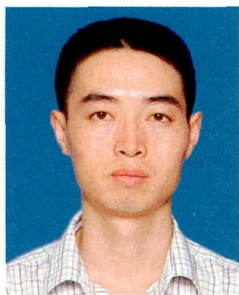


基于数据挖掘的电力设备缺陷分析研究

在线监测技术的成熟与推广为电力设备缺陷数据的分析提供了更宽阔的维度空间。针对传统缺陷数据分析的不足,利用数据挖掘方法对设备缺陷数据进行了多维展示分析、相关性分析和对应分析,并结合在线监测数据和SCADA数据对设备缺陷数据进行了潜在特征规律的发现尝试。基于数据挖掘的数据分析思路对电力设备缺陷数据的应用分析研究具有较强的参考价值。

黄荣辉 李 勋 吕启深/深圳供电局有限公司 刘典安/深圳市康拓普信息技术有限公司



黄荣辉/高级工程师

高质量的电力设备是保证电网安全运行、提高供电可靠性的基础。在实际运行中,有些电力设备虽然能继续使用,但运行状态发生异常或存在隐患,并将影响人身、设备和电网安全,电网和设备的可靠经济运行,设备出力或寿命以及电能质量等。这种异常或隐患称为缺陷。南方电网公司输变电设备缺陷管理标准^[1]定义了缺陷严重程度分类,明确了缺陷处理流程及单位职责,建立了缺陷指标考核体系,并通过建设安全生产管理信息系统来实现对电力设备缺陷的流程化管理。

目前对设备缺陷的分析主要集中在缺陷发生类型和数量、缺陷消缺率和消缺及时率的统计方面^[2, 3]。实际上,缺陷数据包含的内容十分丰富,且随着在线监测技术的不断成熟与大力推广,为设备缺陷数据的分析提供了更宽阔的维度空间。

传统的数据分析工作一般在处理数据时,会预先假定事物之间存在某种因果关系,然后在此因果关系假定的基础上构建模型并验证预先假定的因果关系。在数据规模剧增、数据结构复杂以

及变量错综复杂的大数据背景下,从数据到结论的过程升级为一个非常复杂的人机互动比较和选择的过程。数据挖掘工具能够辅助人们探索和分析数据中一些有启发性的结构,发现数据中有价值的模式和规律^[4]。本文将采用数据挖掘分析方法对电力设备缺陷数据进行深入分析和探讨。

数据挖掘

数据挖掘是从大量的、不完全的、有噪声的、模糊的和随机的实际应用数据中发现隐含的、规律性的信息和知识的技术,是统计学、数据库技术和人工智能技术的综合。

数据挖掘的任务就是发现隐藏在数据中的模式。根据模式特征,可将模式大致细分为如下^[5]:

1) 分类模式。分类就是构造一个分类函数,把具有某些特征的数据项映射到某个给定的类别上。

2) 聚类模式。聚类就是将数据项分组成多个类或簇,类之间的数据差别应尽可能大,类内的数据差别应尽可能小。与分类模式不同的是,聚类中要划分的类别是未知的。

3) 回归模式。回归模式的函数定义与分类模式相似,主要差别在于分类模式采用离散预测值,而回归模式采用连续的预测值。许多问题可以用线性回归解决,对于许多非线性问题可以对变量进行转换,从而转换为线性问题来解决。

4) 关联模式。关联模式是数据项之间存在的关联规则,是在同一事件中出

关键词/Keywords

设备缺陷·
数据挖掘·
多维展示·
相关性分析·

现的不同项之间的相关性。

5) 序列模式。序列模式是描述基于时间或其他序列的经常发生的规律或趋势, 并对其建模。序列模式将关联模式和时间序列模式结合起来, 重点考虑数据之间在时间维度上的关联性。

6) 偏差模式。偏差模式是对差异和极端特例的描述, 如聚类外的离群值。大部分数据挖掘方法都将这种差异信息视为噪声而丢弃, 然而在一些应用中, 罕见数据可能比正常数据更有用。

