

基于电力数据的居民人口流动情况评估

刘畅¹, 滕予非¹, 唐伟¹, 靳旦¹, 李赋兴²

(1. 国网四川省电力公司电力科学研究院, 四川 成都 610041;

2. 国网四川省电力公司, 四川 成都 610041)

摘要: 为克服社区人口人工摸排周期长且准确性不高的问题, 提出一种基于电力数据的居民人口流动情况评估方法。首先, 通过用户档案信息读取用户历史用电量数据, 完成缺失数据补齐和异常数据处理等工作实现数据清洗。然后, 基于聚类方法挖掘用户不同场景下的稳定用电量。在此基础上, 定义了用能突变判据和人口流动判据, 实现人口流动状况的评估, 并利用统计方法实现小区/社区的人口流动情况分析。最后, 以成都市某小区实际数据为例, 验证了算法的有效性。

关键词: 电力大数据; 用能特征; 人口流动评估; 用能突变判据

中图分类号: TM714 **文献标志码:** A **文章编号:** 1003-6954(2020)03-0033-05

Population Mobility Evaluation of Residents Based on Electric Power Data

Liu Chang¹, Teng Yufei¹, Tang Wei¹, Jin Dan¹, Li Fuxin²

(1. State Grid Sichuan Electric Power Research Institute, Chengdu 610041, Sichuan, China;

2. State Grid Sichuan Electric Power Company, Chengdu 610041, Sichuan, China)

Abstract: In view of the long period and low accuracy of manual survey of community population, an evaluation method of population mobility based on electric power data is proposed. Firstly, the user's historical load consumption data are obtained based on user profile, and the missing data completing and abnormal data processing are used to ensure the validity of load data. Then, the clustering method is used to mine the stable power consumption of residents in different scenarios. On this basis, load mutation criterion and population mobility criterion are defined to evaluate population mobility of residents. Finally, statistical method is employed to analyze the population mobility of the community. Taking the actual data of a residential area in Chengdu for example, the effectiveness of the proposed algorithm is verified.

Key words: power big data; electricity consumption feature; population mobility evaluation; load mutation criterion

0 引言

随着互联网技术和大数据技术的不断推广和应用, 发展智能经济、数字经济已成为大势所趋。在电力物联网建设进程不断深入过程中, 数字化转型已经上升为电网企业的重要战略, 激活数据价值、挖掘数据应用场景、开发数据产品、服务社会发展成为热点^[1-2]。

人员流动情况摸查是电力数据应用于提质增效的重要场景。尤其是自新冠肺炎疫情发生以来, 由于疫情时值春节, 走亲访友的情况较多, 加之城市社

区人口密集, 给工作人员本来就有限的社区排查、管控工作带来极大的困难, 导致人员摸排周期长、摸排不准确。基于海量用电数据把握用户的用电规律和用电特性从而辨别居民人口流动情况, 可以及时有效地解决上述问题。

然而, 在用户行为分析方面, 当前研究主要集中在基于电力数据的用户用电行为解析^[3-5], 基于用电信息的人员流动情况评估方法较少。文献[6]提出了一种基于细粒度用能数据的居民家庭活动人口评估方法, 评估结果具有较高的可行性。但上述方法数据要求较高, 需要获取采样间隔为 15 min 的居民分项电器用电数据、空调和电热负荷详细数据。

对于大部分小区尤其是老旧小区而言,分时、分项数据获取基本无法实现,严重限制了方法的应用场景,也难以满足实际需求。

鉴于此,提出了一种基于居民日电量数据的人口流动情况评估方法。首先,通过历史用电数据聚类挖掘居民的稳定用电水平;然后,基于日电量变化情况和稳定用电水平提出了人口流动判据;最后,基于上述判据对小区居民流动情况进行分析和研判,为社区疫情排查工作提供依据,并以成都市某小区居民实际数据为例验证了所提方法的有效性。

1 居民历史用电特征挖掘

1.1 数据清洗

在线采集上传的日电量,真实反映客户当日生产生活用电情况,具有实时性强的特点;但受到信号干扰、软硬件故障、通信异常等情况的影响,数据可能会出现缺失、异常等情况。数据质量直接关系到分析的结果,因此,需要对采集的数据进行校验和清洗等预处理。

