

基于大数据挖掘的新能源电站设备故障诊断方法的研究与应用

孙树敏¹, 李华东¹, 张用¹, 王玥娇¹, 李笋²

(1. 国网山东省电力公司电力科学研究院, 济南 250000 2. 国网山东省电力公司, 济南 250000)

Research and application of fault diagnosis method for new energy power station equipment based on big data mining

SUN Shu-min¹, LI Huan-dong¹, ZHANG Yong¹, WANG Yue-jiao¹, LI Sun²

(1. State grid shandong electric power research institute, Jinan 250000 2. State grid shandong electric power company, Jinan 250000)

ABSTRACT: In view of the diversity, complexity and magnanimity of information data in the operation status of new energy power station, the data mining technology is applied to fault diagnosis. On this basis, the overall process of fault diagnosis methods including clustering analysis, fault rule mining, fault modeling and other mathematical statistical theories is designed. The logical relationship between massive data is redefined at a deeper level, and the most effective fault characterization is extracted. The correctness and validity of the method are proved by comparing the curve of the fault characteristic parameter prediction results with the field measurement results. This method has high data processing efficiency, effectively improves the reliability and accuracy of fault diagnosis and early warning, and provides a strong guarantee for the safe operation of new energy network.

KEY WORD: New energy; fault diagnosis; big data mining; data handling; feature extraction

摘要: 针对新能源电站设备运行状态信息数据多样性、复杂性以及海量性的特点, 将大数据挖掘技术应用于故障诊断领域。在此基础上, 设计了包含聚类分析、故障规则挖掘、故障建模等数理统计理论的故障诊断方法总体流程, 从更深层次上重新定义海量数据间的逻辑关系, 提取出最为有效的故障表征, 并通过故障特征参数预测结果与现场测量结果的曲线对比证明了该方法的正确性和有效性。该方法具有较高的数据处理效率, 有效提高故障诊断预警的可靠性与准确性, 为新能源并网安全运行提供有力保障。

关键词: 新能源; 故障诊断; 大数据挖掘; 数据处理; 特征提取

0 引言

电站设备稳定工作是保证电网安全运行的重要基础。随着当代工业技术的进步, 设备向着多元化方向发展^[1], 其中新能源电站就包括涉网设备、变流器、变电站主设备、电压互感器、电流互感器、开关、消弧线圈、动态无功补偿装置、箱变、光伏电池、多绕组变压器、逆变器、太阳能电池板等多种设备。由于设备类别及数量不断增加, 设备发生故障的可能性也随之升高。将故障诊断技术引入到日常设备监管中是保证设备安全运行和提高运维效率的一种有效途径, 而现代新能源电站设备的发展日趋高效化、复杂化、网络化, 各部分的构造关系愈发复杂与紧密, 随之而来的故障也是多样性的、隐蔽的、复杂的, 传统的故障诊断方法的分析能力已不能满足电站设备诊断的需求, 系统智能化成为故障诊断研究领域的一个重要发展方向和必然趋势。

随着设备数量越来越多、结构越来越复杂, 导致状态参量数据量呈指数上升, 设备运维难度不断增加^[2]。新能源电站设备日常状态监管信息吞吐量巨大, 面对海量繁杂信息的挑战, 数据挖掘技术在计算机科学技术的发展下应运而生, 为设备故障诊断智能化提供了新的技术支持^[3]。数据挖掘技术是否准确有效的关键在于其算法的选择, 只有确定了正确算法, 才能够从不完整的、不明确的并且包含噪声的海量信息中提取出隐含模式。但由于数据存储方式

的不断改变以及数据关系复杂化程度的迅速加剧,故障诊断对数据挖掘技术的要求越来越高,挖掘算法的选择以及挖掘流程的制定仍需进行深入研究,数据挖掘在故障诊断领域的发展任重而道远。

为确保电站设备多源海量状态信息处理的准确性和可靠性,本文根据数据挖掘研究现状,将数理统计、模式识别等方面的理论和工具引入了故障诊断领域,提出基于大数据挖掘的新能源电站设备故障诊断方法,重点挖掘不确定模型条件下被分析因素之间的关联性,删减复杂冗余数据信息的不确定性,并结合实际工程应用,制定合理的诊断流程,实现精准定位、故障判断及故障预警等功能。该诊断方法的应用,使设备运维从传统的计划维修向状态维修转变,保障新能源电站设备的稳定运行,提高了并网安全可靠,减少了并网事故停电带来的经济损失,有助于新能源大规模安全并网的实施,对我国新能源事业的发展具有重大意义。

