



面向“削峰填谷”的电力客户用电行为分析

王炳鑫¹, 侯岩², 方红旺², 陈雨泽², 刘建²

(1. 国网福建省电力有限公司泉州供电公司, 福建 泉州 362000;

2. 北京中电普华信息技术有限公司, 北京 100085)

摘要: 为对海量电力客户实施有针对性的“削峰填谷”措施, 提出了一种面向“削峰填谷”的海量电力客户用电行为分析方法。首先, 利用聚类算法对国网某省公司主网一年的日负荷曲线数据进行聚类分析, 得到不同时期主网的负荷特征。然后, 分别对每个时期下所有电力客户的日负荷曲线数据进行聚类分析, 得到不同主网特征下用户群体的负荷特征, 对比主网和用户的负荷特征得到用户群体的“削峰填谷”模式。最后, 利用动态时间规整算法将未来日期与历史日期进行匹配, 得到未来日期用户群体的“削峰填谷”模式。实证研究表明, 分析结果可以为有序用电、峰谷电价等公司决策提供更有针对性的参考依据, 以更进一步实现配电网负荷的“削峰填谷”和平稳运行。

关键词: 用电行为分析; 削峰填谷; 聚类分析; 动态时间规整

中图分类号: TM732

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017067

Analysis of customers' electricity consumption behavior for peak load shifting

WANG Bingxin¹, HOU Yan², FANG Hongwang², CHEN Yuze², LIU Jian²

1. State Grid Fujian Electric Power Company, Quanzhou Electric Power Supply Company, Quanzhou 362000, China

2. Beijing China-Power Information Technology Co., Ltd., Beijing 100085, China

Abstract: In order to implement well-directed peak load shifting for massive customers, a method for analyzing massive customers' electricity consumption behavior for peak load shifting was proposed. Firstly, clustering algorithm was used to cluster daily load curves of the main power grid in the previous year and get load characteristics in different dates. Then the load curves of all customers under every date cluster were clustered, the peak load shifting method was derived by comparing the load characteristics of the main power grid and customers. Finally, the peak load shifting method in a future day was given by date matching between the future day with a historical day using dynamic time warping (DTW). Empirical study shows that the method is conducive to peak-valley electricity pricing and orderly electricity consumption and can further achieve peak load shifting and stable operation of the main power grid.

Key words: electricity consumption behavior analysis, peak load shifting, clustering analysis, dynamic time warping

收稿日期: 2017-01-25; 修回日期: 2017-03-06

基金项目: 国网信息通信产业集团科技项目“支持售电业务的大数据应用”(No.52680016007G)

Foundation Item: Science & Technology Project of State Grid Information and Telecommunication Group “Big Data Applications for Supporting Electricity Sales Business over Grid”(No.52680016007G)

1 引言

随着电力信息化的深化应用,电力信息数据呈现爆炸式增长趋势,电力行业的大数据时代已经到来。传统的在小规模数据量下对客户用电行为进行统计分析,具备一定的指导价值,但在海量数据情况下的应用效果不是很理想。与传统的客户用电行为分析相比,基于大数据的客户用电行为分析更注重对客户用电价值的挖掘,可以实现对海量客户用电行为的定量分析,提高客户行为定位的准确度,为更有效地开展客户服务、提高客户满意度、降低运营风险提供决策参考。

近年来,已有学者对电力客户的用电行为进行了研究,主要的研究思路有以下几种。

(1) 基于专家经验的方法

根据业务专家经验进行用电行为分析。何永秀等人^[1]通过对居民智能用电情况进行问卷调查,采用模糊综合评价对居民智能用电的态度进行量化分析,该方法采用的算法简单易理解、数据采集容易,但需要凭经验设定权重,主观性强。黄瀚^[2]根据业务经验探讨了智能电网下典型用电客户的用电行为,该方法结合了大量的行业经验,但没有相应的数据支撑。

