

智能用电用户行为分析特征优选策略

陆俊, 朱炎平, 彭文昊, 孙毅

(华北电力大学电气与电子工程学院, 北京市 102206)

摘要: 针对大数据应用背景下用户智能用电行为分类的计算复杂性和特征选择有效性的问题, 提出一种基于特征信息量的特征优选策略。首先, 以用电特征的互信息量与相关系数作为特征有效性和关联性判据, 设计用电特征优选准则。然后, 提出一种该准则下的用电行为特征优选策略, 通过减少特征间的分类信息冗余实现高维特征的降维, 并选取有效独立的特征, 从而构建用户用电行为聚类精简特征集。最后, 基于特征优选策略实现了一种特征自适应的用户用电行为分析方法, 完成优化的用户用电行为分析。通过电网实际用电数据验证了所提策略能提高聚类准确率和减少计算复杂性的有效性。

关键词: 用电行为分析; 特征选取; 互信息; 聚类分析; 智能用电

0 引言

随着智能电网的快速发展, 用电信息采集系统及配电自动化系统逐渐完善^[1], 配用电数据呈现出数据量大、数据类型多、增长速度快等大数据特征, 同时电力大数据也渐渐成为人们关注研究的热点^[2-3]。近年来, 国内外先后对电网大数据展开了研究^[4-7]。在大数据环境下分析用电数据之间的关联性, 挖掘潜藏在用电数据之中的用户用电行为习惯, 研究有效的数据挖掘算法, 对不同类别用电用户进行细分, 高效、准确地挖掘出用电行为、电量消费等大量有价值的信息, 使之有力地支撑用户节能工作、智能化业务分析与决策^[8-10]。电网公司研究适用于用户用电行为分析的模式识别技术、聚类分析方法、数据挖掘算法, 并构建通用的数学模型, 此类用户用电特性的感知是智能配电网态势感知的重要组成部分^[11]。电网公司通过对用户用电行为、用电特性的分析, 得到用户用能情况, 为用户节能改造提供参考和建议。同时, 用户根据电网公司提供的用电报告, 调整用电方式, 使用户参与电网的运行, 从而为用户节能、电网营销决策和多元化用户服务等提供有力支撑^[9-10]。

在用户用电模式识别方面, 文献[12]提出基于负荷曲线的用户分类研究, 但其使用负荷曲线分类

数据计算量大, 不适合大量数据情况下的使用; 文献[13]根据用户实际类别对用户的负荷特性进行了分析, 为用户分类提供了依据, 但并未涉及用户分类策略。在用户分类的基础上分析用户用电行为方面已有学者对此展开一定的研究, 文献[14]主要从用户响应方面构建了用户行为模型, 为用户的分类提供了依据, 但并没有涉及不同种类用户用电行为的差异。文献[15]对居民用户分类展开研究, 采用云计算的方式对居民用电行为进行聚类分析, 但未涉及所选特征有效性评价工作。文献[16]在文献[15]的基础上增加了除负荷特征以外的用户信息特征并改进了所用的 K -means 算法聚类中心的选择。文献[17]使用模糊综合评价分析居民智能用电行为, 但其使用的调查问卷方式难以广泛使用。文献[18]使用用户月平均负荷来表示其用电模式, 对用户异常用电行为进行分析。

在上述对用户用电行为的研究当中, 并未提及对用电行为特征的分析与选择方法, 其使用的用电行为特征集并没有经过数据分析以及优化选择, 故面对所需分析的用户样本, 其用电行为特征集的有效性仍有待验证。而面对大数据环境下用电数据快速增长的情况, 通过优选合适样本用户集的用电特征集, 不仅能大大减少分析时所需计算的数据, 降低计算量, 同时能有效地提高分析性能。针对如何有效选取用于分析特征集的问题, 本文提出一种用电行为特征优选策略, 并通过该策略实现了优化的用电行为分析。

本文首先分析了常用用电行为特征, 并采用了

收稿日期: 2016-06-07; 修回日期: 2016-07-05。

上网日期: 2016-09-20。

国家电网公司科技项目“智能电网用户行为理论与互动化模式研究”。

基于信息量的分析方法对其进行分析。然后在此基础上提出了一种基于用电行为特征的有效性与其关联性评价的特征优选策略,通过对相关指标的定量分析,以较少计算量实现效果较好的特征集优化选择。最后基于此策略本文实现了优化的用户用电行为分析,通过对抽取用户样本依照特征优选策略获取适用于用户数据集的特征集,并以此完成用户用电行为分析。

