

基于改进密度聚类的用气异常检测

张梦园^{1,2}, 彭定涛^{1*}, 胡殿涛²

¹贵州大学数学与统计学院, 贵州 贵阳

²航天智慧能源研究院, 上海

收稿日期: 2021年10月22日; 录用日期: 2021年11月11日; 发布日期: 2021年11月24日

摘要

为了更好地从管网系统中挖掘数据信息, 科学地计量工商业用气规律, 帮助燃气公司对用户异常用气行为进行智能识别, 本文提出了一种基于K-近邻距离图和网格搜索法(Grid search)的密度聚类(DBSCAN)算法, 结合分段聚合近似表示方法(PAA)在包含噪声的数据集中通过寻找工业燃气数据集的内在分布规律和聚类效果的变化来识别异常点。首先以来自SCADA和智能表具采集的南方某陶瓷工厂日负荷数据为例, 使用PAA方法对数据进行降维处理。其次利用改进的DBSCAN算法对案例用户监测时段中的异常数据进行识别。最后将算法在某南方陶瓷行业的325个用户数据上进行了验证。结果表明, 算法的平均准确率在90%以上, 人工智能算法在燃气领域的应用对于燃气经营企业实现精细化管理、以及达到降本增效的效果具有一定的指导意义。

关键词

DBSCAN, PAA, K近邻距离图 - 网格搜索法, 日负荷曲线

Anomaly Detection of Industrial Gas Load Based on Improved Density Clustering

Mengyuan Zhang^{1,2}, Dingtao Peng^{1*}, Diantao Hu²

¹School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

²Aerospace Smart Energy Research Institute, Shanghai

Received: Oct. 22nd, 2021; accepted: Nov. 11th, 2021; published: Nov. 24th, 2021

Abstract

In order to better mine the data information from the pipeline network system, scientific measurement of industrial and commercial gas consumption laws, and help gas companies to intelli-

*通讯作者。

gently identify abnormal gas consumption behavior of users, a density clustering (DBSCAN) algorithm based on k-nearest neighbor distance graph and Grid search is proposed, which identifies outliers in noisy data sets by the method of approximate representation based on piecewise aggregation (PAA) to search for the distribution of industrial gas data sets and the change of clustering effect. Firstly, the daily load data of a ceramic factory in south China collected by SCADA and smart meters are taken as an example, and the data are processed with PAA method. Secondly, the improved DBSCAN algorithm is used to identify the abnormal data in the monitoring period of case users. Finally, the algorithm is applied to 325 users' data in a ceramic industry in south China. The results show that the average accuracy of the algorithm is over 90%, and the effective application of artificial intelligence algorithm in gas field is realized. It has certain guiding significance for gas enterprises to realize fine management and achieve the effect of reducing cost and increasing efficiency.

Keywords

DBSCAN, PAA, K-Nearest Neighbor Distance Graph-Grid Search Method, Daily Load Curve

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 研究背景

大数据时代的繁荣,使得数据成为重要的战略资源,如何运用海量数据提升燃气企业精细化管理、降本增效成为急需解决的关键问题。在工业用气中,对于一些负荷量大的工业用户[1],当用户行为发生较大变化或计量设备损坏时,用户标况瞬时流量就会出现较大波动。燃气公司以标况总量计费,标况瞬时流量和累计流量呈现线性相关。因此标况瞬时流量异常意味着流量计可能发生故障,流量计是燃气公司结算费用的重要仪器,因为缺乏智能化管理,通常总是遭受损失很久后燃气公司才后知后觉,因此如何对燃气数据分析对于推动燃气公司逐步迈向智能化监测具有重要意义[2] [3]。

目前对于异常值检测算法已经有了较为成熟的研究,文献[4]采用GPR模型融合距离相关系数的方法检测卫星数据流;文献[5]采用了基于层次聚类的异常检测算法对数据降维,发现对降维后的数据进行异常检测能保持90%以上的精准率。除了基于统计原理和聚类原理的异常检测方法[4] [5] [6] [7],基于机器学习的异常点检测方法也有丰富的进展[8] [9] [10] [11] [12]。文献[8]在网络运行环境下构造了数据的特征,并搭建了三种基于神经网络和基于语义与空间等网络异常行为检测研究模型。文献[10]采用了基于机器学习的无线传感器网络异常值检测算法,对用聚类算法处理后的数据生成支持向量机并采用鲸鱼优化算法进行优化,提升了模型的准确率和模型的泛化性。在与燃气领域相似的电力负荷异常检测方面,文献[13]提出了基于变分自编码器的负荷检测方法。文献[14]通过基于距离的聚类检测算法确定最佳聚类数目,结合K-means设计了一套用户异常用电嫌疑等级报警算法。文献[15]通过小波变换,过滤掉原始数据的噪音,使异常点在高通滤波器的应用下更为突出,更容易被检测出来。文献[16]通过LOF离群点算法建立用电异常评价指标,将异常用电检测方法推广到反窃电系统中。