(2) 基于统计分析的方法

利用单个指标或组合指标分析用电行为。如Chuan L等人^[3]研究了新加坡不同居民住宅电器的用电量分布,利用不同居民住宅中的电器设备信息和用电信息,构建了月度用电量计算式,得到了不同居民住宅类型电器的用电量。Abreua J M等人^[4]基于实验中获取的智能家电数据,发现和识别了家庭用电的行为模式。这些方法相对简单,考虑了各种用电设备的用电信息,缺点是用电设备信息采集较为困难,不适用于国内智能家电数据难以获取的情况。

(3) 基于无监督学习的方法

主要是指通过聚类算法实现对用电数据的划

分,从而得到不同的用电特征。宗柳等人^[5]利用聚类对用户的用电特征进行了精细化挖掘,但聚类方法的计算复杂度较高。瞿海妮等人^[6]从居民的绝对用电量入手,将居民按日均用电量分为异常用电、低耗电、普通和高耗电用户4类,分析了用户行为的特征。张素香等人^[7]基于云计算平台和并行K-means聚类算法,建立了峰时耗电率、负荷率、谷电系数等用电特征,分析了智能小区的居民用电行为。该方法采取了并行算法,适用于海量数据下的用电行为分析,但聚类维度存在冗余且不完备。

本文在研究电力业务的基础上,提出了一种面向“削峰填谷”的海量电力客户用电行为分析方法,对福建省公司的主网和客户负荷数据进行的分析结果表明,该方法能够为有序用电、峰谷电价等公司决策提供更有针对性的参考依据,以更进一步实现配电网负荷的“削峰填谷”和平稳运行。

2 相关算法

2.1 EM 聚类算法

最大期望算法(expectation maximization algorithm)^[8,9]是一种迭代算法,用于含有隐变量的概率参数模型的最大似然估计或极大后验概率估计。EM算法主要有两个步骤。

(1) E步骤

根据式(1)计算每个样本属于不同分布的概率,选择概率最大的分布作为样本类别。

$$r_{jk} = \frac{\phi(y_j | \mu_k, \sigma_k)}{\sum_{k=1}^2 \phi(y_i | \mu_k, \sigma_k)} \quad (1)$$

(2) M步骤

根据式(2)估计每个类别分布函数的概率参数。

$$\mu_k = \frac{\sum_{j=1}^N r_{jk} y_j}{\sum_{j=1}^N r_{jk}}, \sigma_k^2 = \frac{\sum_{j=1}^N r_{jk} (y_j - \mu_k)^2}{\sum_{j=1}^N r_{jk}} \quad (2)$$



其中, y_j 为第 j 个样本数据, μ_k 、 σ_k 分别为第 k 个分布的均值、标准差, $\bar{r}_{jk}$ 为第 j 个样本隶属于第 k 个分布的概率。通过迭代使用这两个步骤, EM 算法逐步改进模型的参数, 使参数和训练样本的似然概率逐渐增大, 最后终止于一个极大点。EM 算法在依赖于不可观察的隐变量的概率模型中被用于寻找概率模型参数的最大似然估计, 在 EM 聚类算法中, 隐变量就是数据项所属的类。

2.2 K-means 聚类

K-means 算法^[10,11]是一种被广泛应用于科学研究和工业应用的经典聚类算法, 核心思想是把数据点划分为不同类别, 使得每个类别中的数据点到该聚类中心距离的平方和最小。传统 K-means 算法只适用于集中式数据的聚类。随着网络技术的发展, 大量的数据是分布式存储的, 为对分布式存储的数据进行聚类, 李小武等人^[12]在分布式环境下对传统聚类算法进行了扩展和改进, 提出了分布式 K-means 聚类方法, 其基本思想是: 首先每个节点对本节点的数据集进行局部聚类分析, 然后汇总各节点聚类结果得到全局聚类结果。

2.3 动态时间规整

动态时间规整(dynamic time warping, DTW)是由日本学者 Itakura^[13,14]提出的, 是一种衡量两个长度不同的时间序列 $X = (x_1, \dots, x_i, \dots, x_m)$ 和 $Y = (y_1, \dots, y_i, \dots, y_n)$ 相似度的方法。通过把时间序列进行延伸和缩短, 计算两个时间序列之间的相似性。DTW 是一个典型的优化问题, 求解两模板匹配时累计距离最小所对应的规整函数 $D(i, j)$, 它表示序列 X 上点 i 和 Y 上的点 j 之间的规整路径距离:

$$D(i, j) = \text{Dist}(i, j) + \min\{D(i-1, j), D(i, j-1), D(i-1, j-1)\} \quad (3)$$

