文章编号: 1006 - 4729(2012)02 - 0176 - 05

数据挖掘技术在电力设备状态检修中的应用

杨国庆[1], 张 宇[2]

(1. 上海电力学院 电力与自动化工程学院,上海 200090;
2. 上海市电力公司 技术与发展中心,上海 200122)

摘 要: 介绍了数据挖掘的基本过程、方法及在我国电力系统的应用现状;分析了数据挖掘在电力设备状态检修和设备故障诊断中的应用状况,给出了基于数据挖掘的电力设备状态评估模型,并以关联规则方法为例进行了分析.

关键词: 数据挖掘技术;电力设备;状态检修
中图分类号: TP274; TM762.1+4 文献标志码: A

Application of Data Mining Technology in Power Equipment Condition-based Overhaul

YANG Guo-qing[1], ZHANG Yu[2]

(1. *School of Electric Power and Automation Engineering, Shanghai University of Electric Power, Shanghai* 200090, *China*;
2. *Technology and Development Center, Shanghai Municipal Electric Power Co., Shanghai* 200122, *China*)

Abstract: The basic procedures, methods of data mining and its practical application to the power industy in China are introduced, the present situation of data mining used in condition-based maintenance in China is analysed, and the state evaluation model based on data mining of power equipment is given, and the example of association rules is used to analyse the application of data mining to equipment fault diagnosis.

Key words: data mining technology; power equipment; condition-based overhaul

随着信息技术的迅速发展,数据库规模的扩大带来了大量的数据. 但大量的数据往往使人们无法辨别隐藏在其中的能对决策提供支持的信息,而传统的查询和报表工具无法满足挖掘这些信息的需求[1].

数据挖掘技术就是通过处理大量的数据并从中抽取有价值的潜在信息的一种新的数据分析技术. 随着电力系统数据库中的数据也呈爆炸性趋势增长,在电力设备的在线监测系统中引入数据挖掘技术势在必行.

1 数据挖掘

数据挖掘就是从大量的数据中挖掘出有用的信息,即从大量的、不完全的、有噪音的、模糊的、随机的实际应用数据中发现隐含的、规律性的、人们事先未知的,但又是潜在有用的并且最终可理

收稿日期: 2011 - 09 - 14

通讯作者简介: 杨国庆(1984 -),男,在读硕士,河南周口人. 主要研究方向为基于在线监测系统的输电线路动态增容. E-mail: yangguoqing_2004@163.com.

解的信息和知识的非平凡过程[2]. 近年来,数据挖掘技术的研究和应用发展迅猛,在商业和银行领域已有广泛的应用. 随着相关学科的发展出现了各种数据挖掘算法[3].

数据挖掘的一般过程包括以下 5 个方面:

(1) 数据选择　指与要解决的问题有关的数据的选取,当前,数据可以很方便地通过互联网进行采集;

(2) 数据预处理　即通过滤噪进行信息处理的过程;

(3) 数据转换　将定性的数据转换成定量的数据,在某种意义上也叫特征提取;

(4) 数据挖掘　寻找数据库中隐藏的重要模式,该步骤在知识发现的过程中起着关键的作用;

(5) 数据解释　用来评价和解释通过数据挖掘获得的结果,即知识.

从数据挖掘中发掘有关知识的规则需要具备以下 4 个条件:

(1) 有效性　意味着规则或知识的重要性,得到的规则或知识应适用于未知的数据;

(2) 新颖性　即与先验的知识没有关系,在实践中,重要的是要发现新规则;

(3) 有用性　目的在于发现对用户有用和感兴趣的规则;

(4) 简单性　即发现的规则应力求简单,应能够创建和容易解释复杂的数据.

数据挖掘技术与其他的研究领域如统计分析、机器学习、数据库,以及可视化技术等密切相关. 数据挖掘的方法和数学工具包括关联规则、统计学、决策树、神经网络、线性规划、模糊逻辑等.

目前,数据挖掘在我国电力部门的应用还不多,它在电力系统中的应用包括电力设备状态评估、电力系统负荷预测和分类、电力系统的运行模式分类、电力系统运行状态、设备监控、电力设备故障诊断、电力调度优化、电力系统建模等[4].

2　数据挖掘在电力系统中的应用

随着电力设备监测系统中数据规模的不断扩大,电力新技术的发展,以及各种监测设备的使用,将数据挖掘技术与状态检修相结合也是大势所趋. 目前,致力于这方面的研究和应用越来越多,并取得了一定的成果.