1 基于大数据挖掘的设备故障诊断理论依据

1.1 数据挖掘聚类算法

基于大数据挖掘的设备故障诊断方法是通过异常数据参量进行聚类分析来挖掘设备常见的故障模式。本文主要使用的聚类划分方法是目被广泛使用的 k-means 聚类方法。

k-means 聚类算法的基本思想如下:(1)从数据集中的 n 个对象里面随机的选取 k 个质心,分别代表 k 个类簇的初始聚类中心点。(2)计算每个数据对象与各个簇中心的欧式距离,将数据对象划分到距离最近的簇中。(3)对于已经进行完步骤(2)的每个簇,重新进行内部处理,计算簇中对象的新均值,并把新均值作为簇的中心。(4)对比计算聚类中心,如果符合收敛条件,则算法终止,否则返回到步骤(2)继续迭代更新,直到收敛。

聚类准则函数使用的是误差平方和准则函数,其定义为:

$$G_c = \sum_{j=1}^c \sum_{k=1}^{n_j} |x_k^{(j)} - m_j|^2 \quad (1)$$

对于每个数据对象 i, 计算其所属类的函数:

$$c^i = \arg \min_j |x^i - m_j|^2 \quad (2)$$

其中 m_j 的值是聚类中心点;

对于聚类 j 中心的中心点计算函数如下:

$$m_{new_j} = \frac{\sum_{i=1}^n w_{ij} x_i}{\sum_{i=1}^n w_{ij}} \quad (3)$$

其中 x_i 为数据对象, w_{ij} 为准则函数计算数据对象 x 是否处于类簇 j 中, 处于其中则 $w_{ij}=1$, 不处于其中则 $w_{ij}=0$ 。

1.2 信息参量间的规则算法

新能源电站设备种类繁多, 采集到的信息参量数量庞大冗余, 就研究现状而言, 无法对设备参量间的关系进行深入分析, 对设备参量缺乏系统性的认识, 导致故障诊断准确度降低。关联规则的挖掘主要是以规则的形式将数据库中项与项之间所有可能的相关联系表现出来, 进而帮助人们做出管理决策的一种算法。在现有的关联规则挖掘算法中, 影响最深远、应用最广泛的是 Apriori 算法。

Apriori 算法采用了逐层搜索的迭代方法, 算法基本思路如下: 首先从事务数据库 D 中找出频繁 1 项集 L_1 ; 再根据 L_1 搜索产生候选频繁 2 项集 C_2 , 剔除 C_2 中支持度小于最小支持度的项, 得到频繁 2 项集 L_2 ; 然后由 L_2 迭代搜索产生候选频繁 3 项集 C_3 , 如此循环下去, 直至找到最大频繁 k 项集 L_k 。当数据库中的所有频繁项集都找到后, 便可根据这些频繁项集产生即满足关联规则支持度大于等于最小支持度, 又满足关联规则置信度大于等于最小置信度的强关联规则。

1、最小支持度算法如下:

已知事务数据库 D 中包含项集 A 的事务个数有 $CountA$, 那么 A 的项集支持度定义为

$$Support(A) = P(A) = \frac{CountA}{CountD} \quad (4)$$

对于某一个给定的关联规则 $A \Rightarrow B$, 已知事务数据库 D 中同时包含项集 A 和项集 B 的

事务个数有 $Count(A \cap B)$ ，那么项集 A 和 B 同时在事务中出现的概率，即其关联规则置信度为

$$Support(A \Rightarrow B) = P(A \cap B) = \frac{Count(A \cap B)}{CountB} \quad (5)$$

其中 $Support(A \Rightarrow B) \in [0,1]$

定义最小支持度为 $\min Sup$ ，其中 $\min Sup \in (0,1]$

2、最小置信度算法如下：

对于某一个给定的关联规则 $A \Rightarrow B$ ，那么项集 A 和 B 同时在事务中出现的概率，即其关联规则置信度为

$$\begin{aligned} Confidence(A \Rightarrow B) &= P(B|A) \\ &= \frac{P(A \cap B)}{P(A)} = \frac{Count(A \cap B)}{CountA} \end{aligned} \quad (6)$$