日电量是一个累积值,因此,对于缺失数据的处理有两种方法。

1) 用缺失日后1日的日电量作为平均值。计算方法为:

$$\begin{aligned} W_{\text{new}}(N) &= \frac{W(N+1)}{2} + \varepsilon \\ W_{\text{new}}(N+1) &= \frac{W(N+1)}{2} - \varepsilon \end{aligned} \quad (1)$$

式中: N 为数据缺失日期; $W_{\text{new}}(N)$ 为更新数值; $W(N+1)$ 为数据缺失日的后一日的日电量采集数值; ε 为随机噪声信号。该方法适用于对历史数据的处理。

2) 用前一段时间日电量均值代替。计算方法为

$$W_{\text{new}}(N) = \frac{\sum_{i=1}^7 W(N-i)}{7} \quad (2)$$

该方法适用于最新日电量的处理。

对于异常数据的校正,主要用于超出合理范围的数据。一般而言,居民日用电量在0至允许容量之间,若日电量小于0,则用0替代;若日电量大于允许容量,则该日电量用最大容量或历史均值替代,计算方法与式(2)相同。

1.2 居民稳定用电量挖掘模型

聚类是样本分类的常用算法^[7],聚类分析的主要目的是通过分析数据的分布特征实现将数据对象划分为若干数据子集,保证每个子集中数据的相似性。其中,k-means聚类方法因其简单高效而广泛应用于电力系统负荷分类中^[8-9]。

考虑到居民用电情况可分为当日不在家、当日部分时间在家和当日全天在家3种场景,可以通过聚类算法将居民历史日电量分为3类,分别代表3种场景的用电量样本。这里主要采用k-means算法对不同居民用电场景的电量进行聚类。

假设在同一季节,居民的用电量与在家时长呈现正相关,求取各类数据均值,均值数据由大到小分别对应全天在家 W_{part} 、部分时间在家 W_{full} 和不在家 W_{no} 3类情况的稳定用电量值。分别统计各类中的样本数量,获取各类场景的出现概率,从而了解居民的生活习惯。

1.3 模型校正

由于历史数据不足,部分用户用电场景不一定齐全,因此,需要对模型进行校正。校正判据为

$$\frac{d_{\text{min}}}{W_{\text{max}}} < 0.1 \quad (3)$$

式中: d_{min} 为类间距离,是3类样本均值之间的距离; W_{max} 为历史日电量最大值,表示所选样本中最大值,样本一般为同一季节的日电量数据集合。

若上述判据满足,则将距离最近的两类合为一类。

2 单户居民人口流动判据

由于日电量信息颗粒度较大,无法反映实时人口活动情况,因此,日电量信息难以反应具体的人数信息。对于居民流动情况,这里重点关注人口流出后住房空置、由空置到入住的人口流入两种情况。

2.1 居民用电量变化量判据

考虑到居民离开家当天用电量可能介于全天在家场景和全天不在家场景,因此,居民用电变化量考虑次日用电量变化和隔日电量变化两个层次进行指标设计,其计算方法为:

$$\text{次日突变率} = \frac{W(t) - W(t-1)}{W_{\text{max}}} \quad (4)$$

$$\text{隔日突变率} = \frac{W(t) - W(t-2)}{W_{\text{max}}} \quad (5)$$

式中: $W(t)$ 为当日日电量; $W(t-1)$ 为前一日日电量; $W(t-2)$ 为隔日日电量; W_{max} 为历史日电量最大值。

若某居民用户的日电量突变率满足式(6)中的两个判据之一,则该户居民为用能突变用户。

$$\begin{cases} |\text{次日突变率}| > 0.3 \\ |\text{隔日突变率}| > 0.3 \end{cases} \quad (6)$$

对于用能突变用户而言,若突变率大于0,则居民为用能激增用户;若突变率小于0,则该居民为用能骤减用户。

2.2 居民流入流出判据

考虑到居民流入、流出前后用电状态的变化,对于用电量可划分为3类的情况,定义底值系数为

$$K_{cut} = \frac{k_1 W_{no} + k_2 W_{part}}{W_{max}} \quad (7)$$

式中, k_2 和 k_1 分别为两个系数,其关系满足 $k_2 + k_1 = 1$, 且 $k_1 > 2k_2$, 这里取 $k_1 = 5/6$, $k_2 = 1/6$ 。

