

# 高压电网故障信息数据挖掘系统的研究

桂强, 刘意川, 张沛超

(上海交通大学电子信息与电气工程学院, 上海 200240)

**摘要:** 电力公司因多年运行积累了海量的历史故障数据而导致数据过剩。另一方面, 电网运行人员以及电力公司的决策者却又感觉到信息匮乏。近年来数据挖掘技术因能发现隐藏在大型数据集中的有价值的模式而倍受关注。提出了一种高压电网故障信息数据挖掘系统。论文从故障数据的分析提取、数据仓库的构建以及挖掘过程的设计等方面详细论述了该数据挖掘系统。最后以关联规则挖掘为例, 从电网故障信息数据库中挖掘出保护不正确动作模式, 该模式能为电网运行管理提供有益的决策支持。

**关键词:** 电力系统; 故障信息; 数据仓库; 数据挖掘; 故障模式; 保护

## Research on data mining system based on fault information in HV power grid

GUI Qiang, LIU Yi-chuan, ZHANG Pei-chao

(Shanghai Jiaotong University, Shanghai 200240, China)

**Abstract:** Electric power companies have accumulated huge amounts of historical fault data after many years' operation, which have not been made full use of but led to the phenomena called 'rich data, poor information'. The decision-makers have not been greatly supported because of the limitation of knowledge. In recent years, data mining technologies have attracted great attention since they can find interesting patterns from big data sets. This paper proposes a novel data mining system based on fault information in HV power grid. The analysis and extract of fault data, the establishing of data warehouse and the design of mining process are presented in detail. Finally, association rules of fault patterns are obtained, which can greatly help the decision-making for the system operating and management.

**Key words:** power systems; fault information; data warehouse; data mining; fault pattern; protection

中图分类号: TM76      文献标识码: A      文章编号: 1003-4897(2007)10-0037-04

## 0 引言

电力系统在故障时会产生大量的故障录波数据和动作事件信息。经过多年运行, 电力公司积累了海量的历史故障信息。但是若依靠传统的数据处理手段, 只能对数据进行较初步的分析和处理而不能挖掘出其中蕴含着的丰富知识, 无法从时间和空间意义上深层次理解并有效利用这些数据。所收集的数据不仅得不到充分利用, 反而带来越来越严重的数据过剩。同时, 电力公司的决策者正期盼应用隐含在这些数据中的知识支持以做出正确的决策, 却因“知识局限”而守着这些“数据宝库”开采不出有价值的信息。

数据仓库及数据挖掘技术因能有效分析处理海量数据集合、挖掘隐藏知识而倍受重视。目前数据仓库与挖掘技术在电力系统中的应用主要有设备状态监测<sup>[1]</sup>、负荷预测<sup>[2]</sup>、电网规划<sup>[3]</sup>、稳定性分析<sup>[4]</sup>和电力市场分析<sup>[5]</sup>等方面<sup>[6, 7]</sup>, 而对海量历史故障数据的挖掘还没有深入研究。

本文针对电网故障数据的海量特点, 以及电力公司的实际需求, 提出了一种高压电网故障信息数据挖掘系统的框架设计; 从数据挖掘的角度, 利用数据仓库和知识发现的理论和技术, 对电网故障信息进行分析提取并集成所需外部数据来构建数据仓库; 最后用数据挖掘技术挖掘出故障模式等知识。

## 1 故障信息数据挖掘系统的组成

### 1.1 高压电网故障信息的特点

高压电网故障信息主要有两类, 即断路器变位及保护动作等开关量信息, 以及基于故障录波器的模拟量信息。另外, 为进行故障定位以及分析故障与气候条件的关系等, 还引入了地理、气象等外部信息。

开关量信息是离散化事件, 有利于符号推理, 但也存在由于通信干扰或信号接入错误引起的开关量的错报或漏报等问题。另外当保护和断路器误动、拒动时, 开关量不能真实反映电网的故障情况。

故障录波器有较高的采样频率, 因此基于故障录波器的模拟量信息可信度高、容错性好, 能真实

反映电网的故障情况<sup>[8]</sup>。同时它还有标准化的特点。但是模拟量不利于符号推理。

## 1.2 数据仓库

数据仓库是一个面向主题的、集成的、时变的、稳定的数据集，用以支持经营管理决策的制定过程<sup>[9]</sup>。数据仓库的基本体系结构主要有数据源、监视器、集成器、数据仓库、客户应用等组成部分。

