

基于聚类方法的工业电气设备大数据特征识别

马文, 田园

(云南电网有限责任公司信息中心, 云南 昆明 650000)

摘要: 为了提高工业电气设备大数据分析和识别能力, 提出一种基于模糊信息熵特征提取的工业电气设备大数据特征识别方法。通过 34980A 数据采集器获取工业电气设备参数, 结合信息流融合调度方法与期望频繁项 (EFI) 采样方法融合数据参数, 并采用多分布的传感器阵列进行工业电气设备大数据采样, 得到电气设备大数据。结合数据聚类方法, 并根据大数据的个体差异度进行工业电气设备大数据信息流非线性特征重组。利用重组结果进行数学化处理, 进行特征匹配, 为特征识别提供可依基础, 最终实现大数据的多特征识别。通过仿真结果表明, 采用该方法进行工业电气设备大数据特征识别的精度较高, 特征识别过程的收敛性较好, 提高了工业电气设备的信息化管理和监测能力。

关键词: 工业电气设备; 大数据; 特征识别; 信息熵; 特征重组

中图分类号: TP391

文献标识码: A

文章编号: 1673-629X(2020)11-0190-05

doi: 10.3969/j.issn.1673-629X.2020.11.035

Feature Recognition of Big Data of Industrial Electrical Equipment Based on Clustering Method

MA Wen, TIAN Yuan

(Information Center of Yunnan Power Grid Co., Ltd., Kunming 650000, China)

Abstract: In order to improve the ability of big data analysis and recognition of industrial electrical equipment, a feature recognition method based on fuzzy information entropy feature extraction is proposed. The parameters of industrial electrical equipment are obtained by 34980A data collector, which are combined with the fusion scheduling method of information flow and the expected frequent term (EFI) sampling method. Big data sampling of industrial electrical equipment is carried out by using multi-distributed sensor array for the electrical equipment big data. Combined with the data clustering method, and according to the individual difference degree of big data, the nonlinear characteristics of big data information flow of industrial electrical equipment are reorganized. The reconstruction results are used for mathematical processing and feature matching to provide a reliable basis for feature recognition and finally realize big data's multi-feature recognition. According to simulation, the proposed method has high precision in feature recognition of big data of industrial electrical equipment and excellent convergence in feature recognition process, which improves the information management and monitoring ability of industrial electrical equipment.

Key words: industrial electrical equipment; big data; feature recognition; information entropy; feature recombination

0 引言

随着工业自动化控制技术的发展, 大量的工业电气设备得到使用, 过程中需要结合工业电气设备的运行工况特性进行自动化性能监测, 提高工业电气设备的运行可靠性^[1]。而进行自动化性能监测时需要提取运行状态下设备的运行参数大数据特征, 为自动化性能监测提供有效信息, 因此研究相关的特征提取方法在工业电气设备大数据分析和智能化应用方面具有重要意义^[2]。

侯莉莎^[3]利用聚类集成算法, 在大数据组造期间,

通过抽样方法抽取大数据的子集样本, 并进行数冗余特征排除, 结合随机子空间的 miRNA 识别算法, 对排除冗余数据后的数据子集进行融合, 最终实现大数据集合中冗余特征排除的聚类。王玥^[4]提出一种基于均值密度中心估计的大数据网络中用户特征数据准确检测的方法, 通过用户特征数据间的关联度进行计算, 获取特征数据。根据特征数据阈值判断并筛选有效数据, 利用数据不同特征的支持度与置信度对数据进行聚类, 最后以总用户数据点为核心, 实现用户数据中心密度集合, 完成数据检测。丁春晖^[5]为优化大数据分

收稿日期: 2019-10-20

修回日期: 2020-03-10

基金项目: 云南省国家电网省级项目 (5230JY18000H)

作者简介: 马文 (1981-), 男, 高级工程师, 通讯作者, 从事电网信息化相关工作。

类识别效果,提出改进蛙跳的数据分类优化识别方法,同时,结合了流形半监督算法更新并筛选最新数据个体,最终实现大数据的分类及优化识别。但上述方法仅应用于网络信息或图像数据的检测上,未应用至设备中,并未有效分析出电气设备大数据的特征。