本文基于网格寻优的密度聚类方法结合分段聚合近似表示方法对工业燃气用户的标况流量数据进行异常检测。通过对原始数据(标况状态下由流量计记录的瞬时负荷数据)分段聚合绘制日负荷曲线,利用K近邻距离图综合具体条件遍历区间筛选出最优参数,在包含噪声的数据集中通过寻找工业燃气数据的内在分部规律和聚类效果的变化来识别异常曲线,文章的技术路线如图1:

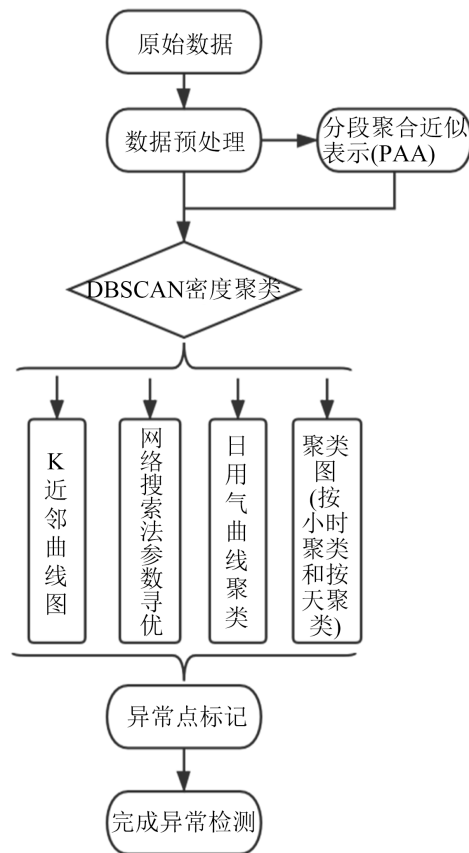


Figure 1. Algorithm flowchart
图 1. 算法流程图

2. 异常负荷曲线检测算法

2.1. 分段聚合近似算法

分段聚合近似表示方法(Piecewise Aggregate Approximation, PAA) [17]是按某种规则把序列分段表示的方法, 利用序列分段后的平均值映射原始信息从而达到降维的目的。降维后的数据保留了原始序列的主要信息。给定时间序列 $T = \{t_1, t_2, \dots, t_n\}$, 将 n 个数据做 m 等距分段, 第 \bar{i} 段的计算方法为:

$$\bar{t}_i = \frac{m}{n} \sum_{j=i\frac{n}{m}}^{\frac{n}{m}(i+1)-1} t_j \quad (1)$$

$0 < i \leq m$, $i \in N$, N 为正整数。

2.2. 密度聚类算法

密度聚类(Density-Based Spatial Clustering of Applications with Noise, DBSCAN)是基于一组邻域参数 $(\varepsilon, \text{MinPts})$ 来描述样本分布紧密程度的方法[18] [19]。

定义[20] [21]: 给定 n 维向量空间记 $X \in R^n$, 将 X 分为 m 组. $X = (X_1^T, X_2^T, \dots, X_m^T)$, $X_i \in R^n$, $i = 1, 2, \dots, m$, 每个组有 n_i 个元素。且 $\sum_{i=1}^m n_i = n$, $n_i \geq 1$ 。

1) ε -邻域(Eps): 对 $X_i \in R^n$, 其 ε -邻域包含 R^n 中与 X_i 的距离不大于 ε 的所有样本。

$$N_\varepsilon(X_{(i)}) = \{X_{(j)} \in R \mid \text{dist}(X_{(i)}, X_{(j)}) \leq \varepsilon\} \quad (2)$$

$$\text{dist}(X_{(i)}, X_{(j)}) = \sqrt{\sum_{m=1}^n (x_m^{(i)} - x_m^{(j)})^2}$$

2) MinPts: ε -邻域内样本个数最小值。

3) 核心对象: 若 $X_{(j)}$ 的 ε -邻域至少包含 MinPts 个样本 $|N_\varepsilon(X_{(j)})| \geq \text{Minpts}$, 则 $X_{(j)}$ 为一个核心对象。

4) 密度直达: 若 $X_{(j)} \in N_\varepsilon(X_{(i)})$, 且 $X_{(i)}$ 是核心对象, 则称 $X_{(j)}$ 由 $X_{(i)}$ 密度直达。