在本文中，首先对故障数据进行数据分析、提取，然后将提取到的故障特征量存入调度中心的数据仓库中，并把其他所需信息也集成进来。数据仓库的构建为数据挖掘提供了有力的支持。

## 1.3 数据挖掘

数据挖掘是采用人工智能的方法对数据仓库中的数据进行分析从而获取知识的过程<sup>[10]</sup>。所提取的知识通常表示为概念、规则、规律和模式等形式，可以被用于信息管理、查询优化、决策支持、过程控制以及数据自身的维护等。数据挖掘的知识模型有分类模型、时间序列模型、聚类模型和关联模型等。

本文基于故障信息数据仓库进行挖掘，旨在发现故障数据中隐含的故障模式等知识，以服务于电网的运行管理及决策支持。本文用到的知识模型主要是关联规则模型。

## 2 数据挖掘系统的架构设计

电网故障信息数据挖掘系统主要包括故障数据收集、数据监视器、故障数据分析、数据集成器、调度中心数据仓库、数据挖掘以及知识库等七个模块，其架构如图1所示。

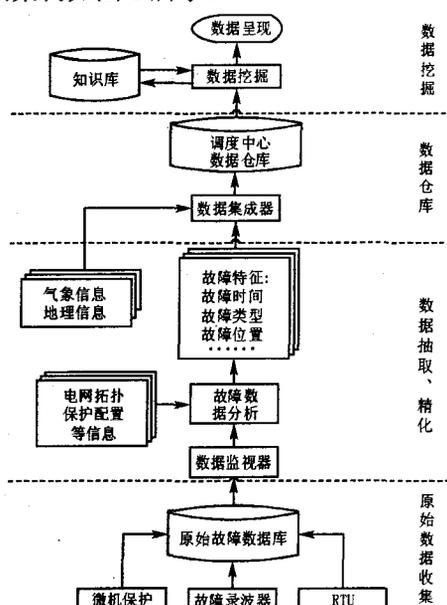


图1 数据挖掘系统的架构图

Fig. 1 Architecture of data mining system

### (1) 故障数据收集模块

收集来自全网的故障录波数据、保护及断路器动作信息，存入原始故障数据库中。原始数据库中的数据是海量的、有噪声的以及冗余的。

### (2) 数据监视器模块

探测变化的故障数据并把其追加到数据库中。

### (3) 故障数据分析模块

从原始故障数据库中取出数据，利用故障分析技术提取故障特征量，包括故障时间、故障类型和故障位置等。从数据挖掘角度，这是对数据进行抽取及精化的阶段。

### (4) 数据集成器模块

将上述处理得到的数据和来自外部的数据（如气象信息和地理信息等）进行集成，为存入数据仓库做准备。另外，数据集成器模块还要在原始数据（如故障录波文件）和处理过的数据之间建立关联，以便在查看分析结果时随时调出原始数据。

### (5) 调度中心数据仓库

储存了经过上述处理的多维度、多粒度的故障信息以便进行数据挖掘。

### (6) 数据挖掘模块

对数据仓库中的数据进行挖掘，以发现隐藏在故障数据背后的故障模式等知识。

### (7) 知识库

指导数据挖掘过程并评估、存储挖掘的结果。

## 3 故障数据分析

故障信息数据仓库和数据挖掘系统能为已有的故障分析、处理技术提供一个良好的技术集成框架。在数据抽取、精化阶段，可以应用目前广泛采用的小波、神经网络、专家系统等故障分析方法进行故障诊断并提取故障特征量，主要有故障时间、故障类型、故障位置等，并对保护及断路器动作行为进行评价。该框架对今后可能出现的新的故障分析技术开放。限于篇幅，本文仅简略讨论本系统采用的故障分析技术（即数据抽象、精化技术）。

基于故障录波数据可以利用提升小波进行故障时刻的提取，误差小于 $0.3\text{ms}$ <sup>[11]</sup>。类似的方法可以得到主保护、失灵保护、后备保护以及重合闸动作时刻等特征量。