为了提升工业电气设备大数据分析和识别能力,该文提出一种基于模糊信息熵特征提取的工业电气设备大数据特征识别方法。内容包括:应用数据采集器采集并分析相关电气设备参数大数据;经计算机处理,实现大数据可视化,并重组其特征;构建特征识别的简化数学模型;实现大数据的多特征识别。最后进行仿真测试分析,得出有效性结论。

1 工业电气设备参数大数据分析

首先需采集电气设备的参数大数据,利用采样结果进行特征重组,整合获取到的分散数据,以此提高数据识别效率。具体实现流程如图 1 所示。

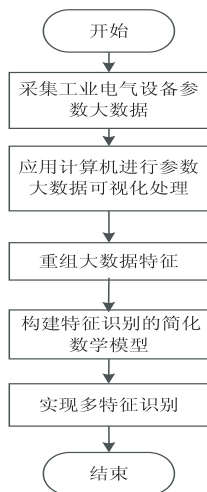


图 1 方法实现流程

1.1 参数大数据采样

1.1.1 采样装置及模块分析

由于数据较多且复杂,因此对其数据采样的要求较高,该文将通过 34980A 数据采集器获取设备参数大数据。整个数据采集器是由三部分共同组成,分别为信号预处理部分、信号调理部分以及通信部分,如图 2 所示。

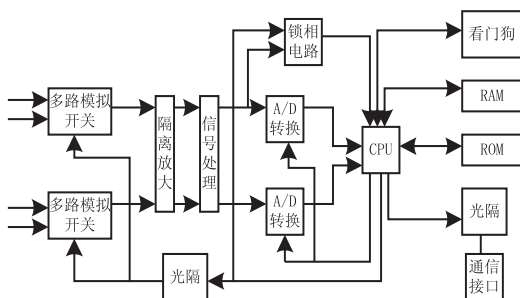


图 2 数据采集器模块

由图 2 可知,信号预处理部分是实现待检测数据切换的准备阶段;信号调理部分是切换数据的周期、信号采集阶段;通信部分是实现与 PC 传输采集信号数据的连接环节。

利用上述装置,对工业电气设备的参数进行采集。

1.1.2 计算机处理过程

为了实现对工业电气设备大数据信息的优化提取和特征识别,需进行工业电气设备大数据信息采样。结合信息流融合调度方法,进行工业电气设备大数据的统计信息模型构建,通过期望频繁项(EFI)采样方法,进行大数据的聚类。而通常情况下聚类算法可以分为两步进行处理,一是目标识别,二是聚类处理。目标识别阶段先要对目标区域进行网络化,通过目标的轮廓或者是数据识别,获取高密度的格网概率密度函数。可以使网格化的目地在于加快目标的识别速度,以便于更加快速的定位。进行聚类处理时,在目标标识中间结果的基础上,通过 DENCLUE 得到概率密度函数局部值定义密度吸引子,以此确定最终的聚类结果。

此密度聚类的算法对于噪声数据有很强的抗干扰能力,另外,与其他的计算方法进行对比,该算法能够发现随意聚类,更好地叙述空间分布模式,以此提取更多的工业电气设备大数据信息聚类特征。同时对监控区域历史视频以及异常事件信息统计,对非正常与正常情况下的工业电气设备大数据信息特征分类统计,建立一个相对应的统计分析模型。其中工业电气设备大数据特征分布式调度集函数为:

$$D(j) = t(x_j(t) + l_j(t)) \quad (1)$$

其中, $x_j(t)$ 表示工业电气设备大数据的模糊平均集 D 中的平均信息熵,描述了在第 j 个聚类中心的样本子集, $l_j(t)$ 表示模糊聚类中心中工业电气设备大数据的统计特征量, t 表示采集时长。

