

基于聚类分析的用电模式判别研究

冉 冉,陈 硕,刘 颖,李 钊

(国家电网辽宁电力公司,辽宁 沈阳 110000)

摘 要:研究用户的用电模型,根据用户的用电习惯,制定相应的调控方案对保证电网安全平稳运行以及对提高用户的服务质量具有重要意义。本文首先针对小数量的负载数据集,比较了不同聚类算法的聚类结果和算法效率,对聚类算法进行了初步筛选,进而比较了这些算法对大量的负载数据集的聚类结果和聚类效率,结果表明基于划分的分析算法对负载数据集有很好的聚类效果;采用划分算法,通过聚类分析,将用户用电行为模式分为八个类别,分别为双峰型、避风型、平稳型、单峰稳定型、单峰型、后峰型、波动型和前锋型;最后针对聚类分析结果,筛选出具有单峰型、单峰稳定型和双峰型并且用电量较大的用电客户为可调控用户,对该类型用户分别制定了不同的错峰调控方案,并比较了不同调控方案给用户带来的经济效益。

关键词:聚类分析;用电模式;错峰效益

文章编号:2096-4633(2019)04-0043-07 **中图分类号:**TM712 **文献标志码:**B

电网安全稳定运行是全社会生产作业的重要保障,电力公司的首要任务是确保能够提供稳定、安全的电力,随着全社会经济发展水平的提高以及用电设备的增多,企业和居民的电力需求越来越大,而电力供应却不能快速跟上发展所需,电力缺口呈现出逐渐扩大的趋势,从而导致了电力的供需不平衡。另外,随着社会的多元化,用户的用电习惯也呈现出较大的不一致性,用电结构正在发生变化,电力负荷特征表现为多种类型,其规律也变得更为复杂。由于各种因素的综合影响,目前电网峰谷差日益扩大、日负荷率较低,高峰时段供电紧张,负荷缺口大,常需要进行拉闸限电,低谷时段负荷无法充分利用。因此需要制定科学合理的负荷调控方案,以实现引导用户调整用电模式,转移高峰时段负荷,降低全网该时段内的用电量,缓解高峰时段的用电紧张情况^[1-2]。

传统的用电模式分析方法主要是按照电力客户的所属行业进行分析^[3-4],该方法需要对各个行业下的用户进行细化分类,随后以定性或定量的方法对用户的用电模式进行具体分析;通过分析不同行业用户的负荷特征可以得出各个行业用电负荷伴随时间的变化规律。按照行业分类进行分析的优点是在整体上对客户的用电特征进行把握,并且可以将该方法推广用于负荷结构不同的地区,但是同一行业下的用电模式一致性并不高,因此该模式的

精确度较差。

随着计算机技术的逐渐发展,近年已有学者尝试将新的技术和理论运用于用电模式识别^[5-7],如灰色关联算法,数据挖掘算法等,通过将以上技术进行综合分析以提高用电模式识别的准确度。同时,考虑到气象因素,经济发展水平,产业结构调整,需求侧管理措施等对用电模式的影响,因此,用电模式的识别需要多方面综合分析用户的用电特征,才能更加全面地掌握用户的用电规律,以更好地辅助电力公司制定科学的管理措施。伴随电网信息化建设水平的提升和智能电网建设的日益深入,电网公司积累了海量的数据,本次研究正是基于这样的背景开展的,可调控大用户用电负荷数据分析是供电企业了解用户的用户负荷模式特性的重要方法,研究用户的用电负荷模式特征,有助于用电企业更深刻地认识用户,并能够根据不同的用户群制定相应的市场策略、提供相应的个性化服务。

1 聚类分析

聚类分析模型是基于多元统计的新兴计算机技术,它是以无监督学习进行模式识别的重要方法,通过模型学习的规则将特征相似的样本归为一类,将特征差异较大的样本划分到不同的类别中,从而进行不同模式之间的界限划分,以达到用

电模式判别的目的。由于电力用户群体的逐步扩大,用户的用电模式也变得更加灵活,因此在制定需求侧的措施之前很难对每个电力用户一一去进行分析研究,通常较为可行的方法是通过聚类分析的算法,将用户和已经确认的具有典型特征的用电模式进行比较,以近似算法将不同负荷点和不同时段中用电负荷特征类似的用户归纳为同一类,并推理该用户的用电模式,从而辅助需求侧制定合理的措施。因此,聚类分析方法在一定程度上提高了工作效率。

