因果关系之别, 通过网络算法强调任何两件事物的关联关系。大数据发展的每个阶段^[4-7]都面临着

基金项目: 国家高技术研究发展计划(863计划)(2012AA050209)。
Foundation item: Supported by the National High Technology Research and Development Program of China (863 Program) (2012AA050209)。

巨大的挑战,如图1所示。要将大数据应用于大电网设备状态评估领域,需要建立设备状态(正常/故障/缺陷)的所有关联网络。

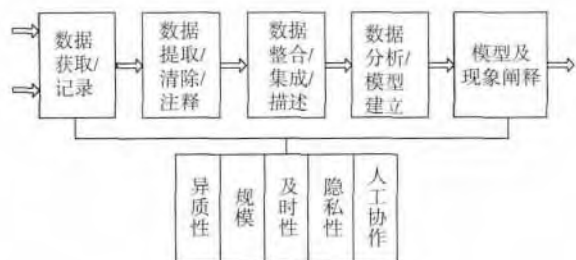


图1 大数据研发阶段系统图

Fig. 1 Phases of Big data

与大电网设备状态紧密相关的数据来源繁杂,包括传感器数据、实验数据、日志记录等。文献[6]提供了一种有效的方法,可以用于解决大数据面临的生数据存储、分析、检索和有效提取等问题。

适用于大数据挖掘设备状态信息的方法目前有机器集群和分布式并行化技术,并行技术主要包括将高性能计算系统使用的信息传递界面^[8](message passing interface, MPI)应用于大规模集群的映射-化简(MapReduce)。MAD^[8](magnetism, agility and depth)表征用户从大数据分析系统期望得到的特征。如大数据分析软件Hadoop^[9]软件堆栈,它是一种MAD系统,包含了可扩展的MapReduce模型^[10],该模型用于处理文本信息,包括映射、结合、分割、分类及简化的算法步骤,十分方便。

大电网设备状态量的获取信息中同样存在着大量的生数据,包括定位、故障类型、故障分析、故障预测等文字信息。通过MapReduce模型可将这些数据转换为计算机可识别操作的数据,然后将大数据网络算法应用于设备状态评估之中即可建立设备状态与故障/缺陷的关联关系,实现准确定位、故障判断及故障预测等功能。

图2是设备状态评估不同方法的关系框架,与传统的打分制和加权计算、设备故障诊断专家系统相比,大数据挖掘分析方法具有更高的技术要求和水平,其结果更加准确、可靠性更高。

2 设备状态评估多元多维度数据源

多元多维度数据源是开展设备状态评估大数据挖掘分析的基础条件。设备状态信息包括设备全生命周期范围内适用于设备健康状态及风险评估的各

种资料、数据、记录等,以更新频率和数据来源为依据,可以将所有数据划分为4大类参数。

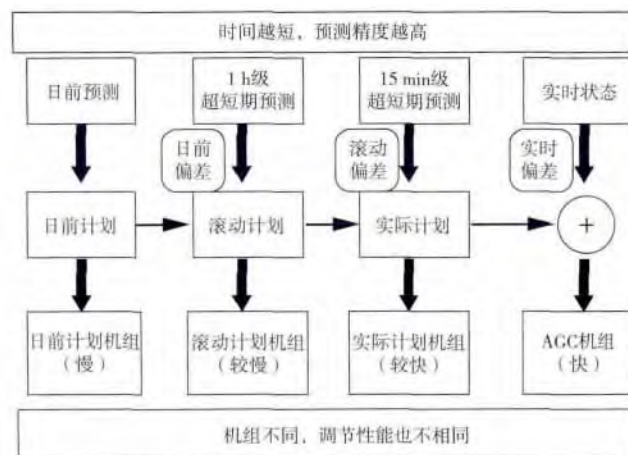


图2 大电网设备状态评估关系框架

Fig. 2 Framework of the Relationship Between Different State Assessment Methods

2.1 静态参数

静态参数包括设备台帐参数和设备投运前的试验参数等。静态参数自记录后不再发生变化,作为某项状态参量评分时的参考值和判断依据,用作初始值、限制阈值等。

2.2 动态参数

动态参数包括运行记录数据、巡视记录参数、带电检测参数、在线监测参数等,定期获取并更新,周期通常以天为单位,最大周期不超过1个月。动态参数数据更新周期较短,时效性较好,是最主要、最关键的设备状态评估数据来源,能最及时地反应设备的健康状态及风险。