在满足约束条件下得到工业电气设备大数据的类信息熵满足:

$$e(D) > x_j(t) \quad (2)$$

采用模糊 C 均值聚类方法,进行工业电气设备大数据的融合处理,使用一个四元组结构来描述工业电气设备大数据的信息关联特征: $F = (X_i, P_i, e(D), T_{kl})$, 其中, X_i 为工业电气设备大数据特征信息流在 T_{kl} 时刻的信息素强度, P_i 为输出优化训练的最优概率。利用谱聚类算法,得到密度的格网概率密度函数为:

$$x_o^i = x_s^i + \frac{1}{\|x_L^i - x_s^i\|} e(D) (x_L^i - x_s^i) \quad (3)$$

其中, x_L^i 和 x_s^i 分别表示不同节点下的工业电气设备大数据值。对于 x_o^i 的计算,提取工业电气设备大数据的

关联维,根据工业电气设备大数据的类标信息进行关联性聚类处理,得到聚类输出结果:

$$L(x) = x_o^i e(D) \quad (4)$$

采用多传感器信息采样方法进行工业电气设备大数据信息采集^[6-7],得到工业电气设备参数大数据的统计分析模型为:

$$P_{i,j}^t = \begin{cases} P_{i-1,j-1}^t \times p_i + P_{i-1,j}^t \times (1 - p_i), & i - 1 = t \\ P_{i-1,j}^t, & i - 1 \neq t \end{cases} \quad (5)$$

其中, p_i 为工业电气设备大数据的谱特征对齐的概率密度函数,其大小取决区域 K 中的概率。根据上述分析,构建了大数据采集模型,为提高特征识别效率,需根据参数特征将采集到的参数大数据进行重组。而大数据特征的重组不仅为电气设备特征识别提供可依数据,还为识别应用奠定可行的理论基础。

1.2 大数据特征重组

采用多分布的传感器阵列进行工业电气设备大数据采样,对采集的工业电气设备大数据进行特征重组^[8-10],得到工业电气设备大数据信息流融合特征聚类的概率分布:

$$\sum_{\omega \in PW, C(\omega) \geq \text{minsup}} P(\omega) > \delta \quad (6)$$

其中, δ 为关联方差。结合分簇聚类融合的方法,构建工业电气设备大数据的分布式融合聚类模型,得到工业电气设备大数据信息的个体差异度函数为:

$$G(x) = ax - bx^2 + P(\omega) \quad (7)$$

其中, a 、 b 为常数, s 为变量。使用有限的频谱资源融合聚类分析的方法进行工业电气设备大数据离散化调度,得到工业电气设备大数据的统计特征检测模型为:

$$f(x) = \frac{G(x) \sum_{i=1}^{l_i} (\alpha_i + \gamma_i) K(x_i, x)}{\delta} \quad (8)$$

其中, α_i 、 γ_i 分别表示工业电气设备大数据信息流融合的高频向量和低频向量, $K(x_i, x)$ 表示相关资源的预测偏好值。结合标量序列分析方法得到工业电气设备大数据信息流的梯度向量分布模型为:

$$A(x) = \frac{1}{m \times n} f(x) \sum_{x=1}^n \sum_{y=1}^m G(x) \quad (9)$$

其中, m 、 n 分别是工业电气设备大数据的嵌入式维数。建立工业电气设备大数据的负荷波动模型,得到工业电气设备大数据信息流非线性特征重组输出为:

$$T(x) = \text{sgn}(z_R^2(k) - \hat{e}_R^2(k)) e(x) \quad (10)$$

其中, $\hat{e}_R^2(k)$ 表示信息流的观测序列, $z_R^2(k)$ 为原始训练集中的信噪比, $e(x)$ 表示不确定数据库。

综合上述,完成工业电气设备大数据信息流非线性特征重组。在前文的基础上,进行大数据特征识别。

2 大数据特征识别

2.1 特征识别的简化数学模型

对采集的工业电气设备大数据进行特征重组,提取元数据结构特征,通过对稳态电压平衡特征量进行统计,结合统计信息分析方法,得到模糊迭代状态方程为:

$$A(x) = AJ(x)a(x) + B(1 - b(x)) \quad (11)$$

令 $A = \{a_1, a_2, \dots, a_n\}$ 为终端用户分配的负载特征集, $B = \{b_1, b_2, \dots, b_m\}$ 为分布属性类别集, $a(x)$ 为负载特征集修正系数, $b(x)$ 为分布属性类别集修正系数, $J(x)$ 为检测幅值。

