

数据挖掘在智能电能表故障分析中的应用

祝宇楠, 徐 晴, 刘 建, 田正其, 周 超

(国网江苏省电力公司电力科学研究院国家电网电能计量重点实验室, 江苏 南京 210019)

摘 要: 自2009年推广以来, 全国范围内已安装运行数亿只智能电能表, 国家电网公司采取集中检定模式对其进行装出前的管理, 并配套了用电信息采集系统监控其现场运行状态, 管理部门可在相应平台上获得智能电能表海量的质量数据。文中尝试在矩阵实验室(MATLAB)环境下, 通过训练神经网络, 建立现场智能电能表故障数据与检定数据之间联系, 讨论采用数据挖掘技术分析检定数据以提前获得智能电能表故障信息的可能性及有效性, 为电能表的质量管控提供另一种工具。

关键词: 智能电能表; 数据挖掘; 故障分析; 神经网络

中图分类号: TM933.4

文献标志码: A

文章编号: 1009-0665(2016)05-0019-05

随着信息技术的发展, 智能电能表突破传统业务, 集成了双向多费率计量、实时互动、负荷控制、能效管理等多种功能, 由此引入了系统复杂性, 可靠性下降。自2009年推广以来, 全国范围内已安装运行智能电能表4亿多只, 不可避免地出现了由设计或工艺问题导致的设备故障。这些电表拆回后虽进行了故障分析, 但由于数量多、故障原因繁杂、缺乏规模化分析手段, 多数分析结果被当成个体、偶然发生的情况来处理, 其反映的深层信息有待进一步研究; 同时, 未开展故障信息与其他质量数据之间的关联性研究。

数据挖掘是从大量数据中发现并提取隐藏信息的过程^[1], 已有数据挖掘在用电及相关信息领域的应用实例。文献[2]采用数据挖掘技术对AMI计量数据的电能误差进行分析, 发现超差电表, 使校准有的放矢。文献[3]研制了一套计量装置在线监测和智能诊断系统, 可运用数据挖掘技术对用户违约用电窃电和计量装置故障进行智能诊断。文献[4]将数据挖掘技术与用电大数据信息化技术结合, 通过台区线损异常智能诊断模型, 对台区线损异常进行管理, 提高了企业的经济效益。文献[5]在建立关口电能表性能状态评估模型的基础上, 采用数据挖掘技术对检测数据进行误差变化趋势分析、状态研究、性能分析等。可见, 采用数据挖掘可以发现传统统计方法难以确立大量数据之间的隐形关系, 为故障诊断等提供有效的分析手段。文中分析了电能表生命周期内的质量数据内容及其来源, 然后通过基于神经网络的数据挖掘技术分析大量智能电能表故障信息与其他质量信息的宏观关联性, 以期通过数据之间内在联系发现故障征兆。

1 智能电能表生命周期质量数据

1.1 智能电能表生命周期

每一只智能电能表都带有唯一的资产编号, 其整

个生命周期中, 通过该资产编号可将时间维度上的所有质量数据联系起来。智能电能表的生命周期按时间先后顺序可分为以下4个阶段。

(1) 设计生产: 智能电能表厂家依据产品标准、需求方特殊要求等进行表计的设计与生产, 为每只电能表分配资产编号, 并在出厂前进行参数设置、老化试验等初步检测;

(2) 检测验收: 供货前质量监督信息的统计分析过程, 包括到货后抽样样品全性能检测和整个到货批次表计的全检验收。质量数据包括准确度、电气要求、电磁兼容、通信功能、一致性、费率控制等试验结果;

(3) 现场运行: 装出后电能表将在现场工作运行, 此过程中记录的各种异常数据或异常事件有时能反映出自身质量缺陷或性能退化;

(4) 退出运行: 电能表不能正常工作被拆回分拣或超出规定运行时间退出现场运行, 可能进行故障分析或分拣, 提取故障质量信息。

电能表的质量数据在整个生命周期内是不断变化和积累的, 任何一个阶段的质量数据都可能与故障信息相关。

1.2 用电信息采集系统与智能电能表的集中检定

现代的电能计量系统是由智能电能表、数据采集设备、主站及连接这些设备的通信网络组成。国家电网公司在此基础上建立了用电信息采集系统(EEDAS)。用电信息采集系统通过对配电变压器及终端用户进行用电数据的监测和采集, 实现远程付费、阶梯电价、负荷管理, 最终达到自动抄表、错峰用电、用电检查、负荷预测和节约用电成本等目的。EEDAS已积累了大量的电能表运行数据, 蕴含着电能表状态和性能信息。

