

基于居民用户用电行为分析的 室内大麻种植检测

毛源军^{1,2}, 陈超强³, 舒一飞⁴, 刘爱国⁴, 苏盛¹

(1. 长沙理工大学电气与信息工程学院, 湖南长沙 410114; 2. 国网江西省电力有限公司宜春供电分公司, 江西宜春 336000; 3. 国网湖南省电力公司邵阳供电分公司, 湖南邵阳 422000; 4. 国网宁夏电力有限公司营销服务中心(计量中心), 宁夏银川 750001)

摘要:在欧美大麻合法化风潮影响下,国内开始批量出现室内种植大麻的违法现象。在此背景下,利用室内种植大麻需大量消耗电能且用电行为具有规律性的特点,提出一种基于用电功率频域分布相对熵的室内大麻种植检测方法。首先,分析室内种植大麻的用电需求规律特征,并搭建实验室仿真环境进行种植大麻用电数据的模拟产生;然后,比对居民用户和室内大麻种植用电行为在时域、频域及具体指标项上的差异,并制定根据用电功率频域分布相对熵等指标识别非法大麻种植的检测流程。基于爱尔兰居民用电数据的测试分析表明,所提方法可有效识别和排除正常居民用户,提高大麻种植异常检测的靶向性。

关键词:大麻种植;用电特性;相对熵;定量分析;靶向性

DOI:10.19781/j.issn.1673-9140.2022.04.023 中图分类号:TM863 文章编号:1673-9140(2022)04-0198-11

Detection of illegal cannabis cultivation based on residential electricity usage behavior analysis

MAO Yuanjun^{1,2}, CHEN Chaoqiang³, SHU Yifei⁴, LIU Aiguo⁴, SU Sheng¹

(1. School of Electrical & Information Engineering, Changsha University of Science & Technology, Changsha 410114, China; 2. Yichun Power Supply Branch, State Grid Jiangxi Electric Power Co., Ltd., Yichun 336000, China; 3. Shaoyang Power Supply Branch, State Grid Hunan Electric Power Co., Ltd., Shaoyang 422000, China; 4. Marketing Service Center (Metrology Center), State Grid Ningxia Electric Power Co., Ltd., Yinchuan 750001, China)

Abstract: Influenced by the cannabis legalization of USA and EU, the illegal indoor cannabis cultivation got popular in China in recent years. Since indoor cultivation of cannabis has a large number of electricity consumption with distinct feature of electricity usage, A detecting method based on the relative entropy of the frequency distribution of electricity usage is developed to identify abnormal users in this paper. Firstly, the electricity usage data of indoor cultivate cannabis is generated from the laboratory simulation. Then, the electricity usage feature difference between the indoor cultivation of cannabis users and the normal residential users are compared in terms of the temporal, frequency do-

收稿日期:2021-06-02;修回日期:2021-09-11

基金项目:国家自然科学基金(51777015)

通信作者:苏盛(1975-),男,博士,教授,博士生导师,主要从事电力系统网络安全防护与大数据技术应用研究;E-mail:eessheng@163.com

main and specific indicators. Thereafter, the identification process of abnormal users of indoor cultivation of cannabis is proposed. In the end, a numerical analysis of electricity usage data from Irish residents suggests that the proposed approach can separate residential users and anomaly users. It is helpful to identify the illegal user with the indoor cultivating cannabis.

Key words: cannabis cultivation; electricity consumption characteristics; relative entropy; quantitative analysis; targeting

大数据时代下,蕴含丰富信息的电力大数据已逐渐发展成为供电企业重要的数据资产。挖掘分析电力大数据,可以为供电企业降本增效、开展新型业务,提供重要支撑^[1]。电力科技工作者从大数据视角出发,围绕电力设备的状态监测和电力用户的用电行为分析开展了大量研究。由于用电行为可有效体现实体企业运营状况,近年来也衍生出了一些基于电力大数据进行复工复产监测、重污染企业环保监察以及金融风险防控等社会化服务^[2]。将电力大数据与不同层面的社会需求结合,还可以衍生出大量创新应用模式。

受欧美国家大麻合法化风潮影响,近年来中国室内种植大麻的事件频发,严重危害社会治安和人民身心健康。据不完全统计,中国沿海和中部地区 10 余个省份均有抓获室内种植大麻的案例报导,有的甚至形成了种植、运输、贩卖的产业链,危害巨大。为避人耳目,室内种植大麻多见于郊区及城中村的出租屋、废旧工厂、郊区别墅房以及毛坯房等空置民房中,环境隐蔽、发现困难。目前,稽查毒品犯罪主要通过高危人群追踪分析、快递检测和群众举报等方式获得犯罪线索,然后逐一排查。因为室内种植大麻需要消耗大量电能,用于补光和调控温、湿度环境,用电行为上具有比较明显的特异性,利用此特异性比较正常居民用户用电行为,可以识别室内种植大麻的异常用户,拓展电力大数据的社会服务范畴。

本文利用室内种植大麻在用电行为上的特性,提出基于用电数据的大麻种植用户检测方法。首先,分析居民用户用电行为在时域和频域上的行为特征;然后,比较分析室内种植大麻的用电规律及其与居民用电行为的差异,进而提出基于用电功率频域分布相对熵的空置房内种植大麻的检测方法;最后,在实验室产生室内种植大麻的用电数据和爱尔兰居民用电数据的基础上,测试验证所提方法的有效性。

1 用电行为分析

1.1 居民用户用电时序分析

居民用户是电力服务的重要对象,作为需求侧资源参与电网互动的潜力同样巨大。因此,通过结合居民侧智能电表积累的海量精细负荷数据和家庭特征信息,分析居民用电行为模式,对于提升电力服务水平和用户满意度、充分唤醒居民用户需求响应资源具有重要意义,除此之外,针对居民用户的异常用电行为分析研究也是当前的热点,居民用户用电主要为家用电器。家用电器可分为常开电器和非常开电器,其中长期处于运行状态的冰箱以及持续待机的电视机顶盒、家庭网络终端等为常开电器,而其他的空调、电视等多为非常开电器。不同类型居民用户根据家电启用时段和使用时长的组合,可分为多种用电模式,但与室内大麻种植用电模式差异较大。本文针对居民用户用电行为分析进行室内大麻种植检测,在居民用户用电与室内大麻种植用电特征差异大的基础上,提出一种基于相对熵的室内大麻种植检测方法。