根据上述分析,结合傅里叶变换分解结果,计算工业电气设备大数据特征识别的简化数学模型,其可以用下述公式进行描述:

$$G_n = b_1 a_1 + b_2 a_2 + \dots + b_n a_n \quad (12)$$

其中, a_n 和 b_n 都具有较强的关联性,分别表示工业电气设备大数据的偏差限制和谐波振荡值。根据上述分析,构建了工业电气设备大数据的分布式特征提取模型。

2.2 多特征识别

对采集的工业电气设备参数大数据进行特征重组,提取其元数据结构特征,采用多模状态重组的方法进行特征匹配,得到特征识别的关联规则量。其中特征识别的统计分析模型为:

$$T(j) = \frac{Gh_i}{\sum_{j \neq i} h_i p_i + \delta^2} \left(\frac{1}{a_i + \gamma_i} - \beta_c \right) \quad (13)$$

其中, G 表示工业电气设备大数据中的主成分, h_i 表示信息流属性值, β_c 表示原始训练集中的信噪比。

结合元数据分析的方法,进行工业电气设备大数据的信息熵检测^[11],得到结构重组信息分量为:

$$\eta_{\text{comm}} = \frac{k_i}{E_{\text{comm}}} \cdot (1 - p_{\text{drop}}) \quad (14)$$

其中, k_i 为负荷容量, E_{comm} 为重组频率, p_{drop} 为能耗分布。任意选取一采样周期进行大数据负荷调度,提取元数据结构特征,根据元数据的结构分布进行模糊信息聚类处理,实现工业电气设备大数据的优化特征识别,输出为:

$$f(x) = \text{sgn} \left\{ \sum_{j=1}^l \alpha^* K(x_i, x) + b^* \right\}, x \in R \quad (15)$$

其中, a^* 、 b^* 分别表示不同负荷下的大数据传输功耗。根据以上方法,提取大数据特征,结合特征分类技术,实现工业电气设备大数据特征识别。根据特征融合结果实现大数据的优化特征识别^[12-14],构建工业电气设备大数据特征识别的模糊聚类分布模型,其表达式为:

$$F_b = \min_{w, \xi} \frac{1}{2} \|w\|^2 + f(x) \sum_{j=1}^l u(x_j) \xi_j \quad (16)$$

其中, w 表示多维大数据流的采样时间延迟; $u(x_j)$ 表示大数据分布范围; ξ_j 表示聚类系数。模糊聚类分布模型的约束条件为:

$$\begin{aligned} \text{s. t. } & y_i(w \cdot u(x_j) + 1) + \xi_j \geq 1 \\ & \xi_j \geq 0, j = 1, 2, \dots, l \end{aligned} \quad (17)$$

根据工业电气设备大数据的相互耦合关系,引入高维相空间重构,得到其核函数 $H(x)$, 则工业电气设备大数据监测的分辨函数为:

$$D_{fb} = \min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i \alpha_i \alpha_j H(x) - \sum_{j=1}^l \alpha_j \quad (18)$$

其约束条件为:

$$\begin{aligned} \text{s. t. } & \sum_{j=1}^l y_j \alpha_j = 0 \\ & 0 \leq \alpha_j \leq u(x_j), j = 1, 2, \dots, l \end{aligned} \quad (19)$$

其中, a_j 表示关联规则集。工业电气设备大数据的实时输出序列表示为 $\{x_i\}_{i=1}^N$, 在约束条件的限制下,取分辨函数的最小值,并利用该值进行工业电气设备大数据的实时流量的特征映射,映射过程为:

$$\Phi: D_{fb} \rightarrow MR^{2d+1} \quad (20)$$

通过映射可获取工业电气设备大数据运行过程中的有效信息,利用信息。将获取信息作为输入值,进行工业电气设备大数据的多特征识别输出,得到最终识别结果为:

$$P_d = \sum_{k=-K}^K c_k e^{-2\pi} \quad (21)$$

其中, c_k 表示工业电气设备大数据的多元信息特征分布集, e 表示特征识别强度。为检验该方法的有效性及其可行性,需进行仿真实验。

3 仿真测试分析

通过 Matlab 7 测试该方法在实现工业电气设备大数据特征识别中的应用性能,给出工业电气设备的大数据分布式采样的电压参数为 100 kV,对工业电气设备大数据搜索的维数为 8,关联信息熵为 0.35,其他仿真参数设定见表 1。