另外, 国家电网公司还建立了省级集中检定制度, 在供货初期即对所有智能电能表进行计量检定, 只有符合检定规范要求的电能表才能配送至现场运行。此过程中, 装出前误差等反映其性能的检定数据均记录

在省级计量中心生产调度平台(MDS)上,这些信息目前仅代表检定时电能表的某些性能,与电能表的长期性能、质量有无关联尚无研究。

2 基于神经网络的数据挖掘算法

传统的数据挖掘技术适用于线性、结构化的数据关系,由于智能电能表质量数据结构不统一,且非线性、持续性及噪音是普遍存在的,因此需要一种不同于传统的理论和方法来解决大量质量数据关联分析的问题^[6]。在有足够的数据且网络足够大的时候,神经网络可高效地学习带有非线性分类的结构模式^[7]。基于神经网络的数据挖掘算法步骤如下:

- (1) 生成矩阵形式的数据表;
- (2) 确定输出变量,生成输出量向量表;
- (3) 确定神经网络类型、内部算法参数、神经元数量等;
- (4) 将输入数据表划分为训练集、验证集、测试集三部分,初始化三者比例,验证集、测试集中的示例数远少于训练集中的示例数;
- (5) 用训练集的输入输出量训练神经网络;
- (6) 将验证集提供给训练好的网络进行判别;
- (7) 将测试集提供给训练好的网络进行测试;
- (8) 若验证、测试结果符合预期,则(9);否则,重新修改神经元数目、或更改训练集、验证集、测试集三者比例,重复(5—8);
- (9) 结束。

3 实例分析

3.1 故障数据预处理

将智能电能表故障进行分类并导入系统的数据表中备用。智能电能表的故障可分为硬件、软件、通信3个方面的故障,其中硬件故障又可追溯至整机、模块、元器件等3层级。国家电网公司在电能表质量管控相关管理办法中总结了智能电能表的49种故障类型,在剔除外观、安装环境等原因明显的故障类型后,典型的硬件故障按照模块层级分类如表1所示。

表1 典型的模块层级故障分类

故障描述	所属模块
脉冲输出故障	计量模块
屏幕不亮	显示模块
误差超差	计量模块
RS-485 通信失败	通信模块
电池失压	电源模块
数据读取异常	存储模块

表1中几种故障类型均是在现场运行或退出运行环节中发现的,数据均记录在EEDAS中。为了了解故障发生前,是否能从已有的检测检验数据中发现故障

隐患,需要将检测检验数据进行整理。需要说明的是,若故障与检测检验数据含有某种隐形关联,这种关联只能通过大量数据的趋势体现出来。文中根据江苏省级计量中心现有检测检验手段,提取检测环节中可获得的63个实验结果进行数据关联分析,这63个试验结果包括“制造商”、“检定台号”、各种平衡负载下的“相对误差”、“平均误差”、“采样值”、“校核计度示数”、“时钟日计时误差”等各类数据。

3.2 数据挖掘算法实现

数据挖掘的目的是找到输入变量与每种故障的联系,以及每个输入变量(因素或特征)对每种故障的影响量有多大。将EEDAS中导出的所有故障电能表进行预处理后,分别通过唯一的资产编号从记录检定检验数据的MDS中找到对应的63类数据,生成数据表,作为数据挖掘算法的输入(训练项之一)。

为表述清楚起见,下文示例只选取“误差超差”及“RS-485 通信失败”2种故障,这2种故障均为智能电能表的主要失效模式,示例重点研究上述63种检测数据结果与这2种故障的关联。

采用矩阵实验室(MATLAB)神经网络模式识别工具箱对输入数据进行训练。由于训练过程中不仅需要上述故障信息,还需要非故障表的相关信息,因此,以1:2的比例在EEDAS与MDS中随机提取非故障表的63类数据,与故障数据表合成总数据表,作为模式识别工具箱的输入。以三维向量 $[1,0,0]$, $[0,1,0]$, $[0,0,1]$ 分别表示非故障表、通信模块失效、计量模块失效3种情况,即输出量1,2,和3。输出向量维数=失效模式数+1。