国外有关数据挖掘方法在电力系统中应用的论文统计结果如图 1 和图 2 所示[5]. 从图 1 和图 2 可以看出,在电力系统领域的数据挖掘方法中,决策树是主流的方法,占 88.6%; 使用数据挖掘解决的具体问题主要有安全评估(48.8%)、故障诊断(11.6%)、电力系统控制(9.3%)、负荷预测(6.9%)、负荷拟合(6.9%)等. 由于电力设备的状态检修就是以状态在线监测和故障诊断(占 11.6%)为基础的,因此数据挖掘技术在电力设备状态检修领域的应用也占据不小的比重,发展空间广阔. 此外,应用于故障诊断的主要方法有决策树、人工神经网络、统计分析和粗糙集等.

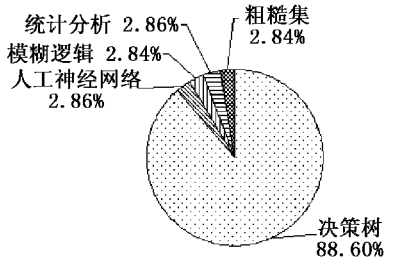


图 1　国外各种数据挖掘方法应用于电力系统领域统计

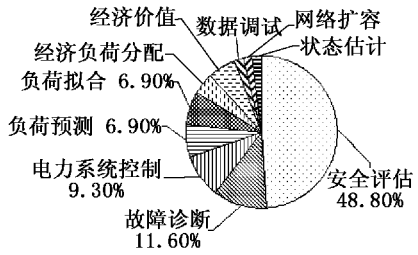


图 2　数据挖掘解决的主要具体问题

国内数据挖掘在状态检修中的应用并不普遍,仅在个别企业或部门少量使用; 技术也不成熟,总体还处于初级阶段. 目前,数据挖掘技术在发电厂设备、变压器、配电网设备及高压输电线路等的状态检修中已有研究和应用,如: 采用模糊聚类和粗糙集理论分析汽轮机轴系振动的数据,得出相关规则对机组进行故障诊断[6]; 将粗糙集技术应用于高压输电线路的故障诊断[7]; 将决策树算法应用于变电站的电气设备在线监测系统[8]等.

电力系统中,数据挖掘的处理过程要求有较深厚的电力系统知识,而该方面专业人才较为缺乏; 目前电力设备的监测设备不够完善,只能从不完备的状态数据中挖掘规律; 状态数据收集分布

于各单位，缺少公共数据平台. 这些都制约了状态检修的发展.

国家电网公司已经提出统一输变电设备状态监测系统的要求，以满足坚强智能电网建设及生产精益化管理工作要求. 目前福建电力公司、华东电网公司和华北电网公司都相继建立了状态监测系统，并与电力生产管理系统(PMS) 实现数据共享，个别单位还开发了检修辅助决策系统. 但在数据处理技术上普遍实力不足，这样便造成海量数据的利用率不高，且缺乏专门的机构负责技术的研发和推广.

以上海为例，目前上海市电力公司正在大力研究和推广在线监测，为状态检修打好基础. 公司正在建设一个输变电设备状态监测中心，已经构造了状态监测的系统框架，中心建设的技术方案也已编制完成. 建设工作在不断进行中. 接下来的工作将涉及对在线监测数据的研究分析、数据挖掘技术的应用研究.

图 3 为状态检修管理流程，其中的数据管理和专家诊断两个环节将涉及数据挖掘技术的应用研究.

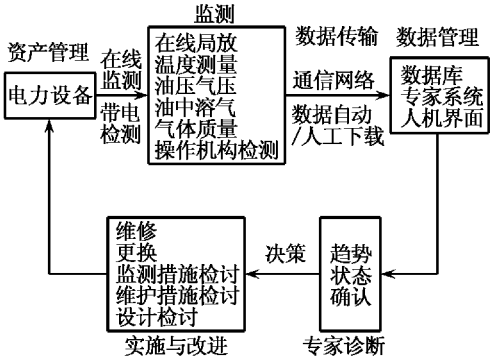


图 3 状态检修管理流程示意

在专家诊断和状态评估的过程中，将构建基于设备故障模式分析结果的诊断知识库，采用人工诊断和基于人工神经网络、模糊逻辑等数据挖掘技术相结合的方法，以实现对设备缺陷和故障的诊断.

目前电网中的部分运行设备已有比较成熟的状态监测参数及监测手段，能够对设备的状态进行明确量化阐述. 对此国家电网公司也发布了相关标准，提出了几大类电力设备的状态评价导则和状态检修导则. 若电力设备出现导则中明确的缺陷，其问题会迅速显现. 但部分设备状态监测参数或监测手段还未成熟，不能对设备状态进行明确量化阐述，这就需要设备专家通过数据中心的数据挖掘功能，充分利用不能量化阐述的设备状态参数，对设备进行综合诊断，以便发现设备的隐性问题.