表 1 参数设定

参数	取值
工业电气设备分布区域	500×1 200
工业电气设备大数据采样 Sink 的位置	(0,0)
工业电气设备大数据规模	120 kBps
统计频率	1 000 kHz
耦合系数	0.34
数据采集软件接口	16 通道
衰减系数	0.34 pJ/(bit · m ⁴)
权重系数 w	0.73

根据上述参数设定,进行工业电气设备大数据特征识别,得到采集的工业电气设备大数据时域分布如图 3 所示。

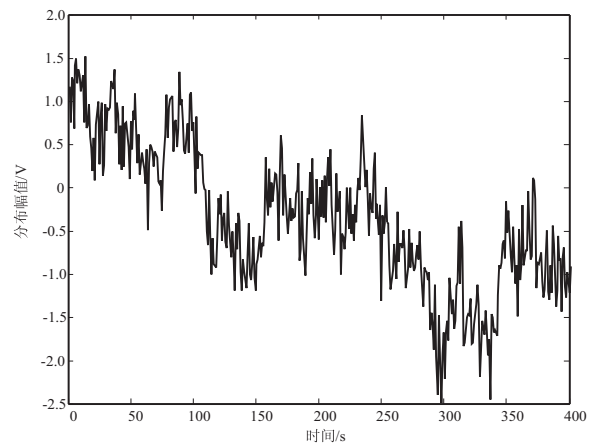


图 3 工业电气设备大数据采集时域分布

分析图 3 可知,工业电气设备大数据在单位时间内分布幅值大小较为稳定。利用上述过程获取的数据,建立工业电气设备大数据的多分布特征检测和属性聚类模型,采用模糊融合方法进行工业电气设备大数据的多元信息重构,根据信息熵特分布进行大数据特征识别^[15-17],得到的特征识别结果如图 4 所示。

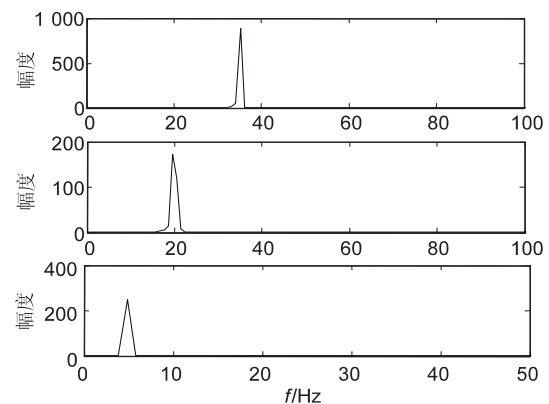


图 4 工业电气设备大数据特征识别结果

分析图 4 可知,采用文中方法能有效实现工业电气设备大数据特征识别,工业电气设备大数据的特征识别准确性较高^[18-20]。测试工业电气设备大数据特征识别的精度,得到的对比结果见表 2。

表 2 工业电气设备大数据特征识别的精度对比

迭代次数	文中方法	文献[3]	文献[4]
100	0.943	0.837	0.845
200	0.974	0.865	0.876
300	0.992	0.921	0.942
400	0.998	0.947	0.967

分析可知,文中方法进行工业电气设备大数据采集的精度较高。这是由于文中方法在采集数据的过程中进行了数据预处理,排除了冗余数据的干扰,故而提

高了特征识别的精度^[21-24]。

4 结束语

工业电气设备大数据特征识别是实现工业设备的运行工况监测的关键技术。提出了一种基于模糊信息熵特征提取的工业电气设备大数据特征识别方法。采用模糊 C 均值聚类方法,进行工业电气设备大数据的融合处理,建立工业电气设备大数据的负荷波动模型,采用多模状态重组的方法,进行工业电气设备大数据特征匹配,结合信息熵特分布进行大数据特征识别。分析可知,利用该方法进行工业电气设备大数据特征识别的精度较高,在工业电气设备的信息化管理和监测中具有很好的应用价值。

参考文献:

- [1] 李永攀,刘正江,郑中义. 基于时空密度的船载 AIS 数据聚类分析方法研究[J]. 重庆交通大学学报:自然科学版, 2018,37(10):117-122.
- [2] 庞人铭,王波,叶昊,等. 基于 PCA 相似度和谱聚类相结合的高炉历史数据聚类[J]. 山东大学学报:工学版, 2017,47(5):143-149.
- [3] 侯莉莎. 大数据集中冗余特征排除的聚类算法设计[J]. 现代电子技术,2018,41(14):48-50.
- [4] 王玥. 大数据网络中用户特征数据准确检测仿真[J]. 计算机仿真,2017,34(6):415-418.
- [5] 丁春晖. 关于大数据网络中数据分类优化识别研究[J]. 计算机仿真,2018,35(8):307-310.
- [6] 王惠杰,李鑫鑫,许小刚,等. 基于电厂工况划分的模糊 C-均值聚类算法研究[J]. 电力科学与工程,2016,32(11):55-59.
- [7] 申兆慕,张健钦,王胜开,等. 基于 Hadoop 的交通大数据的可视化 Web GIS 平台的设计与实现[J]. 北京建筑大学学报,2018,34(3):33-41.
- [8] 姚富光,钟先信,周靖超. 粒计算:一种大数据融合智能建模新方法[J]. 南京理工大学学报,2018,42(4):503-510.
- [9] 宋人杰,余通,陈宇红,等. 基于 MapReduce 模型的大数据相似重复记录检测算法[J]. 上海交通大学学报,2018, 52(2):214-221.
- [10] 汪星,黄小瑜,刘瑄璞,等. 面向工业大数据的多层增量特征提取方法[J]. 西安电子科技大学学报:自然科学版, 2018,45(4):106-111.
- [11] 杨霁,曾现均,王晋宇,等. 基于激光检测及大数据分析的电力设备损伤研究[J]. 激光杂志,2018,39(12):78-82.
- [12] 晋远,孙红三,叶紫,等. 基于大数据 Bayes 分类的家电设备识别算法[J]. 建筑科学,2017,33(4):31-38.
- [13] 毛国君,胡殿军,谢松燕. 基于分布式数据流的大数据分类模型和算法[J]. 计算机学报,2017,40(1):161-175.
- [14] 陈建峡,朱季骐,张月. 基于 Spark 的输变电路实时故障监测研究[J]. 计算机工程与应用,2018,54(5):265-270.
- [15] 王昀海,何利力. 物化查询表和多维群集在大数据查询中的应用[J]. 工业控制计算机,2015(1):122-124.
- [16] 笃峻,张海宁,柏杨,等. 智能变电站设备状态监测系统通信关键技术及实现[J]. 电力自动化设备,2016,36(4):151-156.
- [17] 高云峰,李江,王新宇,等. 大型变电站分布式自动化检修系统设计[J]. 计算机测量与控制,2017,25(10):255-258.
- [18] 张巧霞,贾华伟,叶海明,等. 智能变电站虚拟二次回路监视方案设计及应用[J]. 电力系统保护与控制,2015,43(10):123-128.
- [19] 张一茗,余亚东,李少华,等. 智能变电站断路器与二次设备融合技术深化研究及装置研制[J]. 电子产品世界, 2017,24(7):66-70.
- [20] 廖永福,林磊,李傲,等. 移相串联谐振高压电容器充电电源谐振参数设计方法及其电流控制策略[J]. 电工技术学报,2016,31(16):83-92.
- [21] 何晓阳,王小英,王铮. 高压输电线自具电源的设计[J]. 电子技术应用,2015,41(3):137-140.
- [22] 陈晓科,钟榜,李钙,等. 三相不平衡配电系统不同补偿方式下的线路损耗[J]. 广东电力,2017,30(7):96-100.
- [23] 陈子辉,王硕君,刘勇浩,等. 基于数据质量的电能质量监测装置远程可靠性评估[J]. 广东电力,2017,30(1):73-78.
- [24] 周鑫,田兵,许爱东,等. 基于 CYMDIST 的配电网运行优化技术及算例分析[J]. 电网与清洁能源,2015,31(2):91-97.