

基于电能质量监测数据的企业环保异常工况识别

张逸, 姚文旭, 邵振国, 张良羽

(福州大学电气工程与自动化学院, 福建省福州市 350108)

摘要: 针对目前污染企业环保工况异常监测实施困难、识别误差大、结果易被篡改等问题,提出了一种基于电能质量监测数据的环保异常工况识别方法。区别于对每个设备安装分表进行用电监测的现有方案,使用企业设备公共用电入口处非侵入式负荷监测所得的多维电能质量数据进行工况分类模型训练,实现异常工况识别。首先,对表征生产情况的特征数据进行时序变点检测与聚类计算,实现企业生产工况的划分;然后,结合环保设备运行情况得到用于分类的环保工况类别;进而,采用 Stacking 集成学习模型对环保相关的工况场景进行分类学习;最后,利用所训练的分类模型识别出企业存在的环保异常工况。利用仿真测试数据与实际企业数据验证了所提方法的有效性。

关键词: 环保工况; 电能质量; 监测数据; 非侵入式负荷监测; 异常工况识别; 环保监测; 突变点检测; 工况分类; Stacking 模型

0 引言

在中国生态文明建设不断推进过程中,大数据、云计算等技术手段已逐步应用于污染治理、气候预测等生态环境领域,成为推进环境治理体系现代化的重要方式^[1-2],数字环保与智慧环保体系在环境治理应用与环保产业转型上将发挥重要作用^[3-4]。

在企业的污染与环保工况监测问题上,传统方案常采用监测末端水、气的方式,判定企业污染排放是否超标,但许多企业仅在定期监管检查时应付治理,监测缺乏实时性。在智慧环保的背景下,相关人工智能计算方法也应用到了重点企业污染排放物的监测分析中^[5],而这些方案成本较高、适用范围较小。近年来,考虑到电力数据具有实时性好、可挖掘价值高等特点^[6],相关部门运用电力数据来开展环保与污染治理工作,文献[7-8]通过在污染源企业中的产污、治污、排污等设施处分别安装电力监测装置,利用用电数据判断各个设备是否启用,进而判定环保情况。但此类方案需要对企业中与污染相关的设备一一监测,是一种侵入式的监测方式。当企业内部设备众多时,用电监测设备的数量也随之增加,用电线路改造也相应复杂,对生产影响较大,需要企业配合较多。

随着大数据分析机器学习等技术发展,非侵

入式负荷监测(non-intrusive load monitoring, NILM)受到了广泛的研究与应用^[9-10]。NILM是一种在用户用电入口处监测用电数据,通过事件检测、特征提取等方法分解出各用电负荷情况的监测方法^[11-12]。本文研究拟借鉴NILM思路,使用产污、治污设备所在生产单元电气入口处的用电数据,识别企业环保情况,以克服侵入式监测的不足。

针对工业企业的NILM场景,由于生产环节复杂,涉及用电设备种类与数量众多,难以实现完整的负荷分解过程,相关的研究与应用中常是根据实际需求合理简化。文献[13]中将完整的企业负荷分解,缩小至以厂房为单位,实现厂房内各个设备的能耗监测;文献[14]将工业负荷投切时的用电波形与模板进行匹配,识别不同的负荷投切事件。考虑到污染企业的环保监测任务是针对“生产排污时环保环节正常投入”“生产排污时环保环节未投入”等直接反映环保正常与否的工况进行识别,因此,本文在实际污染企业场景,将完整的负荷功率识别任务转换为对环保工况类别的识别。

关于数据需求方面,在传统NILM中,要想较精准地识别各负荷情况,常需要利用1 kHz以上的高频采样波形数据,并从中提取谐波、电压-电流轨迹等特征进行计算^[15],但高频的数据又面临测量采集要求与长期存储成本较高的问题。随着对电能质量问题的重视,电网与用户两侧均安装了大量在线电能质量监测装置,其监测数据具有特征类型多、频率相对较高、在长期应用过程中监测相对稳定可靠等

收稿日期: 2021-12-03; 修回日期: 2022-04-28。

上网日期: 2022-08-18。

国家自然科学基金资助项目(51777035)。

特点^[16-18],故本文采用企业的电能质量监测数据作为后续工况分析识别的数据源。

综上,本文针对环境污染企业的环保监测问题,采用非侵入监测方案,将企业环保监测问题转化为对环保工况的分类与识别。基于电能质量监测数据,首先,对表征生产工况的特征数据进行时序变点检测与聚类,实现企业总体生产工况的划分,进而提取环保工况类别;然后,采用 Stacking 分类模型,对环保相关的工况进行分类学习,识别出环保环节异常的工况;最后,利用所训练的工况分类模型识别出环保异常工况,并在仿真算例与实际企业数据中验证。

1 污染企业环保工况提取

污染企业环保工况与生产工况及环保设备运行情况有关,二者可通过人工现场记录或安装设备状态采集终端的方式获取。但实际中生产设备类型和数量均较多,现场记录和安装终端成本较大。同时,不同生产工况下用电数据常表现出不同特征且差异明显,使得通过数据分析以区分生产工况成为可能。因此,对于企业生产工况,本文通过数据挖掘方式获取,减少实际应用过程中对多种生产设备单独监测的依赖;对于环保设备运行情况,因设备较少,为保证准确,仍以人为记录或安装采集终端的形式进行运行状态获取。生产工况判断基于变点检测理论,以下进行详细说明。

1.1 时序数据突变点检测

突变点是时序中存在的引起数据前后均值、方差等统计量出现明显变化的点,也可称作变点,突变前后数据特征的变化常反映出实际物理状态的改变^[19]。在用电场景中,突变点反映出负荷用电过程的变化,因此,可通过计算时序突变点,寻找用电工况变化的时刻。具体地:对于用电时序: $Y = \{y_1, y_2, \dots, y_n\}$,假设存在一个由 ω 个突变点构成的突变点集合 $T = \{\tau_1, \tau_2, \dots, \tau_\omega\}$,其中,每个突变点的位置均为1至 n 之间的整数,并定义 $\tau_0 = 0$, $\tau_{\omega+1} = n$,则这些突变点将原时序分割为 $\omega + 1$ 段子序列,其中,第 i 段可表示为: $Y_i = \{y_{\tau_{i-1}+1}, y_{\tau_{i-1}+2}, \dots, y_{\tau_i}\}$ 。