1 用户用电行为特征评价准则设计

1.1 用户用电行为特征归纳

用户用电行为,具体到用户是指其什么时候使用了什么样的电器产生了多少负荷,如文献[19-20]就根据此类特征调度用户负荷,实现较好的用户用电收益。但受采集数据频率低以及技术手段不足等因素影响难以进行如此细致的分析,故通常以用户整体用电负荷为数据源选择特征,表示用户用电行为特点^[13-17],典型特征包括峰时耗电率^[15,21]、谷电系数^[15,21]、日负荷率^[15,22]和日最大负荷^[16,22]等。

文献[15]从峰谷特性的角度建立了包含峰时耗电率和谷电系数等指标的时间序列特征。文献[16]不仅使用了日用电量、日最大负荷、日最小负荷等表征负荷特性的特征,还使用了房屋面积、家庭成员数等关联特征;文献[22]从负荷特性的角度选取典型日负荷曲线、用户平均负荷率以及最大负荷率作为电力用户分类的依据;文献[23]从时间特征量的角度选取年度用电数据、季节用电数据、日用电类型数据等组成聚类特征向量;文献[24]使用各时间点的具体用电量来描述用户的智能用电特性。

以上工作从用电负荷数据入手,根据用电特性的差异,可将不同用户的用电行为划分为多种用电模式,进而基于不同用电模式对用户进行分类,研究各类用户的用电量变化规律,说明负荷类特征选择用于用户用电行为分析具有一定的合理性。

1.2 特征选择必要性分析

用电模式的判别方法和结果与用电特性指标的选取密切相关,因此,需要定义合理的用电特性评价指标以辅助用电模式识别和用户分类^[25]。

大部分用电特征都从时间与负荷这两方面反映了用户用电行为习惯,但由于用户用电行为的多样化,对于不同的用户样本来说,不同用电特征组成的用电行为特征集的分析效果也不同。通过选择有效的用电行为特征集来映射用电数据,能去除特征空间含有的多余信息,提高分析性能,使分析过程更简单,故分析用电特征的有效性是十分有必要的。

同时由于用电特征与用户用电行为密切相关,

故各种用电特征之间关联较大,在分析过程中提供的信息存在重叠和冗余的现象。因此各种有效的用电特征组成的用电特征集合可能由于信息冗余导致分析效果较差,在分析用电特征有效性时有必要综合考虑用电特征间的关联性。

1.3 特征集评价准则设计

为了选取合适有效的用电特征集合,需要综合考虑用电特征的有效性以及各用电特征间的关联性,选用互信息来评价特征的有效性。同时本文主要以用户用电行为的聚类分析为例,故下文主要针对聚类分析进行论述。

以特征全集作为初始候选特征集 X ,由 n 个特征 x 组成,优选特征子集为 Y 。

用电特征 x 与用户类别 c 的互信息 $I(x;c)$ 表示已知用电特征 x 的情况下所能得到的关于用户类别 c 的信息量的多少,即已知特征 x 时类别 c 的不确定度的减少程度,其值越大则表示对应特征 x 越有效。使用概率密度函数 p 进行熵估计,计算用电特征 x 与用户类别 c 的熵。

用电特征 x 的信息熵 $H(x)$ 为:

$$H(x) = - \sum_x p(x) \log_2 p(x) \quad (1)$$

用户类别 c 的信息熵 $H(c)$ 为:

$$H(c) = - \sum_c p(c) \log_2 p(c) \quad (2)$$

使用用电特征 x 与用户类别 c 的联合信息熵 $H(x,c)$ 衡量用电特征为 x 、用户类别为 c 的不确定度:

$$H(x,c) = - \iint p(x,c) \log_2 p(x,c) dx dc \quad (3)$$

则互信息 $I(x;c)$ 为:

$$I(x;c) = H(x) + H(c) - H(x,c) \quad (4)$$

同时本文选用相关系数衡量两个用电特征间的相关程度,作为用电特征的相关性补充。用电特征 x 与优选特征子集 Y 中的特征 y 的相关系数 ρ_{xy} 取值范围为 $-1 \sim 1$,故其绝对值越接近 1,则相关性越大,反之越接近 0 则关联越小,当优选特征集为空时, ρ_{xy} 默认为 0,其计算公式为:

$$\rho_{xy} = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y} \quad (5)$$

式中: $\text{cov}(x,y)$ 为用电特征 x 与优选特征子集中的特征 y 的协方差; σ_x 和 σ_y 分别为 x 和 y 的标准差。

基于以上分析,本文综合考虑用电信息特征对分析性能的有效性和互补性所构建的特征优选的评价函数为:

$$J(x) = I'(x;c) \prod_y (1 - |\rho_{xy}|) \quad (6)$$

$$J(Y)=\sum_y J(y) \tag{7}$$

式中： $J(x)$ 为针对单个用电特征 x 的评价值，其值域为 $[0,1]$ ； $I'(x;c)$ 为用电特征 x 与用户类别 c 的归一化互信息； $J(Y)$ 为优选特征子集 Y 的评价函数； $J(y)$ 为优选特征子集中特征 y 的评价函数。