可调控大用户用电行为分析在用户负荷特征分析的基础上,按照聚类分析算法建立模型,对用户的用电行为模式进行自动划分^[8-10]。用户用电负荷数据大多数是不平衡的时间序列数据。不平衡数据是指数据集中各个类别的成员数量差异明显,这将

提升聚类算法错误判断数据类别的几率^[11-13];时间序列数据会导致平方以上的高次方运算出现显著的截断误差,这将导致时间序列数据经过聚类算法运行得出的结果不稳定。随着数据集的不断增大,聚类结果愈来愈趋向不理想。

针对上述问题,应找出能对用户用电负荷数据聚类有一定效果的聚类算法出来。初步探索性聚类分析的思路是选取不同数量级的用户用电负荷数据集,使用不同的聚类分析算法,考察各个聚类算法在不同数量级下的用户用电负荷数据集的分类效果。

在本次实验中,选取数据集样本数量分别为 50、80、200、500、1 200 五个数据集。评判聚类算法优劣是以聚类运行结果展示图和运行时间两个角度衡量。选取的聚类算法如表 1 所示。

表 1 原始聚类分析算法集合

Tab. 1 Primitive clustering analysis algorithm set

类别	算法名称	类别	算法名称	类别	算法名称
基于划分	kmeans	基于层次	hclust(canberra 距离, ward 方法)	基于模型	FunCluster
	clara		hclust(canberra 距离, average 方法)		mocca
	pam		hclust(canberra 距离, complete 方法)		nnclust
	diana		hclust(Minkowski 距离, ward 方法)		Mclust
	agens		hclust(Minkowski 距离, average 方法)		clustconst
	CCFKMS		hclust(Minkowski 距离, complete 方法)	基于网格	apcluster
	bclustvs		hclust(Euler 距离, ward 方法)		distable
	hclustvar		hclust(Euler 距离, average 方法)		MixSim
	compHclust		hclust(Euler 距离, complete 方法)		cluster. test
	cmeans		cluster. Sim	基于密度	em
	cclust		hybridHclust		
	bootFlexclust		NbClust		
	kml				
	ewkm				
	fgkm				
	SKMEANS				

在数据集为 50 以及 80 条时,所有模型的运行速度都十分快,且聚类分析算法运行效果较佳,但是已经有部分聚类算法无法达到较好区分效果。

图 1 为利用主成分方法对用户用电行为聚类分析效果的可视化效果展示。图中,一个点代表一个用户,该点所在位置是由用户的用电行为决定的。

一个圈代表一个簇,可以看出,该批用户的用电行为较为明显地被划分为四个簇,并且其中三个簇有较好区分度。然而,该方法受到数据是不平衡数据的影响,有一个簇的区域较其他小,并与另一个簇有部分的重合,这也同时凸现出用户用电负荷数据的时序特性对聚类分析算法的影响。

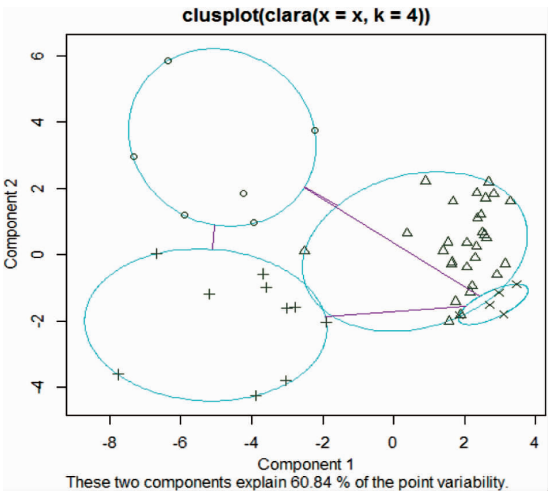


图1 样本量为 50 的聚类分析效果

Fig. 1 Cluster analysis effect with sample size of 50

经过用户用电负荷数据集样本为 50 和 80 的情况下聚类分析,剔除部分不适用于用户用电负荷的聚类分析算法后,分别以数据量为 200 条的用户用电负荷数据运行以下算法,表 2。

表 2 经过数据量为 50 的数据集测试后剩余聚类分析算法

Tab. 2 Residual clustering analysis algorithm after data set test with data volume of 50