以上多变点检测问题常常转化为求取最小目标函数来求解^[20-21],目标函数 $F(n)$ 的核心是一个能表征数据同质性的代价函数 $C(\cdot)$,如式(1)所示。当代价函数值越小时,表明数据越呈现出同一种统计性质;反之,在一组性质相同的数据中引入一个突变点时,数据的代价函数值将增大。

$$F(n) = \min \left(\sum_{i=1}^{\omega+1} C(Y_i) + \beta f(\omega) \right) \quad (1)$$

式中: $C(Y_i)$ 为第 i 段子序列所计算的代价函数,为提高实际应用时对不同企业、不同类型用电数据的适用性,不同分布类型的数据对应不同计算方式,详见附录A式(A1); $\beta f(\omega)$ 为目标函数优化的附加惩罚项,用于平衡所计算的突变点数量,减少出现变点过多或过少的现象,其中 β 为惩罚因子, $f(\omega)$ 为一个随变点数量增多而增大的惩罚函数。

令惩罚项 $f(\omega) = \omega + 1$,则求取 $F(n)$ 的迭代式可由式(2)推导得出,进而将式(1)中求解全局最优解的问题转化为不断寻找前一个变点的过程。

$$\begin{aligned} F(n) = F(\tau_{\omega+1}) &= \min \left(\sum_{i=1}^{\omega+1} C(Y_i) + \beta(\omega + 1) \right) = \\ &= \min \left(\sum_{i=1}^{\omega+1} (C(Y_i) + \beta) \right) = \\ &= \min \left(\sum_{i=1}^{\omega} (C(Y_i) + \beta) + C(Y_{\omega+1}) + \beta \right) = \\ &= \min (F(\tau_\omega) + C(Y_{\omega+1}) + \beta) \end{aligned} \quad (2)$$

设置初值 $F(0) = -\beta$,在1至 n 中循环计算,将满足目标函数最小的时间点作为新的变点,不断迭代求解出所有的突变点。此外,在每次搜寻过程中,去除不可能成为变点的数据索引,减少计算量。引入如下的判断条件:假设对于数据时间点 $t, s, l (t < s < l)$,存在一个常数 K 满足式(3),则当式(4)成立时,认为 t 不可能成为 l 之前的最后一个变点,将其排除^[20-21]。

$$C(y_{t+1}, y_{t+2}, \dots, y_s) + C(y_{s+1}, y_{s+2}, \dots, y_l) + K \leq C(y_{t+1}, y_{t+2}, \dots, y_l) \quad (3)$$

$$F(t) + C(y_{t+1}, y_{t+2}, \dots, y_s) + K \leq F(s) \quad (4)$$

1.2 变点整合与环保工况类别提取

基于变点检测方法划分生产工况,并结合环保设备运行情况得到环保工况,步骤如下:

1) 选择 m 个能够反映企业生产工况的特征数据 $\{X_1, X_2, \dots, X_c, \dots, X_m\}$,例如:在生产设备容量占比较大时,使用能直接体现总体用电量的功率型数据,如图1实例中,选择各相有功功率(P_A, P_B, P_C)与三相总有功功率 P_{3ph} 数据;又如主要生产设备是谐波、三相不平衡等干扰源时,也可选择电流总谐波含量、不平衡度等电能质量指标。

2) 对 m 类特征时序分别计算变点,得到式(5)中的总变点集合 O ,时序波形示意图见图1(a)至(d),其中,平行于纵轴的直线即为变点所在位置。

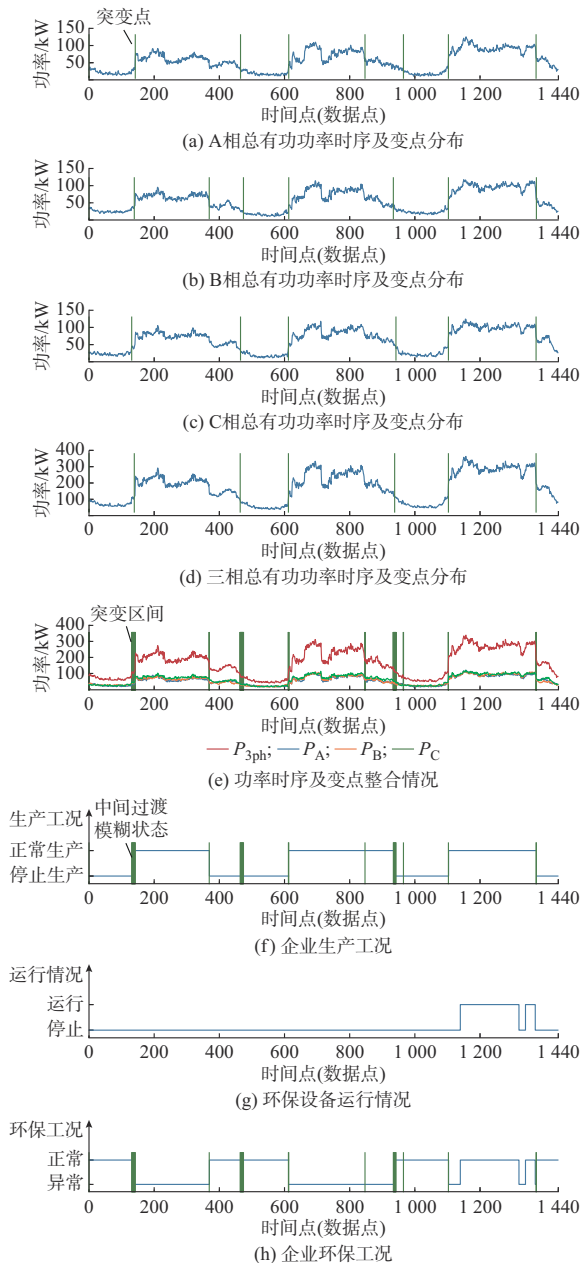


图1 污染企业环保工况提取示例图

Fig.1 Example of extraction of environmental protection operation condition in polluting enterprises

$$\begin{cases}
 O = \{O_1, O_2, \dots, O_m\} \\
 O_1: \{t_{X_1,1}, t_{X_1,2}, \dots, t_{X_1,n_1}\} \\
 O_2: \{t_{X_2,1}, t_{X_2,2}, \dots, t_{X_2,n_2}\} \\
 \vdots \\
 O_m: \{t_{X_m,1}, t_{X_m,2}, \dots, t_{X_m,n_m}\}
 \end{cases} \quad (5)$$

式中： X_1 至 X_m 表示 m 类时序数据； O_1 至 O_m 表示 m 类时序数据各自的变点集合； $t_{X_m,i}$ 为集合 O_m 中的第 i 个变点； n_1 至 n_m 为各集合中的变点数量。