5) 密度可达: 对 $X_{(i)}$ 与 $X_{(j)}$, 若存在样本序列 a_1, a_2, \dots, a_n , 其中 $a_1 = X_{(i)}$, $a_n = X_{(j)}$, a_1, a_2, \dots, a_{n-1} 均为核心对象且 a_{i+1} 从 a_i 密度直达, 则称 $X_{(j)}$ 由 $X_{(i)}$ 密度可达, 且密度可达关系满足传递性。

6) 密度相连: 对与 $X_{(j)}$, 若存在 $X_{(k)}$ 使得 $X_{(i)}$ 与 $X_{(j)}$ 均有 $X_{(k)}$ 密度可达, 则称 $X_{(i)}$ 与 $X_{(j)}$ 密度相连。

密度聚类的簇里面可以有多个核心对象。如果只有一个核心对象, 则簇里面所有样本都应在这个核心对象的 ε -邻域里; 这些核心对象 ε -邻域里的所有样本的集合组成一个 DBSCAN 簇。和传统的 K-means 算法相比, DBSCAN 最大的不同就是不需要输入类别数 k , 当然它最大的优势是可以发现任意形状的聚类簇, 同时它在聚类的时候还可以找出异常点。对于一个数据点若新数据是一个核心对象, 且不属于任何一个已存在的簇内和此数据的密度可达点中没有已存在簇的核心点, 则组成一个新的簇; 如果在领域内没有核心点, 且邻域内最小样本个数小于阈值, 则将此数据标记为异常点[22] [23] [24]。

异常检测的准确率计算公式为:

$$R = \left(1 - \frac{f}{N}\right) \times 100\% \quad (3)$$

其中 R : 准确率; f : 未检测出的样本与检测错误的样本之和, N : 总样本。

2.3. 参数寻优

DBSCAN 算法不需要提前给定簇类别只需要计算 eps 参数和 MinPts, 这两个参数在通过 DBSCAN 密度聚类算法创建正确的簇中起着重要的作用。在以往的研究中这两个参数大都根据经验或比较试错得出有很大的主观性。由于文中采用的是曲线聚类, 计算复杂度随之提升。因此提出以下改进方案: 结合 K 近邻曲线图和网格搜索法(Grid Search)遍历区间内的所有参数, 并将各个参数的可能取值进行排列组合, 综合燃气日负荷曲线原始曲线筛选最佳参数组合, 参数确定的主要流程如表 1:

Table 1. Parameter optimization algorithm based on K-nearest neighbor distance graph and grid search method

表 1. 基于 K-近邻距离图和网格搜索的参数寻优算法

参数寻优算法:

Input: 样本集 日负荷曲线 $D = \{X_1, X_2, \dots, X_m\}$

过程:

原始数据按第 i 列归一化: $x = \frac{x - \min(X_{(i)})}{\max(X_{(i)}) - \min(X_{(i)})}$

For X_1, X_2, \dots, X_m do 前 K 近邻 $L(X)_k = \{L(X_1)_k, L(X_2)_k, \dots, L(X_m)_k\}$

Continued

$$\text{令 } D(X_{(i)}, X_{(j)}) = \sqrt{(X_{(i)} - X_{(j)})^2}, X = [x_1, x_2, \dots, x_m]$$

For $i = 1, 2, \dots, m$

For $j = 1, 2, \dots, k$

$$L(X_i)_j = \min_j \left\{ \sqrt{(X_i - X_1)^2}, \sqrt{(X_i - X_2)^2}, \dots, \sqrt{(X_i - X_m)^2} \right\}, \text{ 其中 } \{X_1, X_2, \dots, X_m\} / X_i$$

对距离 $L(X_i)_j$ 降序排列, 并将近邻距离映射到 (0,1)

Output: K 近邻曲线图 and 参数范围

Import 网格搜索工具

将需要遍历的参数定义为字典

定义 DBSCAN 模型 and 参数范围

设置步长 $\alpha = 5$

拟合数据

Output: 最优参数

end

3. 实验分析

3.1. 异常检测

以广州某陶瓷工厂燃气标况流量数据为例进行实验分析。实验数据共3720个, 大约10分钟左右采集一次。由于数据的采集时间是不规律的, 对数据做如下预处理: 首先把时间间隔转化为10分钟一次, 缺失的以相邻数据填补。根据工业原理, 工厂在一小时的用气量是相对稳定的, 经过24段聚合后的流量曲线如图2所示。绿色线段是分段聚合近似算法处理后的结果, 相比于处理之前, 处理后的曲线降低了曲线平滑度, 使得段内特征减弱, 段与段之间的差异则更加凸显。分段聚合近似算法实现了日负荷曲线分段降维, 同时增加权重在后续的相似性分析中使得相似度精度更高。