类别	算法名称	类别	算法名称
基于划分	kmeans	基于层次	hclust(canberra 距离, ward 方法)
	clara		hclust(canberra 距离, average 方法)
	CCFKMS		hclust(Minkowski 距离, ward 方法)
	cmeans		cluster. Sim
	kml		hybridHclust
	ewkm		NbClust
	fgkm		cluster. test
基于模型	SKMEANS	基于密度	em
	FunCluster		apcluster

在数据集为 200 以及 500 条时,基于模型,基于密度,基于网格的聚类分析算法运行速度明显增长,特别是当数据集为 1 200 条时,基于贝叶斯先验信息的 bayesclust 聚类分析算法运行 6 小时仍不能运行完毕;同时,当数据集显著增大,异常数据也明显增多,数据的不平衡性显现,部分聚类分析算法抗噪声能力不强,不平衡性聚类效果差,导致分类效果不佳。

以 clara 聚类分析算法,样本容量 200 的聚类结

果为例,基于主成分的聚类分析效果可视化图如图 2 所示,由于受到数据的不平衡性以及时序特性的影响,clara 聚类分析算法的运行效果,虽然用户用电行为还能较明显的区分为 4 簇,但是各簇之间有较多重合,区分效果明显不如样本量为 50 的时候佳。

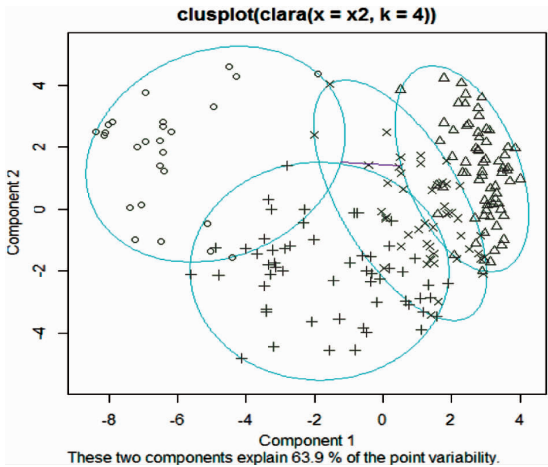


图2 样本量为 200 的 clara 聚类算法运行效果

Fig. 2 Running effect of clara clustering algorithm with sample size of 200

通过查看聚类算法的各个簇的区域重叠效果图(图 3),明显出现有两个聚类区域重叠的现象,并且位于重叠区域的数据点不少于总体样本数量的四分之一,模式识别效果并不理想。因此 clara 聚类分析算法在数据量小时运行效果良好,但是当数据量显著增大时,聚类运行效果不佳,clara 聚类分析算法失效。这说明 clara 并不适用于数据量大的用户用电负荷数据分析。

以基于层次的 hclust (canberra 距离, ward 方法)聚类分析算法的运行效果图如图 4 所示。图中最底一层的节点代表 200 个用户的用电行为,通过计算两个用户用电行为间距离,将距离最近的两个用电类别进行合并,形成新的节点。新的节点再与离该节点距离最近的节点合并,如此逐层合并,直至合并为一个节点为止,这就形成了一棵“树”状的用电行为关系图。从下图可以看出,基于“canberra”距离,“ward”方法生成的聚类“树”,易于分割,且各主干间较为“聚集”,各数据点较为平均地归入各自的主干。其聚类效果明显优于基于“Euler”距离,“average”方法生成的层次“树”。于是,可以得出结论,基于“canberra”距离,“ward”方法的层次算法适用于用户用电负荷数据。