文献[3]针对居民用户电力日负荷曲线数据,提出一种基于加权表决的集成聚类方法。将 4 种常用聚类算法视作选民成员进行投票表决,并根据聚类有效性指标赋权从而集成成员算法的聚类结果,以结合不同算法的性能优势。提取负荷曲线特性指标对居民负荷曲线加权表决聚类得到 6 种典型用电模式;文献[4]基于云计算平台和并行 k -means 聚类算法,建立了峰时耗电量、负荷率、谷电系数等时间序列特征,并采用熵权法计算各类特征权重,将智能小区的居民用户分成空置房、上班族、上班族+老人、老人家庭、商业 5 类用户。

由文献[3-4]中的实验结果分析可知,文献[3]中的 6 类用户中有 4 类用户是两两类似的,相当于分为 4 类。本文在文献[4]所提方法所提方法的基

基础上对爱尔兰居民用电数据集的居民用户进行聚类,文献中将居民用户聚成了5类,其中一类为商业用户。本文研究目的在于居民用户中的异常行为检测,因此剔除了爱尔兰数据集中的商业用户。本文将居民用户分为上班族、老人、其他类型(老人+上班族家庭)以及空置房等不同类型,从该4类用户中选出典型的用户并将其周负荷曲线绘制如图1所示。

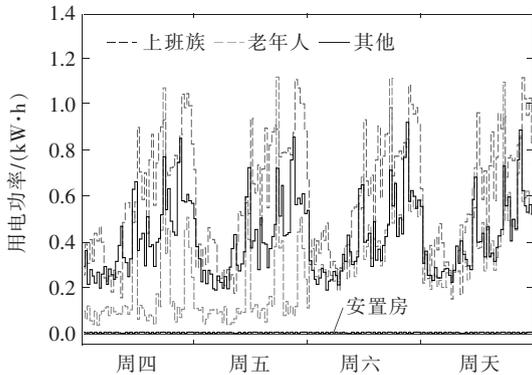


图1 居民用户周用电量曲线

Figure 1 Weekly load curve of residential users

由图1可见:

1)上班族用户用电量有明显的高峰和低谷,昼间除早餐时段和午餐时段有负荷,基本维持在和夜间相近的极低负荷水平;夜间下班后将迎来一天的负荷高峰;周末休息,全天负荷水平明显高于工作日;

2)老人家庭一般全天候家中有人,昼间一直有波动性负荷;由于在生活上没有明显的周末和工作日的差异,周末和工作日的用电曲线差异不明显;

3)其他居民用户与前述用户略有差异,但也同样具有夜间负荷小,昼间负荷波动较大及周末负荷较高等类似规律。

4)空置房因为无人居住,无负荷或仅维持少量常开负荷,在日周期和周周期上的用电量一直维持在极低水平,几乎为零。

1.2 居民用户用电频域分析

不同类型居民用户的负荷曲线在时域上呈现出较大差异性,难以统一描述其共性规律。将时域数据转换到频域,是进行数据分析的常用方法。按用电功率的出现频率分析居民用电数据,可能呈现出其他视角的规律性特征^[5]。将图1中典型居民用户1个月的用电数据按各自单位时间用电量的最大值

为基准进行标准化,对不同频域分布区间数量(6、8、10、12、14、16、18和20)进行计算分析,结果表明,频域区间数量大于8时区间数量对异常筛选结果无影响,所以将用电量等分为10个区间,典型用户可绘制单位时间用电量在不同区间出现频率如图2所示。由于空置房仅有常开电器用电,用电功率集中在较低数值的功率区间,图2中仅绘制其他3个典型用户的用电量频率曲线。

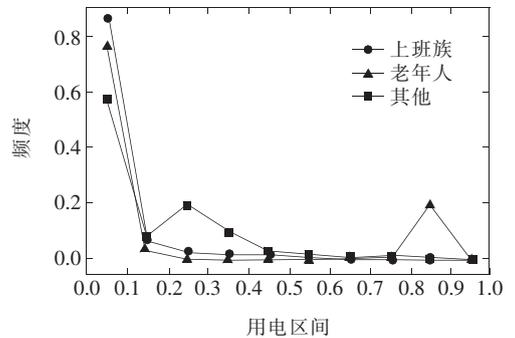


图2 典型居民用户统计分布

Figure 2 Frequency distribution of typical residential users

1)居民用户单位时间用电量取决于投用家电的组合。绝大多数居民用户在夜间仅保留常开负荷,用电负荷会降低到全天的最低水平并持续较长时间,因此,图2中不同类型用户在低功率区段的出现频次最高。

2)其他时间段用电功率伴随非常开电器投用会有所增加。由于每种非常开电器投用时间不一,随着功率区间的上升,出现的频率会逐次降低,呈现出在低功率区间频率最高的单峰状态。当特定非常开电器持续投用时,也可能造成在较高功率区间的频率升高。

3)对于长期有人在家的退休家庭而言,如持续开启电视,可能造成较高功率区间的频率上升;而寒冷或炎热季节需持续开启的温度调节负荷,也可能造成重负荷条件下的频率上升,从而形成另一个频段高峰。

综上,除空置房以外的居民用户用电在频域上具有明显的低用电区间的单峰或整体上的双峰特征^[6-9]。

1.3 居民用户日尖峰电量特征

居民用户有多种家用电器设备,由这些家用电器组合决定的日尖峰电量可能存在比较明显的波动

性。同样以前文 3 个典型用户为例,绘制其 1 个月中每天单位时间尖峰电量曲线如图 3 所示,可见居民用户日尖峰电量具有较明显的波动性。与之相比,大麻种植采用明确而固定的用电设备按设定的方式工作,每天单位时间内的最大用电量可能只有轻微的波动,利用这种用电数据上的差异,也有可能识别异常的用电行为。