选取前馈神经网络进行训练,其具有1个隐藏层和1个输出层。该神经网络采用的内部算法为反向传播共轭梯度法,停止准则选取均方差,当其小于给定阈值时,训练停止。为得到更好的训练效果,输入输出的向量顺序在每次计算时都会重新调整。上述过程流程如图1所示,选取的神经网络如图2所示。

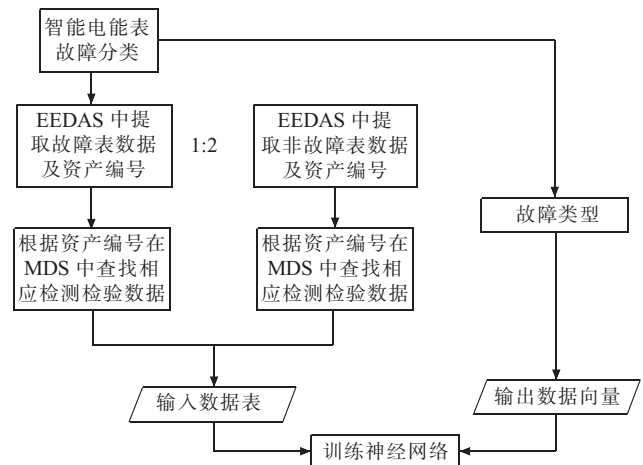


图1 用电能表数据训练神经网络过程流程

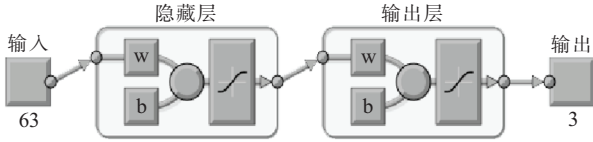


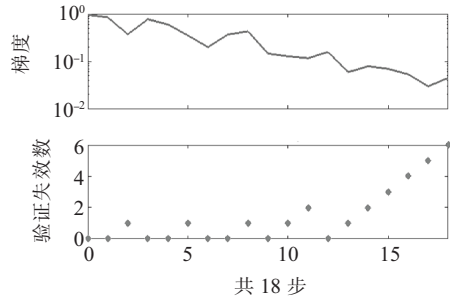
图2 两层前馈神经网络

3.2 训练步骤及结果

为得到最佳训练效果,设计以下2步训练步骤。

(1) 第I步。固定训练集、验证集和测试集的比例,调整神经网络隐藏层中的神经元数量;设定训练集70%、验证集15%、测试集15%,神经网络隐藏层中的神经元数量范围为10-1000。按照本文第2小结介绍的算法步骤对神经网络进行训练。图3—5表示了神经元数量分别为10,100,1000时各自最好的训练结果。图4混淆矩阵中3个训练结果正确率达到了100%,高于图3和图5。由此可见,固定训练集、验证集、测试集的比例时,100神经元可达到最好的训练效果。

(2) 第II步。固定神经网络隐藏层中的神经元数量,调整训练集、验证集和测试集的比例;固定神经网络



(d) 梯度与验证失效数

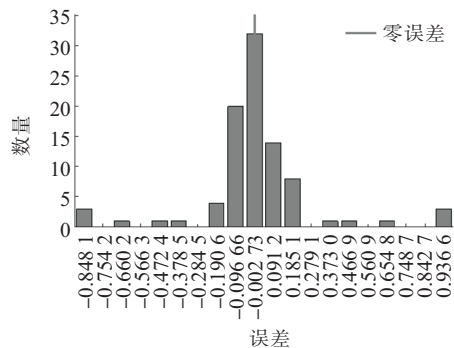
图3 训练集70%验证集15%测试集15%及10神经元时的训练结果

	1	2	3	
1	14 46.7%	2 6.7%	2 6.7%	77.8% 22.2%
2	0 0.0%	6 20.0%	0 0.0%	100.0% 0.0%
3	0 0.0%	0 0.0%	6 20.0%	100.0% 0.0%
	100.0% 0.0%	75.0% 25.0%	75.0% 25.0%	86.7% 13.3%
	1	2	3	