2.3 准动态参数

准动态参数包括检修试验参数、缺陷/故障参数等。准动态参数通常定期或不定期获取并更新,周期通常以月为单位。相比动态参数,其时效性相对较差,但对于设备状态评估的准确性起到关键作用。采用多维度设备状态评估方法分析设备潜在的具体缺陷/故障时,准动态参数的作用尤为重要。

2.4 外部参数

外部参数主要包括:温度、风力、降水等气象因素,地震、洪水等环境因素,以及用电量、经济态势等社会经济因素。

3 设备状态评估大数据挖掘方法

设备状态评估大数据挖掘分析的基本思路是利

用大数据挖掘技术,如聚类、关联等算法,对设备良好运行状态下的大量的历史数据进行挖掘分析,以得到对应于设备在不同的操作模式下正常运行的各种参数指标值;并对基于各种故障与缺陷状态下的全景历史数据进行深度的关联分析,得到设备各种监测参数与设备不同健康状态之间的对应关系。其中各参数指标的关联规则,以及各状态量的关联度及其权重是挖掘分析的核心基础。

3.1 关联规则

关联规则在设备故障诊断中至关重要。关联规则挖掘是数据挖掘领域的一个重要方面,通过关联规则对设备的故障、缺陷状态与可能反映设备各方面状态的全景参数进行相关性分析,可以实现设备多个状态量的有效组合、关键状态参量提取与合并、状态权重的精细化选择等有益的结果。

但是,在大数据的时代背景下,现有的基于数据挖掘的电力设备状态评估思路的实用性并不一定可以得到保证。大数据背景下,电力设备的状态监测数据也趋于结构多样化,出现了越来越多的非结构化数据如视频、音频、图像等。此类数据的建模和分析技术尚未能充分满足评估需求。

此外,面对海量数据,部分算法如经典 k-Means 算法^[11]无法在合理的时间范围内挖掘出有效数据,需要用模糊理论加以改进以适用于大数据背景下的聚类分析需求。同样,经典的频繁项集生成算法——Apriori 算法需要多次扫描事务数据库,其 I/O 负载开销及响应处理能力尚无法达到要求^[10]。

因此,各种适用于大数据分析的深度关联规则分析便应运而生,如基于粒计算的关联规则挖掘算法^[12]、基于压缩矩阵的 Apriori 算法^[13]、基于云计算平台 Hadoop 的快速关联规则增量更新算法(C-FUP)^[14]等等。

3.2 状态量关联度分析

在传统的有限数据设备状态评估系统中,状态量的选取往往是单一的,故障/缺陷情况也是单一的,两者对应的关系规律性弱。采用设备状态评估大数据挖掘分析方法,基于多元多维度数据进行状态评估,重点考察设备状态量(包括设备权限、故障的结果性状态量)之间的关联性。

“状态量关联度”概念的核心是与部件/设备等具体对象相关的多元多维度状态量之间存在一定程度的相关性,数学上定义“状态量关联度”为状态向

量之间的协方差矩阵:

设 $X = [x_1, x_2, \dots, x_n]$, 其中每个元素代表一个状态向量, $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T, i = 1, 2, \dots, n$ 。则“状态量关联度”可定义为 $\Omega = \text{corr}(X)$ 。

通常,反映较多信息的不是“状态量相关度”而是“状态量变动关联度”。它反映了当前部件/设备的运行异常情况,是对设备或电网的某种调节实施效果的评价依据。因此令 $\Delta X = [\Delta x_1, \Delta x_2, \dots, \Delta x_n], i = 1, 2, \dots, n$, 则“状态量变动关联度”可定义为 $\Sigma = \text{corr}(\Delta X)$ 。