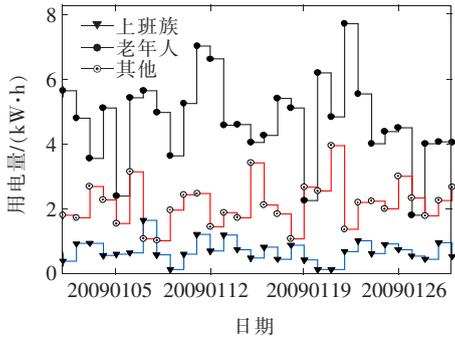


图 3 典型居民用户日尖峰电量

Figure 3 Daily peak electricity of typical residential users

2 室内种植大麻的用电行为特性

2.1 大麻种植耗能需求分析

大麻为一年生草本植物,叶、苞片和花朵中含四氢大麻酚,吸入后易导致上瘾,长期吸食会摧残人的身体和意志,是中国明令禁止种植的毒品原植物。大麻是一种生长于温带和热带地区的植物,室内种植时需要用大功率聚光灯照射以促进植物光合作用,生长期与成熟期还要用不同光色(温)促进植株生长。房间需安装调温调湿器进行保温保湿。由于大麻植株会散发出浓郁难闻的气味,有时还需要安装排气管道、烘干机等设备祛除气味。室内种植大麻全过程需要消耗较高强度的电能,相关研究认为每平米种植面积年耗电可达 $1\ 615\ \text{kW}\cdot\text{h}$,其中,通风、冷却和除湿耗能约占 51%,照明占 38%,室内加热占 5%,其他占 6%^[10]。

大麻种植可分为幼苗期、营养期和开花期 3 个阶段,各阶段持续时间及所需湿度和光照时间如表 1 所示。根据温、湿度条件控制的要求,室内种植大麻涉及的主要是空调、排风扇和替代阳光照射的植物补光灯等电器设备。由于涉及的电器设备类型和数量有限,不同设备启停状态的组合以及在单位时间内对应的用电量也仅有有限的几种。因此,种植

大麻的用电曲线在时域上可能呈现为阶梯波形式,对应功率区间的出现概率将集中在少数几个区段,在时域和频域上与一般居民用户的用电行为存在显著差异。

表 1 不同阶段大麻生长所需环境

Table 1 Environment required for different stages of cannabis growth

阶段	相对湿度/%	日照明时长/h	持续天数/d
幼苗期	70~80	18~24	3~10
营养期	60~70	18~24	28~56
开花期	40~60	12	42~70

2.2 大麻种植用电数据模拟实验

根据大麻种植手册,在实验室搭建 $1.2\ \text{m}\times 1.2\ \text{m}$ 大麻种植大棚,并配置所需的空调、风扇和植物补光灯等电器和温、湿度传感器。是否启用空调对大麻种植用电有突出影响,因此,在春季和夏季分别进行 2 组模拟大麻种植的用电试验,其中,春季室温约在 $20\ ^\circ\text{C}$ 左右无需开启空调,夏季气温持续在 $32\ ^\circ\text{C}$ 时,需持续启动空调进行温度、湿度调节。每次试验持续 30 d,对幼苗期、营养期和开花期各按 10 d 进行试验模拟。其中植物补光灯按表 1 所述要求,在幼苗期和营养期每天开启 20 h,开花期开启 12 h;夏季试验中空调按 $25\ ^\circ\text{C}$ 持续开启;各期室温高于 $25\ ^\circ\text{C}$ 时开启风扇。由于实验所在地区湿度未达到种植手册标注的湿度上限,故未按湿度控制电器启停。

参与实验的电器采用具有电量计量和数据远传功能的智能云插座进行电量计量。以 1 min 间隔记录实时电量数据,将数据汇总后可形成 30 min 间隔用电量数据。实验在 $1.2\ \text{m}\times 1.2\ \text{m}$ 大棚中配置 1 盏 $1\ 000\ \text{W}$ 的植物补光灯和 1 台 $50\ \text{W}$ 的风扇,实验环境配置空调为 $2\ 400\ \text{W}$ 。

2.3 大麻种植用电时域特性分析

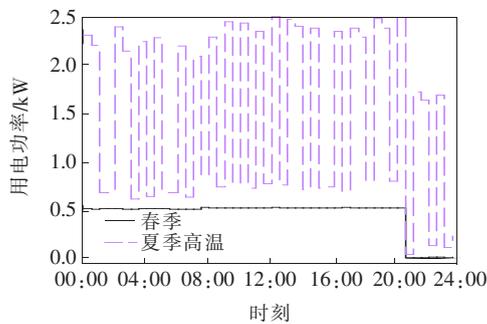
大麻种植各阶段 1 d 的用电负荷曲线如图 4 所示。

1) 春季气温正常,仅在昼间气温上升后需启动风扇,此时负荷为植物补光灯与风扇构成的功率恒定负荷;夜间低温时段为植物补光灯恒定负荷;夜间关灯时段为零负荷。全天负荷大致呈现为三级的阶梯波。

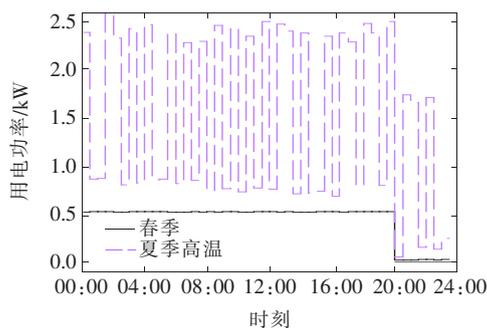
2) 夏季高温条件下需全天持续启动风扇和空调降温,而植物补光灯仅在昼间开启。因空调按目标气温进行启停控制,会造成用电负荷曲线整体形态上的阶梯波性质;夜间关闭补光灯后负荷阶梯波将出现明显下降。

3) 幼苗期、营养期和开花期主要影响补光灯的启停时间,不会改变负荷曲线的整体形态。

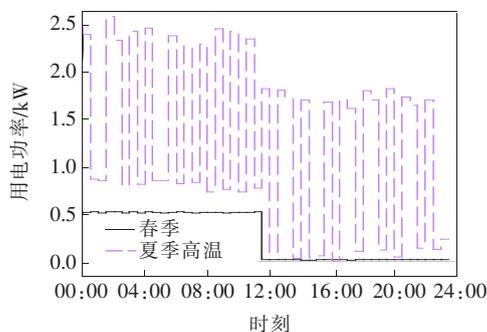
4) 用电数据按 30 min 为间隔进行累加计量。理想情况下如在 30 min 间隔中,空调始终处于启动或停机状态,用电负荷将表现为幅值基本恒定的阶梯波。由于空调并不一定在整个 30 min 间隔中都处于启动或停机状态,故阶梯波的幅值会有一定波动,但从实际数据来看,基本上稳定在一定区间内。