3)由于实际中各相用电数据非完全平衡,各时序提取的突变点并非一一对应,且部分工况转换过

程需持续一定时间而非瞬时完成。因此,考虑对式(5)中不同特征时序的变点进行归并,并将时间相近的变点整合为区间,如图1(e)所示。具体过程为:首先,将 O_1 至 O_m 做并集处理,并将得到的结果按照从小到大排列,得到 O_{all} ,如式(6)所示;然后,设置过渡时间阈值 ϵ (本文将该阈值设定为30 min,在实际应用时,可视环保监管要求的严格程度调整所容许的过渡环节时长);最后,对 O_{all} 中的变点逐一判断,当满足 $t_{i+1} - t_i \leq \epsilon$ 时,将前后两变点构成的区间 $[t_i, t_{i+1}]$ 作为变化时段,得到新的变点集合 O'_{all} ,如式(7)所示。

$$O_{all} = O_1 \cup O_2 \cup \dots \cup O_m = \{t_1, t_2, \dots, t_i, \dots, t_n\} \quad (6)$$

$$O'_{all} = \{t_1, [t_2, t_3], \dots, [t_k, t_{k+1}], \dots, t_{n-1}, t_n\} \quad (7)$$

式(6)中 $t_1 < t_2 < \dots < t_i < \dots < t_n$, i 为1到 n 之间的整数;式(7)中 $[t_2, t_3]$ 与 $[t_k, t_{k+1}]$ 均表示整合后的变点区间。

4)采用k-means聚类方法,对由变点分段后的各子段均值数据进行聚类,聚类数选为2,分别对应企业正常生产工况与停止生产工况,具体过程见附录B。此外,对于第3)步中计算的变点及变点区间,将其作为工况间切换的中间过渡模糊状态。最终,得到企业生产工况时序分布,如图1(f)所示。

5)根据环保相关工况提取规则(详见附录A图A1),将第4)步中所得的生产工况与环保设备运行情况组合,可得到环保工况,对应于图1(f)至(h)。其中,过渡环节由于数量少、数据特征存在变化,在后续环保异常识别与监管过程中不作考虑。

由于在环保工况正常与异常2类工况下,企业用电数据特征存在差异,如附录A图A2所示,图中展示的是A相电流总谐波含量、B相5次谐波电流方均根值以及三相总有功功率数据,数据为图1(h)中各工况时段的均值。基于这种差异性,可将多维电能质量监测数据作为特征,环保工况作为标签训练分类模型,识别异常环保工况,具体分析见第2章。

2 基于Stacking模型的环保工况识别

2.1 Stacking集成分类原理

不同的机器学习模型由于原理与训练方式不同,存在各自的适用性与特点,如支持向量机(support vector machine, SVM)对高维数据与非线性问题的处理上更加稳健且对极值不敏感;逻辑回归(logistic regression, LR)的概率原理清晰、实现方式简单高效,在线性问题上处理更具优势;K最邻近分类算法(K-nearest neighbor, KNN)不对数据进行假设,易于实现;随机森林(random forest, RF)不易出现过拟合现象,可以处理高维度的数据;梯度提升

决策树(gradient boosted decision tree, GBDT)对数据异常值的鲁棒性较强。但实际中不同污染企业设备类型和用电情况多变、监测数据特征存在一定差异,单一的模型常难以适用于后续不明确的污染企业场景。为提高对实际数据的适用性,本文选择Stacking集成模型作为污染企业环保工况识别的分类模型,通过对不同分类器的学习结果进行堆叠组合的方式,从多个模型不同的角度挖掘数据特征,取长补短,提高模型泛化能力与工况识别准确度^[22-23]。

如附录A图A3所示,Stacking模型结构通常由2层构成。第1层由若干个基分类器构成,其中,考虑到不同模型对数据的适用程度不同,基分类器的确定通常是分别利用SVM、LR、RF、KNN、GBDT等不同模型,在不同参数或限定条件下,对数据集进行初步测试并从中选取效果较优的3~5个分类模型,训练的特征与标签分别为多维电能质量监测数据与环保工况。将原始数据在各个基分类器下得到的分类结果(概率得分形式)作为第2层元分类器的输入特征数据,元分类器模型与基分类器类似,可选择为常用的单一机器学习分类模型,但由于其训练的特征仅余第1层计算输出的结果,需适当降低模型复杂度以减少过拟合现象,如在LR中增加正则化,在树模型中减少数量与深度等。最后,在第2层元分类器中对基分类器计算所得的特征进行训练学习,计算出环保工况类别。

此外,Stacking模型在各个分类模型训练的过程中采用 k 折交叉验证的方法,以减少样本不平衡与过拟合等影响,提高模型稳定性。如附录A图A3中上半部分所示, k 折交叉验证的具体操作为:在划分好训练集数据与测试集数据后,再将训练集数据平分为 k 份不交叉的数据,依次将其中1份设置为验证集,训练余下 $k-1$ 份数据。将 k 次训练数据预测的得分求取均值,作为最终训练的结果。

2.2 环保工况异常识别流程

本文基于电能质量监测数据进行环保工况识别的主要流程如图2所示,其主要步骤如下:

1)原始数据源获取。考虑企业用电设备数量众多,存在不同的用电行为与电能质量特性,为了在后续分类模型的训练中更好地地区分不同用电特征,选择的数据类型涵盖电流、功率这类基本电气参数以及谐波畸变率、不平衡度等体现企业电能质量特性的多维电能质量指标数据。此外,在具体场景中,还需考虑背景电网电能质量扰动情况以及其他干扰源负荷的影响,在中低压场景可筛选与用户发射特性相关性更高的电流类电能质量等特征参数。

2)环保工况标签获取。按第1章所述方法获得环保工况标签。

3)环保工况分类模型训练。将第1)步中得到的电能质量监测数据作为输入特征,结合第2)步所得环保工况标签,对一定时段的训练数据进行Stacking分类模型训练。

4)异常环保工况识别。对企业后续数据进行环保工况分析,识别异常环保工况。

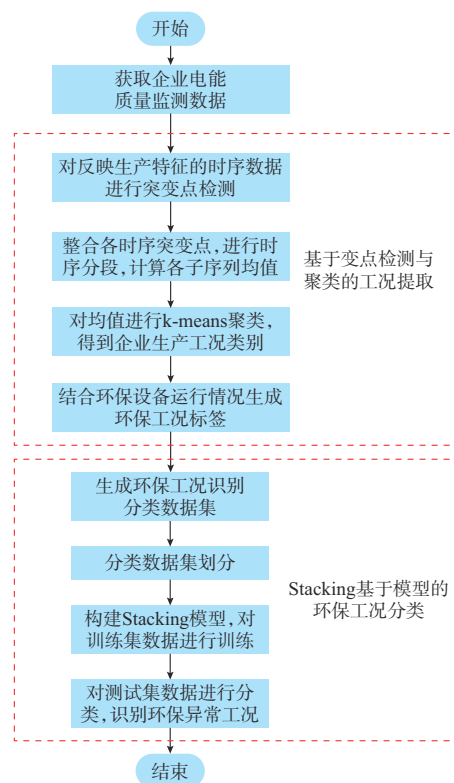


图2 环保工况分类识别流程图
Fig. 2 Flow chart of classification and identification of environmental protection operation conditions