多维度状态量的实质是在测定了各种状态量的基础上,对部件/设备整体当前或潜在故障/缺陷的评估。对于大电网而言,状态量与设备状态之间有非常复杂的非线性函数关系,此处不关注输入输出的具体传递函数关系,记作 $S(X)$ 并二阶泰勒展开: $S(Y) = S(Y_0) + \Delta S(Y - Y_0) + (Y - Y_0) H(Y - Y_0)^T$ 。(3) 式中, H 为传递函数 $S(Y)$ 的 Hessian 矩阵, $Y = [x_{1t}, x_{2t}, \dots, x_{nt}]$ 表示部件/设备在时刻 t 的状态向量,上式变形可写作:

$$S(Y) = \alpha_0 + \sum_{i=1}^n \alpha_i (x_{it} - x_{i0}) + \sum_{i=1}^n \sum_{j=i}^n \beta_{ij} (x_{it} - x_{i0}) (x_{jt} - x_{j0}) \quad (4)$$

式中, $m = 1, 2, \dots, 0.5n(n+1)$ 。由于状态向量是可以获取的,待设定参数 α_i, β_i 共有 $0.5(n+1)(n+2)$ 个后,即可确定传递函数,对整体部件/设备状态进行评价。以上给出的是 n 维度状态量相关性的数学模型。

3.3 权重分析

在设备状态评估大数据挖掘分析过程中,设备状态参量的权重系数是建立分析和决策模型的重要参数,需要构建合理的权重系数与修正模型。各权重数据按以下原则确定。

依照调度规程、反事故措施、风险预控措施、特维方案、预试方案等对设备重要性的分析和划分,考虑设备缺陷/故障记录统计情况,设定设备/部件的重要性权重系数。基于设备缺陷/故障记录统计情况,考虑设备缺陷/故障的影响程度及其后果严重性,设定设备/部件的缺陷/故障严重性权重系数;根据设备/部件权重、状态变量能够反应的缺陷/故障类型、相应的缺陷/故障严重性权重,设定状态参量重要度权重;结合历史数据,根据监测技术获得的状态监测量的准确及有效程度,设定监

测技术成熟度权重系数。

利用历史数据及新进样本,对设备综合评估成果进行长期的动态跟踪和分析,利用回归拟合、数据挖掘等技术手段,对权重系数进行不断的修正,可以根据设备状态监测参数分析得到更及时、准确的设备健康状态和风险指标。

4 设备状态评估大数据挖掘应用前景

随着电网运行监测设备的大量使用,电力设备数据海量化的趋势将越来越明显。面对如此海量的数据,实现大数据挖掘分析在电力设备状态评估领域成功应用的关键点,在于开发诸如基于云计算平台 Hadoop 的快速关联规则增量更新算法(C-FUP)的大数据挖掘算法^[14-15]。

美国纽约爱迪生电力公司对高危井盖的分析处理是电力设备状态大数据挖掘分析的典型案例。该公司针对每年多起电缆井盖爆炸造成的严重事故,采取“大数据”手段统计出106种预警先兆,预测2009年可能出事的井盖并严加监控。结果位列前10%的高危井盖中,预测准确率达44%^[10]。

在我国,一些技术较为领先的风力发电机制造厂商,也已投入研发力量,进行基于海量设备状态信息的大数据分析工作。风机厂商安装在风力发电机各部位的传感器,收集诸如叶片载荷、叶片温度、叶片结冰、机舱及塔筒晃动等数据,并通过大数据挖掘分析方法,开展风机各部件的状态分析、预测故障、制定检修策略的工作,在秒级的采样频率下,已达到每天10亿个数据的海量数据规模^[16]。

目前,在国内的电网企业中,尚未有大数据挖掘分析在电力设备状态评估领域的具体应用实例,但有理由相信,随着电网设备状态信息的爆炸式增长,基于大数据挖掘方法实现电力设备状态精细化评估是未来必然的发展趋势。