图 4 和图 5 分别是设备的显性问题和隐性问题的触发处理流程. 对比图 4 和图 5 可以看出，两类问题的处理过程存在差别. 设备隐形问题的相关处理步骤可以作为处理显性问题不足时的补充，而对于隐形问题，目前设备专家可以利用数据挖掘技术对设备进行缺陷确认或故障诊断.

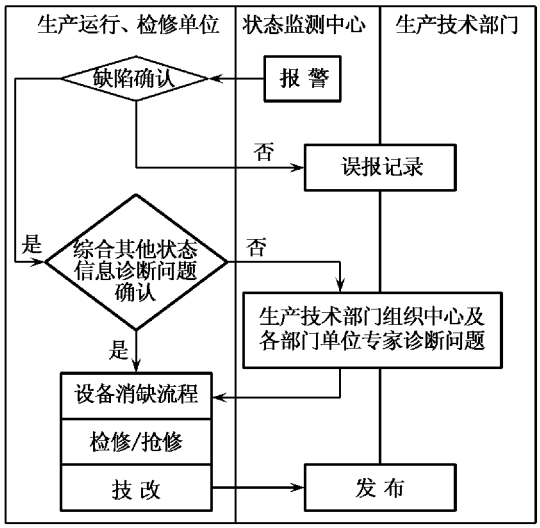


图 4 设备显性问题触发处理流程

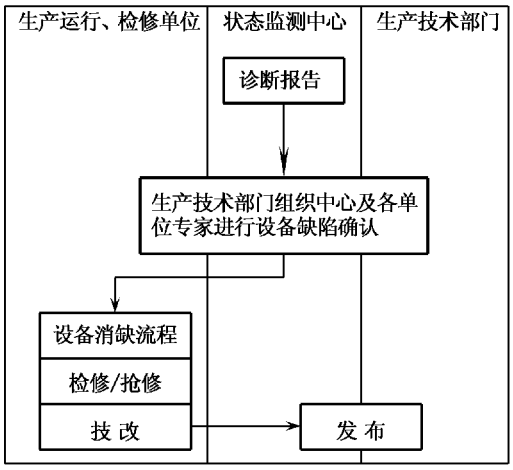


图 5 设备隐性问题触发处理流程

3 基于数据挖掘的电力设备状态评估系统

状态检修实质上就是建立一套设备的运行状

态分析系统来判断设备的健康状况,以决定是否对其进行检修. 基于数据挖掘的电力设备状态评估模型如图 6 所示[9].

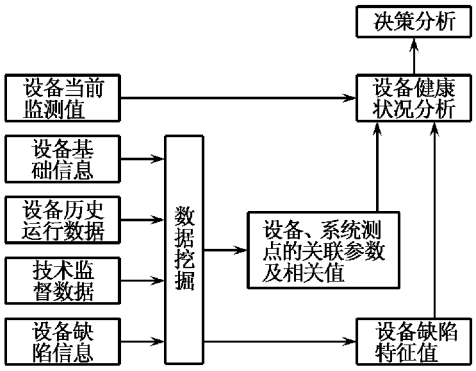


图 6 基于数据挖掘的电力设备状态评估模型

对电力设备状态进行评估,首先要收集设备的相关信息,包括设备的基础信息、历史运行数据及设备缺陷信息等. 通过对设备历史运行数据和缺陷信息进行挖掘,得出设备缺陷情况下的特征值和设备关联参数值,方便用户对设备的历史情况进行查询和使用; 设备当前监测值从实时数据库或试验中获得,是被分析的对象; 设备健康状况分析就是以数据挖掘的结果为依据,对设备当前监测值进行对比分析,判断当前设备运行状态是否正常.

关联规则是数据挖掘方法中的一个重要分支,通过分析数据库中不同数据属性之间存在的潜在关系,找出满足给定支持度和置信度的关系规则,对设备进行在线故障诊断[10,11]. 表 1 是某电厂汽轮机振动报警记录.

表 1 参数报警历史记录

报警时间	事件记录	事件标志
2003 - 02 - 01 01:13:23	A 振动报警	A
2003 - 02 - 01 01:33:02	C 振动报警	C
2003 - 03 - 12 08:40:30	A 振动报警	A
2003 - 03 - 12 08:42:16	D 振动报警	D
2003 - 03 - 12 10:00:38	C 振动报警	C
2003 - 05 - 28 19:35:12	B 振动报警	B
2003 - 05 - 28 19:49:52	A 振动报警	A
2003 - 05 - 28 20:01:06	C 振动报警	C
2003 - 06 - 12 05:12:52	A 振动报警	A
2003 - 06 - 12 05:20:06	B 振动报警	B

利用关联规则挖掘算法,选择最小支持度为 20%,最小置信度为 20%(关于关联规则的算法描述及支持度、置信度的计算公式参见文献[2]和文献[10]). 对表 1 进行分析后的故障现象记录如表 2 所示.