利用小波神经网络，结合开关量和模拟量信息可以进行电网故障类型识别<sup>[12]</sup>。

由于开关量信息的不确定性，可以采用以开关量为主用模拟量进行冗余纠错的故障分析方法。如文献[13]提出的基于时态逻辑的高压电网故障诊断方法，能处理简单故障和复杂故障，判断故障位

置, 并对保护和断路器动作行为做出评价。

经过上述方法处理后, 数据得到了精化, 最后得到关于此次故障的特征性描述。主要特征信息如表 1 所示。表中作为示例列出了 Line1 发生的 2 次故障。记录 1 说明线路 Line1 在时间 T1 发生 A 相接地, 故障位置距线路首端为全长 60% 的地方, 主保护在故障发生后 20 ms 动作, 保护动作评价为动作正常。这里, 主保护的動作评价有动作正常、动作缓慢、误动和拒动等。记录 2 主保护动作缓慢。

表 1 故障特征信息

Tab.1 Fault characteristic

设备	时间	类型	距离	接地电阻/Ω	主保护动作时间/ms	动作评价
Line1	T1	AG	60%	5	20	正常
Line1	T2	AB	80%	-	40	缓慢

### 4 故障信息数据仓库的设计

与传统数据库不同, 数据仓库的设计要满足多维度、多粒度数据分析的要求。这样, 在横向可以组合不同的维度、在纵向可以在不同粒度上进行数据挖掘, 以发现有价值的知识。本文将数据仓库数据模型分为三层, 分别是信息包图、星型图和物理数据模型。

信息包图是数据仓库数据模型的第一层。在信息包图中, 要确定数据仓库的事实、维度和粒度。事实即所分析的目标数据; 维度即事实信息的属性; 粒度是对数据仓库中的数据综合程度高低的一个度量。如表 2 所示, 事实是指具体的故障情况。每一种故障情况用多个维度来描述, 有故障时间维、故障位置维、故障类型维、接地电阻维和动作情况等。每个维度又有多个粒度, 如时间维有年、月、日、时、分和秒六种粒度。

表 2 故障信息数据仓库的信息包图

Tab.2 Information package diagram of fault information

粒度\维度		data warehouse			
	时间	位置	故障类型	接地电阻	保护与断路器动作情况
	年	厂站	是否接地	高、中、低	动作情况
	月	设备	具体类型	电阻值	
	日	相别			
	时				
	分				
	秒				
	事实: 故障前后电压、电流相量, 主保护动作信息, 后备保护动作信息, 故障时刻, 保护动作信息, 重合闸时刻, 气象信息, 故障录波文件索引……				

星型图是数据仓库数据模型的第二层, 如图 2 所示, 是向最终的数据结构添加某些细节的逻辑模

型。与传统的关系模型相比, 星型图模型简化了用户分析所需的关系, 从支持决策的角度去定义数据实体, 更适合大量复杂查询。

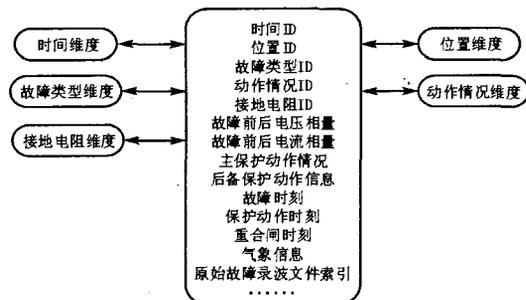


图 2 故障信息数据仓库的星型图模式

Fig. 2 Star scheme of fault information data warehouse

物理数据模型是数据仓库数据模型的第三层, 是星型图模型在数据仓库中的实现, 如物理的存取方式、数据存储结构等。按数据的重要程度、使用频率以及响应时间的要求进行分类, 并将不同类的数据分别存储在不同的存储设备上。