(a) 幼苗期



(b) 开花期



(c) 成熟期

图 4 不同外界环境下大麻种植户单日用电模式

Figure 4 Daily load curve under different seasons

除此以外,由于用电设备构成明确、固定并按设定条件启停,因此,大麻种植的日尖峰电量保持稳定。

2.4 大麻种植用电频域特性分析

由大麻种植用电数据模拟实验及其大麻种植用电时域特性分析可见,大麻种植用电时间序列具有突出的阶梯波特征,单位时间的用电量密集地分布在少数几个区间上,如将其转换到频域上,则有可能更好地凸显其与一般居民用户用电数据的差异特性。为方便不同规模用户间的对比,需对大麻种植用电数据标幺化后再进行对比分析。用户用电负荷时间序列时序 $x = \{x_1, x_2, \dots, x_n\}$, 取其单位时间内用电量最大值 X_{\max} 和最小值 X_{\min} , 采用最大—最小标准化方法对每半小时电量数据进行标准化^[11]:

$$x'_i = \frac{x_i - X_{\min}}{X_{\max} - X_{\min}} \quad (1)$$

式中 x_i 为 30 min 的用电量; x'_i 为标准化后 30 min 的用电量。

根据以上步骤,在春、夏季高温时段各 30 d 的大麻种植实验数据标准化后,将功率区间分为 10 份,计算每个功率区间的负荷持续时间,并绘制直方图,如图 5 所示。

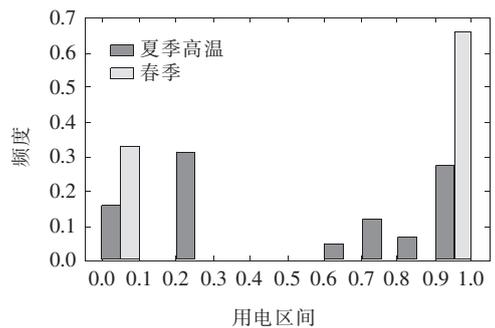


图 5 大麻种植用电功率区间分布直方图

Figure 5 Statistical distribution of cannabis growers in different external environments

由图 5 可见:

1) 春季只有植物补光灯和风扇开关切换对应的 4 种启停组合;由于风扇功率较小,开启补光灯、风扇开关与否,对应的用电功率都落在频率直方图的不同区间,因此,用电功率主要根据补光灯是否开启而落在不同区间,对应在频率直方图中落在 2 个区间;

2) 夏季高温时空调、风扇和补光灯 3 种设备共

有 8 种不同的电器启停组合;受 30 min 内空调启停持续时间不同的影响,且风扇功率相较于空调和补光灯可以忽略不计,因此,对应用电功率落在频率直方图的 6 个区间;

3)春季用电设备全停时间较长,对应零功率区段出现频率较高;其他时段多为风扇与补光灯同时工作,满功率区段出现频率最高;

4)在夏季高温时段,植物生长灯关闭后空调任然有间断性启动,零功率区段出现频率明显低于春季;其他时段受不同电器组合影响,与春季相比,在其他功率区段出现的频率相对平整;

5)对比图 3、5,可发现居民用户用电在功率区间分布上具有明确的低功率区间频率高、高功率区间频率低的特点,整体上呈现单峰特性;在个别区段,由于部分用户可能频繁使用特定高功率电器设备,使得高功率区间出现频率的第 2 个凸起,具有双峰特性^[12];与之相比,大麻种植使用的电器数量少,运行特性简单,在频率直方图中的表现简单且用电量落在高耗能区的频率大,与正常居民用户的频率直方图差异明显。

3 大麻种植用电异常检测方法

由居民用电行为及其室内种植大麻的用电行为特性分析表明,大麻种植与正常居民用户在时序上有明显差异,但这种差异很难用指标量化评价。将用电数据转换到频域后,能更明确地凸显两者间的差异性特征,有可能通过频域分析来检测大麻种植导致的用电异常行为。

3.1 基于相对熵的大麻种植用电异常检测

受用电设备组合投用规律的影响,转换到频域上的大麻种植用电数据具有集中在少数区间的特征,与居民用电存在显著差异。利用这一特点,有可能将居民用户用电数据与空置房屋内种植大麻的用电数据都转换到频域后进行对比,并将在与种植大麻用电数据频域特征相似的居民用户判断为异常。

相对熵(relative entropy)是一种衡量 2 个时间序列概率分布差异性的非对称性指标,可以定量评价 2 个时间序列在频域上的差异。相对熵在故障检测和异常检测等领域得到了广泛应用。本文利用相

对熵来评价待测居民与空置房大麻种植用电行为在频域中的差异,识别出在用电数据的频域分布上具有空置房大麻种植特性的异常用户。

对于随机变量的 2 个概率分布 P 和 Q ,若 X 是离散随机变量,则分布函数 $P(x)$ 对 $Q(x)$ 的相对熵定义为

$$D(P \parallel Q) = \sum_{i=1}^X P(x_i) \log \frac{P(x_i)}{Q(x_i)} \quad (3)$$

由式(2)可知,相对熵具有不对称性和非负性,即 $D(P \parallel Q) \neq D(Q \parallel P)$,且 $D(P \parallel Q) \geq 0$ ^[13-16]。为了使相对熵公式计算有意义,需保证 $P(x) \neq 0$ 、 $Q(x) \neq 0$ 。因为将用电数据时间序列转换到频域后,可能用电数据样本集中在少数几个区间而部分功率区间没有样本,这时将会出现概率为 0 的情形。为避免该类异常影响,参照文献[13]中方式,规定在转化到频域的用电数据离散随机变量中,将用电数据在每个区间的频域样本数量加 1,从而消除 $P(x) = 0$ 或 $Q(x) = 0$ 的影响。