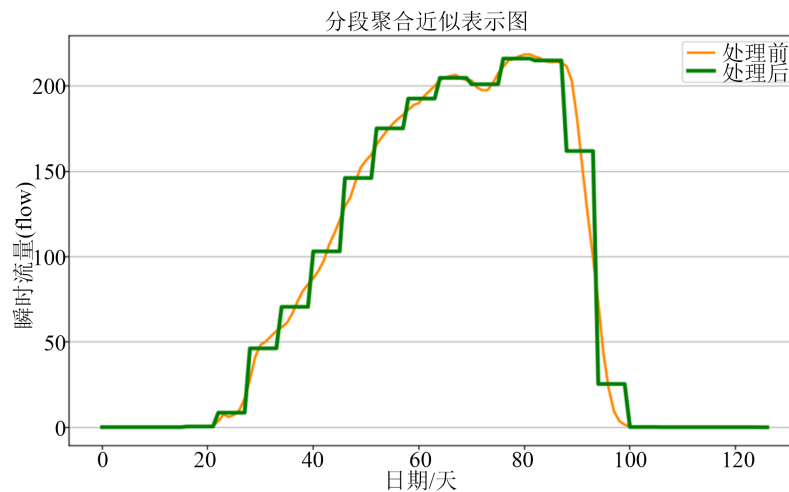


Figure 2. Piecewise aggregate approximation representation graph

图 2. 分段聚合近似表示图

通过 K-近邻曲线图和网格搜索法来确定密度聚类的参数, 图 3 为日负荷曲线的近邻距离图可以看出数据集于第 5 条曲线开始出现突变。因此可以初步得出 MinPts 为 5, 然后根据拐点处的 y 值可以得出 ε 为 88。再根据天然气使用规律可知日负荷曲线在纵向和横向上表现出聚类特性, 对于一个工业用户来讲, 其每天的用气量应与之前的用气量具有相似性, 且由于实际中工厂在休息日和工作日的负荷曲线具有差异性, 因此排除异常天的情况下聚点应至少为两类, 且异常点应属于少数点, 利用网格搜索法寻优的结果如表 2 所示, 从中自动筛选出的红线框住部分为合适的参数, 可以观察到异常值 $n_clusters$ 处于突变中, $stats$ 样本聚类数分部较平均。因此综合考虑选择的参数为 $(\varepsilon, \text{MinPts}) = (100, 4)$ 。

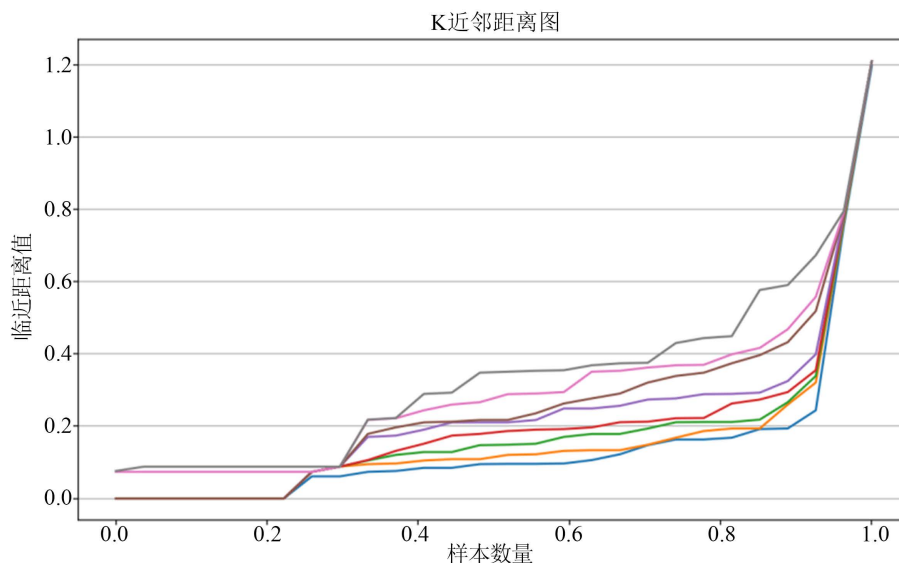


Figure 3. K-nearest neighbor distance graph
图 3. K-近邻距离曲线图

Table 2. Grid search method parameter optimization
表 2. 网格搜索法参数寻优

参数的网格搜索寻优					
	eps	min_samples	n_clusters	outliners	stats
0	40.00	2	4	5	[9 7 4 3]
			⋮		
36	100.00	2	2	2	[17 9]
37	100.00	4	2	2	[17 9]
38	100.00	6	2	2	[17 9]
			⋮		
71	155.00	6	2	2	[17 9]