### 5 数据挖掘

基于故障信息数据仓库来挖掘隐含的故障模式等知识, 这些模式可以用关联规则 (Association Rule) 表示。关联规则是形如  $A \Rightarrow B$  的蕴涵式<sup>[10]</sup>, 其中 A 为条件, B 为结论。规则的强度可用支持度 s 和可信度  $\alpha$  度量, 公式如下<sup>[10]</sup>:

$$s(A \Rightarrow B) = \Pr(A \cup B) \quad (1)$$

$$\alpha(A \Rightarrow B) = \Pr(B|A) \quad (2)$$

可挖掘的故障模式有故障发生与天气因素的关系; 在某一特定时间段内的某类设备易发的故障类型; 特定线路在特定位置发生各类故障的概率; 特定线路保护的動作性能评价, 等等。这些故障模式知识有助于辅助决策, 合理安排检修计划, 减小气候和负荷变化对产生故障的影响, 提高供电质量, 改善保护性能等。

Apriori 算法<sup>[10]</sup>是最有影响的挖掘单维关联规则频繁项集的算法。但由表 2 所示的信息包图及图 2 所示的星型图可见, 故障信息不但是多维的, 而且每一维还具有多个粒度, 因此本文遇到的是一个多维、多层关联规则挖掘问题。本文采用改进的 Apriori 算法处理这类挖掘问题。限于篇幅, 对此将另文阐述。

下面以本文系统所挖掘的关于保护动作评价的关联规则为例来进行说明。表 3 列出了三条关联规则及相应支持度和可信度。需要说明的是, 若支持度和可信度的阈值过小, 将导致产生大量弱关联规

则并延长挖掘时间;反之,则可能遗漏有价值的关联规则。实用中应先在较小样本中通过修改阈值进行交互式挖掘,以确定合适的阈值,然后再对海量数据进行挖掘,并辅以按支持度或可信用度排序技术。本文支持度和可信用度阈值分别取 3%和 60%。

表 3 保护不正确动作模式

Tab.3 Association rules of mis-operation patterns of protections

保护	关联规则	支持度	可信用度
P1	在 50%~85%处经高过渡电阻单相接地=>拒动	6.7%	93%
P1	在 70%~85%处单相接地=>动作缓慢	4.0%	60%
P2	在 50%~85%处相间故障=>动作缓慢	9.0%	100%

规则 1 说明,当保护 P1 在保护范围末端发生单相短路且高阻接地时,保护会拒动,规则可信用度为 93%, 该种情况占有不正确动作情况的 6.7%; 规则 2 说明在线路 70%到 85%的位置发生单相接地时,保护 P1 会动作缓慢,规则可信用度为 60%, 该情况占有不正确动作情况的 4%; 规则 3 说明,在线路 50%到 85%范围内发生相间故障时,保护 P2 会动作缓慢,规则可信用度为 100%, 该情况占有不正确动作情况的 9.0%。决策者在得到这样的知识支持后就能更好地掌握保护性能与故障模式的关系,从而可以优化保护配置。

## 6 结论

电力系统故障往往在偶然性之后隐藏着规律性。本文针对电力公司多年运行积累下来的海量历史故障数据,提出了一种故障信息数据挖掘系统,并从故障数据的分析提取、数据仓库的构建以及挖掘过程的设计等方面详细论述了该数据挖掘系统,最后以保护不正确动作模式为例,挖掘出关联规则。利用数据仓库和数据挖掘技术,可以构成多维度、多粒度、专业化的故障信息处理系统,可以将日积月累的故障数据转化为电力企业的知识,值得继续深入研究。本文重点在于提出故障信息数据仓库和数据挖掘的系统架构和设计技术,其具体技术实现将另文介绍。

## 参考文献

[1] Ogilvie, Swidenbank E, Hogg B W. Use of Data Mining Techniques in the Performance Monitoring and Optimization of a Thermal Power Plant[J]. IEE Colloquium (Digest), 1998.

[2] 崔旻, 顾洁. 基于数据挖掘的电力系统中长期负荷预测新方法[J]. 电力自动化设备, 2004, 24 (6): 18-21.

CUI Min, GU Jie. Mid-long Term Load Forecast of Power System Based on Data Mining[J]. Electric Power Automation Equipment, 2004, 24 (6): 18-21.

[3] 伍力, 吴捷, 周乐荣. 面向规划的广东电网数据库的建立及其应用[J]. 电力系统自动化, 1999, 23 (13): 50-52.

WU Li, WU Jie, ZHOU Le-rong. A Newly Designed Database and its Application to Guangdong Power Network Planning[J]. Automation of Electric Power Systems, 1999, 23 (13): 50-52.