因为居民用户用电行为的频域特性相近且与大麻种植存在显著差异,两者的相对熵将可能形成 2 个类簇,其中,绝大多数正常居民用户的相对熵数值较大,形成一簇;而用电频域特性与大麻种植相近的用户的相对熵数值较小,作为离群点形成另一簇。

利用相对熵,可以将与种植大麻用电数据频域特征相似的居民用户识别为异常,恰当的阈值设置是准确检出异常用户的重要基础。本文根据箱线图选取阈值。箱线图是一种显示数据分布情况的方法,具体结构如图 6 所示。

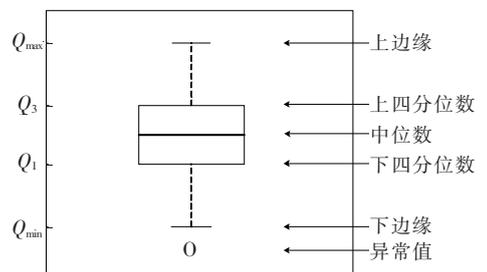


图 6 特征指标箱线

Figure 6 Characteristic indicator box plot

将所有样本的数值从小到大排列后分成四等分,其中处于 3 个分割点位置的数值称为四分位数,按数值从小到大的顺序依次是下、中、上四分位数

Q_1 、 Q_2 、 Q_3 ，上、下四分位数的差值称为四分位数差 (interquartile range, IQR)。基于异常样本远离正常样本群体的假设，可利用箱线图来检测异常值^[15-16]。一般将大于上四分位数 1.5 倍四分位数差的值以及小于下四分位数 1.5 倍四分位数差的样本视为异常样本。本文中相对熵小的为用电行为类似大麻种植用户，可选取相对熵小于下四分位数 1.5 倍四分位数差的样本视为异常样本。由于四分位数受异常值影响较小，因此，箱线图法具有较高的鲁棒性^[17-21]。

3.2 大麻种植用电异常检测的辅助判据

供电企业服务数量庞大的居民用户，仅以用电数据频域相对熵判断异常，即便很少比例的误判都会造成大量用户被视为嫌疑对象。为满足实际应用需求，需要设置辅助判据以尽可能消除误判。

大麻种植用电在频域上分布较为均匀，没有明显的频率单峰或双峰特性。居民用户在特定场景下可能出现类似的频域特征并被识别为异常。以空置房用户为例，当家中仅有 1、2 件常开电器投用时，对应负荷的频域特征有可能和规律性投用少量电气设备进行大麻种植形成的用电负荷频域特征具有相似性。为筛除空置房用户造成的误判，设置规则将日均电量低于 $3 \text{ kW} \cdot \text{h}$ 的用户剔除，不进行异常检测。

当居民用户像大麻种植一样长时间启用大功率电器时，也可能造成混淆误判，还需要进一步选取辅助判据降低误判概率。对居民用户而言，尽管用电在时域曲线形态上具有规律特性^[22]，但每天的用电行为还是有比较明显波动性的，而大麻种植用电高度规律，每天的最大负荷基本维持不变。因此，有可能利用用户每天单位时间最大电量是否具有波动性来判断是否为大麻种植用户。

采用变异系数评价用户单位时间最大电量的波动性。变异系数 C_v 为标准差与平均数的比值，是衡量资料中各观测值变异程度的统计量，定义：

$$C_v = \frac{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}}{\bar{x}} \quad (4)$$

其中， x_i 、 n 、 \bar{x} 分别为时间序列 X 的值、数据个数、平均值。本文采用变异系数量化评价用户日单位时间最大用电量的波动。根据春、夏季高温时的

大麻种植模拟用电实验数据，将变异系数阈值设置为 0.1。

3.3 大麻种植用电异常检测流程

结合分析所提异常检测判据，可制定大麻种植用电异常检测流程，如图 7 所示，具体步骤如下。

1) 数据预处理。获取居民用户用电数据，剔除日均用电量小于 $3 \text{ kW} \cdot \text{h}$ 的用户，归一化处理用电数据后把时序数据转换为在各功率区间出现概率的频域数据。

2) 相对熵计算。计算被测用户用电数据频域分布与实验产生大麻种植用电数据频域分布之间的相对熵。

3) 阈值求取。根据箱线图求取用户用电数据相对熵的阈值。

4) 异常用户检测。获取居民用户中相对熵小于阈值的用户用电时序数据，计算异常用户日尖峰电量的变异系数，将变异系数小于 0.1 的判断为异常用户。

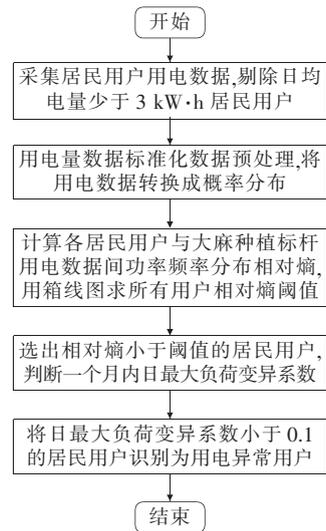


图 7 室内大麻种植检测流程

Figure 7 Flowchart to detect the indoor cannabis growers

4 算例测试分析

爱尔兰智能用电实验记录有 2009—2010 年 (535 d) 间按 30 min 间隔存储的用电时序数据^[23]，本文采用其中 4 225 户居民用户用电数据作为测试数据集。

4.1 居民用户相对熵分析示例

正常居民用户尽管类型不同，在用电数据时间

序列形态上存在差异,但在频域分布上较为相似,相互间的相对熵较小。为弄清正常居民用户间及其与大麻种植用电数据间在相对熵上的差异,选取文 1.1 中 3 类非空置房的典型居民用户各 10 户,结合 2009 年 1 月的用电数据进行示例分析。

为方便呈现不同类型正常居民用户间及其与大麻种植用电数据相对熵的差异,计算 30 个用户及春、夏季高温时种植大麻用电数据间的相对熵,并绘制热力图,如图 8 所示,每个小方块代表某用户与其他用户的相对熵,如第 i 行第 j 列色块表示第 j 个用户相对第 i 个用户的相对熵。热力图共包含 32×32 个小方块,其中第 1~10 个为上班族类型用户,第 11~20 个为老人家庭用户,第 21~30 个为其他用户,第 31、32 个分别代表夏、春季高温时的大麻种植用电数据。