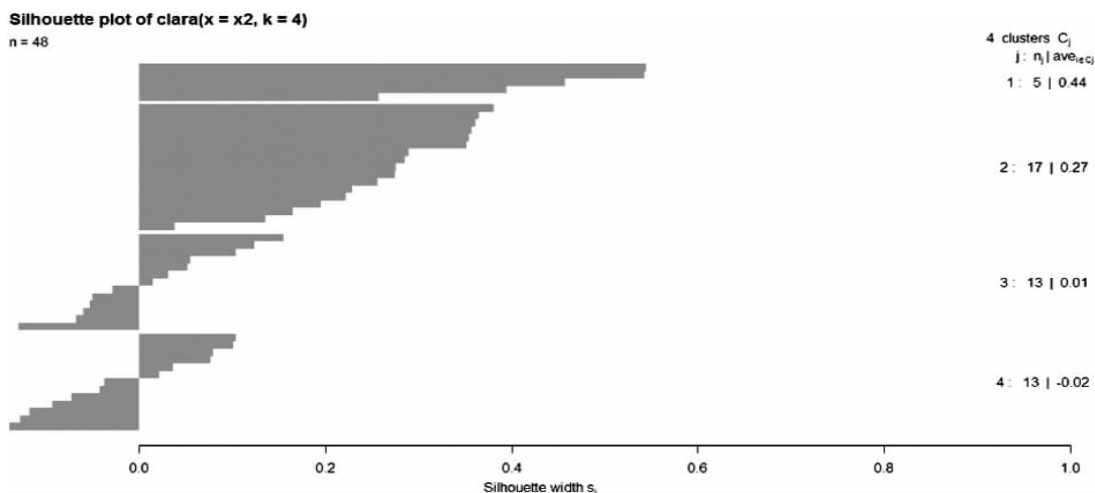


图 3 样本量为 200 的 clara 聚类算法类区域重叠图

Fig. 3 Overlap graph of clara clustering algorithm class area with sample size 200

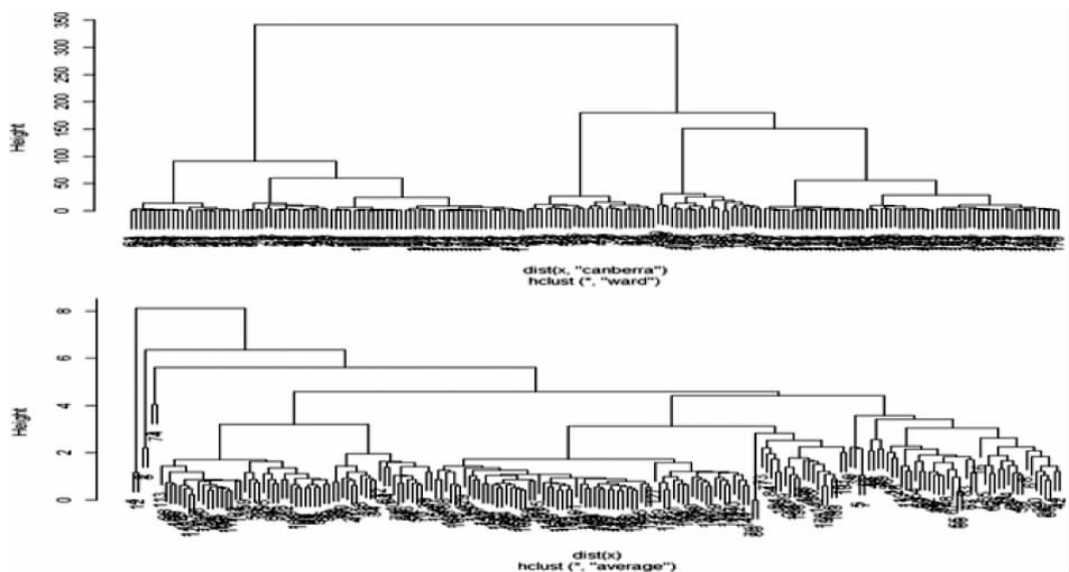


图 4 用户用电负荷数据生成层次“树”

Fig. 4 Hierarchical "tree" of users' electrical load data generation

最后,运行数据量为 1 000 以上的用户用电负荷数据后,以下算法较能适应用户用电负荷的时序特性以及不平衡性(表 3)。

表 3 经过数据量为 200 的数据集测试后
剩余聚类分析算法

Tab. 3 Residual clustering analysis algorithm after data
set test with data volume of 200

类别	算法名称
基于划分	CCFKMS
	cmeans
	SKMEANS
基于层次	hclust(canberra 距离, ward 方法)
	hybridHclust

上述五种聚类分析算法较能适应用户用电负荷数据的特性,比其他算法运行时间短,分类的效果也较佳。但是,当数据集内数据更换,数据容量进一步扩大,上述算法的分类效果也差强人意,仅仅通过调整参数对聚类结果进行修正对结果的影响并不显著。并且,在聚类分析过程中,需要大量的人工参与其中。