[4] 于之虹, 郭志忠. 基于数据挖掘理论的电力系统暂态稳定评估[J]. 电力系统自动化, 2003, 27 (8): 45-48.

YU Zhi-hong, GUO Zhi-zhong. A Novel Approach of Power System Transient State Stability Assessment Based on Data Mining Theory[J]. Automation of Electric Power Systems, 2003, 27 (8): 45-48.

[5] 马瑞, 颜宏文. 基于判定树的电力市场中发电商竞标能力数据挖掘的新思路[J]. 电力系统自动化, 2002, 26 (15): 22-26.

MA Rui, YAN Hong-wen. A New Data Mining Framework Based on Decision-Tree for Competitive Bidding Assessment in Power Marketing[J]. Automation of Electric Power Systems, 2002, 26 (15): 22-26.

[6] 廖志伟, 孙雅明. 数据挖掘技术及其在电力系统中的应用[J]. 电力系统自动化, 2001, 25 (11): 62-66.

LIAO Zhi-wei, SUN Ya-ming. Data Mining Technology and Its Application on Power System[J]. Automation of Electric Power Systems, 2001, 25 (11): 62-66.

[7] 路广, 张伯明, 孙宏斌. 数据仓库与数据挖掘技术在电力系统中的应用[J]. 电网技术, 2001, 25 (8): 55-58.

LU Guang, ZHANG Bo-ming, SUN Hong-bin. Application of Data Warehouse and Data Mining Techniques to Power Systems[J]. Power System Technology, 2001, 25 (8): 55-58.

[8] 高翔, 张沛超, 章坚民. 电网故障信息系统应用技术[M]. 北京:中国电力出版社, 2006.

GAO Xiang, ZHANG Pei-chao, ZHANG Jian-min. Applied Technologies of Fault Information System in Electric Power System[M]. Beijing:China Electric Power Press, 2006.

[9] Inmon W H. Building the Data Warehouse[M]. New York: John Wiley & Sons, 1996.

[10] HAN Jia-wei, Kamber M. Data Mining: Concepts and Techniques[M]. Morgan Kaufmann Publishers, 2001.

[11] 郑华珍, 乐全明, 郁惟镛, 等. 基于小波理论的超高压电网故障时刻提取[J]. 电网技术, 2005, 29 (19): 33-38.

ZHENG Hua-zhen, YUE Quan-ming, YU Wei-yong, et al. Location of Fault Moment by Wavelet in Ultra-High-Voltage Network Based on Fault Record Data[J]. Power System Technology, 2005, 29 (19): 33-38.

(下转第 44 页 continued on page 44)