对于模型校正后只有两类的样本,底值系数为

$$K_{cut} = \frac{0.5 W_{no}}{W_{max}} \quad (8)$$

底值系数主要用于区分用户的居家状态,若该日日电量与历史日电量最大值的比值小于或等于底值系数,则判定用户不在家的概率较大。

2.3 居民流入流出判据

1) 疑似人口流入的判据

判据1: 该用户为用能激增用户。

判据2: 电量激增前的日电量小于 $W_{max} K_{cut}$ 。

同时满足判据1和判据2,则认为该户为人口流入客户,标志位记为1。

2) 人口流出的判定判据

判据3: 该用户为用能骤减用户。

判据4: 电量骤减后的日电量小于 $W_{max} K_{cut}$ 。

同时满足判据3和判据4,则认为该户为人口流出客户,标志位记为-1。

值得注意的是,若隔日突变率和次日突变率均大于1/3时,可能存在连续两天判定外出的情况。因此,需要对上述情况进行校核,选取中间一天作为流入或流出的时间,保证流入和流出的准确性。

由于在流出到流入之间的时段,房间处于空置状态,用电量变化一般很小,利用该特征可以进一步校核判定结果的有效性。即: 人口流出后到下一次

人口流入前,(日电量变化量/历史日电量最大值)小于0.01,则认为该户短时空置,人流测算有效。

根据流入流出判据标记位按日累加,可以获得该户居民居家状况曲线。若居民居家状况的最小值为-1,则该曲线整体加1。居家状况为1,代表该户当日非空置,否则代表该户当日空置。

3 小区总体居民人口流动状况统计

小区总体居民人口流动状况可以反映总体的人口流动状态,为社区防疫工作提供参考信息。具体包括:

1) 小区持续空置户数

持续空置客户包含两类: 1) 电表表底数据持续为空的未开户用户; 2) 表计已开户但变化量持续为零的空置住房用户。定义两类用户户数的总和为持续空置户数。

2) 居民用电量突变户数

定义除持续空置用户外的居民中,居民次日日电量突变率或隔日电量突变率绝对值大于1/3的居民户数,为当日居民用能突变户数。其中,次日突变率数值为正的居民户数为用户用电激增户数,突变率数值为负的居民户数为用户用电骤减户数。

3) 稳定用能户数

当日稳定用能户数定义为除持续空置用户外的居民中,居民次日日电量突变率或隔日电量突变率绝对值小于1/3的居民户数。

4) 短期空置户数

小区内短时空置的用户数目。

5) 非空置住宅数目

小区内当日有人居住的住宅数目。

4 算例分析

以成都市某小区2020年1月14日至2月13日一个月的电量数据为例进行分析。该小区总电表户数为180户,其中,表计未计数用户数为48户,统计期间未入住空置住宅户数为36户,空置率为46.67%。由于该小区为成都市内的新小区,入住率不高。

剩余96户的用能曲线如图1所示。由于该小区为别墅区,部分负荷较高。从负荷曲线可以看出,当人员离开时负荷特征明显。

居民的典型判定曲线如图2所示。其中,居家情况为0表示该户不在家,为1表示该户在家。如图2(a)所示,该居民2020年2月3日返回家中,居民电量在当日出现上升,但次日上升电量较小,不满足用能突变判据,但隔日突变率满足要求,可判定该户用能突变。可见,仅利用次日突变率判据可能出