在此基础上,由于基于划分的算法的运行速度较基于层次的算法更快,特别是在数据量较大时,基于划分的算法优势更加明显。其次,基于层级的算法采用逐层合并的方式形成聚类树,若期间有一两个数据合并错误,这将可能导致整个类错误。因此,选定基于划分的聚类分析算法作为

聚类成员生成算法集(表4)。

表4 生成聚类成员聚类分析算法集合

Tab.4 Clustering analysis algorithm set for generate cluster member

类别	算法名称
基于划分	CCFKMS
	cmeans
	SKMEANS

2 模型结果分析

根据选定的聚类分析算法对负荷数据进行分析,将用户用电模式划分为8种典型模式,分别为双峰型、避风型、平稳型、单峰稳定型、单峰型、后峰型、波动型和前锋型,如图5所示。

双峰型负荷特征:该类型用电特征表现为一天24小时内有两个用电高峰,两个峰值发生时间点在10时和16时附近,有时也存在一定的时间差偏移,两个峰值的大小不一致,且两个用电高峰时间长度和最大峰值大小也不一致。具有该类型的用电行为一般是政府、学校、写字楼等性质的单位。

避风型负荷特征:该类型用电特征表现为在白天7-18时为低负荷特征,用电量非常小,在其余的时段为高负荷特征。具有该用电特征的用户主要是一些在全网负荷较低的时段进行生产用电的专变用户。

平稳型负荷特征:平稳型负荷特征表现为在一

天24小时内负荷较为稳定,没有较为明显的用电高峰和用电低谷。具有平稳型用电负荷特征的用户负荷率较高,该类型用户一般是处于连续生产、用电状态,该类型用户一般属于优保企业。

单峰稳定型负荷特征:该类型用电特征总体表现为从6时左右开始负荷曲线逐步上升,到10时左右负荷达到高峰,并且随后高峰期表现较为稳定,为持续的高负荷特征,到21时左右负荷下降到用电低谷期。具有单峰稳定型负荷特征的用户多属于生产型企业,该类型企业在白天进行生产,且生产时间较长,在进行错峰用电时,应适当降低错峰用电比例,以保证企业的正常运营。

单峰型负荷特征:该类型用户的用电模式特征表现为在上午10时左右出现一个用电高峰,不同于单峰稳定型的是该类型负荷特征的用电高峰持续时间较短,为尖峰特征。

后峰型负荷特征:该类型用户的用电模式特征在负荷曲线上表现为在17时左右开始出现用电高峰,用电高峰通常都在一天内的最后一段时间,在其余的时间,从0时到17时负荷均较低。

波动型负荷特征:该类型特征表现为在一天内负荷持续波动,为不稳定型,该类型负荷没有规律的用电高峰和用电低谷。

前锋型负荷特征:该类型特征表现为在0时到7时为用电高峰,其余时间用电负荷较低,为前高后低的用电类型。该类型用户多为高耗能生产企业,其用电时段主要集中在全网的用电低谷期。

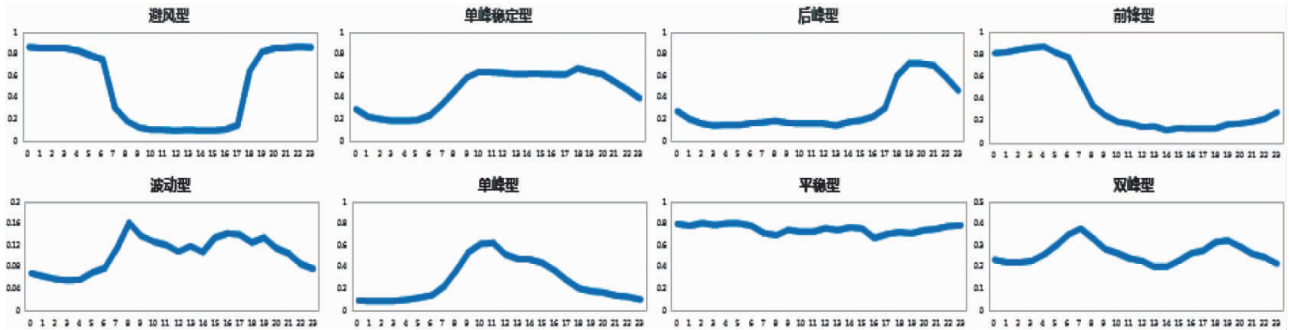


图5 典型用电行为特征

Fig.5 Characteristics of typical electrical behavior

3 错峰调控分析

在对用户用电模式的特征进行挖掘分析后,筛选出可调控的用户,主要针对具有单峰型、单峰稳定型和双峰型并且用电量较大的用电客户,具有这种

特征的用户主要采用一班制、二班制生产制度,同时因为其具有较大的用电量,相比于小用电量的用户,其在进行实际错峰时可以发挥较大的作用,以各用户聚类后的典型负荷曲线来代表各用户的负荷特征,分别计算每个用户的总电量^[14]。在业务基础