5 结语

本文以电力设备状态评估作为研究对象,在传统的打分制和加权计算、设备故障诊断专家系统基础上,将大数据挖掘方法引入到电力设备状态评估中,从更深层次上重新评估电力系统海量数据的逻辑关系和使用价值。建立了电力设备状态评估的大数据分析架构,展望了适用于电力设备状态评估的相关大数据挖掘技术的发展趋势。对电力设备状态评估中的具体问题则

有待进一步开展相关的研究与分析。

参考文献:

- [1] 张金萍,刘国贤,袁泉,等. 变电设备健康状态评估系统的设计与实现[J]. 现代电力,2004(8): 45-49.
ZHANG Jinping, LIU Guoxian, YUAN Quan, et al. Design and Realization of Evaluation of Health Condition System for Transformation Equipment [J]. Modern Electric Power, 2004 (8): 45-49.
- [2] 黄建华,全零三. 变电站高压电气设备状态检修的现状及其发展. 电力系统自动化,2001,25(16): 56-61.
HUANG Jianhua, QUAN Lingsan. Current Status and Development of Condition-Based Maintenance of High-Voltage Electric Power Equipment in Substation [J]. Automation of Electric Power Systems, 2001, 25(16): 56-61.
- [3] MADDEN S. From Databases to Big Data [J]. IEEE Internet Computing, 2012(5/6): 4-6.
- [4] BHATIA A, VASWANI G. Big Data - A Review [J]. International Journal of Engineering Science & Research Technology, 2013, 2(8): 2102-2106.
- [5] MAYER-SCHÖNBERGER V. 大数据时代[M]. 盛杨燕,周涛,译. 杭州: 浙江人民出版社,2012.
- [6] DAS T K, KUMAR P M. Big Data Analytics: A Framework for Unstructured Data Analysis [J]. International Journal of Engineering and Technology, 2013, 5(1): 153-156.
- [7] KRANZLMÜLLER D, KACSUK P, DONGARRA J. Recent Advances in Parallel Virtual Machine and Message Passing Interface [C]//11th European PVM/MPI Users' Group Meeting Budapest, Hungary, September 19-22, 2004, Hungary: EPUG, 2004.
- [8] COHEN J, DOLAN B, DUNLAP M, et al. New Analysis Practices for Big Data [C]//VLDB '09, Lyon, France, August 24-28, 2009, Lyon, France: VLDB, 2009: 1481-1492.
- [9] HERODOTOU H, LIM H, LUO G, et al. Starfish: A Self-tuning System for Big Data Analytics [C]//CIDR 2011, 5th Biennial Conference on Innovative Data Systems Research (CIDR '11), January 9-12, 2011: 261-272.
- [10] KATSOV I. MapReduce Patterns, Algorithms, and Use Cases [EB/OL]. (2012-02-01) [2014-03-10]. <http://highlyscalable.wordpress.com/2012/02/01/mapreduce-patterns/>.
- [11] 赵洪昌. 云计算下的关联分析和模糊聚类研究[D]. 南京: 南京信息工程大学, 2013.
- [12] 刘小霞. 面向海量数据的关联规则挖掘算法研究[D]. 北京: 北京邮电大学, 2007.
- [13] 苗苗苗. 数据挖掘中海量数据处理算法的研究与实现[D]. 西安: 西安建筑科技大学, 2012.
- [14] 朱晓峰. 面向海量数据的快速挖掘算法研究[D]. 南京: 南京邮电大学, 2012.
- [15] 车斌. 基于Hadoop海量数据处理关键技术研究[D]. 成都: 电子科技大学, 2013.
- [16] 黄树红, 李建兰. 发电设备状态检修与诊断方法[M]. 北京: 中国电力出版社, 2008.

收稿日期: 2014-03-18

作者简介:

宫宇(1963),男,吉林吉林人,高级工程师,学士,从事大电网交直流工程设计、建设、运行等电力专业技术管理工作;

吕金壮(1974),男,河南商丘人,高级工程师(教授级),博士,从事直流输电及高电压技术研究与应用工作(e-mail) ljzcomcom@163.com。