对该工厂日负荷曲线做密度聚类(DBSCAN)的聚类结果如下图 4 所示, 可以发现: 日用气流量曲线分为两类, 一类是橙色曲线描述了工厂正常运行时的日流量走势, 另一类是绿色曲线描述了工厂停工时的少量用气曲线, 剩余的紫色曲线为异常曲线。再由每日用气曲线制作的标签表 3 中可以看出异常曲线对应的日期为 8 月 8 号和 8 月 13 号, 而实际中陶瓷工厂开始运转后用气量应该是平稳上升一段时间后经过中午短暂的下降后再回升再下降为零的, 异常天的负荷曲线呈现出大幅的波动, 因此不符合工业用气

机理，具体的异常原因还需结合工厂实际情况进行分析。另外 26 为正常天，正常天又分为两种用气特征曲线，正常工业用气和停业小幅用气。由日负荷曲线聚类结果可以看出该工厂的用气大致呈图中橙色曲线和绿色曲线两种类型，其中出现了 2 次异常曲线，异常曲线和正常曲线的对比如图 5 所示。

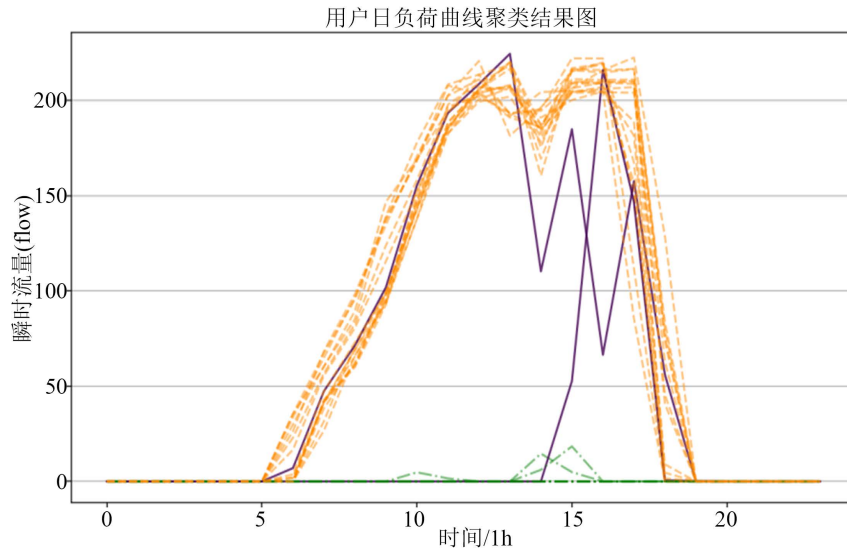


Figure 4. Daily load curve clustering result graph
图 4. 日负荷曲线聚类结果图

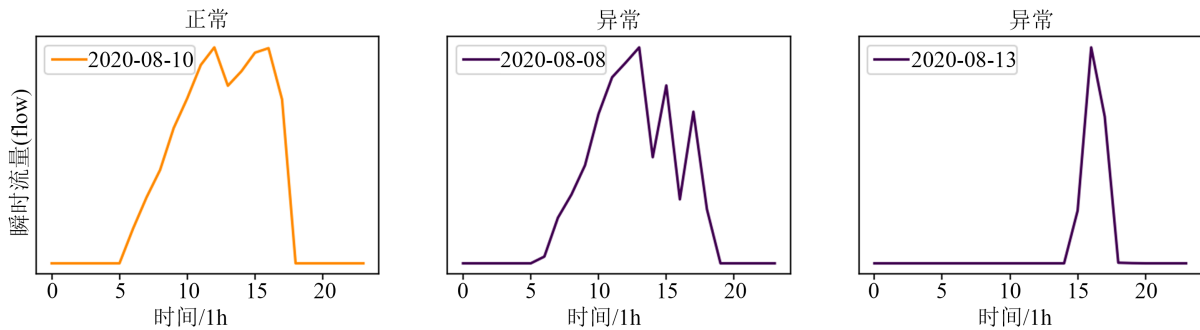


Figure 5. Comparison chart of daily load curve
图 5. 日负荷曲线对比图

Table 3. Daily load curve classification table
表 3. 日负荷曲线分类表

日期	8.1	8.2	8.3	8.4	8.5	8.6	8.7	8.8	8.9	8.10	8.11	8.12	8.13	8.14
类别	0	1	0	1	1	0	1	-1	0	1	1	0	-1	1
	8.15	8.16	8.17	8.18	8.19	8.20	8.22	8.23	8.24	8.25	8.26	8.27	8.28	8.29
	1	1	0	1	1	0	0	1	1	1	0	1	1	1