3 仿真分析

3.1 仿真算例基本情况说明

在MATLAB/Simulink环境中搭建污染企业用电工况仿真模型,验证本文方法对不同场景的适用性。考虑实际企业中一个含生产、治污等各类用电设备的完整生产单元通常由单段/条母线供电,仿真中以一段10 kV母线进行模拟,并设置多个污染企业中常用的典型负荷,仿真电气接线示意图如图3所示。其中,在环保设备的选择上,通过调研可知:在水、气等污染治理过程中,变频、变流及自动控制技术得到广泛使用,如鼓风机、除尘风机、水泵以及成套污水处理设备均已采用相关自动控制技术进行转速、振动频率的控制,静电除尘设备还应用到升压与变流技术生成高压直流电场等^[24-25],因此,将变频负荷4与5视作2台环保治理设备;在生产设备与其他厂务设备的选择上,设置了线性负荷、变频负荷以

及部分单相负荷等不同特性的设备,用于后续组合出不同的用电场景。

参考实际电能质量监测设备的测量与数据统计方法^[26],在10 kV进线处模拟一个电能质量监测装置,进行特征指标计算与统计,通过等比例时间缩放,得到与实际监测数据相同的3 min间隔的每个仿真天(24 h)共480个点的95%概率值数据,数据类型包括各相电压与电流有效值、各相功率与总功率(有功功率、无功功率、视在功率)、功率因数、各相电压与电流的基波有效值、各相电压与电流的2至49次谐波含有率与有效值、各相电压与电流的直流分量、各相电压与电流的0.5至15.5次间谐波含有率、电压偏差以及电压与电流负序不平衡度。

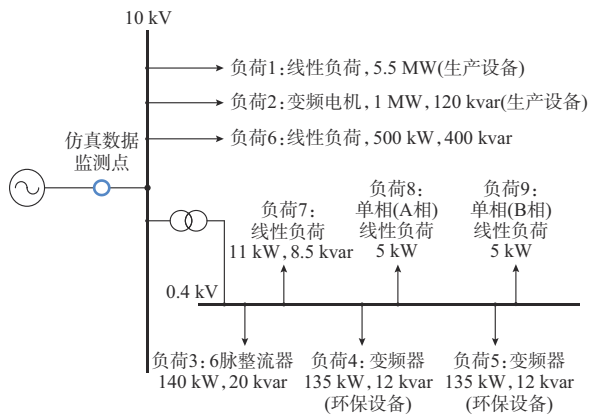


图3 仿真算例电气接线示意图

Fig. 3 Schematic diagram of electrical connection for simulation case

3.2 不同生产工况场景设置说明

仿真算例中,对于负荷1至5,在每个仿真天内设置不同随机数控制其启停时间,模拟环保设备与生产设备的不同运行情况;对于负荷6至9,设置为持续运行负荷,模拟实际企业生产过程中部分不间断运行的设备。共进行24个仿真天的测试,得到不同的用电场景,详细的负荷接入时间见附录A表A1。

对于生产工况,分别将仿真中负荷特性不同、运行时间不定的生产负荷1、2作为主要生产设备,设置3类用电场景:

场景1:线性生产负荷1运行时,视作正常生产工况;

场景2:变频生产负荷2运行时,视作正常生产工况;

场景3:线性生产负荷1与变频生产负荷2同时运行时,视作正常生产工况。

以上场景中,主要生产设备未运行时均视为停止生产,且对主要生产负荷外的其余负荷是否运行

不进行单独考虑,仅作为工况分类的干扰与不确定因素。

在生产工况划分的基础上,按照实际中环保部门要求的环保工况异常判定与监管策略,设置了2类环保工况判定规则:

规则1:生产正常时,仅当2台环保设备同时启动才视作环保工况正常,其中1台环保设备关闭则为异常。

规则2:生产正常时,至少1台环保设备启动,即可视作环保正常工况,2台环保设备均关闭时为环保异常工况。

2个规则下,生产工况为关停时均视为环保工况正常。

3.3 仿真算例结果与分析

首先,进行变点检测与工况聚类,获取不同场景中的生产用电工况。考虑主要生产负荷中负荷1的容量占比最大、负荷2为主要谐波干扰源,在场景1选择各相有功功率与三相有功功率数据,在场景2选择各相电流总谐波含量数据进行计算,各特征数据变点识别结果如附录A图A4、图A5所示。为说明本文变点检测算法的适用性与准确性,将本文算法以及其他几类常用变点检测方法的计算结果与仿真设置的生产负荷切换事件(2个场景中单一负荷切换事件均为48次)进行对比,如表1所示。

表1 不同场景下的变点计算结果
Table 1 Calculation results of change points in different scenarios

生产用电场景	方法	变点评价指标			
		识别变点总数	预设变点识别数	预设变点识别率/%	无效识别数
场景1	本文方法	64	48	100.00	16
	CUSUM ^[27]	105	48	100.00	57
	BIC ^[27]	108	48	100.00	60
	改进BS ^[28]	84	47	98.00	36
场景2	本文方法	105	45	93.75	60
	CUSUM ^[27]	321	48	100.00	279
	BIC ^[27]	177	45	91.66	133
	改进BS ^[28]	29	14	29.17	15

注:表中各变点评价指标的计算方式见附录A式(A2)至式(A4)。

由表1可以看出,本文方法在2个场景中均能够较为准确地识别出预设变点,仅在场景2中出现3个漏识别的变点。同时,由于存在其他相似负荷的投切事件干扰,例如同为谐波干扰源的环保设备投切对谐波数据变点识别存在影响等,导致各方法都额外识别出了部分无效变点,但本文方法的敏感度较基于累积和(cumulative sum, CUSUM)与基于

贝叶斯信息准则 (Bayesian information criterion, BIC) 的方法低,无效变点数量相对较少。此外,考虑后续工况分类时,将去除变点时刻数据,本文所得变点数量在合理范围内(场景1、2占比分别为0.57%、0.91%)。