热力图中以色阶表示相对熵数值,条状中颜色在上面的表示相对熵越高,即热力图中颜色越浅则相对熵越低,越深则相对熵越高。

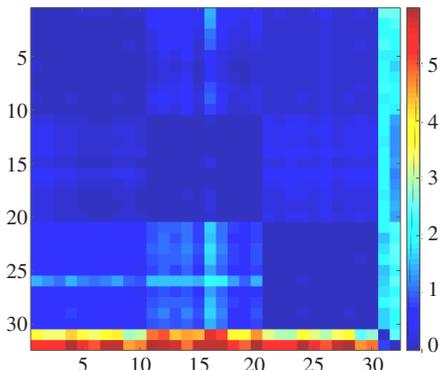


图 8 典型居民用户相对熵热力图

Figure 8 Heat figure of relative entropy of typical resident user

由图 8 可见:

- 1) 对角线上的方块表示用户与自身的相对熵,因此,数值全为零且对应条状中为零的颜色;
- 2) 同类典型用户之间的相对熵普遍较小,对应到热力图中表现为对角线上 10×10 的大色块颜色明显更深;
- 3) 同类型 10 个用户与其他类型 10 个用户的相对熵数值较大,对应到热力图中表现为相互间形成 10×10 的大色块,而且大色块之间有比较明显的差异色边界;
- 4) 热力图中第 i 行第 j 列为第 j 个用户对第 i

个用户的相对熵,因为相对熵不具有对称性,对应的热力图色块并不具有对称性;热力图底部的 2 行是正常居民用户对大麻种植用电数据的相对熵,数值较大,而热力图最右边的 2 列为大麻种植用电对居民用户的相对熵,数值较小;在对应颜色上也有所差异;

5) 居民用户对春、夏季高温时大麻种植数据的相对熵存在明显差异,对春季时的相对熵更大而对夏季时的相对熵更小。

4.2 异常用户识别测试

选取数据集中 4 225 个居民用户 2009 年 1 月的数据,按图 7 所示流程进行运算分析。剔除 6 户日均负荷小于 $3 \text{ kW} \cdot \text{h}$ 并可能为空置房的用户后,计算剩余 4 119 户居民对春、夏季高温时大麻种植用电数据的相对熵,并绘制箱线图,如图 9 所示,可知夏季高温时段大部分用户的相对熵落在区间 $[2, 3]$ 中,箱线图的 Q_1 、 Q_3 分别为 2.01 和 2.69,对应的相对熵下界阈值为 0.99,低于该阈值的用户有 2 户,分别为 3244 和 3683 号;春季时大部分用户的相对熵落在区间 $[3, 4]$ 中,箱线图的 Q_1 、 Q_3 分别为 3.21、3.95,对应的相对熵下界阈值为 2.1,低于该阈值的用户有 7 户,分别为 2497、2601、2603、2856、3177、4871、6399 号用户。

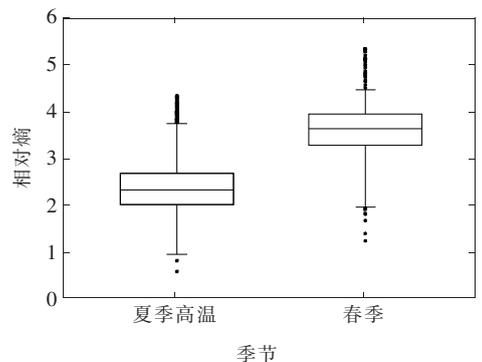


图 9 不同外界环境居民用户与大麻种植户相对熵阈值

Figure 9 Relative entropy thresholds of residents and cannabis growers in different external environments

由图 9 可知,夏季高温时低于下界阈值的 2 户在 1 月份的日最大单位时间用电曲线如图 10(a)所示(仅绘制 3 d 用电数据,以便观察),可见前者日尖峰电量波动较大,1 月份日最大单位时间用电负荷的变异系数达 0.325,后者也达 0.108,明显超过 0.1 的阈值,可排除该 2 户的嫌疑;春季时低于下界

阈值的7户在1月份的时序用电曲线如图10(b)所示,其日最大单位时间用电负荷的变异系数分别为0.03、0.33、0.33、0.14、0.26、0.08、0.25,除2497号用户外均超过设置的阈值,可排除大部分用户嫌疑。

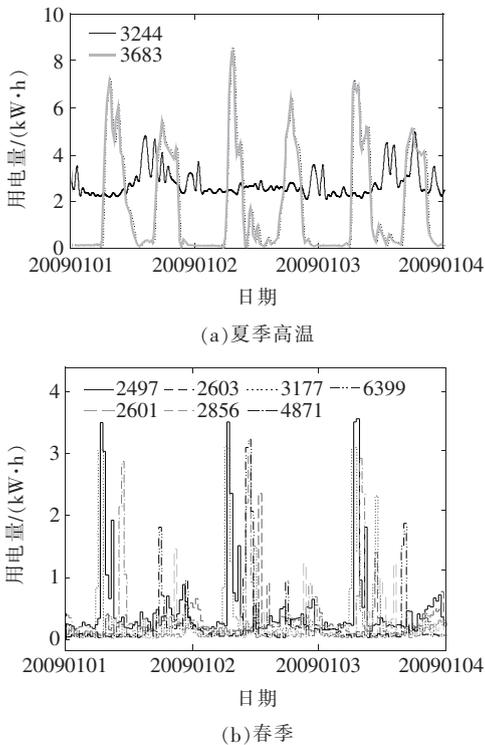


图10 相对熵筛选被测为异常用户的时序用电曲线

Figure 10 Time series electricity consumption curve of users tested as abnormal

根据图7所示流程可最大程度排除正常居民用户,显著提高毒品稽查工作的靶向性。从测试数据来看,还是有一部分正常居民数据,如2497号用户(变异系数仅0.03)会被误判。需要指出的是,根据与公安部门的交流,与电力行业窃电稽查相比,公安部门毒品稽查可承受更高的误报率,尽管仍存在少量误报,但实际工作中是可以接受的。