在数据挖掘过程中, 每个模型和算法具有不同的适用场景, 最终评价的结果一定是以业务价值为导向, 而非模型自身的评价指标。

缺陷数据内部分析

设备缺陷自身包含的数据信息很丰富, 以生产管理系统中的一条缺陷数据为例, 数据内容包括: 缺陷设备、设备类型、缺陷设备生产厂家、投运日期、出厂日期、所属站线、电压等级、所属部门、缺陷部件、缺陷描述、缺陷严重等级、缺陷类别、缺陷来源、缺陷状态、发现时间、发现人员、消缺人员、计划消缺时间、实际消缺时间、是否停电、缺陷原因、处理措施、处理结果和验收时间等。针对如此多的数据内容, 如果仅凭数据表格和表单方式, 想要清晰准确地掌握缺陷信息比较困难。

本文以深圳供电局2007~2013年变压器缺陷数据为例, 首先采用可视化技术实现缺陷数据的多维展示, 然后分别采用相关性分析、对应分析方法对主变压器缺陷各主要参量的内在关系进行了描述性研究。因为数据样本有限, 数据挖掘分析结果可能与实际有差别, 本文仅涉及对数据挖掘方法的应用研究。

1. 缺陷数据多维展示

数据挖掘的第一个工作目标是增进对复杂数据内容的了解。可视化技术强调数据便捷形象的展现, 可帮助对数据进行准确的描述, 并找到进一步研究和解释的途径。

(1) 缺陷概貌

分析缺陷历史数据, 一般需要对缺陷数据的

整体情况有一个定性的概念, 如哪个部门发生的缺陷多, 哪个厂家的设备发生的缺陷多, 哪类缺陷发生的次数多等。图1采用散点矩阵图对缺陷数据中的所属部门、设备生产厂家、设备投运年限和缺陷类别4个维度信息进行了展示。图1中密度的大小直接反映了缺陷发生次数的多少。

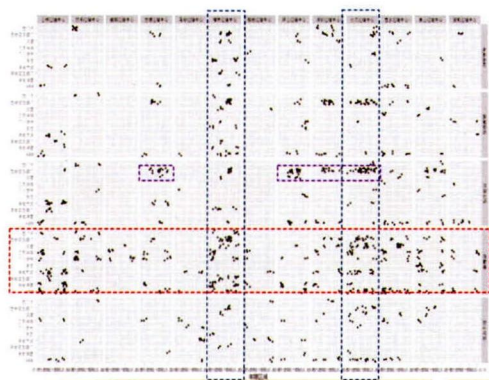


图1 缺陷概貌展示

由图1可知, 红色框内的数据表示渗漏油缺陷发生的次数最多, 紫色框内的数据表示某变压器厂生产的主变压器容易发生冷却系统缺陷, 而蓝色框的数据表示这两个部门的变压器设备发生缺陷的次数更多。

(2) 缺陷主题河

在了解缺陷整体情况的基础上, 可能需要了解某些历史数据在时间维度上的统计情况。采用主题河 (ThemeRiver) 技术来可视化展示缺陷部件和缺陷类别的统计结果, 如图2所示。

通过图2可直观地发现风机和油枕是最常发生缺陷的设备部件, 而渗漏油则是发生次数最多的缺陷类型。

(3) 缺陷脸谱图

脸谱图是用脸谱来表达多变量的样品, 是将观察的各变量分别用脸的某一部位的形状或大小来表示。图3是采用脸谱图来表示所属部门、设备生产厂家、设备投运时间和缺陷类别4个维度的统计信息。