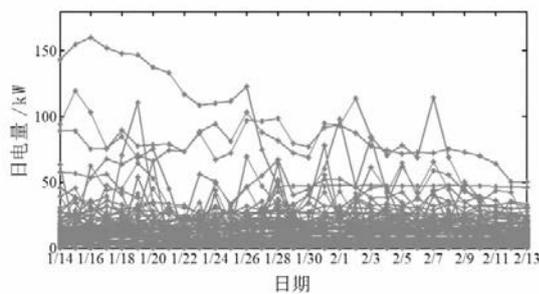
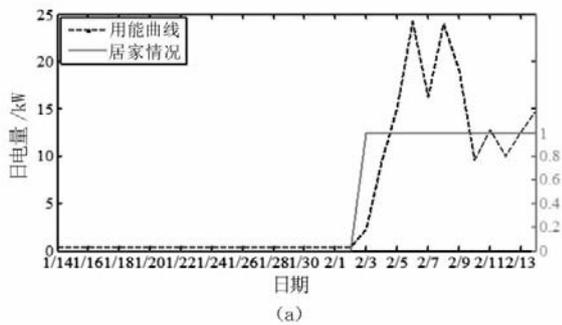
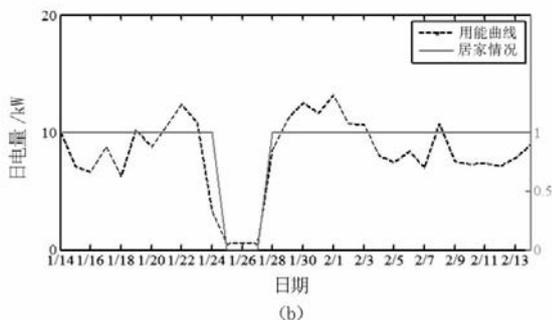


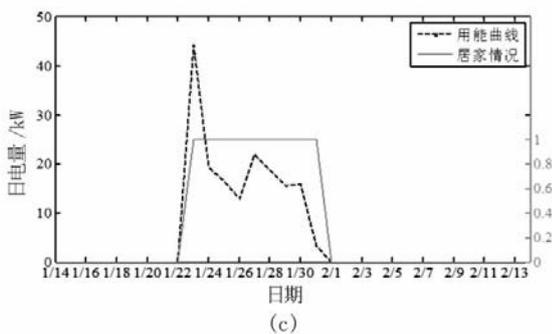
图1 居民用能曲线



(a)



(b)

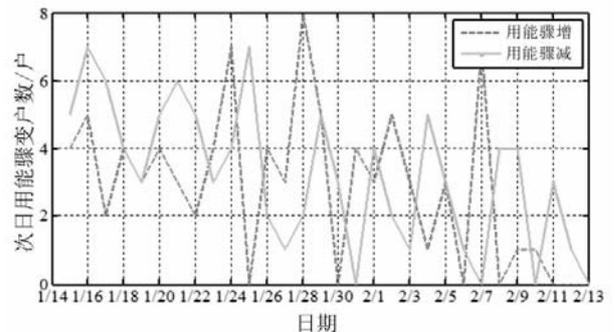


(c)

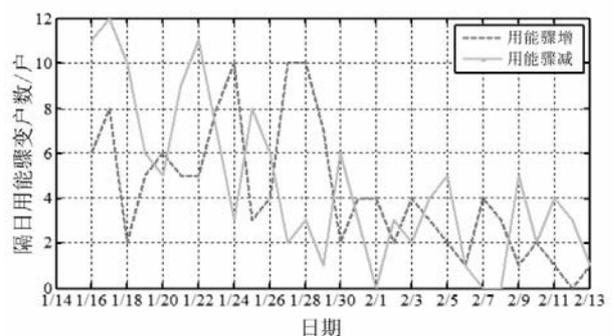
图2 居民居家情况典型判定

现漏判。根据居家情况判定结果可知,该算法能准确反映人口流动情况。图2(b)所示为居民1月24日离开,1月28日返回;图2(c)所示为居民1月23日返回,2月1日离开,居家判定结果准确。

从小区整体来看,用能骤增和用能骤减的居民户数随时间的变化曲线如图3所示,其中图3(a)为次日用能骤变居民户数,图3(b)为隔日用能骤变居民户数。



(a)



(b)

图3 小区用能骤变户数

小区疑似人口流动的户数如图4所示。由图可见,1月22日(春节前)疑似流出人口居民户数大于当日用能骤变数目或隔日用能骤变数目,说明所采用的两个指标结合的方法性能更好。