除居民用户外,爱尔兰数据集中还包含有中小型企业与工商业用户485户,其他类别1735户。根据流程(图7)所示方法,通过箱线图设定相对熵阈值可判别:

1)在中小型企业与工商业用户中,与大麻种植用户春季用电数据比较时,低于阈值被判别为异常的用户有30户,其中7户的日最大负荷变异系数低于阈值0.1;与大麻种植用户夏季高温用电数据比较时,低于阈值的用户有7户,其中3户的日最大负荷变异系数低于变异系数阈值0.1,占该类用户的

0.62%;

2)在其他类别用户中,与大麻种植用户春季用电数据比较时,低于阈值被判别为异常的用户有31户,其中8户的日最大负荷变异系数低于阈值0.1;与大麻种植用户夏季高温用电数据比较时,低于阈值的用户有7户,其中1户的日最大负荷变异系数低于变异系数阈值0.1,占该类用户的0.06%;

3)中小型企业 and 商业用户以及其他类别用户的工作经营模式可能与大麻种植用户的类似,如用户1333,其工作模式与春季工作时的的大麻种植户接近,因而产生误判。

需要指出的是,种植大麻会散发出浓郁难闻的气味,公安人员本身又具有毒品稽查的执法权,在稽查难度和便利性上都明显优于窃电稽查,少量的误判在实际工作中是可以接受的。

5 结语

针对室内非法种植大麻发现困难的问题,利用空置房中种植大麻需消耗大量电能且用电行为具有明确规律的特点,提出了基于用电负荷频率相对熵的异常检测方法,根据居民用户用电数据与空置房中大麻种植用电数据概率分布的相对熵识别用电行为类似大麻种植的异常用户。基于实验产生的大麻种植用电数据和4000多户实际居民用电数据进行的测试表明,所提方法可以筛除正常居民用户,提高异常用户检测的靶向性和工作效率。

在有限的用户范围和少量测试数据基础上,本文初步验证了所提方法的有效性。目前正由供电企业和公安部门开展合作研究,进行实际验证。未来将结合抓捕到的实际大麻种植用户用电数据,进一步探索非空置房中种植大麻的异常用电行为检测。

参考文献:

- [1] 乔黎伟,王静怡,郭炜,等. 基于随机森林算法的中短期用电量预测[J]. 电力科学与技术学报, 2020, 35(2): 150-156.
QIAO Liwei, WANG Jingyi, GUO Wei, et al. Medium and short-term electricity demand prediction based on random forests algorithm[J]. Journal of Electric Power Science and Technology, 2020, 35(2): 150-156.

- [2] 张然,孙晓璐,何仲潇,等. 基于异常点检测和改进 K-means 算法的台区用户相别辨识方法[J]. 智慧电力, 2020,48(1):91-96.
ZHANG Ran, SUN Xiaolu, HE Zhongxiao, et al. Phase identification method for distribution area users based on outlier detection and improved K-means algorithm [J]. Smart Power, 2020,48(1):91-96.
- [3] 严强,李扬,樊友杰,等. 基于加权表决集成聚类的居民用电行为回归分析[J]. 电网技术, 2021,45(11):4435-4446.
YAN Qiang, LI Yang, FAN Youjie, et al. Regression analysis of residential electricity consumption behavior based on weighted voting ensemble clustering[J]. Power System Technology, 2021,45(11):4435-4446.
- [4] 张素香,刘建明,赵丙镇,等. 基于云计算的居民用电行为分析模型研究[J]. 电网技术, 2013,37(6):1542-1546.
ZHANG Suxiang, LIU Jianming, ZHAO Bingzhen, et al. Cloud computing-based analysis on residential electricity consumption behavior [J]. Power System Technology, 2013,37(6):1542-1546.
- [5] HINO H, SHEN H, MURATA N, et al. A versatile clustering method for electricity consumption pattern analysis in households[J]. IEEE Transactions on Smart Grid, 2013,4(2):1048-1057.
- [6] 郑思达,梁琪琳,彭鑫霞,等. 基于模糊聚类的异常用电行为识别研究[J]. 电测与仪表, 2020,57(19):40-44.
ZHENG Sida, LIANG Qilin, PENG Xinxia, et al. Research on abnormal power consumption behavior identification based on fuzzy clustering[J]. Electrical Measurement & Instrumentation, 2020,57(19):40-44.
- [7] 张洁,夏飞,袁博,等. 基于特征优选策略的居民用电行为聚类方法[J]. 电力系统自动化, 2022,46(6):153-159.
ZHANG Jie, XIA Fei, YUAN Bo, et al. Clustering method for residential electricity consumption behavior based on feature optimization strategy[J]. Automation of Electric Power Systems, 2022,46(6):153-159.
- [8] 王德文,周昉昉. 基于无监督极限学习机的用电负荷模式提取[J]. 电网技术, 2018,42(10):3393-3400.
WANG Dewen, ZHOU Fangfang. Extraction of electricity consumption load pattern based on unsupervised extreme learning machine[J]. Power System Technology, 2018,42(10):3393-3400.
- [9] 周明,宋旭帆,涂京,等. 基于非侵入式负荷监测的居民用电行为分析[J]. 电网技术, 2018,42(10):3268-3276.
ZHOU Ming, SONG Xufan, TU Jing, et al. Residential electricity consumption behavior analysis based on non-intrusive load monitoring[J]. Power System Technology, 2018,42(10):3268-3276.
- [10] Growace. Growing marijuana indoors[EB/OL]. <https://growace.com/blog/growing-marijuana-indoors/>, 2018-08-01.
- [11] 栾文鹏,余贻鑫,王兵. AMI 数据分析方法[J]. 中国电机工程学报, 2015,35(1):29-36.
LUAN Wenpeng, YU Yixin, WANG Bing. AMI data analytics[J]. Proceeding of the CSEE, 2015,35(1):29-36.
- [12] 金晟,苏盛,薛阳,等. 数据驱动窃电检测方法综述与低误报率研究展望[J]. 电力系统自动化, 2022,46(1):3-14.
JIN Sheng, SU Sheng, XUE Yang, et al. Review on data-driven based electricity theft detection method and research prospect for low false positive rate[J]. Automation of Electric Power Systems, 2022,46(1):3-14.
- [13] SINGH S K, BOSE R, JOSHI A. Entropy-based electricity theft detection in AMI network[J]. IET Cyber-Physical Systems: Theory & Applications, 2018,3(2):99-105.
- [14] 袁宏俊,杜康,胡凌云. 基于相对熵的最优区间型组合预测模型[J]. 统计与决策, 2020,36(11):26-31.
YUAN Hongjun, DU Kang, HU Lingyun. Optimal interval combination forecasting model based on relative entropy[J]. Statistics & Decision, 2020,36(11):26-31.
- [15] 吴亮,王谊,谢岳,等. 基于异常事件故障关联度的电能表可靠性评价方法[J]. 电测与仪表, 2020,57(20):147-152.
WU Liang, WANG Yi, XIE Yue, et al. Reliability evaluation method of electricity meter based on fault correlation degree of abnormal event[J]. Electrical Measurement & Instrumentation, 2020,57(20):147-152.
- [16] 林顺富,田二伟,符杨,等. 基于信息熵分段聚合近似和谱聚类的负荷分类方法[J]. 中国电机工程学报, 2017,37(8):2242-2252.
LIN Shunfu, TIAN Erwei, FU Yang, et al. Power load classification method based on information entropy piecewise aggregate approximation and spectral clustering[J]. Proceeding of the CSEE, 2017,37(8):2242-2252.
- [17] 方安然,李旦,张建秋. 异常值和未知观测噪声鲁棒的