其中, $\text{Dist}(i, j)$ 为 X 上点 i 和 Y 上的点 j 之间的欧式距离。

3 用电行为分析方法

本文提出的用电行为分析方法是了解决海量电力客户的“削峰填谷”问题, 因此需要分析电力客户的用电行为特征与主网负荷特征的匹配

关系。首先基于 EM 算法对主网历史 1 年的日负荷曲线数据进行聚类分析, 得到不同时期主网的负荷特征。再分别利用分布式 K-means 算法对每一时期的海量电力客户的日负荷曲线数据进行聚类分析, 得到不同主网特征对应电力客户的负荷特征。对比主网负荷特征曲线与客户负荷特征曲线可以得到电力客户的用电行为类型, 如迎峰型(用户负荷特征曲线与主网负荷特征曲线走势相同, 应该采取“削峰填谷”措施)、逆峰型(用户负荷特征曲线与主网负荷特征曲线走势相反, 应该鼓励客户保持用电习惯)等, 从而建立主网和电力客户群体的“削峰填谷”匹配模式。最后利用 DTW 对未来目标日和历史日进行日期匹配, 分析未来目标日主网与用电客户的“削峰填谷”匹配关系, 以确定电力客户群体在目标日的“削峰填谷”模式, 进而对不同模式的客户群体采取不同的“削峰填谷”措施。

EM、K-means 等划分型聚类算法初始化时存在聚类个数需要人为确定的问题, 一般而言可以通过业务经验对聚类个数进行设定。本文根据聚类结果类内距离最小化和类间距离最大化两个准则, 提出了最优聚类评价式, 实现了最佳聚类个数的自动识别:

$$POC = \frac{\sum_{i=1}^k \sum_{m>i}^k \|V_i - V_m\|^2}{C_k^2 \sum_{i=1}^k \left(\frac{1}{n_i} \sum_{j=1}^{n_i} \|x_{ij} - V_i\|^2 \right)} \quad (4)$$

其中, k 表示聚类个数, V_i 、 V_m 分别为第 i 类和第 m 类的类中心, n_i 表示第 i 类的样本数量, x_{ij} 表示第 i 类的第 j 个样本, $\|\cdot\|$ 为欧式距离。指定 k 的最大值和最小值(如 $k \in [1, 5]$), 对数据集进行聚类, 然后根据式(4)计算 POC 值, POC 值最大的为最优聚类。

(1) 主网负荷特征分析

某区域历史 1 年的主网负荷曲线数据见表 1, 利用 EM 聚类算法将主网的日负荷曲线数据分成不同群体, 并利用最优聚类评价公式自动选择最佳聚类结果。通过聚类分析得到主网在不同日期

群体的负荷特征, 每个群体内的日期具备相似的负荷曲线特征, 而不同群体之间的日期的负荷曲线存在比较大的差异。

表 1 某区域主网的日负荷曲线数据

日期	t_1 /万千瓦时	t_2 /万千瓦时	...	t_n /万千瓦时
day1	370	410	...	390
day2	378	385	...	400
...	...	...	...	...
day365	390	340	...	413

(2) 电力客户群体用电行为分析

针对每一类主网的日期群体, 利用并行 *K-means* 聚类算法对该日期群体下所有日期的所有电力客户的日负荷曲线数据进行聚类分析, 电力客户的日负荷曲线数据与表 1 所示数据格式相同, 得到不同主网类别下不同电力客户群体的负荷特征。通过比较主网负荷特征曲线和电力客户群体的负荷特征曲线, 得到具有不同“削峰填谷”模式的电力客户群体, 如迎峰型、逆峰型等。

(3) 日期匹配

上述主网及客户负荷的聚类分析是对电力客户群体的历史用电行为进行分析, 如果要得到未来某一天客户群体的“削峰填谷”模式, 还需要进行未来目标日期与历史日期的匹配, 然后将历史日期电力客户群体的“削峰填谷”模式作为未来目标日期的模式, 进而得到目标日期需要进行“削峰填谷”的客户群体。