表 2 分析表 1 后的故障现象记录

TID	事务项	TID	事务项
1	A,C	3	B,A,C
2	A,D,C	4	A,B

注: I —所有项目的集合,即数据库的所有字段; D —所有事务的集合,即数据库; T —部分项目的集合,T 包含在 I 中,每个事务可以用唯一的标识符 TID 来表示.

通过分析可获得以下的强关联规则:

A⇒C,支持度 = 0.75,置信度 = 0.75.

表明在 A 振动报警后不久就会出现 C 也振动报警. 该规则可以向运行人员提供早期故障预警和故障原因分析,从而在故障发生前将其排除,以确保电力设备的安全运行.

4 结 语

近年来随着状态检修的蓬勃发展,以及各大电网公司状态监测中心的建设,为数据挖掘技术的研究和应用提供了基础平台. 同时电力设备监测系统中数据规模的增大,也推动了数据挖掘技术在电力设备状态检修中的研究和应用. 但数据挖掘在电力设备状态检修中的应用还处于初级阶段,仍存在许多问题,如: 面临多维超大量数据集的挑战; 如何正确处理冗余信息和噪音数据的问题; 挖掘结果的无效性问题等. 因此,电力设备状态检修中的数据挖掘技术还需作进一步的深入研究.

参考文献:

[1] 陈超金. 基于数据挖掘的电力设备状态检修技术研究综述[J]. 广东电力,2009,22(9):21-24.

[2] 张云涛,龚玲. 数据挖掘原理与技术[M]. 北京: 电子工业出版社,2004:1-2.

[3] 王光宏,蒋平. 数据挖掘综述[J]. 同济大学学报,2004,32(2):246-252.

[4] 何友全. 数据挖掘方法及其在电力系统故障诊断中的应用研究[D]. 成都: 西南交通大学,2004.

[5] HIROYUKI Mori. State-of-the-art overview on data mining in power systems [C]//Power Engineering Society General Meeting,2006:1-4.

[6] 张勤. 基于数据挖掘技术的汽轮机性能分析系统的研究与实现[D]. 北京: 华北电力大学,2007.

[7] 廖志伟,孙雅明. 基于数据挖掘模型的高压输电线系统故障诊断[J]. 电力系统自动化,2001,25(15):15-19.
[8] 周茜. 数据挖掘技术在电气设备在线监测系统中的应用[J]. 科技情报开发与经济,2006,16(8):209-210.
[9] 成永强. 基于数据挖掘的设备状态检修[D]. 北京:华北电力大学,2005.
[10] 梁志瑞,陈鹏. 关联规则挖掘在电厂设备故障监测中的应用[J]. 电力自动化设备,2006,26(6):17-19.
[11] LI Jian-qiang, WANG Song-ling, NIU Cheng-lin, *et al*. Research and aplication of data mining technique in power plant[C]// International Symposium on Computational Intelligence and Design,2008:250-253.

(编辑 胡小萍)

(上接第 170 页)

(2) 流量控制保证每个关键业务独立占用 10% ~30% 的基本带宽;

(3) 必须保证日常应用中,将 CPU 的利用率控制在 50% 以内,内存利用率控制在 50% 以下.

2.6.2 广域网故障处理要求

(1) 广域网异常故障必须在 8 min 内反映到监控系统上;

(2) 运行维护人员 24 h 对广域网监控系统进行监控;

(3) 运行维护人员在 15 min 内查明异常情况,并及时上报至网络负责专工;

(4) 网络专工 1 h 内抵达现场或远程安排技术人员实施应急方案;

(5) 2 h 内故障问题仍然发生,需上报运行维护专家,并上报省公司.

3 结 语

随着供电企业信息化程度越来越高,广域网承载的业务应用日渐增多,其安全防护管理体系将更加多样化、复杂化. 因此供电企业在吸取广域网管理理论精华的同时,更应该结合企业自身特点,有选择、有步骤地将最新成果应用到实际项目中去,这样才能不断提高项目管理的效率.

参考文献:

[1] O'Reiuy 公司. 广域网络[M]. 陈明,译. 北京:中国电力出版社,2000:67-68.
[2] 崔海涛,王童童,贺靖民. IP 网络管理系统功能需求分析和描述[J/OL]. [2007-02-12] http://wenku.baidu.com/view/39e0b33143323968011c9213.html.
[3] BAIDU 文库网. 北塔 BTNM 技术白皮书 3.6.3. [EB/OL]. [2011-12-02] http://wenku.baidu.com/view/ec99a60d4a7302768e99396b.html.
[4] 曾明,李建军. 网络工程与网络管理[M]. 北京:电子工业出版社,2003:734-736.

(编辑 苏 娟)