对聚类结果做一些转换可以得到密度聚类的以小时为单位的的结果如下图 6 所示，黄色和绿色分别为两类不同的簇，紫色为异常点，从图中可以看出小时图更清晰的反映了该工厂运营时出现的异常用气行为。

通过对陶瓷工厂进行研究发现：陶瓷产品必须要达到规定的温度状态下才能烧制成功，否则做出来的就是残次品；因此工厂处于工作状态时对温度有着严格的把控，而该工厂在 8 月 8 日和 8 月 13 日的下

午使用燃气时流量数据不足以产生达到制成产品所需的最佳温度，但根据回访工厂当天生产的产品却是合格的，因此我们判断该工厂存在异常用气行为。

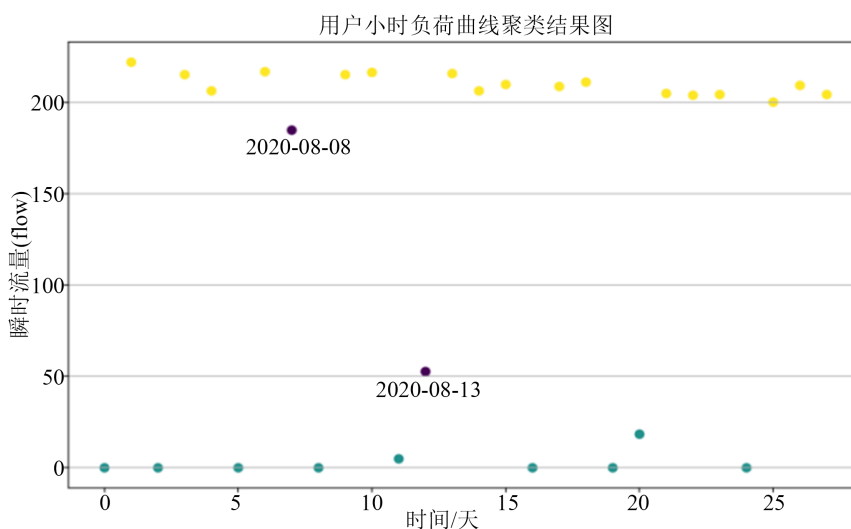


Figure 6. Clustering results of hourly load curve
图 6. 小时负荷曲线聚类结果图

3.2. 多用户大规模比较分析

为了验证算法的有效性，选取了 325 家燃气工业用户(2020.1.1~2021.4.30)进行大规模测试，其中随机选取四家工业用户的日负荷数据聚类结果如下图 7 所示，相关参数如表 4 所示。