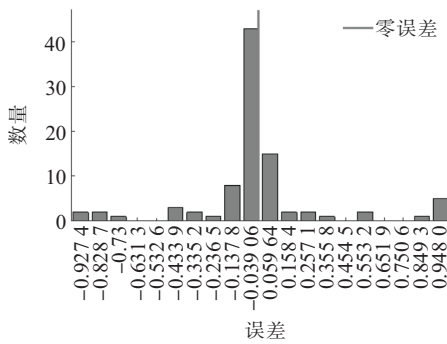
(a) 融合矩阵

	1	2	3	
1	12 40.0%	1 3.3%	2 6.7%	80.0% 20.0%
2	1 3.3%	6 20.0%	0 0.0%	85.7% 14.3%
3	1 3.3%	1 3.3%	6 20.0%	75.0% 25.0%
	85.7% 14.3%	75.0% 25.0%	75.0% 25.0%	80.0% 20.0%
	1	2	3	

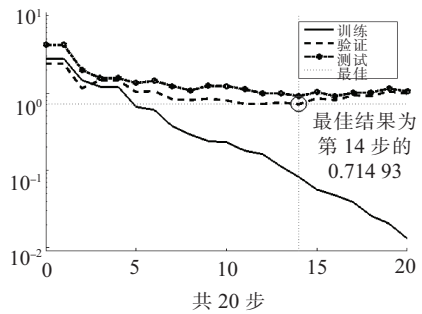
(a) 融合矩阵



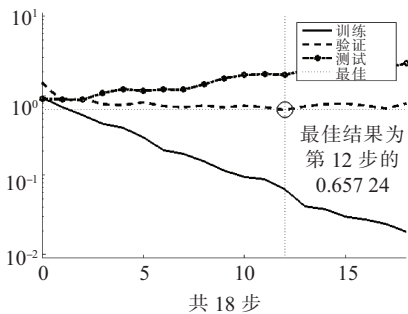
(b) 误差柱状图



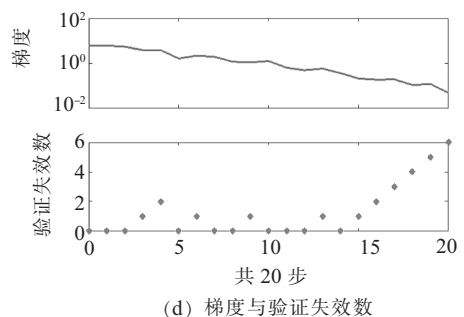
(b) 误差柱状图



(c) 验证执行结果



(c) 验证执行结果



(d) 梯度与验证失效数

图4 训练集70%验证集15%测试集15%及100神经元时的训练结果

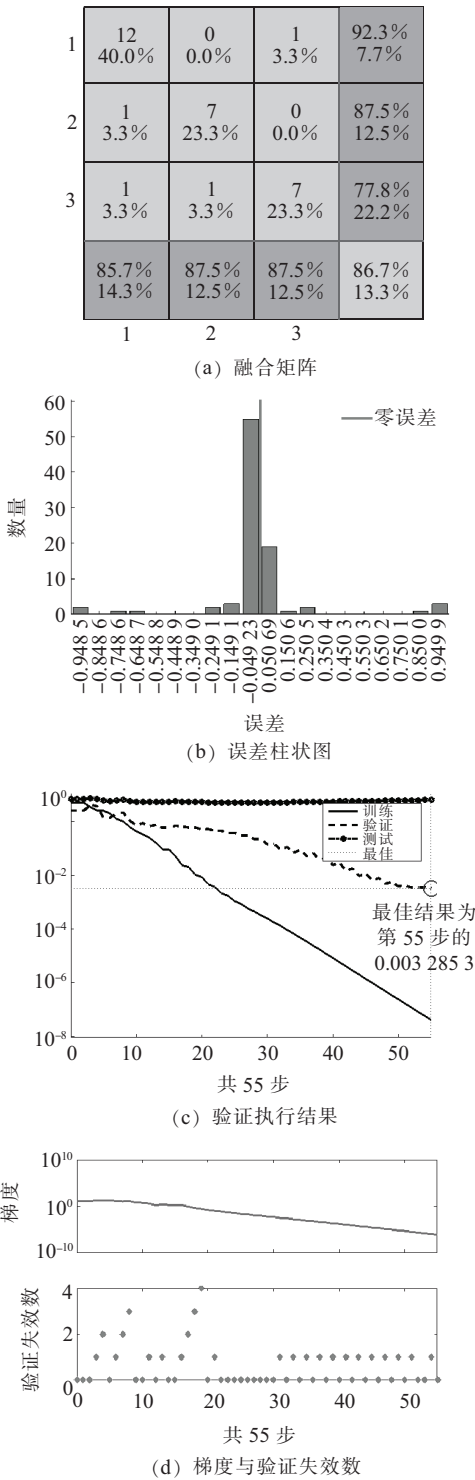


图 5 训练集 70%验证集 15%测试集 15%及 1000 神经元时的训练结果

络隐藏层神经元数 100,调整训练集、验证集和测试集比例,图 6 中为 50%,25%,25%;图 7 中比例调整为 60%,20%,20%。训练该神经网络,图 6、图 7 所示分别为各自情况中效果最好的一次,然而均不如比率为 75%,15%,和 15%时的情况(如图 4 所示)。