上,制定不同等级的错峰方案,分别按照企业用电总量的5%、10%、15%、20%和25%等五个等级进行错峰效益分析。在本方案中,主要考虑将峰期负荷错峰移至谷期,经过错峰,在每个用户总用电量不变的情况下,能够将部分的负荷推移至全网负荷的谷期,有助于减小全网负荷峰谷差,提高全网负荷的稳定。同时在实行峰谷电价策略的基础上,企业将生产运行时间部分调至用电低谷期,从大用电量企业自身的角度进行了一定程度的成本节约,能够给用户带来实际的经济效益^[15-16]。

4 结论

(1)本文探索了适用于大规模数据集的用电行为模式判别算法,从不同聚类算法的聚类运行结果图和运行时间两方面对各算法进行评估优选,最终选择基于划分的算法CCFKMS、CMEANS和SKMEANS作为聚类成员生成算法集。

(2)通过聚类分析,将用户用电行为分为8类,分别为双峰型、避风型、平稳型、单峰稳定型、单峰型、后峰型、波动型和前峰型。

(3)针对聚类分析结果,筛选具有错峰调控可行性的用户,对该类型用户分别制定不同的调控方案,并比较不同方案产生的效益,以供企业选择自主选择。

参考文献:

- [1] 李鹏, 窦鹏冲, 李雨薇, 等. 微电网技术在主动配电网中的应用[J]. 电力自动化设备, 2015, 35(04): 8-16.
LI Peng, DOU Pengchong, LI Yuwei, et al. Application of microgrid technology in active distribution network [J]. Electric Power Automation Equipment, 2015, 35(04): 8-16.
- [2] 王璨, 冯勤超. 基于价值评价的电力用户分类研究[J]. 价值工程, 2009, 28(05): 64-67.
WANG Can, FENG Qinchao. The research of power customers classification based on value assessment [J]. Value Engineering, 2009, 28(05): 64-67.
- [3] 彭显刚, 赖家文, 陈奕. 基于聚类分析的客户用电模式智能识别方法[J]. 电力系统保护与控制, 2014, 42(19): 68-73.
PENG Xiangang, LAI Jiawen, CHEN Yi. Application of clustering analysis in typical power consumption profile analysis [J]. Power System Protection and Control, 2014, 42(19): 68-73.
- [4] 苏适, 李康平, 严玉廷, 等. 基于密度空间聚类和引力搜索算法的居民负荷用电模式分类模型[J]. 电力自动化设备, 2018, 38(01): 129-136.
SU Shi, LI Kangping, YAN Yuting, et al. Classification model of