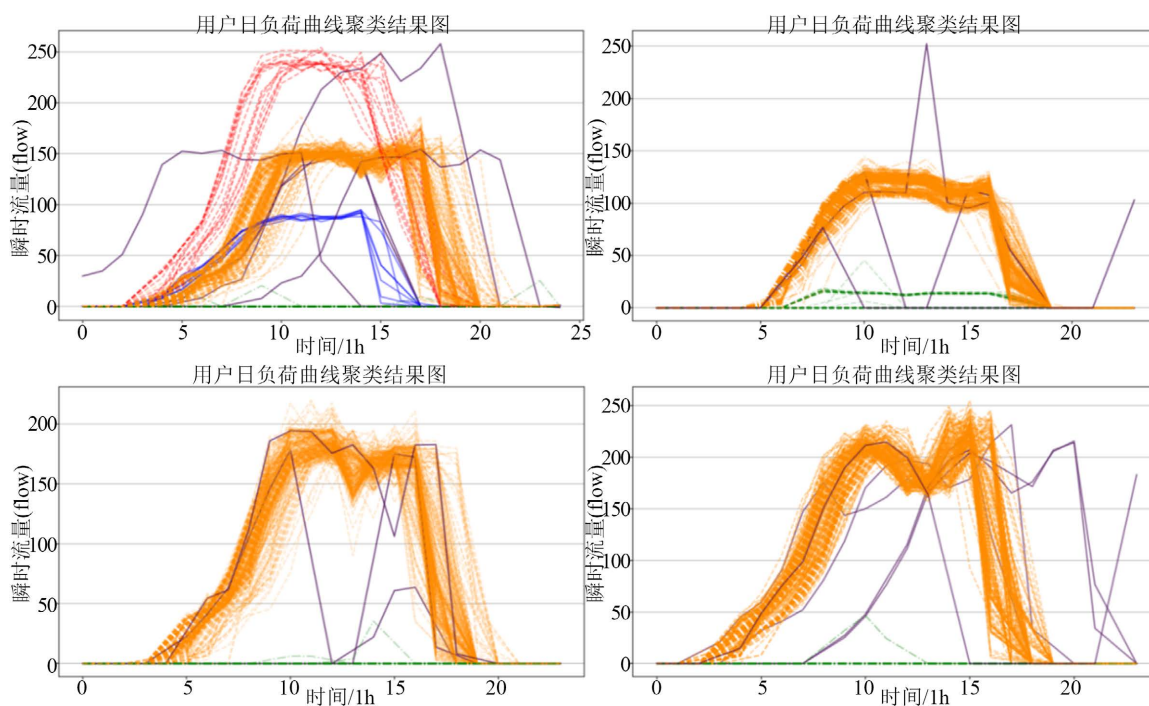


Figure 7. Clustering results of multi-user load curves
图 7. 多用户负荷曲线聚类结果图

Table 4. Parameter statistics table**表 4.** 参数统计表

用户	样本个数(天)	簇数(个)	$(\epsilon, \text{MinPTs})$	异常天数	检测准确率
A	465	3	(3,160)	11	98.06%
B	452	2	(2,60)	15	96.56%
C	469	2	(4,100)	8	98.93%
D	476	2	(5,100)	7	99.78%

对 325 家工业用气用户进行测试, 而且异常检测准确率达到 90% 以上, 其中四个案例用户的密度聚类结果如上图 7 所示, 其中红色、橙色、绿色、蓝色曲线为正常用气行为, 紫色曲线为异常用气行为; 密度聚类的参数通过改进的 K -近邻距离图和网格搜索法计算如表 4 中 $(\epsilon, \text{MinPTs})$ 所示, 用户 A 主要分为 3 簇, 3 种主要用气行为, 2 种稳定的正常用气行为, 1 种少量用气或几乎不用气行为, 簇外的紫色线段为检测出的异常用气行为, 检测准确率为 98.06%; 用户 B 主要分为 2 簇, 检测出的异常样本有 3 个, 准确率为 97.56%; 用户 C 主要分为 2 簇, 检测出的异常样本为 3 个, 异常样本检测准确率为 98.93%; 用户 D 主要分为 2 簇, 检测出的异常样本为 5 个, 异常样本检测准确率为 99.78%。

4. 结论

本文通过分析用户的日负荷曲线聚类效果的变化来识别异常曲线。对识别燃气用户用气异常提供了一些思路, 且在包含噪声的异常检测问题中 DBSCAN 算法具有较好的优势, 而且通过 K -近邻距离图和网格搜索法确定 DBSCAN 的各项参数提高了算法的可信度, 得到的参数更加精确。提升了模型的泛化性。为识别燃气用户用气异常提供了一些思路。

本文的改进算法在 325 家用户进行验证时发现, 基于改进密度聚类算法的异常检测准确率在 90% 以上, 说明利用密度算法进行异常检测具有一定的普适性。对智能化燃气行业异常检测有一些参考性, 尤其是对于规律用户模型的训练效果和测试效果良好。图 7 展示四个陶瓷工业用户的用气行为大致分为 3 类: 大规模用气行为(日用气量超过 1500 m^3)、中等用气行为以及停业小幅用气或几乎不用气行为。除此之外, 通过对 325 家陶瓷工业用户的异常曲线做分析发现, 日负荷曲线存在一定的规律, 相对于单一的用气负荷点来说一维曲线更能体现工业用户的用气特征。同时对于文中聚类结果的第三类停业小幅用气行为, 我们也不应该忽视, 因为天然气是一种危险气体, 如果发生泄露极易导致火灾甚至是爆炸等危险情况发生。

随着数字化产业的发展, 燃气公司走向智能化转型是必然趋势, 因此对于燃气行业的流量异常的检测有待更多的研究, 未来会考虑使用图像识别处理技术把异常曲线转化为图片, 通过大规模训练学习异常图片, 对正常用气图片和异常用气图片做二分类, 来实现异常用气行为检测。

基金项目

国家自然科学基金项目(11861020)、贵州省高层次留学人才创新创业择优资助重点项目([2018]03)、贵州省科技计划项目(ZK[2021]009)和贵州省青年科技人才成长项目([2018]121)。

参考文献

- [1] 赵然. 燃气涡轮流量计的在线监测及故障诊断实验研究[D]: [硕士学位论文]. 北京: 北京建筑大学, 2018.
- [2] 张彤, 徐晓钟, 王晓霞, 等. 基于改进 LMD 与 GRU 网络的短期天然气负荷预测[J]. 计算机系统应用, 2019, 28(6): 31-39.