- 非线性滤波器[J]. 航空学报, 2021, 42(7): 532-545.
- FANG Anran, LI Dan, ZHANG Jianqiu. A nonlinear filter robust to outlier and unknown observation noise [J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(7): 532-545.
- [18] 袁翔宇, 张蓬鹤, 熊素琴, 等. 基于逻辑回归算法的异常用电辨识方法研究[J]. 电测与仪表, 2021, 58(12): 81-87.
- YUAN Xiangyu, ZHANG Penghe, XIONG Suqin, et al. Research on identification method of abnormal power consumption based on logistic regression algorithm [J]. Electrical Measurement & Instrumentation, 2021, 58(12): 81-87.
- [19] 李远松, 高博, 须琳, 等. 基于差分序列方差与 CPS 融合的数字变电站数据异常检测方法[J]. 电网与清洁能源, 2021, 37(2): 30-41.
- LI Yuansong, GAO Bo, XU Lin, et al. An anomaly detection method for digital substation abnormal data based on fusion of difference sequence variance and CPS[J]. Power System and Clean Energy, 2021, 37(2): 30-41.
- [20] 万轶婧, 王东海, 梁钊明, 等. 华南暖区暴雨环境参量的统计分析[J]. 中山大学学报(自然科学版), 2020, 59(6): 51-63.
- WAN Yijing, WANG Donghai, LIANG Zhaoming, et al. Statistical analysis of the environment parameters of warm-sector heavy rainfall in South China[J]. Acta Scientiarum Naturalium Universitatis Sunyatseni, 2020, 59(6): 51-63.
- [21] 郑宝芬, 苏宏业, 罗林. 无监督特征选择在时间序列数据挖掘中的应用[J]. 仪器仪表学报, 2014, 35(4): 834-840.
- ZHENG Baofen, SU Hongye, LUO Lin. Application of unsupervised feature selection in time series data mining[J]. Chinese Journal of Scientific Instrument, 2014, 35(4): 834-840.
- [22] 陆俊, 朱炎平, 彭文昊, 等. 智能用电用户行为分析特征优选策略[J]. 电力系统自动化, 2017, 41(5): 58-63+83.
- LU Jun, ZHU Yanping, PENG Wenhao, et al. Feature selection strategy for electricity consumption behavior analysis in smart grid[J]. Automation of Electric Power Systems, 2017, 41(5): 58-63+83.
- [23] 张梦瑶, 李磊, 刘彦文, 等. 基于灰色物元的变压器多指标绝缘状态评估方法[J]. 电力科学与技术学报, 2020, 35(1): 137-143.
- ZHANG Mengyao, LI Lei, LIU Yanwen, et al. Multi-indices insulation condition assessment method for transformer based on grey matter-element method[J]. Journal of Electric Power Science and Technology, 2020, 35(1): 137-143.
-
- (上接第 142 页 Continued from page 142)
- [20] 郜宁, 张慧媛, 王子琪, 等. 区域电网分布式储能选址定容规划[J]. 高压电器, 2020, 56(8): 52-58.
- GAO Ning, ZHANG Huiyuan, WANG Ziqi, et al. Planning for site selection and capacity determination of distributed energy storage in regional power grid[J]. High Voltage Apparatus, 2020, 56(8): 52-58.
- [21] RABIEE A, FARAHANI H F, KHALILI M, et al. Integration of plug-in electric vehicles into microgrids as energy and reactive power providers in market environment[J]. IEEE Transactions on Industrial Informatics, 2016, 12(4): 1312-1320.
- [22] SANJARI M J, KARAMI H, GOOI H B. Analytical rule-based approach to online optimal control of smart residential energy system [J]. IEEE Transactions on Industrial Informatics, 2017, 13(4): 1586-1597.
- [23] MEHTA R, SRINIVASAN D, KHAMBADKONE A M, et al. Smart charging strategies for optimal integration of plug-in electric vehicles within existing distribution system infrastructure[J]. IEEE Transactions on Smart Grid, 2018, 9(1): 299-312.
- [24] WEI H, SASAKI H. An interior point nonlinear programming for optimal power flow problems with a novel data structure [J]. IEEE Transactions on Power Systems, 1998(3): 870-877.
- [25] 任大伟, 金晨, 肖晋宇, 等. 计及灵活性基于时序的“十四五”储能需求分析[J]. 中国电力, 2021, 54(8): 190-198.
- REN Dawei, JIN Chen, XIAO Jinyu, et al. Demand analysis of energy storage for the 14th five-year plan period based on time series considering power system flexibility[J]. Electric Power, 2021, 54(8): 190-198.
- [26] QUAN H, SRINIVASAN D, KHAMBADKONE A M, et al. A computational framework for uncertainty integration in stochastic unit commitment with intermittent renewable energy sources [J]. Applied Energy, 2015, 152: 71-82.