其中脸高度所代表的某变压器厂在水贝的高度最高, 即表明该变压器厂的变压器设备在水贝发生的缺陷次数最多; 而鹏城、马坳的脸形状小

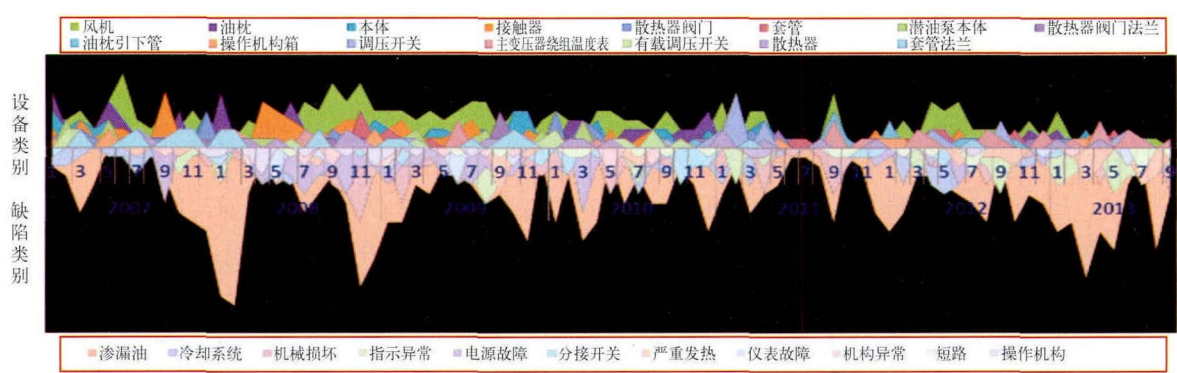


图2 缺陷主题河展示



图3 缺陷脸谱图

且嘴形小，即表明这两个地方的变压器各项缺陷指标数都较小。

2. 缺陷相关性分析

相关分析就是探索现象之间相关关系的密切程度和表现形式，利用相关系数反映B随A的变化情况，其值的大小代表关系的强弱关系。采用相关性分析对缺陷类型之间、缺陷类型与投运年限之间的关系进行研究，可以增进对缺陷发生原因的理解。

采用可视化技术展示缺陷类型间的相关分析，结果如图4所示。蓝色代表正相关，颜色越深

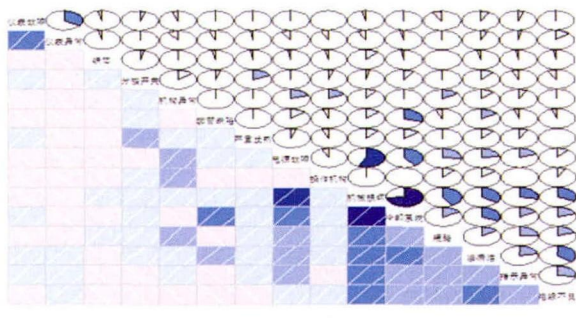


图4 缺陷类型间相关性分析结果

代表正相关性越强；粉红色代表负相关，颜色越深代表负相关性越强。由图4可知，以油杯和风扇破裂为代表的机械损坏与冷却系统故障、机械损坏与电源故障之间有较强的关联关系，即表明某些变压器在发生机械损坏缺陷时可能也会出现冷却系统故障或电源故障。

图5是缺陷类型与投运年限之间的相关分析结果。投运年限10~15年与冷却系统具有很强的相关性，说明主变压器在投运10年以上容易发生冷却系统的缺陷；而投运15年以上与机械损坏之间也有强相关性，说明主变压器在投运15年以上发生的机械损坏次数会增多。

	渗漏油	冷却系统	机械损坏	指示异常	电源故障	分接开关	严重发热	
投运1~5年	0.37	-0.05	-0.02	0.02	0.22	0.04	0.15	
投运5~10年	0.37	0.54	0.30	0.18	0.28	0.23	0.08	
投运10~15年	0.47	0.82	0.64	0.25	0.38	0.15	0.15	
投运15年以上	0.52	0.46	0.71	0.38	0.40	0.02	0.13	
	仪表故障	机构异常	短路	操作机构	仪表异常	套管缺陷	铁心	绝缘不良
投运1~5年	0.08	0.21	-0.09	0.17	0.21	0.05	0.02	0.09
投运5~10年	0.19	0.16	0.09	0.06	0.15	0.39	0.02	-0.07
投运10~15年	0.03	0.07	0.26	-0.02	-0.04	0.29	-0.01	0.18
投运15年以上	0.02	0.12	0.40	0.02	-0.07	0.02	0.01	0.43

图5 缺陷类型与投运年限间相关性分析结果

3. 缺陷对应分析

对应分析也是一种用来研究变量与变量之间联系紧密程度的分析技术。对应分析是在因子分析的基础上发展起来,同时把变量与样本反映到相同的坐标轴的一张图上,变量与样本距离越近,说明越相似。

缺陷原因与缺陷部件之间的对应分析结果如图6所示。其中,红色表示缺陷发生的原因,黑色表示变压器设备的部件。由图6可知,散热器阀门可能常因漏油、松动等原因发生缺陷;风机因卡涩、烧损等原因易发生缺陷;而操作机构箱因元器件故障易发生缺陷。