然后,对变点划分后各段数据的均值进行聚类,结果在附录A图A6中以三维示意图体现,分别对应正常生产工况与停止生产工况,可以看出二者存在差异。同时,将计算的生产工况与预设生产工况进行对比,如附录A图A7所示,其中,场景3的生产工况计算则由场景1、2组合而成,计算工况与预设工况基本一致,3个场景准确度分别达99.44%、99.07%与98.72%,表明了本文方法准确有效。

最后,对3个生产用电场景分别采用2类环保工况判定规则,得到环保工况类别,取前18个仿真天的数据作为训练集,模型参数设置见附录A表A2,后6个仿真天的数据作为测试集,在附录A图A8中对各场景与规则下的模型预测值与真实值进行对比,其中,1与0分别表示环保工况正常与异常。在数据选择上,考虑实际的低压配电网中监测点谐波电压、电压总谐波畸变率、电压偏差、电压不平衡度等电压质量参数受背景中其他电能质量干扰源的影响较大,并非与用户自身行为完全相关^[29-30],本算例中选择与用户自身运行关联更密切的电流类型电能质量监测数据进行训练。从附录A图A8中可以看出,各场景与规则的分类效果均较好,真实值与预测值基本一致,仅少部分时段存在错误(如场景3中红框所示)。

计算混淆矩阵进行量化分析,混淆矩阵表示模型识别的2类环保工况结果相对真实值的情况,其4个元素分别记为 T_P 、 F_N 、 F_P 、 T_N ,其中: T_P 、 T_N 分别为分类器准确识别出的环保工况异常与正常的数量, F_N 、 F_P 表示对2类真实工况类别误判的数量。进一步,如式(8)至式(11)所示,从混淆矩阵中计算分类结果评价指标准确度 S_{ACC} 、精确率 S_{PRE} 、召回率 S_{REC} 与F1得分 S_{F1} ,其中: S_{ACC} 是对整体工况分类的准确度评价;精确率 S_{PRE} 与召回率 S_{REC} 表示对“环保异常”的识别精度与完整度,F1得分 S_{F1} 是精确率与召回率的调和平均值,体现分类器的综合特性。

$$S_{ACC} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \quad (8)$$

$$S_{PRE} = \frac{T_P}{T_P + F_P} \quad (9)$$

$$S_{REC} = \frac{T_P}{T_P + F_N} \quad (10)$$

$$S_{F1} = \frac{2S_{PRE}S_{REC}}{S_{PRE} + S_{REC}} \quad (11)$$

限于篇幅,仅列出各个方案的混淆矩阵以及衡量综合分类效果的 S_{ACC} 与F1得分 S_{F1} ,如表2所示。

表2 不同场景和规则下的分类结果
Table 2 Classification results in different scenarios with different rules

生产用电场景	环保工况提取规则	T_P	F_N	F_P	T_N	$S_{ACC}/\%$	$S_{F1}/\%$
场景1	规则1	553	2	0	230 9	99.93	99.82
	规则2	20	0	0	284 4	100.00	100.00
场景2	规则1	615	0	0	223 9	100.00	100.00
	规则2	219	0	0	263 5	100.00	100.00
场景3	规则1	340	15	0	248 9	99.47	97.84
	规则2	20	0	0	282 4	100.00	100.00

由表2分析可知,不同场景与规则下,环保工况正常与异常的数量不同,但均能得到超过95%的 S_{ACC} 与 S_{F1} ,表明本文方法能够较好地适用于不同类型和数量的生产设备与环保设备所组合的用电场景。此外,为说明本文方法数据源选取的合理性,与下列数据特征选择方式下的结果进行对比,包括:1)选择全量电能质量监测数据;2)仅选择电压、电流、功率及功率因数4类基本数据;3)仅选择电能质量指标数据;4)仅选择电流质量指标数据;5)参考文献[31]利用主成分分析法降维,选取贡献率为前90%的特征。在图4中比较各方式下的分类 S_{ACC} 与 S_{F1} 值。

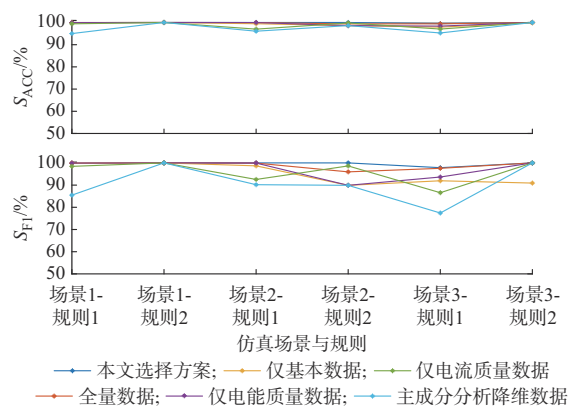


图4 不同数据特征选择下的分类效果对比
Fig. 4 Comparison of classification effect with different data feature selections

由图4分析可知,按本文所使用的去除电压质量等参数的方案以及选择全量特征时,在不同场景和规则下的分类效果均较好;仅选择基本数据、电流质量数据与电能质量数据时,在场景2、3中的分类效果出现一定波动;而使用主成分分析降维输入数

据时, S_{F1} 值在各场景与规则下的波动最大, 稳定性较差。具体分析误判情况为: 在第 23 天(4~6 h)时段, 几类方案易将异常环保工况误判为正常, 可能原因在于这一时段 2 台环保设备虽均未启用, 但与环保设备容量相近且存在电能质量干扰发射特性的整流负荷 3 正在运行, 引起干扰。以上情况表明, 当企业用电设备多、工况复杂时, 将电能质量数据与基本用电数据结合, 引入更多特征进行训练, 能够覆盖更多用电特性, 有利于分类效果提升。

4 实例验证

4.1 污染企业基本状况与监测数据情况

以中国福建某制鞋企业为试点, 进行环保工况识别, 该企业电气接线示意图如附录 A 图 A9 所示, 生产设备主要包括加热与冷却装置、异步电机、紫外线杀菌光源等; 环保设备主要为抽风机和 UV 光解废气处理设备组成的排污和治污系统。企业在生产与环保过程中存在谐波与不平衡等电能质量发射特性, 故在企业的电气入口处采集电能质量监测数据, 用于后续环保工况识别。但考虑到该污染企业所处电压等级较低, 频率以及电压类的电能质量易受电网背景波动与其他电能质量干扰源影响, 因此, 最终选择反映污染企业自身用电特性的电流类电能质量以及功率指标用于后续环保工况分类, 指标详见表 3, 数据类型为 3 min 的均值数据。