其中 $Confidence(A \Rightarrow B) \in [0,1]$

定义最小置信度为 $\min Conf$ ，其中 $\min Conf \in (0,1]$

1.3 数据模型的挖掘构建

设备在运行过程中会产生大量数据，导致原始特征计算量大，样本分布稀疏，传统的物理模型无法快速地找出异常状况并作出准确诊断评估，因此提出基于数据的建模方法。在对数据集进行处理分析，得到最为有效的特征参数的基础上，数据模型构建过程如下：

选取 Db10 小波基对数据进行 3 级小波多分辨分析，获得低高频段的 3 层小波系数；每个故障数据集要计算一个 3 维的特征向量，表达式如公式 (7) 所示：

$$f = [f_1, f_2, f_3], \quad f_i = \sum_{i=1}^n w_i^2 \quad (7)$$

其中 i 为小波分解层数， w 为每层小波分解系数；

应用混合高斯模型对多维特征数据的复杂分布进行描述，生成故障模式数学模型，其中高斯混合模型表达式如公式 (8) 所示：

$$p(z) = \sum_{j=1}^M w_j p_j(z) = \sum_{j=1}^M w_j N(z; \mu_j; \Sigma_j) \quad (8)$$

式中， Z 为特征数据向量， M 为模型混合数， w_j 为混合模型的权重系数计算

$$N(z; \mu; \Sigma) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(z-\mu)^T \Sigma^{-1}(z-\mu)} \quad (9)$$

式中， μ 为此密度函数中心点， Σ 为此密度函数的协方差矩阵。

2 新能源电站设备故障诊断方法流程分析

2.1 故障诊断方法整体设计

研究设计的新能源电站设备故障诊断方法具备数据预处理、数据聚类分析、数据模型故障诊断、故障数据预测等模块，诊断方法的具体流程图如下图 1 所示。

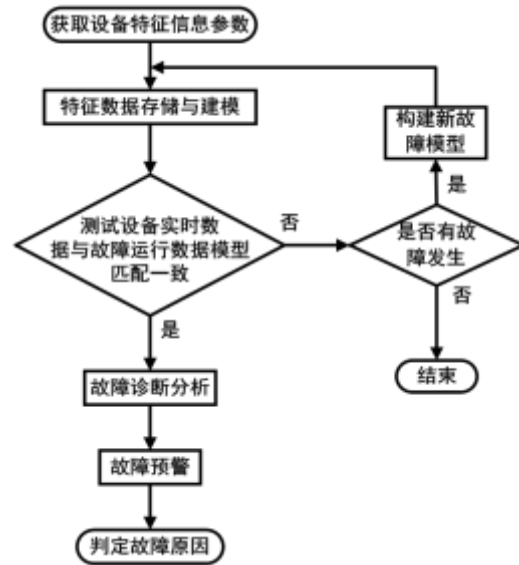


图 1 故障诊断方法整体框图

Fig. 1 Overall block diagram of fault diagnosis method

针对各类新能源电站设备，按一定周期提取设备状态信息，对其进行预处理后获得设备特征信息参数；然后将特征数据存储保存至设备运行历史数据库，针对异常状态数据的故障类型进行标注分类并构建数据模型；接下来对设备实时监测数据与故障运行数据模型相匹配，若匹配不成功且无实际故障发生，则结束

分析,若匹配不成功且有故障发生,则经过专家诊断后将新的故障模式加入故障模型库,若匹配成功,则确定故障类型并将故障定位至具体故障元件,计算显性故障在整个故障库发生的概率;最后将故障状态的关键因素作为已训练好的 BP 神经网络模型的输入值,得到故障元件隐性故障影响值并确定其所属值域,判断隐性故障发展程度,预测隐性故障发展趋势,给出隐形故障预警检测结果,并及时发出警报。

2.2 故障诊断数据处理流程

针对上述新能源电站设备故障诊断方法,本文以三相逆变器故障诊断为例详述数据挖掘过程,首先收集逆变器的异常状态数据,主要包括故障、缺陷的案例数据。三相逆变器电路接线如图 2 所示,在此重点研究逆变器最常见的开关功率管的开路 and 短路故障,特征信息参量取电压电流信号,将其作为已知量输入故障模型,模型输出阻抗与实际观测到的输出阻抗

