

The Feasibility Analysis Report of Data Mining Technology in the Secondary Device Status Evaluation

Shishuang Wang, Bin Lu, Qingjie Zhou, Jianliang Wei, Mingzhu Yuan, Chengyong Yuan

School of Computer Science and Technology, The North China Electric Power University, Baoding
Email: chenshuang_1023@163.com

Received: Oct. 19th, 2012; revised: Oct. 29th, 2012; accepted: Nov. 13th, 2012

Abstract: Face the characteristics of power system data that mixed type, poor quality, high requirements, real-time, the extraction and processing of the valid data is becoming imperative research topic. For the state evaluation of the power system secondary equipment, the validity and accuracy of the data is more important, it will directly affect the evaluation results of the state evaluation, as well as the formulation of the overhaul plan. In this paper, through the analysis of the problem and inadequacy of basic data processing in existing equipment condition evaluation, and the feature of rough set and neural network algorithm of data mining, put forward a kind of ideas that combine the two algorithms for processing basic data, so as to achieve the purpose of more efficient and more objective for providing effective quantity of state for the state assessment. As there is currently no examples that using data mining technology in the secondary equipment condition evaluation, hope for providing a very good reference value for the research of data mining technology in the secondary device status evaluation in the future.

Keywords: Secondary Equipment; State Evaluation; Data Mining; Rough Sets; Neural Network

数据挖掘技术在二次设备状态评价中的可行性分析报告

王师霜, 鲁 斌, 周庆捷, 魏建亮, 袁明珠, 袁成勇

华北电力大学计算机科学与技术学院, 保定
Email: chenshuang_1023@163.com

收稿日期: 2012年10月19日; 修回日期: 2012年10月29日; 录用日期: 2012年11月13日

摘 要: 面对电力系统数据种类混杂、数据质量差、要求高、实时性等特征, 有效数据的提取和处理就成为了势在必行的研究课题。而对于电力系统二次设备的状态评价来说, 数据的有效性和准确性更加重要, 将直接影响到状态评价的评价结果以及检修计划的制定。本文通过分析现有对二次设备状态评价基础数据处理的方法存在的问题和不足, 以及数据挖掘算法中粗糙集与神经网络算法的特征, 提出了一种将这两种算法结合用于处理基础数据的思想, 从而达到更高效、更客观的为状态评价提供有效状态量的目的。由于目前还没有将数据挖掘技术运用于二次设备状态评价中的例子, 因此希望能为今后在二次设备状态评价中的数据挖掘技术研究提供很好的参考价值。

关键词: 二次设备; 状态评价; 数据挖掘; 粗糙集; 神经网络

1. 引言

随着国民经济的不断发展, 电网规模的不断扩大, 输变电设备在数量、质量上的迅速提升, 以及用户对供电可靠性要求的逐步提高, 原先的周期性定

检, 容易引起设备过度检修或检修不当, 造成装置损坏和拒动、误动, 对电网安全带来了极大的影响, 检修工作量剧增和人员不足矛盾也日益突出, 难以适应精益化管理和电网发展的需要。为此国家电网公司提

出了供电设备的状态检修，并制定了部分设备的状态检修标准及评价导则。

为了做好二次设备状态检修工作，真正达到“应修必修，修必修好”的目的，就必须规范和处理好状态检修工作中的各个环节。二次设备状态检修的流程主要包括以下几个环节：数据采集数据处理→对状态量进行状态评价→检修计划制定→审核并实施检修。可以看出，要想制定出合理的检修计划，就必须得出正确的状态评价结果，而状态评价过程的第一步就到数据采集及处理。因此，数据的采集和处理过程就显得尤为重要。在现有二次设备状态检修系统中，状态评价模块中状态量的确定是根据以往经验对基础数据的分析判断人为敲定的，费时费力，当状态量信息不够充分的时候，还需要以往经验人为的增加较为合理的数据以确保最终评价结果的准确性，以上这些势必会造成人为因素对结果的影响过大，降低了结果的准确性。因此，急需寻求一种合理的数据处理方法来弥补现有方法的不足，从而得到更加准确的状态评价信息，提高状态检修的可靠性。

2. 数据挖掘算法在国内的应用现状

数据挖掘(Data Mining)是从大量数据中发现潜在规律、提取有用知识的方法与技术^[1]。因为与数据库密切相关，因此又被成为数据库知识发现(Knowledge Discovery in Databases, KDD)。数据库知识发现是在1989年8月于美国底特律市举行的第十一届国际人工智能会议上正式形成的。目前，KDD国际研讨会的研究重点逐渐从发现方法等理论研究转向系统应用研究，注重多种发展策略和技术的集成，以及多种学科之间的相互渗透^[2]。近年来，数据挖掘技术在国内也更加被重视、进行了进一步研究和实际应用。