- residential power consumption mode based on DBSCAN and gravitational search algorithm [J]. Electric Power Automation Equipment, 2018, 38(01): 129-136.
- [5] 苟港益. 基于聚类分析与随机森林的短期负荷滚动预测[J]. 智能城市, 2018, 4(09): 9-11.
XUN Gangyi. A short-term load rolling prediction model based on cluster analysis and random forest (K-RF) [J]. Intelligent City, 2018, 4(09): 9-11.
- [6] 王炳鑫, 侯岩, 方红旺, 等. 面向“削峰填谷”的电力客户用电行为分析[J]. 电信科学, 2017, 33(05): 164-170.
WANG Bingxin, HOU Yan, FANG Hongwang, et al. Analysis of customers' electricity consumption behavior for peak load shifting [J]. Telecommunications Science, 2017, 33(05): 164-170.
- [7] 王堃, 杨飞, 李斌. 用电采集大数据的用户用电行为分析研究[J]. 贵州电力技术, 2017, 20(10): 1-6.
WANG Kun, YANG Fei, LI Bin. Analysis and research of user electricity consumption behavior based on big data collected by electricity [J]. Guizhou Electric Power Technology, 2017, 20(10): 1-6.
- [8] 陈伟, 周峰, 韩新阳, 等. 国家电网负荷特性分析研究[J]. 电力技术经济, 2008, 20(04): 25-29+39.
CHEN Wei, ZHOU Feng, HAN Xinyang, et al. Analysis on load characteristics of state grid [J]. Electric Power Technologic Economics, 2008, 20(04): 25-29+39.
- [9] 朱文俊, 王毅, 罗敏, 等. 面向海量用户用电特性感知的分布式聚类算法[J]. 电力系统自动化, 2016, 1(12): 21-27.
ZHU Wenjun, WANG Yi, LUO Min, et al. Distributed clustering algorithm for awareness of electricity consumption characteristics of massive consumers [J]. Automation of Electric Power Systems, 2016, 1(12): 21-27.
- [10] 赵腾, 王林童, 张焰, 等. 采用互信息与随机森林算法的用户用电关联因素辨识及用电量预测方法[J]. 中国电机工程报, 2016, 3(36): 604-614.
ZHAO Teng, WANG Lintong, ZHANG Yan, et al. Relation factor identification of electricity consumption behavior of users and electricity demand forecasting based on mutual information and random forests [J]. Proceedings of the CSEE, 2016, 3(36): 604-614.
- [11] 王扬, 吴凡, 姚宗强, 等. 基于正则化矩阵分解的用户用电行为分析[J]. 计算机应用, 2017, 8(37): 2405-2409.
WANG Yang, WU Fan, YAO Zongqiang, et al. Residential electricity consumption analysis based on regularized matrix factorization [J]. Journal of Computer Applications, 2017, 8(37): 2405-2409.
- [12] 张铁峰, 顾明迪. 电力用户负荷模式提取技术及应用综述[J]. 电网技术, 2016, 40(03): 804-811.
ZHANG Tiefeng, GU Mingdi. Overview of electricity customer load pattern extraction technology and its application [J]. Power System Technology, 2016, 40(03): 804-811.
- [13] LIU W, NIU S, XU H. Optimal planning of battery energy storage

considering reliability benefit and operation strategy in active distribution system [J]. Journal of Modern Power Systems and Clean Energy,2017,5(02):177-186.

[14] SHI B,JING Z,YU H E, et al. Electricity consumption behavior analysis of single-phase power consumers in distribution network based on dynamic game theory[J]. Automation of Electric Power Systems,2017,41(14):87-91+139.

[15] 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.
CAI Qiuna, LIU Sijie, LU Qiuyu. Identification method for user industry classification based on GMM clustering and SVM [J]. Guangdong Electric Power,2017,30(12):91-96.

[16] 崔立卿,贺伟军,田晶,等. 基于 K 均值聚类算法的大客户用

电行为分析[J]浙江电力. 2017,36(12):47-52.

CUI Liqing, HE Weijun, TIAN Jing, et al. Analysis on power consumption behavior of large customers based on K-means clustering algorithm[J]. Zhejiang Electric Power,2017,36(12):47-52.

收稿日期:2019-01-13

作者简介:



冉 冉(1980),女,硕士,高级工程师。主要从事电力信息化项目管理及大数据研究等工作。

(本文责任编辑:范 斌)

Research of electricity consumption pattern recognition
based on cluster analysis

RAN Ran, CHEN Shuo, LIU Ying, LI Zhao

(State Grid Liaoning Electric Power Co., Ltd., Shenyang 110000 Liaoning, China)

Abstract: It is of great significance to study the user's electricity consumption model and formulate corresponding control schemes according to the user's electricity consumption habits to ensure the safe and stable operation of power grid and improve the service quality of users. Firstly, the clustering results and efficiency of different clustering algorithms are compared for a small number of load data sets, and the clustering algorithms are initially screened. and then the clustering results and efficiency of these algorithms are compared for a large number of load data sets. The results show that the algorithm based on partitioning has a good clustering effect on load data sets. By using partitioning algorithm and clustering analysis, the user behavior patterns are divided into eight categories, which are bimodal, Shelter type, stationary type, single peak stable type, unimodal type, rear peak type, wave type and forward type; finally, for cluster analysis results, screened with unimodal type, single peak stable type and bimodal type and used electricity A large number of power customers are controllable users, and different types of peak-to-peak control schemes have been developed for this type of users, and the economic benefits brought by different control schemes to users have been compared.

Key words: cluster analysis; effectelectricity consumption mode; peak load shifting effect