相比较得到残差(γ),残差信号包含的故障信息用于故障决策。

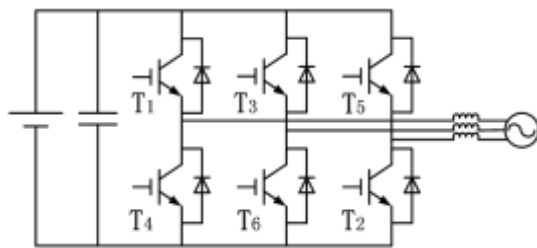


图 2 光伏并网逆变器

Fig. 2 PV grid connected inverter

收集多条逆变器运行数据,以此为基础采用 k-means 算法对原始数据进行聚类分析。为适应挖掘算法将残差离散化,残差值在阈值范围内记为“0”,阈值范围外记为“1”,可理解为该残差有无引起此类故障的可能。处理后的故障集合与残差集合的关系如下表 1 所示,其中, $\gamma_1, \dots, \gamma_7$ 为获得的残差, $R_{on1}, \dots, R_{on6}, R_{off1}, \dots, R_{off6}$ 为故障参数。

表 1 故障集合与残差集合关系表

Tab. 1 Fault set and residual set relation table

	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7
R_{on1}	1	1	0	0	1	0	0
R_{on2}	0	0	0	0	0	0	1
R_{on3}	1	0	1	0	0	1	0
R_{on4}	0	0	0	0	1	0	0
R_{on5}	1	0	0	1	0	0	1
R_{on6}	0	0	0	0	1	1	0
R_{off1}	1	1	0	0	1	0	0
R_{off2}	0	0	0	1	0	0	1
R_{off3}	1	0	1	0	0	1	0
R_{off4}	0	1	0	0	1	0	0
R_{off5}	1	0	0	1	0	0	1
R_{off6}	0	0	1	0	0	1	0

从上表可看出故障数据维数较高且含有较多冗余信息,利用 SIE 算法对高维数据进行非线性维数约简,并采用自适应极大似然估计方法估计出维数约简的目标维数,得到约简结果如下表 2 所示。数据维数降低,除去了大量无用信息,为后续故障规则的挖掘提供前提条件。

为提高故障诊断的全面性与准确性,利用 Apriori 算法对约简后的故障数据信息进行故

障规则挖掘。为挖掘出有价值、有意义的强关联规则,需要合理设置置信度和支持度值。由于设备存在多种状态参量,因此支持度的阈值不宜过大,在此设置 0.1 为最小支持度。同时,为保证获得的关联规则的高可信度,设置 0.85 为最小置信度。经上述加工处理后,最终得到的最小故障表征如表 3 所示,用以故障决策。

表 2 约简后的故障表征

Tab. 2 Fault characterization after reduction

	γ_1	γ_2	γ_5	γ_7
Ron1	1	1	1	0
Ron2	0	0	0	1
Ron3	1	0	0	0
Ron4	0	0	1	0
Ron5	1	0	0	1
Ron6	0	0	1	0
Roff1	1	1	1	0
Roff2	0	0	0	1
Roff3	1	0	0	0
Roff4	0	1	1	0
Roff5	1	0	0	1
Roff6	0	0	0	0

表 3 整合后的最小故障表征

Tab. 3 Minimum fault characterization after integration

	γ_1	γ_2	γ_5	γ_7
Ron1	1	1	1	0
Ron2				1
Ron3	1		0	
Ron4			1	
Ron5	1			1
Ron6			1	
Roff1		1		0
Roff2	0		0	
Roff3	1	0		
Roff4		1	1	
Roff5	0			1
Roff6				0

3 应用实例

新能源电站设备故障诊断方法已于 2017 年 6 月在光伏电站上线测试,对三相逆变器的关键特征信息参量——电压、电流信号进行了数据预测,同时在现场采用 5000Hz 的速率获取了晴天工况下逆变器的 9~18h 的数据有效值,每隔 6min 测一组,每组测量 40s。预测结果以及现场测量结果的分析对比曲线如图 3、图 4 所示。

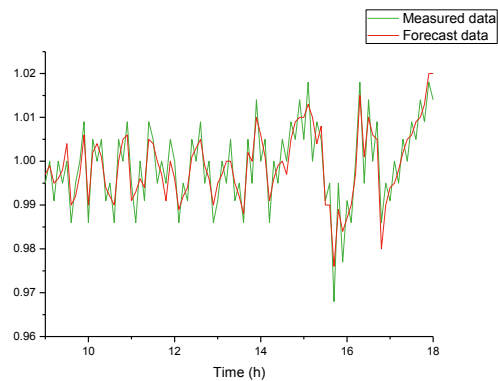
图 3 电压预测结果与实测结果的比较曲线
(标么值)