目前，国内的数据挖掘技术应用主要集中在当前数据挖掘应用主要集中在电信(客户分析)、农业(行业数据预测)、零售(销售预测)、网络日志(网页定制)、电力、银行(客户欺诈)、生物(基因)、天体(星体分类)、医药、化工等方面^[3]。它能解决的问题主要在于：数据库营销(Database Marketing)、客户群体划分(Customer Segmentation & Classification)、背景分析(Profile Analysis)、交叉销售(Cross-Selling)等市场分析行为，以及客户流失性分析(Churn Analysis)、客户信

用记分(Credit Scoring)、欺诈发现(Fraud Detection)等等，并且在许多领域都得到了成功的应用。

然而，看似其涉及的领域非常广泛，但实际应用的普及度还远远不够。就电力行业来说，数据挖掘技术在作为现代化生产主力的工业部门的应用并不多，只在电力部门火力发电方面有一些系统的应用，其他领域只是少数零散的应用。图1是数据挖掘在电力系统方面应用的示意图。

通过这个图可以看出，对于电力系统这样一个数据信息非常繁多复杂的、地域涉及非常广的系统来说，数据挖掘技术的应用程度还远远不够。而本文所要探讨的是其在二次设备状态评价中对基础数据处理方面的一个应用。

3. 粗糙集及神经网络概述

粗糙集和神经网络是数据挖掘领域中最常用的两种技术。本文所提到的方法就是将这两种技术有效融合于二次设备状态评价的基础数据处理过程中，希望通过这个方法更高效、更可靠的把基础数据处理成状态评价所需要的状态量。

3.1. 粗糙集理论

粗糙集理论，是继概率论、模糊集、证据理论之后的又一个处理不确定性的数学工具。它是由波兰数学家 Z. Pawlak^[4]在1982年提出的。粗糙集理论与其他处理不确定和不精确问题理论的最显著的区别是：它无需提供问题所需处理的数据集合之外的任何先验信息，所以对问题的不确定性的描述或处理可以说是比较客观的。粗糙集可以实现对信息系统的知识约简，去除冗余的信息，减少输入信息的空间维度，提高处理效率。粗糙集理论有着许多显著的优点：

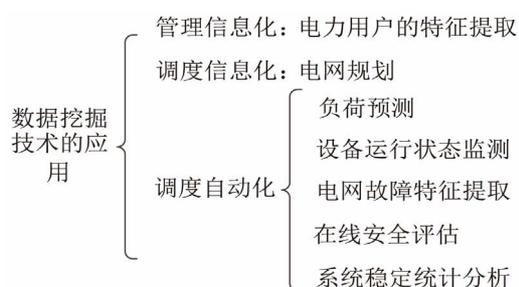


Figure 1. Application of data mining in power system
图1. 数据挖掘在电力系统中的应用

1) 粗糙集理论有一整套处理数据分类问题的数学方法, 当数据不确定、不完整和不精确的时候尤为适用。

2) 基于粗糙集的计算方法非常适合并行处理, 适用于大量计算和编程。

3) 糙集理论和神经网络、模糊逻辑、概率推理、信度网络、连接计算、遗传算法等一起形成了软计算方法的基础, 为问题的处理提供了鲁棒性强和成本较低的解决方案。

4) 粗糙集的分析方法无须任何先验知识, 仅利用数据本身提供的信息即可。

5) 粗糙集理论的实用性非常强, 具有较强的容错能力。

6) 粗糙集理论能够发现数据中隐含的模式和关系, 从而对数据进行属性约简, 评价数据的重要性, 从数据中产生易于理解的规则。

虽然粗糙集的优点很多, 但是也存在一些缺陷和不足, 如抗干扰能力较差, 对于噪声较为敏感, 在噪声较大的环境中表现得不尽如人意。

3.2. 神经网络

神经网络是在现代神经生物学研究成果的基础上发展起来的一种模仿人脑信息处理机制的网络系统^[5]。1943年 McCulloch(心理学家)和 Pitts(数理逻辑学家)发表文章提出了 M-P 模型, 提出了神经元的数学描述和网络的结构方法, 标志着神经计算时代的开始。随后又有人提出了赫布规则、感知器等概念, 掀起了神经网络研究高潮。1986年由 Rumelhart 和 McClelland 为首的科学家小组^[6]提出了一种前项反馈神经网络——Back Propagation 网络(简称: BP 网), 它是一种按误差反向传播算法训练的多层前馈网络。BP 网络能学习和存贮大量的输入 - 输出模式映射关系, 而无需事前揭示描述这种映射关系的数学方程。它的学习规则是使用最速下降法, 通过反向传播来不断调整网络的权值和阈值, 使网络的误差平方和最小。BP 神经网络模型拓扑结构包括输入层(Input)、隐层(Hide Layer)和输出层(Output Layer)。

神经网络的特点就是通过训练和学习产生一个非线性的映射, 模拟人的思维方式, 具有很好的自适应性, 可以实现有监督和无监督的学习, 并能够对信

息进行并行处理; 同时, 它具有很好的抑制噪声的能力。但是神经网络的缺陷是它无法对输入的信息进行有用性或冗余性的判断, 因此不能对输入的信息进行简化, 这使得它在处理空间维数较大的信息时会很困难和低效。