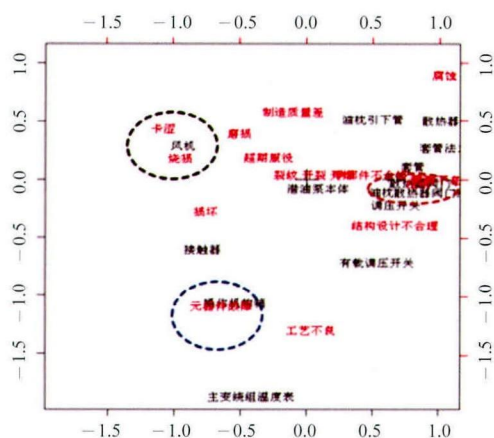


图6 缺陷原因与缺陷部件间对应分析结果

缺陷与其他数据的挖掘分析

在线监测技术的成熟使得电力设备的在线监测应用越来越广泛,如GIS局部放电在线监测和变压器油色谱在线监测基本成为220 kV及以上变电站的标准配置。在线监测为设备管理提供了良好的数据来源,加上电力自动化系统采集的实时电气量数据,以及生产管理系统中不断累积的工单和表单数据,逐步形成了设备管理的大数据。

本文将以变压器设备为例,以大数据观念分析变压器缺陷数据与油色谱在线监测数据、变压器SCADA数据之间的关系,尝试对设备状态与数据特征的关系进行重新发现。分析结果仅针对样本数据有效,不具备通用参考价值,但分析方法在实际应用中具有借鉴作用。

1. 缺陷与在线监测数据的分析

油色谱在线监测主要对变压器等油浸电力设备进行在线监测,并及时准确检测出绝缘油中溶解的各种故障特征气体浓度及变化趋势。这些气体包括氢气、一氧化碳、甲烷、乙烷、乙烯和乙炔等。目前常用的三比值法、大卫三角法等故障分析方法都是通过经验总结出来的,预先设定了气体数据特征与设备状态的映射关系,然后通过在线监测数据进行相应的分析来确定变压器是否发生故障及其发生故障原因。

针对设备缺陷与在线监测数据的大量样本,采用决策树方法来分析设备不同状态下油色谱气体含量,以是否发生缺陷为因变量,油色谱各气体含量为自变量,建立分类模型。决策树分析模型如图7所示。

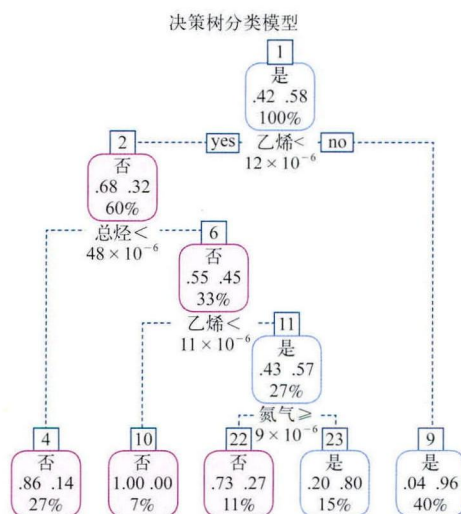


图7 决策树分析模型

在决策树分析模型中,乙烯含量是否大于 12×10^{-6} 是一个比较关键的节点,这在以后对变压器状态的评估时,可以对乙烯含量值的这个指标更为关注些。

2. 缺陷与在线监测数据、SCADA数据的分析

变压器SCADA电气量数据包括有功功率、无功功率、电流和电压的时序历史值。SCADA数据和油色谱在线监测数据都是实时采集的,并按照时间顺序进行历史保存,但两者的采集周期不一致。由于关注对象及应用场景的不同,以前的研究很少将SCADA数据、在线监测数据和变压器缺