本文利用动态时间规整算法, 根据天气、假日等特征将目标日与历史日期进行匹配, 得到与目标日最相似的历史日期。将历史和未来日期划分为 3 个日期集合: 节假日、周末以及工作日。当目标日为节假日时, 直接用历史相同节假日作为目标日的历史匹配日期。当目标日为周末或者工作日时, 利用动态时间规整算法根据温度分别在去年同期前后两个月的周末集合和工作日集合中寻找历史相似日, 得到目标日的历史匹配日期。基于日期匹配方法, 每个目标日都可以找到与之相似的历史日期, 进而得到不同客户群体在目标日期的“削峰填谷”模式。若客户群体在目标日的“削峰填谷”模式为逆峰型, 此类客户有利于主网安全运行, 不是开展“削峰填谷”措施的客户; 若客户群体在目标日的模式为迎峰型, 则此类客户是开展“削峰填谷”措施的重点客户。

4 实验结果

本文采用福建省电力公司主网和福建全部用电客户 2014 年 1 月 1 日至 12 月 31 日的负荷数据, 采样间隔为 15 min, 每条日负荷曲线有 96 个采样点。按本文提出的用电行为分析方法, 首先对 2014 年福建省电力公司主网的日负荷曲线数据利用 EM 算法进行聚类分析, 本文将主网的日负荷数据划分为 3 类, 3 个类别的负荷特征曲线如图 1 所示, 从特征曲线可以看出, 每个类别的特征曲线都呈现白天负荷高、晚上负荷低的特征, 但峰

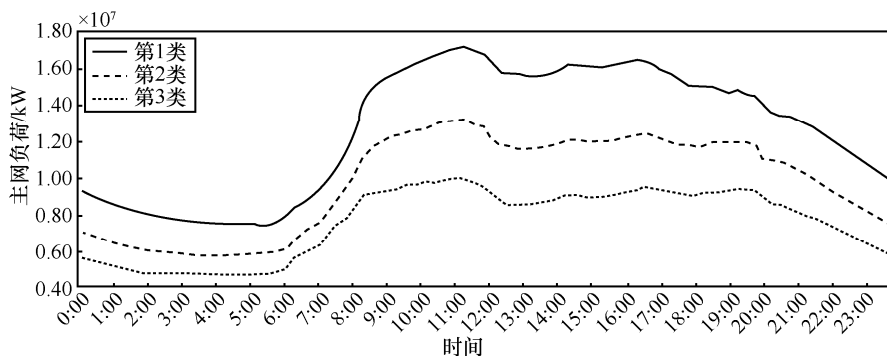


图 1 主网负荷聚类特征曲线



值大小有所差异。按季节和节假日对聚类结果进行统计分析的结果见表2,可以看出负荷特征曲线峰值最高的第1类包括58.7%的夏季日期和52.22%的冬季日期,这些日期由于夏季高温和冬季低温导致负荷较高,因此该类可以称为夏冬高峰型。而对于峰值最低的第3类,从表2看出该类包括了大部分的节假日、春季日期和51.65%的秋季日期,该类日期的负荷由于放假及温度较温

和等原因较低,因此该类可以称为假日低谷型。

针对每一类主网日负荷类别,利用并行K-means聚类算法对该类对应日期的所有客户负荷曲线数据进行聚类分析。本文将3类主网类别对应的电力客户分别划分为4类、3类和5类。图2、图3和图4分别显示了每类主网对应电力客户类别的负荷特征曲线,可以看出,3个主网类别下的第1类客户的负荷特征曲线都分别与对应

表2 主网聚类结果节假日和季节特征

类别	类别名称	假日占比	春季占比	夏季占比	秋季占比	冬季占比
1	夏冬高峰型	0	0	58.70%	6.59%	52.22%
2	普通型	18.52%	15.22%	40.22%	41.76%	33.33%
3	假日低谷型	81.48%	84.78%	1.08%	51.65%	14.45%