- [3] 邓春宇, 吴克河, 谈元鹏, 胡杰. 基于多元时间序列分割聚类的异常值检测方法[J]. 计算机工程与设计, 2020, 41(11): 3123-3128.
- [4] 孙宇豪, 李国通, 张鹤. 距离相关系数融合 GPR 模型的卫星异常检测方法[J]. 北京航空航天大学学报, 2021, 47(4): 844-852.
- [5] 蹇诗婕, 卢志刚, 姜波, 等. 基于层次聚类方法的流量异常检测[J]. 信息安全研究, 2020, 6(6): 474-481.
- [6] 马国鑫. 时间序列流数据的建模预测和异常检测方法研究[D]: [硕士学位论文]. 西安: 西安理工大学, 2020.
- [7] 王婷, 王其兵, 燕争上, 等. 基于改进的快速密度峰值聚类的调控系统故障定位算法[J]. 太原理工大学学报, 2020, 51(6): 860-866.
- [8] 王得福. 基于人工智能的网络异常行为检测研究[D]: [硕士学位论文]. 北京: 北京邮电大学, 2020.
- [9] 钟勇, 林冬梅, 秦小麟. 一种基于查询密度聚类的异常检测算法[J]. 系统工程与电子技术, 2007(4): 640-646.
- [10] 何翼, 刘云. 基于机器学习的无线传感器网络异常值检测算法[J]. 电子技术与软件工程, 2021(2): 40-41.
- [11] 孙友强. 时间序列数据挖掘中的维数约简与预测方法研究[D]: [博士学位论文]. 合肥: 中国科学技术大学, 2014.
- [12] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
- [13] 秦佳奇, 许源源. 基于变分自编码器的异常负荷检测算法研究[J]. 电子技术与软件工程, 2021(7): 183-184.
- [14] 王桂兰, 周国亮, 赵洪山. 大规模用电数据流的快速聚类 and 异常检测技术[J]. 电力系统自动化, 2016, 40(24): 27-33.
- [15] 宋国明, 王厚军, 姜书艳, 刘红. 基于小波分析和分层决策的模拟电路故障识别方法[J]. 计算机应用研究, 2010, 27(6): 2057-2060+2064.
- [16] 黄悦华, 郭思涵, 鲍刚, 等. 基于用电特征分析的异常用电检测方法[J]. 三峡大学学报(自然科学版), 2021, 43(1): 96-101.
- [17] 李宁, 尹小明, 丁学峰, 蔡慧, 汪伟. 一种融合聚类和异常点检测算法的窃电辨识方法[J]. 电测与仪表, 2018, 55(21): 19-24.
- [18] 肖勇, 郑楷洪, 余忠忠, 等. 基于三次指数平滑模型与 DBSCAN 聚类的电量数据异常检测[J]. 电网技术, 2020, 44(3): 1099-1104.
- [19] Mahesh Kumar, K. and Rama Mohan Reddy, A. (2016) A Fast DBSCAN Clustering Algorithm by Accelerating Neighbor Searching Using Groups Method. *Pattern Recognition*, **58**, 39-48. <https://doi.org/10.1016/j.patcog.2016.03.008>
- [20] Livee, C. and Prateek, G. (2014) Improved DBSCAN Clustering Algorithm Using SR-Tree. *International Journal in IT & Engineering*, **2**, 2321-1776.
- [21] Anant, R., Sunita, J., Anand, S.J., et al. (2010) A Density Based Algorithm for Discovering Density Varied Clusters in Large Spatial Databases. *International Journal of Computer Applications*, **3**, 1-4. <https://doi.org/10.5120/739-1038>
- [22] Artur, S., Piotr, G. and Meng, J.E. (2020) A New Method for Automatic Determining of the DBSCAN Parameters. *Journal of Artificial Intelligence and Soft Computing Research*, **10**, 209-221. <https://doi.org/10.2478/jaiscr-2020-0014>
- [23] Zhao, Z., Yang, A., Guo, P., et al. (2020) A Density Clustering Algorithm for Simultaneous Modulation Format Identification and OSNR Estimation. *Applied Sciences*, **10**, 1095. <https://doi.org/10.3390/app10031095>
- [24] Chen, S., Liu, X., Ma, J., et al. (2019) Parameter Selection Algorithm of DBSCAN Based on K-Means Two Classification Algorithm. *The Journal of Engineering*, **2019**, 8676-8679. <https://doi.org/10.1049/joe.2018.9082>