陷数据结合起来分析。

本文采用大数据观念尝试对变压器缺陷数据、油色谱气体含量和SCADA数据间的关系进行分析。油色谱气体含量与SCADA数据的相关性分

析结果如图8所示。油色谱中的乙烷、甲烷、一氧化碳、二氧化碳及氢气与电流、功率间无明显相关性；乙烯、总烃与电流、有功及无功间存在明显的相关性。

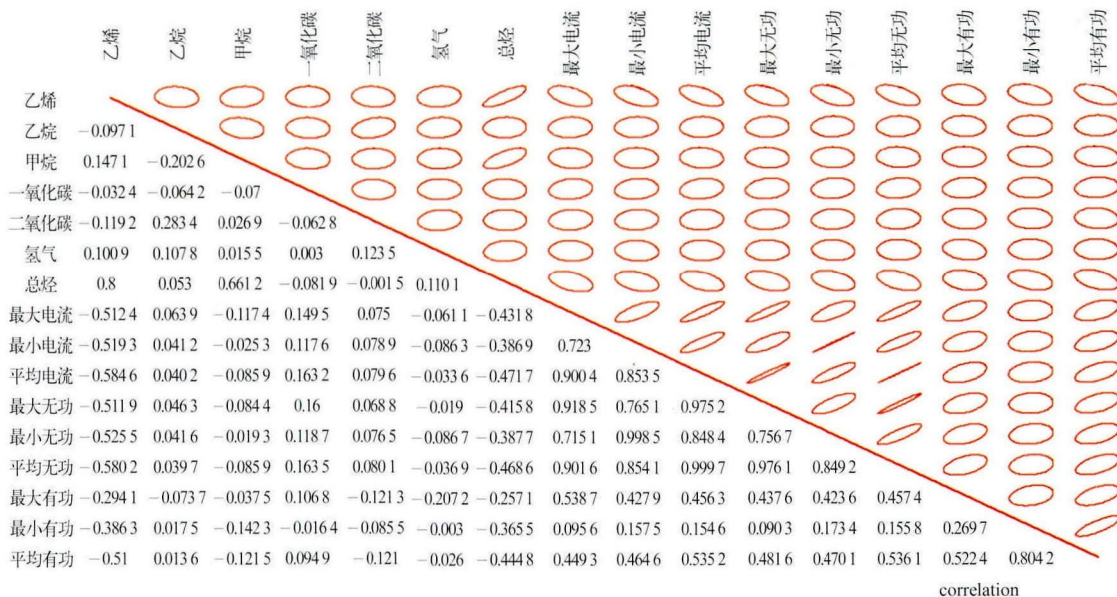


图8 在线监测与SCADA数据的相关性分析

以乙烯含量、平均无功为例，设备缺陷数据、油色谱数据及SCADA数据的相关统计结果如图9所示。结果表明，在乙烯大于 15×10^{-6} ，而平均无功在(30, 47)区域内缺陷发生的可能性较大。

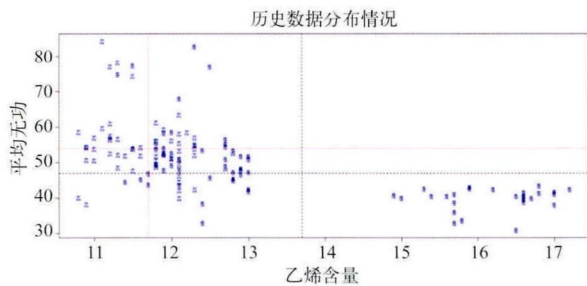


图9 缺陷、油色谱及SCADA数据统计分析结果

结束语

本文采用数据挖掘方法对电力设备缺陷数据的进行了多维度分析，对主变压器缺陷主要参量的内在关系进行了描述性研究，并尝试对设备状

态与各类数据的特征关系进行了重新发现。本文的分析思路对大数据背景下的设备缺陷数据应用研究具有较强的参考价值，后续将在深入分析数据特征原理的基础上对各类电力设备进行更有针对性的分析研究。

参考文献

- [1] 南方电网公司. Q/CSG 10701—2007 输变电设备缺陷管理标准[S]. 广州: 南方电网公司, 2007.
- [2] 王虎. 变电设备缺陷分析与管理[J]. 技术与市场, 2013, 20(10): 55-56.
- [3] 张颖才. 浅谈变电设备缺陷分析和应用[J]. 科技创业, 2011(12): 151-152.
- [4] 王星. 大数据分析: 方法与应用[M]. 北京: 清华大学出版社, 2013.
- [5] 王光宏, 蒋平. 数据挖掘综述[J]. 同济大学学报, 2004, 32(2): 246-252.

(收稿日期: 2014-09-24) EA