通过以上对粗糙集和神经网络特征的分析可以看出, 两者在优缺点上具有非常好的互补性。因此, 将粗糙集作为神经网络的前端处理器, 去除冗余信息, 再根据需要设计多层神经网络, 对处理过的数据进行学习和训练, 这样得到的结果将更加有效、合理, 更适用于实际工业生产。

4. 数据挖掘技术在二次设备状态评价中的应用分析

近年来, 很多学者采用不同的数学方法尝试建立发电设备状态评估模型, 但是这些方法多数用于一次设备的状态评价, 对于二次设备状态评价来说这一块还是空白。

现有对二次设备状态评价基础数据的处理主要是对从现场采集以及其他系统直接或间接获得的基础数据, 通过一定的分析方法和数据处理方式进行的数据信息的处理过程。通过分析对象原始数据集, 根据评价业务需求进行必要的过滤、换算、组合等数据的加工和处理过程, 使其成为能够反映设备健康状态的状态量数据, 以供设备状态评价使用。数据处理过程如图 2 所示。

在整个的处理过程中, 对基础数据进行加工的方法都存放于方法库中。建立处理方法库是根据状态评价要求, 建立满足现有业务并具有可扩充性和可外挂的方法库, 以供数据过滤、换算、组合等数据加工和处理过程使用。这些方法主要都是对每一个基础数据处理判断得到一个分值从而得到信息点数据(例如:

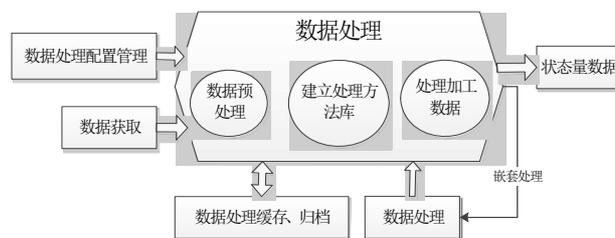


Figure 2. The data processing
图 2. 数据处理过程

温度是否符合要求值),然后根据共同确定某一状态量的几个信息点的值得到一个状态量的分值,从而表征该设备在这一阶段在此状态量上的性能好坏。该处理方法主要存在以下几点问题:

1) 影响设备运行状态的属性的重要程度没有通过科学理论的方式得到认证,采集的某些设备属性或环境状态属性对设备运行状态影响不大;

2) 每一次处理数据都是对每一条数据先进行分析得到信息点,再通过处理得到状态量,运算步骤较多,不能一步到位,并且当数据量很大的时候运算效率较低;

3) 对于每一类原始数据的处理算法都不同,需要编写多个算法程序,一旦算法有所改变程序就得相应变化。

综上所述,综合现有算法的不足之处以及对粗糙集和神经网络算法的分析可知,粗糙集与神经网络的互补性正好弥补了现有算法中存在的种种问题,既能去除冗余信息提高算法效率,又能通过自学能力一步到位的将原始数据转化成所需要的状态量信息,而且整个过程只需要最初对神经网络进行训练即可,无需人为干预,无需过多的先验经验即可得到较为客观的、合理的、准确的结论。因此,基于粗糙集与神经网络的数据挖掘算法对于二次设备状态评价基础数据的处理来说是可行的。

5. 结束语

电力系统的正常运转与对其可靠有效的保护息

息相关,而对一次设备的保护目前已经做的比较完善,二次设备保护方面还有很多欠缺的地方。为了保障二次设备状态检修工作顺利进行,状态评价环节非常关键。因此,本文通过分析现有的对二次设备状态评价基础数据处理方法存在的问题,以及粗糙集和神经网络两种数据挖掘算法的特征,提出了一种把数据挖掘技术运用到二次设备状态评价中的思想。该方法不仅能够解决现有方法中存在的不足,还能增强状态评价结果的客观性和可靠性。在这个信息技术腾飞发展的时代,作为与人们生活密切相关的庞大的电力系统来说,数据挖掘技术的引入势必会成为二次设备状态评价的一个创新点和突破点,也为二次设备状态评价各类处理方法的改进和研究提供了一个较为合理的理论依据。

参考文献 (References)

- [1] 高宁宁. 基于 Excel 的数据建模挖掘[J]. 中央民族大学学报(自然科学版), 2008, 1: 49-52.
- [2] 滕广青, 毛英爽. 国外数据挖掘应用研究与发展分析[J]. 统计研究, 2005, 12: 68-70.
- [3] 数据挖掘技术在零售业领域中的应用[D]. 西南大学, 2007.
- [4] 吴静, 邹海. 基于属性重要性的属性约简算法[J]. 计算机应用与软件, 2010, 2: 255-257.
- [5] 李旭. 基于改进神经网络的WEB数据挖掘研究[J]. 计算机仿真, 2008, 6: 99-102.
- [6] 任国华, 白建云, 邵加晓, 侯鹏飞. BP神经网络在直接空冷系统中的应用[J]. 电站系统工程, 2011, 6: 16-18.