Fig. 3 Comparison curve between voltage prediction and measured results (standard value)

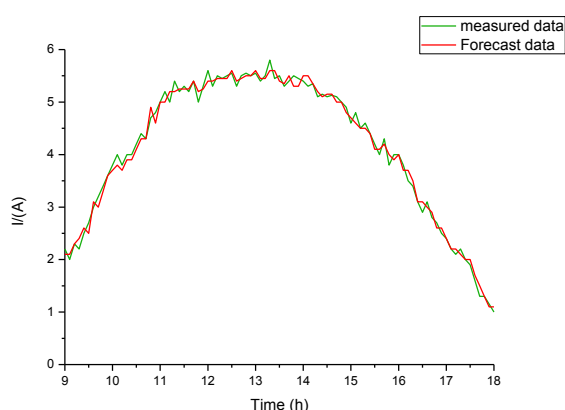


图4 电流预测结果与实测结果的比较曲线
Fig. 4 Comparison curve between current predicted and the measured results

通过对比上图中实测与预测模型的电压和电流信号,可知电压电流数据的预测结果与现场实际测量结果数值相近,曲线变化趋势近乎一致,说明本文所设计的故障诊断方法具有较为准确的故障数据预测能力,能够及早发现新能源电站设备隐性故障,判断其发展程度并及时制止故障扩散,证明了该故障诊断方法的正确性与有效性。

4 结论

新能源电站设备能否保持正常运行将直接影响到新能源并网的安全性及可靠性,对其进行状态诊断和预警评估是当前新能源并网领域关注的重点问题。为此,本文提出一种基于大数据挖掘的新能源电站设备故障诊断方法,通过聚类分析、故障规则挖掘、故障数据建模等一系列处理模式对故障特征信息以及与设备的关联关系进行提取,减少了诊断信息的冗余性,实现了设备故障准确判别以及故障数据精准预测。本文最后通过故障特征信息预测结果与现场测量结果的对比,验证了故障数据预测能力的可靠性。该方法的设计与成功应用对于新能源电站设备具有实用价值,可有效提高其运行可靠性,延长设备使用寿命,为减轻经济损失,提高新能源并网安全稳定性起到了积极作用。

参考文献

- [1] 于博文. 数据挖掘技术在故障诊断中的应用[D]. 大庆: 东北石油大学, 2015
- YU Bo-wen. Application of data mining technology in fault

diagnosis[D]. Daqing: Northeast Petroleum University, 2015

[2] 周丽丽, 孙晶, 王涛. 面向设备故障诊断的数据挖掘技术方法的研究[J]. 自动化技术与应用, 2015, 34(2): 18-22

ZHOU Li-li, SUN Jing, WANG Tao. Research on data mining technology and methods for equipment fault diagnosis[J]. Techniques of automation and applications, 2015, 34(2): 18-22

[3] 袁静. 面向设备故障诊断的数据挖掘关键技术研究及实现[D]. 西安: 西安电子科技大学, 2012

YUAN Jing. Research and implementation of key technology of data mining for equipment fault diagnosis[D]. Xian: Xidian University, 2012

[4] 孟建良, 刘德超. 一种基于Spark和聚类分析的辨识电力系统不良数据新方法[J]. 电力系统保护与控制, 2016, 03: 85-91

MENG Jian-liang, LIU De-chao. A new method for identifying bad data in power system based on Spark and cluster analysis[J]. Power system protection and control, 2016, 03: 85-91

[5] Zhao J, Zhang G, La Scala M, et al. Multistage Phasor-Aided Bad Data Detection and Identification[J]. IEEE Power&Energy Society General Meeting, 2015(GHTCE)

[6] Chen W, Song Y, Bai H, et al. Parallel spectral clustering in distributed systems. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(3): 568-586

[7] 姜碧波. 基于数据挖掘的故障诊断方法研究[D]. 天津: 天津理工大学, 2011

JIANG Bi-bo. Research on fault diagnosis method based on Data Mining[D]. Tianjin: Institutes Of Technology Of Tianjin, 2011