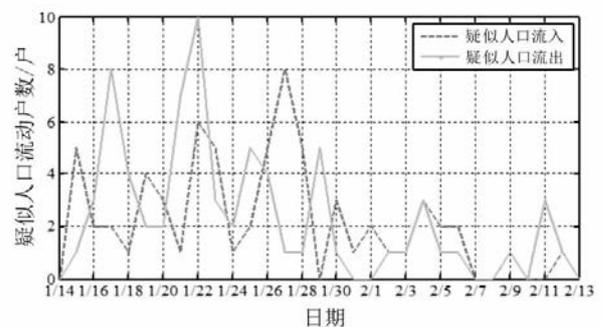


图4 小区疑似人口流动户数

该小区非空置住宅户数如图5所示。由图可见,1月15日至1月22日(春节前)流出户数略大于流入户数,非空置住房数持续下降,整个小区人口流出明显。1月23日(除夕)有所回升,但随后继续下降,至1月25日达到最低点。随后,非空置住房数持续上升,到2月1日后基本保持稳定。分析结果与实际情况基本相符。1月28日后非空置住房数回落疑似与国家将复工时间由1月30日调整至2月3日有关。

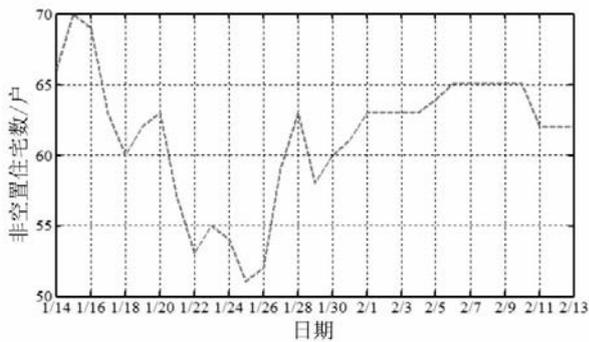


图5 小区非空住宅数

由上述分析可知,虽然日电量数据颗粒度不高,但采用所提出的算法后仍能较好地监测人口流动情况。

5 结语

鉴于城市社区人口密集,采用人工排查人口情况的方式工作量大、工作周期长,且受节假日等因素的影响其数据准确性不高等问题,提出了一种基于电力大数据的居民人口流动情况评估方法,通过构建用能突变判据和人口流动判据实现了居民人口流入、流出情况的动态监测。利用所提方法应用于某小区的人口流动分析,得到结论如下:

1) 根据居民数据分析结果与实际流入流出情况的对比可知,采用居民日电量数据可以较好地反映居民人口流动情况。

2) 采用隔日突变率和次日突变率相结合的方法,可以更好地反映用户的实际用能变化情况,有效提高算法的准确性。

3) 由算例分析结果可知,小区居民流动情况和非空置户数情况受节假日(春节)影响很大,假期前

半段流出数量较多,后半段流入较多。分析结果与实际相符。

由于日电量维度较低,载有的信息量较少,因此,在用户用能习惯变化极大的情况下,所提方法可能出现误判或漏判。因此,未来将进一步研究在有限信息下精确挖掘用户习惯、提高算法准确性的方法。

参考文献

- [1] 王继业,季知祥,史梦洁,等. 智能配用电大数据需求分析与应用研究[J]. 中国电机工程学报, 2015, 35(8): 1829-1836.
- [2] 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.
- [3] 罗滇生,杜乾,别少勇,等. 基于负荷分解的居民差异化用电行为特性分析[J]. 电力系统保护与控制, 2016, 44(21): 29-33.
- [4] 孙毅,崔灿,陆俊,等. 基于遗传优化的非侵入式家居负荷分解方法[J]. 电网技术, 2016, 40(12): 3912-3917.
- [5] 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): 3268-3274.
- [6] 张洋瑞,陶鹏,李翀,等. 基于细粒度用能数据的居民家庭活动人口评估方法研究[J]. 河北电力技术, 2019, 38(6): 48-51.
- [7] 蔡秋娜,刘思捷,陆秋瑜. 基于GMM聚类和SVM的用户负荷行业分类辨识方法[J]. 广东电力, 2017, 30(12): 91-96.
- [8] Al - Wakeel A, Wu J, Jenkins N. K - means Based Load Estimation of Domestic Smart Meter Measurements [J]. Applied Energy, 2017, 194: 333-342.
- [9] 程明畅,刘友波,张程嘉,等. 基于分位数半径动态K - means的分布式负荷聚类算法[J]. 电力自动化设备, 2019, 47(24): 15-21.

作者简介:

刘畅(1991), 硕士, 工程师, 研究方向为电力系统分析与大数据应用。

(收稿日期: 2020-03-09)