表 3 用于环保工况分类的实测特征数据
Table 3 Measured characteristic data for classification of environmental protection operation condition

特征数据基本类别	特征数据名称
基本电气量数据	各相电流有效值
	各相有功功率与三相总有功功率
	各相无功功率与三相总无功功率
	各相视在功率与总视在功率
电能质量监测数据	各相基波电流方均根值
	各相 2 至 31 次谐波电流方均根值
	各相电流总谐波含量
	正负零序电流
	三相电流不平衡度

4.2 企业环保相关工况提取

取 2021 年 1 月 9 日至 25 日共 17 d 的 8 160 条数据进行分析。首先, 利用变点检测方法, 由有功功率数据计算出生产工况, 结果如附录 A 图 A10 所示, 经核实, 与实际企业生产工况基本相同。随后, 由环保设备处安装的监测终端获取环保设备运行情况, 共同确定各时段的环保工况类别, 如表 4 所示。

表 4 环保相关工况类别情况
Table 4 Category of operation conditions related to environmental protection

企业总体生产工况	环保设备运行情况	环保工况	样本数量
生产正常	环保启用	环保正常	920
	环保关停	环保异常	3 843
生产停止	环保启用	不合理场景	0
	环保关停	环保正常	3 148
中间过渡环节		不考虑该场景	249

4.3 工况分类识别与结果分析

对数据集进行缺失数据插值补齐, 去除数值全为 0 的特征, 在附录 A 表 A3 中对这部分数据进行了示意。按 8:2 的比例将 2021 年 1 月 22 日 14:00 以前的数据进行模型训练, 后续数据进行测试, 在分类过程中设置基分类器为 LR、SVM、RF 与 GBDT, 详细参数见附录 A 表 A2。采用 5 折交叉验证进行训练, 计算上述 4 个基分类器单独作用以及在 4 种集成方式下测试集的结果混淆矩阵, 如附录 A 图 A11 所示。可以看出, Stacking 分类器的混淆矩阵非对角元数量相对于单一分类模型均有所减少。其中, 以 SVM 为元分类器的 Stacking 模型下, F_N 、 F_P 数量最少, 分类效果最优。进一步计算分类结果评价指标, 如表 5 所示。

表 5 分类结果的评价指标
Table 5 Evaluation index of classification results

分类模型	$S_{ACC}/\%$	$S_{PRE}/\%$	$S_{REC}/\%$	$S_{F1}/\%$
LR	87.49	99.20	76.23	86.21
SVM	92.04	99.14	85.22	91.66
RF	92.80	93.09	92.86	92.97
GBDT	93.37	97.07	89.78	93.28
(LR, SVM, RF, GBDT)+LR	93.37	99.30	87.68	93.13
(LR, SVM, RF, GBDT)+SVM	96.21	98.96	93.60	96.20
(LR, SVM, RF, GBDT)+RF	93.18	98.35	88.18	92.99
(LR, SVM, RF, GBDT)+GBDT	94.06	99.18	89.16	93.90

注: (LR, SVM, RF, GBDT) 表示 Stacking 分类模型的第 1 层中所选择的基分类器。

从表 5 中可以看出, 集成模型下, 分类结果的各项得分值有所提高, 其中, 第 2 种集成方案下的指标得分最高, S_{ACC} 与 S_{F1} 值均超过 96%, 相对各种单一基分类器提高了 3% 以上, 表明基于 Stacking 集成的工况分类模型可以在误判和漏判较少的情况下, 识别污染企业存在的环保异常工况。最终, 选择效果最好的以 SVM 为元分类器的 Stacking 模型作为该企业的后续工况识别模型。

4.4 实际应用效果分析

以下分析本文方法在实际污染企业环保监测系

统中长期应用的适应性。利用4.3节中选择的模型,从2021年1月22日开始,以每日24:00为时间节点,动态计算环保工况分类识别的累计 S_{ACC} 与 S_{F1} 值,直至2021年3月22日,得到如图5所示的曲线。

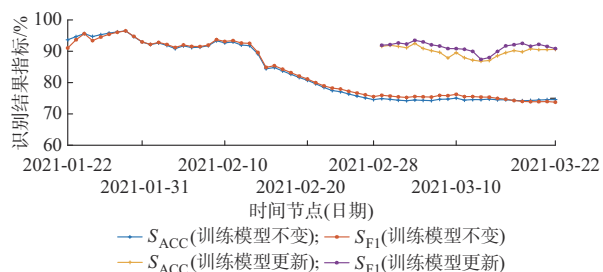


图5 环保工况长期累计识别效果
Fig. 5 Long-term cumulative identification effect of environmental protection operation conditions

由图5可以看出,截至2021年3月22日,2个月的长期数据中,识别效果最终稳定至75%左右。其中前24 d的测试数据中,两指标累计值均能达到90%以上,但在2月中下旬,指标结果有一定幅度的下降,经过现场核实发现,2月中旬为春节后企业复工初期,企业生产模式进行了一定调整,导致训练模型未能较好识别该时段的工况。为提高识别效果,将2月的数据归入原始训练集中进行训练,最终在3月测试的平均 S_{ACC} 与 S_{F1} 分别达91.34%、89.74%,如图5中更新训练模型后的结果曲线所示。可见,本文方法在实际应用中具备一定效果,但在实际用电场景模式变更后,模型识别精度有所下降,需要考虑使用现场核实过的数据更新识别模型等方法来提升结果准确度。

5 结语

本文将污染企业环保监测问题转化为对环保环节相关用电工况的识别问题,利用电能质量监测数据进行学习训练,有效识别环保异常现象,得到结论如下:

1) 相比于对企业生产、环保设备进行一一监测的传统用电监测方式,本文方法有效减少了所需监测终端成本以及对正常生产的影响,更易于实际工程实施。

2) 本文通过滤除过渡环节、引入多维电能质量监测数据以及利用不同分类模型集成的优势,达到较好的准确度与适用性。

3) 本文方法充分利用现有电力监测数据资源,是电力数据的跨行业应用,其思想与方法可拓展到更多电力以外的行业应用场景中。

本文方法应用过程中也存在一定的限制:本文仅考虑正常生产与停止生产以及环保工况正常与异常2类,在未来需进一步开展细化研究工作,且在长期应用过程中,也需综合考虑模型更新成本、风险与识别结果准确性之间的关系,进一步优化企业用电状态不断变化时的模型更新策略。

本文相关实测和仿真电能质量数据已共享,可在本刊网站支撑数据处下载(<http://www.aeps-info.com/aeps/article/abstract/20211203007>)。