[8] 栗然, 张烈勇, 顾雪平. 采用粗糙集联合规则挖掘算法的分布式电网故障诊断[J]. 中国电机工程学报, 2010, 30(4): 28-34

LI Ran, ZHANG Lie-yong, GU Xue-ping. Distributed power system fault diagnosis based on rough set and joint rule mining algorithm[J]. Proceedings of the CSEE, 2010, 30(4): 28-34

[9] Thong Y, Liu D. The application of K-means clustering algorithm based on Hadoop[C]//Cloud Computing and Big Data Analysis (ICCCBDA), 2016 IEEE International Conference on. IEEE, 2016: 88-92

[10] Mohamed Abubaker, Wesam Ashour. Efficient Data Clustering Algorithms: Improvements over K-means[J]. International Journal of Intelligent Systems and Applications(IJISA), 2013, 5(3)

[11] 黄彦浩, 于之虹, 等. 电力大数据技术与电力系统仿真计算结合问题研究[J]. 中国电机工程学报, 2015, 01: 13-22.

HUANG Yan-hao, YU Zhi-hong, et al. Research on the combination of power big data technology and power system simulation calculation[J]. Proceedings of the CSEE, 2015, 01: 13-22

[12] 熊卫红, 谢志成, 李兴东, 等. 基于贝叶斯网络的变压器潜在故障概率评估[J]. 电气应用, 2013, S2: 706-710.

XIONG Wei-hong, XIE Zhi-hong, LI Xing-dong, et al. Probabilistic assessment of transformer potential faults based on Bayesian network[J]. Electrical application, 2013, S2: 706-710

[13] Masrur MA, Chen Z, Murphey Y. Intelligent diagnosis

of open and short circuit faults in electric drive inverters for real-time applications[J]. IET Power Electronics, 2010, 3(2): 279-291

[14] Choi U M, Jeong H G, Lee K B, et al. Method for detecting an open-switch fault in a grid-connected NPC inverter system[J]. IEEE Transactions on Power Electronics, 2012, 27(6): 2726-2739

[15] 吴越赢. 基于数据挖掘的电力系统不良数据检测与辨识算法研究[D]. 南京:南京理工大学, 2017

WU Yue-ying. Data mining based algorithm for detecting and identifying bad data in power system[D]. Nanjing: Nanjing University of Science and Technology, 2017

[16] D'Antona G, Perfetto L. Bad data detection and identification in power system state estimation with network parameters uncertainty[C]// International Conference on Knowledge-Based Engineering and Innovation. 2015:26-31

[17] Mate; Zaharia. An Architecture for Fast and General Data Processing on Large Clusters [J]. Electrical Engineering and Computer Sciences University of California at Berkeley. 2013:1-17

[18] 许晓东, 李柯, 朱士瑞. Web 使用挖掘中 Apriori 算法的改进研究 [J]. 计算机工程与设计, 2010, 31(3): 539-541.

XU Xiao-dong, LI Ke, ZHU Shi-rui. Research on the improvement of Apriori algorithm in Web usage mining[J].

Computer Engineering and Design, 2010, 31(3): 539-541

[19] 张旭, 魏娟, 赵冬梅, 等. 电网故障诊断的研究历程及展望 [J]. 电网技术, 2013, 10: 2745-2753.

ZHANG Xu, WEI Juan, ZHAO Dong-mei, et al. Research process and Prospect of fault diagnosis of power grid[J]. Power grid technology, 2013, 10: 2745-2753

[20] Calderaro V, Hadjicostis C N, Piccolo A, et al. Failure Identification in Smart Grids Based on Petri Net Modeling[J]. Industrial Electronics IEEE Transactions on, 2011, 58(10): 4613-4623

[21] 杨超, 张霖. 数据挖掘技术在电力设备故障诊断中的应用 [J]. 电气开关, 2016, 2: 83-86

YANG Chao, ZHANG Lin. Application of data mining technology in power equipment fault diagnosis[J]. electrical switching, 2016, 2: 83-86

收稿日期:

作者简介:

孙树敏(1968-), 男, 硕士, 高级工程师, 主要从事电力系统研究工作;

李华东(1973-), 男, 高级工程师, 从事电厂热工控制理论的研究及应用工作;

张用(1983-), 男, 硕士, 工程师, 从事新能源及电力电子方面的研究工作。