分析训练集 70%、验证集 15%、测试集 15%、100 神经元时的训练结果可知,智能电能表“误差超差”及“RS-485 通信失败”2 种故障与检测检验环节中记录

的 63 个检测结果有一定的关系。按输出结果显示,根据检测检验环节预测 2 种故障的正确性均为 75%。

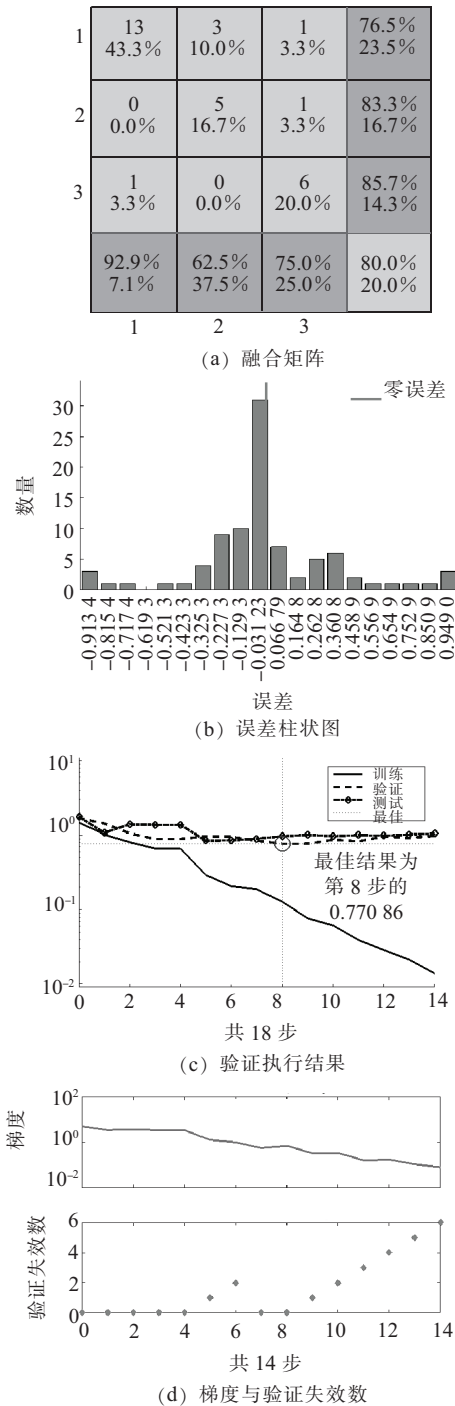
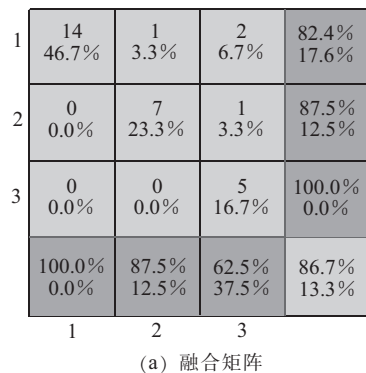
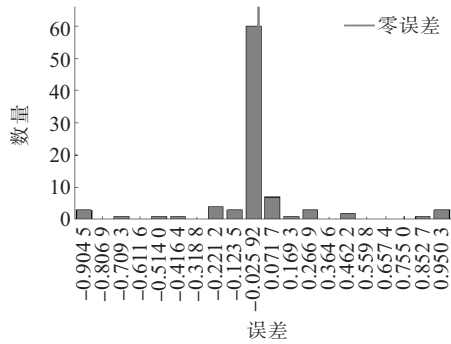
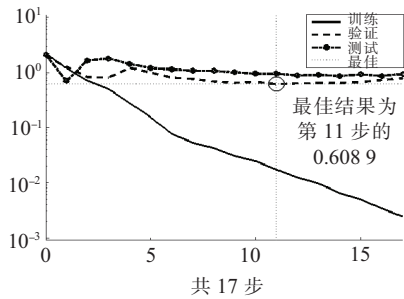