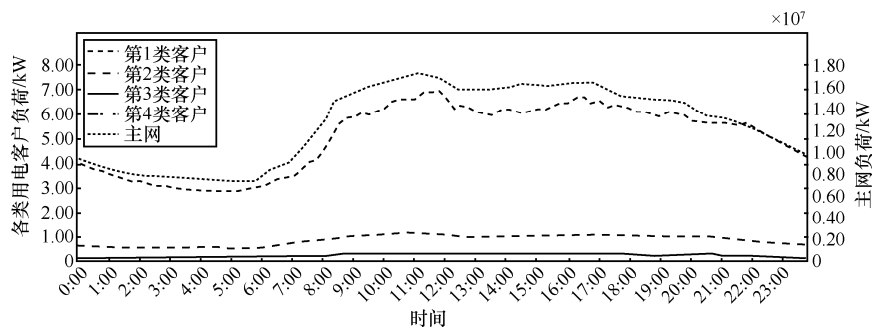


图2 第1类主网类别下电力客户负荷聚类结果

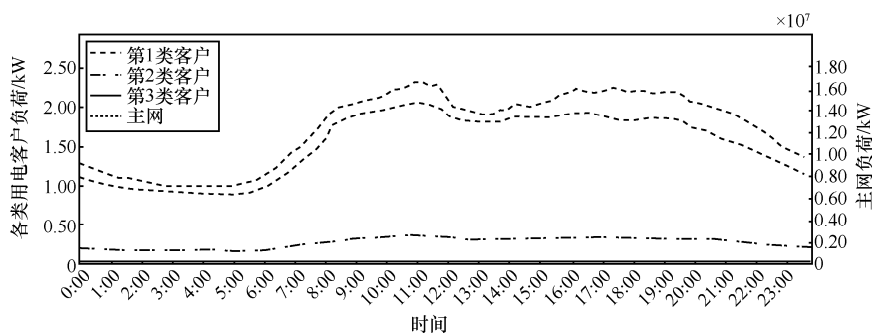


图3 第2类主网类别下电力客户负荷聚类结果

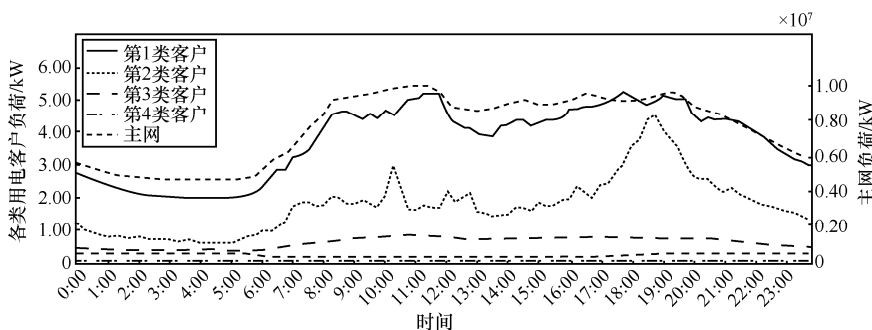


图4 第3类主网类别下电力客户负荷聚类结果

的主网负荷特征曲线非常相似,即这3类电力客户的用电行为类型为迎峰型,这3个客户群体为需要进行“削峰填谷”的群体。

通过对电力客户用电行为的分析,得到了客户在不同历史时期的“削峰填谷”模式。为了得到未来日期电力客户的“削峰填谷”模式,利用本文提出的日期匹配方法得到未来日期的历史相似日。图5给出了2015年5月1-10日的日期匹配结果,可以看出劳动节、工作日、周末分别对应2014年的劳动节、工作日和周末。

从图5可以看出,2015年5月10日的历史匹配日期为2014年5月25日,即可利用2014年5月25日客户群体的“削峰填谷”模式作为目标日期的2015年5月10日的模式。图6和图7分别为

2014年5月25日两类客户群体和主网的特征曲线对比。从图6可以看出,该类电力客户的特征曲线与主网特征曲线的走势相反,因此预计该类客户在2015年5月10日为逆峰用电模式,对该类客户不需采取“削峰填谷”措施,鼓励其保持用电习惯即可。而对于图7,该类电力客户的特征曲线与主网特征曲线的走势几乎相同,因此预计该类客户在2015年5月10日为迎峰用电模式,为“削峰填谷”的目标客户,可采取有序用电、峰谷电价等措施进行“削峰填谷”。