附录见本刊网络版(<http://www.aeps-info.com/aeps/ch/index.aspx>),扫英文摘要后二维码可以阅读网络全文。

参考文献

- [1] 赵苗苗,赵师成,张丽云,等.大数据在生态环境领域的应用进展与展望[J].应用生态学报,2017,28(5):1727-1734.
ZHAO Miaomiao, ZHAO Shicheng, ZHANG Liyun, et al. Applications of eco-environmental big data: progress and prospect [J]. Chinese Journal of Applied Ecology, 2017, 28(5): 1727-1734.
- [2] 秦宇,孙宇,邢可佳,等.基于生态环境大数据建设的环境信息安全监管体系研究[J].环境保护,2018,46(21):56-60.
QIN Yu, SUN Yu, XING Kejia, et al. Investigation on the supervision system of environmental information security in the construction of ecological environment big data [J]. Environmental Protection, 2018, 46(21): 56-60.
- [3] 胥彦玲,李纯,闫润生.中国智慧环保产业发展趋势及建议[J].技术经济与管理研究,2018(7):119-123.
XU Yanling, LI Chun, YAN Runsheng. Trends and suggestions for the development of smart environmental protection industry in China [J]. Journal of Technical Economics & Management, 2018 (7): 119-123.
- [4] 李信茹,周民,米屹东,等.智慧环保体系在环境治理中的应用[J].环境工程技术学报,2021,11(5):992-1003.
LI Xinru, ZHOU Min, MI Yidong, et al. Application of smart environmental protection system in environmental management [J]. Journal of Environmental Engineering Technology, 2021, 11 (5): 992-1003.
- [5] 唐振浩,柴向颖,曹生现,等.考虑时延特征的燃煤锅炉 NO_x 排放深度学习建模[J].中国电机工程学报,2020,40(20):6633-6644.
TANG Zhenhao, CHAI Xiangying, CAO Shengxian, et al. Deep learning modeling for the NO_x emissions of coal-fired boiler considering time-delay characteristics [J]. Proceedings of the CSEE, 2020, 40(20): 6633-6644.
- [6] 刘科研,盛万兴,张东霞,等.智能配电网大数据应用需求和场景分析研究[J].中国电机工程学报,2015,35(2):287-293.
LIU Keyan, SHENG Wanxing, ZHANG Dongxia, et al. Big data application requirements and scenario analysis in smart

- distribution network [J]. Proceedings of the CSEE, 2015, 35(2): 287-293.
- [7] 杨清,周进.用电数据用于污染源企业工况监控技术的研究[J].污染防治技术,2019(3):44-48.
YANG Qing, ZHOU Jin. Research on application of power consumption data in condition monitoring [J]. Pollution Control Technology, 2019(3): 44-48.
- [8] 苗刚松,马雪梅.环保用电监控系统初探[J].现代信息科技,2020,4(7):156-157.
MIAO Gangsong, MA Xuemei. Discussion on environmental protection electricity monitoring and control system [J]. Modern Information Technology, 2020, 4(7): 156-157.
- [9] 周东国,张恒,周洪,等.基于状态特征聚类的非侵入式负荷事件检测方法[J].电工技术学报,2020,35(21):4565-4575.
ZHOU Dongguo, ZHANG Heng, ZHOU Hong, et al. Non-intrusive load event detection method based on state feature clustering [J]. Transactions of China Electrotechnical Society, 2020, 35(21): 4565-4575.
- [10] 雷怡琴,孙兆龙,叶志浩,等.电力系统负荷非侵入式监测方法研究[J].电工技术学报,2021,36(11):2288-2297.
LEI Yiqin, SUN Zhaolong, YE Zhihao, et al. Research on non-invasive load monitoring method in power system [J]. Transactions of China Electrotechnical Society, 2021, 36(11): 2288-2297.
- [11] PARK S W, BAKER L B, FRANZON P D. Appliance identification algorithm for a non-intrusive home energy monitor using cogent confabulation [J]. IEEE Transactions on Smart Grid, 2019, 10(1): 714-721.
- [12] DEVLIN M A, HAYES B P. Non-intrusive load monitoring and classification of activities of daily living using residential smart meter data [J]. IEEE Transactions on Consumer Electronics, 2019, 65(3): 339-348.
- [13] 陈小闽.非侵入式负荷分解技术及应用研究[D].广州:华南理工大学,2018.
CHEN Xiaomin. Research on non-intrusive load disaggregation technology and application [D]. Guangzhou: South China University of Technology, 2018.
- [14] 武昕,于金莹,彭林,等.基于用户边缘侧事件解析的工业电力负荷非侵入式感知辨识[J].电力系统自动化,2021,45(4):29-37.
WU Xin, YU Jinying, PENG Lin, et al. Non-intrusive perception and identification of industrial power load based on analysis of event on user edge [J]. Automation of Electric Power Systems, 2021, 45(4): 29-37.
- [15] 郭红霞,陆进威,杨莘,等.非侵入式负荷监测关键技术问题研究综述[J].电力自动化设备,2021,41(1):135-146.
GUO Hongxia, LU Jinwei, YANG Ping, et al. Review on key techniques of non-intrusive load monitoring [J]. Electric Power Automation Equipment, 2021, 41(1): 135-146.
- [16] 张逸,林森,吴丹岳.电能质量监测系统研究现状及发展趋势[J].电力系统保护与控制,2015,43(2):138-147.
ZHANG Yi, LIN Yan, WU Danyue. Current status and development trend of power quality monitoring system [J]. Power System Protection and Control, 2015, 43(2): 138-147.
- [17] 张逸,杨洪耕,叶茂清.基于分布式文件系统的海量电能质量监测数据管理方案[J].电力系统自动化,2014,38(2):102-108.
ZHANG Yi, YANG Honggeng, YE Maoqing. A data management scheme for massive power quality monitoring data based on distributed file system [J]. Automation of Electric Power Systems, 2014, 38(2): 102-108.
- [18] 胡文曦,肖先勇,汪颖,等.现代电网电力扰动数据分析与主动应用[J].电力系统自动化,2021,45(4):1-11.
HU Wenxi, XIAO Xianyong, WANG Ying, et al. Analysis and proactive application of power disturbance data in modern power grid [J]. Automation of Electric Power Systems, 2021, 45(4): 1-11.
- [19] 刘洁颖,刘俊勇,黄媛,等.基于面板数据有序聚类的主动配电网规划场景降维技术[J].电网技术,2017,41(4):1132-1138.
LIU Jieying, LIU Junyong, HUANG Yuan, et al. Scene dimensionality reduction technology in active power distribution network planning based on orderly clustering of panel data [J]. Power System Technology, 2017, 41(4): 1132-1138.
- [20] KILLICK R, FEARNHEAD P, ECKLEY I A. Optimal detection of change points with a linear computational cost [J]. Journal of the American Statistical Association, 2012, 107(500): 1590-1598.
- [21] 陆蔚.线性调频连续毫米波雷达的抗干扰技术研究[D].广州:广东工业大学,2020.
LU Wei. Research on interference suppression technology based on FMCW millimeter wave radar [D]. Guangzhou: Guangdong University of Technology, 2020.
- [22] 李永刚,王月,刘丰瑞,等.基于Stacking融合的短期风速预测组合模型[J].电网技术,2020,44(8):2875-2882.
LI Yonggang, WANG Yue, LIU Fengrui, et al. Combination model of short-term wind speed prediction based on Stacking fusion [J]. Power System Technology, 2020, 44(8): 2875-2882.
- [23] 魏书荣,张鑫,符杨,等.基于GRA-LSTM-Stacking模型的海上双馈风力发电机早期故障预警与诊断[J].中国电机工程学报,2021,41(7):2373-2383.
WEI Shurong, ZHANG Xin, FU Yang, et al. Early fault warning and diagnosis of offshore wind DFIG based on GRA-LSTM-Stacking model [J]. Proceedings of the CSEE, 2021, 41(7): 2373-2383.
- [24] 冯汛,石健,汪兴兴,等.自动化技术在环保设备中的应用与发展前景[J].工业安全与环保,2016,42(1):100-102.
FENG Xun, SHI Jian, WANG Xingxing, et al. Application and development prospects of automation technology in environmental protection equipment [J]. Industrial Safety and Environmental Protection, 2016, 42(1): 100-102.
- [25] 林祥勇,王秋霞.PLC与变频器在工业污水处理系统中的应用[J].淮海工学院学报(自然科学版),2015,24(3):15-19.
LIN Xiangyong, WANG Qiuxia. Application of PLC and inverter in industrial wastewater treatment system [J]. Journal of Huaihai Institute of Technology (Natural Science Edition), 2015, 24(3): 15-19.
- [26] 国家电网公司.电能质量监测终端技术规范:Q/GDW 650—2011[S].北京:中国电力出版社,2012.
State Grid Corporation of China. Technical specification for power quality monitoring device: Q/GDW 650—2011 [S]. Beijing: China Electric Power Press, 2012.