图 6 训练集 50%验证集 25%测试集 25%及 100 神经元时的训练结果

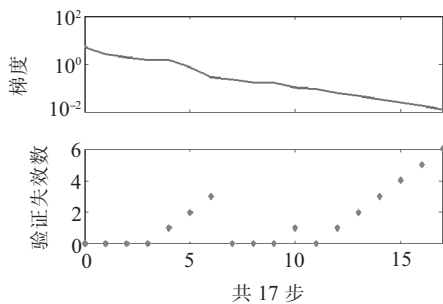




(b) 误差柱状图



(c) 验证执行结果



(d) 梯度与验证失效率

图7 训练集 60%验证集 20%测试集 20%及 100 神经元时的训练结果

4 结束语

采用基于神经网络的数据挖掘技术对故障智能电能表故障分析数据与检测检验环节的记录数据进行了关联性分析,验证了通过前期的检验检测数据预测运行环节中 2 种故障类型发生的可能性。值得注意

的是,文中所训练的神经网络是采用 6000 只故障表和 12 000 只非故障表组成的数据集,按照全国电网数以亿计的智能表来算,文中所采用的数据远远不够,要想得到准确性更高的预测结果,需将故障表的全集作为输入数据对神经网络进行训练。分析结果可以帮助改进智能电能表的制造过程,制定更有针对性的批次轮换计划,提高智能电能表整体质量。

参考文献:

- [1] 李欣. 基于神经网络的数据挖掘方法研究[D]. 大庆:大庆石油学院,2003.
- [2] 郭景涛. 面向智能电网 AMI 的网络计量关键技术与用户用电数据挖掘研究[D]. 天津:天津大学计算机科学与技术学院,2011.
- [3] 肖坚红,严小文,周永真,等. 基于数据挖掘的计量装置在线监测与智能诊断系统的设计与实现[J]. 电测与仪表,2014,51(14):62-67.
- [4] 宋玉,郑海雁,尹飞. 基于智能用电大数据分析的台区线损管理[J]. 电力信息与通信技术,2015,13(8):132-135.
- [5] 张建伟,曹敏,王昕,等. 基于数据挖掘技术的关口电能表检测数据分析统计平台设计[C]//2011年云南电力技术论坛-2011年云南电力技术论坛论文集(入选部分).昆明:云南省电机工程学会,2011.
- [6] 郭庆琳,祖向荣. 基于神经网络与遗传算法的汽轮机组数据挖掘方法[J]. 电力自动化设备,2008,28(3):41-44.
- [7] ABU-MOSTAFA Y. S, MAGDON-ISMAIL M, LIN H. T. Learning from Data, AMLBook.com, 2012.

作者简介:

- 祝宇楠(1983),女,吉林人,工程师,从事电测量及电测仪器仪表工作;
- 徐晴(1973),女,江苏南通人,研究员级高工,从事电测量及电测仪器仪表工作;
- 刘建(1981),男,河南信阳人,高级工程师,从事电测量及电测仪器仪表工作;
- 田正其(1987),男,江苏南通人,工程师,从事电测量及电测仪器仪表工作;
- 周超(1987),男,江苏宜兴人,工程师,从事电测量及电测仪器仪表工作。

Data Mining Application in Smart Meter Fault Analysis

ZHU Yunan, XU Qing, LIU Jian, TIAN Zhengqi, ZHOU Chao

(State Grid Key Laboratory of Electric Energy Metering,

State Grid Jiangsu Electric Power Company Electric Power Research Institute, Nanjing 210019, China)

Abstract: Since 2009, hundreds of millions of smart meters have been installed nationwide. State Grid Corporation of China manages the asset by centralized inspection in large scale. Meanwhile, a system called Power User Electric Energy Data Acquire System was built up to monitor the operation status of these meters. Therefore, vast amounts of life-time data of the meters can be obtained from the centralized inspection system as well as the data acquire system. This paper proposed a way of establishing the connections between fault data and inspection data by training neural network under MATLAB environment. And then the possibility and effectiveness to obtain fault information of smart meters in advance by data mining is discussed, to provide an alternative way of quality control of smart meters.

Key words: smart meter; data mining; fault analysis; neural network