5 结束语

本文提出了一种面向“削峰填谷”的海量电力客户用电行为分析方法。首先,对历史主网负

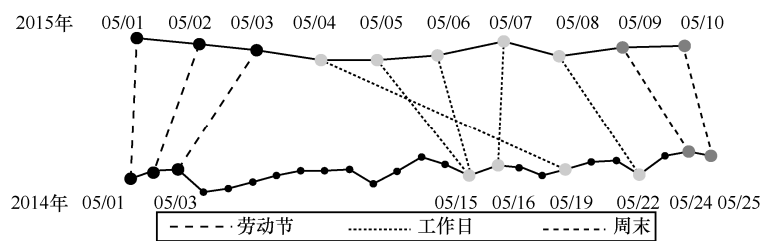


图5 2015年目标日和2014年历史日的匹配结果

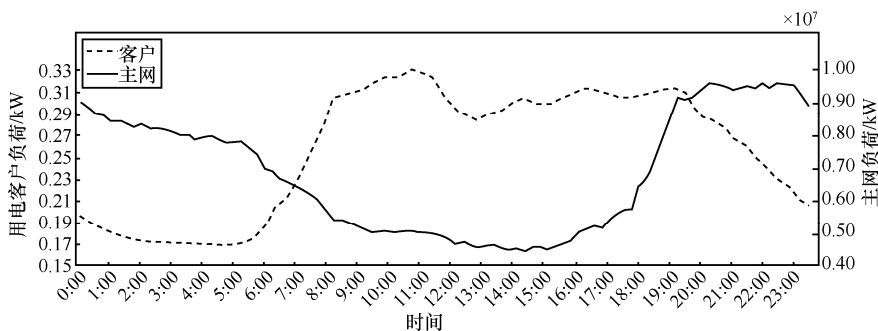


图6 2014年5月25日逆峰型电力客户和主网特征曲线

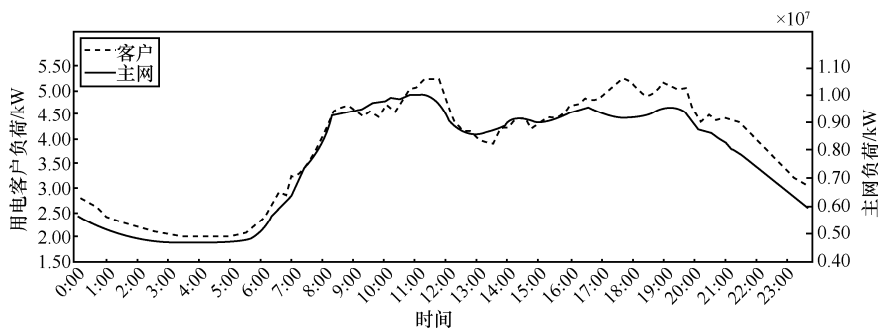


图7 2014年5月25日迎峰型电力客户和主网特征曲线



荷曲线数据进行聚类分析。然后,在考虑不同时期主网负荷特征的基础上,利用并行聚类算法对不同时期海量电力客户的用电行为进行分析,并通过比较主网及客户群体的负荷特征得到不同客户群体的“削峰填谷”模式。最后,利用日期匹配方法找到未来目标日期的历史相似日,将客户群体在历史相似日的“削峰填谷”模式作为目标日期的模式,得到了目标日不同客户群体的“削峰填谷”模式。对福建省电力公司的主网和海量客户的负荷数据分析结果表明,该方法能够得到目标日期不同客户群体的“削峰填谷”模式,并识别需要进行“削峰填谷”的客户群体。本文提出的用电行为分析方法利用聚类分析、并行计算框架实现了对海量客户的用电行为分析,分析结果可以为电网企业开展有序用电、峰谷电价等“削峰填谷”措施提供有针对性的依据,对提高电网需求侧能效管理水平具有建设性意义。

参考文献:

- [1] 何永秀,王冰,熊威,等.基于模糊综合评价的居民智能用电行为分析与互动机制设计[J].电网技术,2012,36(10):247-252.
HE Y X, WANG B, XIONG W, et al. Analysis of residents' smart electricity consumption behavior based on fuzzy synthetic evaluation and the design of interactive mechanism [J]. Power System Technology, 2012, 36(10): 247-252.
- [2] 黄瀚.智能用电中的客户行为分析[J].电气时代,2014(14):28-32.
HUANG H. Analysis of smart electricity consumption behavior[J]. Electric Age, 2014(14): 28-32.
- [3] CHUAN L, UKIL A. Modeling and validation of electrical load profiling in residential buildings in Singapore [J]. IEEE Transactions on Power Systems, 2015, 30(5):2800-2809.
- [4] ABREUA J M, PEREIRA F C, FRRAO P. Using pattern recognition to identify habitual behavior in residential electricity consumption[J]. Energy and Buildings, 2012, 49(2):479-487.
- [5] 宗柳,李扬,王蓓蓓.计及需求响应的多维度用电特征精细挖掘[J].电力系统自动化,2012,36(20):54-58.
ZONG L, LI Y, WANG B B. Fine-mining of multi-dimension electrical characteristics considering demand response [J]. Automation of Electric Power Systems, 2012, 36(20):54-58.
- [6] 瞿海妮,张鹏飞,凌平,等.基于海量数据的电网需求侧用户行为分析研究[J].华东电力,2014,42(12):2922-2925.
QU H N, ZHANG P F, LING P, et al. Analysis of consumption behavior in power demand side based on massive data[J]. East China Electric Power, 2014, 42(12):2922-2925.
- [7] 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 S X, LIU J M, ZHAO B Z, et al. Cloud computing-based analysis on residential electricity consumption behavior [J]. Power System Technology, 2013, 37(6):1542-1546.
- [8] HIDOT S, SAINT-JEAN C. An expectation-maximization algorithm for the wishart mixture model: application to movement clustering [J]. Pattern Recognition Letters, 2010, 31(14): 2318-2324.
- [9] JOLLOIS F X, NADIF M. Speed-up for the expectation-maximization algorithm for clustering categorical data [J]. Global Optimization, 2007, 37(4): 513-525.
- [10] HOFMANS J, CEULEMANS E, STEINLEY D, et al. On the added value of bootstrap analysis for K-means clustering [J]. Classification, 2015, 32(2): 268-284.
- [11] GARCIA M L L, GARCIA-RODENAS R, GOMEZ A G. K-means algorithms for functional data [J]. Neurocomputing, 2015(151):231-245.
- [12] 李小武,邵剑飞,廖秀玲.一种基于 K-means 的分布式聚类算法[J].桂林电子科技大学学报,2011,31(6):460-463.
LI X W, SHAO J F, LIAO X L. A distributed clustering algorithm based on K-means algorithm[J]. Journal of Guilin University of Electronic Technology, 2011, 31(6): 460-463.
- [13] ARRIBAS-GIL A, HANS-GEORM. Pairwise dynamic time warping for event data [J]. Computational Statistics & Data Analysis, 2014(69): 255-268.
- [14] RASHIDI S, FALLAH A, TOWHIDKHAH F. Similarity evaluation of online signatures based on modified dynamic time warping[J].Applied Artificial Intelligence, 2013, 27(7): 599-617.

【作者简介】



王炳鑫(1982-),男,国网福建省电力有限公司泉州供电公司工程师,主要研究方向为电力企业优质服务、电力营销数据处理、营销电力大数据分析、数据挖掘、决策支持。

侯岩(1982-),女,北京中电普华信息技术有限公司助理经济师,主要研究方向为电力营销数据挖掘。

方红旺(1974-),男,北京中电普华信息技术有限公司电力营销客服事业部副总经理,主要研究方向为电力营销大数据分析。

陈雨泽(1988-),男,北京中电普华信息技术有限公司工程师,主要研究方向为电力营销大数据分析。

刘建(1989-),男,北京中电普华信息技术有限公司中级工程师,主要研究方向为电力营销大数据分析。