- Distribution Networks for Resistive Line Losses Reduction[J]. IEEE Trans on Power Delivery, 1989, 4(2).
- [6] 邓佑满, 张伯明, 相年德. 配电网络重构的改进最优流模式算法[J]. 电网技术, 1995, 19(7): 47-50.  
DENG You-man, ZHANG Bo-ming, XIANG Nian-de. An Improved Optimal Flow Pattern Algorithm for Distribution Network Reconfiguration[J]. Power System Technology, 1995, 19(7): 47-50.
- [7] 梁勇, 张焰, 候志俭. 遗传算法在配电网重构中的应用[J]. 电力系统及其自动化学报, 1998, 10(4): 29-34.  
LIANG Yong, ZHANG Yan, HOU Zhi-jian. Application of Genetic Algorithm in Power Distribution System Reconfiguration[J]. Proceedings of the EPSA, 1998, 10(4): 29-34.
- [8] 周青山, 向铁元, 罗亚, 等. 基于改进遗传算法的配电网重构[J]. 电力科学与生命, 2004, (1): 20-24.  
Zhou Qing-shan, XIANG Tie-yuan, LUO Ya, et al. Electric Power Science and Engineering 2004, (1): 20-24.
- [9] 陈根军, 李继洸, 唐国庆. 基于 Tabu 搜索的配电网重构算法[J]. 中国电机工程学报, 2002, 12(10): 28-33.  
CHEN Gen-jun, LI Ji-guang, TANG Guo-qing. A Tabu Search Approach to Distribution Network Reconfiguration for Loss Reduction[J]. Proceeding of the CSEE, 2002, 12(10): 28-33.
- [10] 左飞, 周家启. TS 算法在配电网重构中的应用[J]. 电力系统及其自动化学报, 2004, 16(1): 66-69.  
ZUO Fei, ZHOU Jia-qi. Application of Tabu Search Algorithm in Distribution Network Reconfiguration[J]. Proceedings of the EPSA, 2004, 16(1): 66-69.
- [11] 卢耀川, 廖迎春, 陈星莺, 等. 基于模拟退火法的网络重构技术[J]. 电力自动化, 2003, 23(1): 28-31.  
LU Yao-chuan, LIAO Ying-chen, CHEN Xing-ying, et al. Distribution Reconfiguration Technology Based on GA and SA[J]. Electric Power Automation Equipment, 2003, 23(1): 28-31.
- [12] 张步涵, 沙立华, 曾次玲. 基于随机生成树的配电网重构模拟退火算法[J]. 华中科技大学学报(自然科学版), 2005, 33(6).  
ZHANG Bu-han, SHA Li-hua, ZENG Ci-ling. Simulated Annealing Algorithm Based on Random Spanning Tree in Distribution Network Reconfiguration[J]. Journal of Huazhong University of Science & Technology(Natural Science Edition), 2005, 33(6).
- [13] 王秀丽, 吴宏晓, 别朝红, 等. 以提高系统可靠性为目标的配电网重构[J]. 中国电力, 2001, 34(9): 40-43.  
WANG Xiu-li, WU Hong-xiao, BIE Zhao-hong, et al. Distribution System Reliability and Network Reconfiguration[J]. Electric Power, 2001, 34(9): 40-43.
- [14] 刘健, 毕鹏翔, 董海鹏. 复杂配电网简化分析与优化[M]. 北京: 中国电力出版社, 2002. 273-277.  
LIU Jian, BI Peng-xian, DONG Hai-peng. Simplification and Optimization on Complicated Distribution Network[M]. Beijing: China Electric Power Press, 2002. 273-277.
- [15] 葛少云, 刘自发, 余贻鑫. 基于改进禁忌搜索的配电网重构[J]. 电网技术, 2004, 28(23): 23-27.  
GE Shao-yun, LIU Zi-fa, YU Yi-xin. An Improved Tabu Search for Reconfiguration of Distribution Systems[J]. Power System Technology, 2004, 28(23): 23-27.

收稿日期: 2006-11-05; 修回日期: 2007-01-17

作者简介:

张忠会(1962-), 男, 副教授, 研究方向为电力系统运行与控制;

李曼丽(1982-), 女, 硕士, 研究方向为电力系统运行与控制、配网自动化; E-mail: civilaviation@hotmail.com

熊宁(1975-), 男, 硕士, 研究方向为电力系统运行。

(上接第 40 页 continued from page 40)

- [12] 杨光亮, 乐全明, 郁惟镛, 等. 基于小波神经网络和故障录波数据的电网故障类型识别[J]. 中国电机工程学报, 2006, 26(10): 99-103.  
YANG Guang-liang, YUE Quan-ming, YU Wei-yong, et al. A Fault Classification Method Based on Wavelet Neural Networks and Fault Record Data[J]. Proceedings of the CSEE, 2006, 26(10): 99-103.
- [13] 乐全明, 董志赞, 郑华珍, 等. 基于时态逻辑技术的高压输电线系统故障诊断[J]. 电力系统自动化, 2006, 30(6): 38-43.

YUE Quan-ming, DONG Zhi-yun, ZHENG Hua-zhen, et al. Fault Diagnosis Based on Temporal Logic Technology in HV Transmission System[J]. Automation of Electric Power Systems, 2006, 30(6): 38-43.

收稿日期: 2006-10-18; 修回日期: 2006-12-04

作者简介:

桂强(1982-), 男, 硕士研究生, 主要研究方向为电力系统继电保护; E-mail: guiqiang@sjtu.edu.cn

刘意川(1983-), 男, 硕士研究生, 主要研究方向为电力系统继电保护;

张沛超(1970-), 男, 博士, 副教授, 从事电力系统及其自动化的教学和研究工作。