- Beijing: China Electric Power Press, 2012.
- [27] 张露, AUGER F, 荆朝霞, 等. 基于非侵入式的事件检测方法统计评估[J]. 电测与仪表, 2020, 57(1): 106-112.
ZHANG Lu, AUGER F, JING Zhaoxia, et al. Statistical assessment of abrupt change detections for NILM[J]. Electrical Measurement & Instrumentation, 2020, 57(1): 106-112.
- [28] 庄丹. 基于形态识别的多变点检测方法研究[D]. 成都: 西南财经大学, 2019.
ZHUANG Dan. Research on multiple change-points detection method based on shape recognition [D]. Chengdu: Southwestern University of Finance and Economics, 2019.
- [29] 刘子腾. 考虑背景谐波电压波动的配电网谐波责任评估[D]. 北京: 华北电力大学, 2021.
LIU Ziteng. Harmonic contribution determination of distribution network considering background harmonic voltage fluctuation [D]. Beijing: North China Electric Power University, 2021.
- [30] 龚华麟, 肖先勇, 刘亚梅, 等. 基于主导波分量筛选原理的用户谐波发射水平估计方法[J]. 中国电机工程学报, 2010, 30(4): 22-27.
GONG Hualin, XIAO Xianyong, LIU Yamei, et al. A method for assessing customer harmonic emission level based on the dominant fluctuation filtering principle [J]. Proceedings of the CSEE, 2010, 30(4): 22-27.
- [31] 刘嘉蔚, 李奇, 陈维荣, 等. 基于在线序列超限学习机和主成分分析的蒸汽冷却型燃料电池系统快速故障诊断方法[J]. 电工技术学报, 2019, 34(18): 3949-3960.
LIU Jiawei, LI Qi, CHEN Weirong, et al. Fast fault diagnosis method of evaporatively cooled fuel cell system based on online sequential extreme learning machine and principal component analysis [J]. Transactions of China Electrotechnical Society, 2019, 34(18): 3949-3960.
- 张逸(1984—), 男, 通信作者, 博士, 副教授, 主要研究方向: 电能质量、主动配电网及电力数据分析。E-mail: zhangyi@fzu.edu.cn
- 姚文旭(1998—), 男, 硕士研究生, 主要研究方向: 电能质量、电力数据分析。E-mail: wenyuyao@163.com
- 邵振国(1970—), 男, 教授, 博士生导师, 主要研究方向: 电网不确定性理论与方法、电能质量分析与治理、电力大数据理论与方法等。E-mail: shao.zg@fzu.edu.cn

(编辑 蔡静雯)

Identification of Abnormal Operation Condition for Environmental Protection in Enterprises Based on Power Quality Monitoring Data

ZHANG Yi, YAO Wenxu, SHAO Zhenguo, ZHANG Liangyu

(College of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China)

Abstract: Aiming at the problems of abnormal operation condition monitoring for environmental protection in polluting enterprises at present, such as difficult implementation, large identification errors and easy tampering with the results, this paper proposes an identification method of abnormal operation conditions for environmental protection based on power quality monitoring data. The multi-dimensional power quality data obtained from non-intrusive load monitoring at the public power entrance of enterprise equipment are used to train the model of condition classification, to realize abnormal condition identification, which is different from the existing scheme of power consumption monitoring with separate meters installed for each device. First, the time series change-point detection and the clustering calculation for the characteristic data representing the production conditions are carried out to realize the division of production operation conditions of enterprises. Then, combined with the operation of environmental protection equipment, the categories of environmental protection operation conditions for classification are obtained. Furthermore, the operation condition scenarios related to environmental protection are classified and learned by the Stacking learning model. Finally, the trained classification model is used to identify the abnormal operation conditions for environmental protection in the enterprise. The effectiveness of the proposed method is verified by the simulation test data and the actual enterprise data.

This work is supported by National Natural Science Foundation of China (No. 51777035).

Key words: environmental protection operation condition; power quality; monitoring data; non-intrusive load monitoring; abnormal operation condition identification; environmental monitoring; change point detection; operation condition classification; Stacking model