该评价函数对用电特征所算得的评价值越大则认为该用电特征对于用户用电行为的分析越有效。故对优选所得用电行为特征子集 Y 的评价 $J(Y)$ 为子集中所有特征的评价值之和。

2 基于特征优选的用电行为分析

2.1 用电行为特征优选策略

用电行为特征的选择流程如图 1 所示。

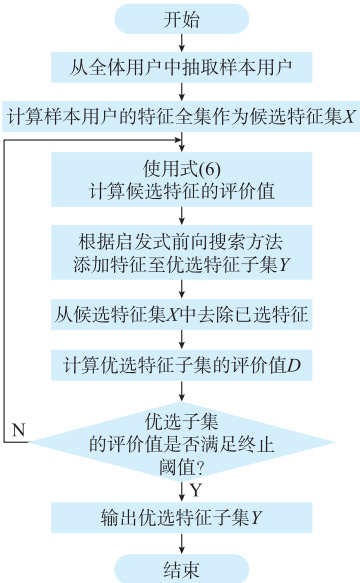


图 1 用电行为特征优选策略
Fig.1 Optimal selection strategy of electricity consumption behavior feature

优选策略原理如下：首先从所分析的总体用户中抽取一定数量的样本用户，提取样本用户的全体特征，然后根据评价准则对样本用户的各种特征进行评价，使用启发式序列前向搜索方法逐个选取特征，即从空集开始，每次搜索会从当前候选子集取出评价值最大的特征放置到已选特征集中，直至已选特征集的性能达到要求。即所选特征 y 为：

$$y=\arg \max \{J(x)\} \tag{8}$$

此方法能同时兼顾计算效率与选择效果，实现以较小的计算量实现较好的筛选效果。本策略运算终止的性能判别条件定义如式(9)所示，即当剩余特征所能提供的信息远低于其造成的冗余时停止选取。

$$D=\frac{\max \{J(x)\}}{J(Y)} \leqslant T \tag{9}$$

式中： D 为候选特征中最优特征的评价值与优选特征子集的评价值的比值，当其小于所设阈值 T 时选择结束，本文所选阈值为 0.1。

该判别条件判别有效，计算量较小，且无需事先设定特征个数。策略筛选流程描述如下。

以初始候选特征集 X 作为输入，此时优选特征集 Y 为空集，首先对于任何属于 X 中的特征 x ，按照式(6)计算其单个特征的评价值，然后根据式(8)挑选最佳特征作为优选特征，将其放置于优选特征集 Y 中，并将其从候选特征集 X 中剔除，最后根据式(9)对优选特征集进行评判，若满足终止条件则输出优选特征集 Y ，若不满足，则重复以上步骤直至满足条件。

2.2 特征自适应的用户用电行为分析方法

特征自适应的用电行为分析方法主要包括针对用户用电数据的特征选择和用户用电行为聚类分析两部分，实现流图见附录 A 图 A1。

本文方法具体实现步骤如下。

1)从用户用电数据中抽取样本，并提取样本的所有用电行为特征构建特征库；对样本进行聚类训练，然后对用电行为特征进行统计分析；依据统计信息，使用本文优选策略优选用电行为特征并构建优选特征集；当满足要求时停止筛选，输出优选特征集。

2)根据优选特征集，提取全部电力用户的对应特征，使用这些优选特征进行聚类分析，对聚类所得的各类用户的用电行为进行分析。

在第一阶段中本文方法针对用户数据定量分析了用电特征的有效性，所用特征空间为算法根据用户数据自适应地从总体特征库中选择得出，故在第二阶段的用电行为分析过程中取得良好的效果。同时从时间复杂度上进行分析，以 K -means 算法为例， K -means 算法的时间复杂度为 $O(mnkt)$ ，其中， m 为聚类对象个数， n 为对象维数， k 为聚类中心个数， t 为迭代次数，当用户数以及聚类中心相同时，本文所提方法的迭代次数与特征个数之积较小，故本文方法的时间复杂度远小于传统算法。

3 实验与分析

基于现有实验数据为某电网 750 个用户的日用电数据，故从常用用电特征中选出峰时耗电量、谷电系数、平段用电百分比、日负荷率、日平均负荷、日峰谷差、日峰谷差率、日最大负荷（依次以编号 1, 2, ..., 8 表示）作为特征全集用以验证优选策略的有

效性。选择 K -means 算法作为聚类分析方法,以输入为不提取特征的原始负荷曲线所得的聚类结果作为对比的用户原始类别。本文算例在单台 CPU 为 2.0 GHz,内存为 6 GB 的个人计算机上完成,同时本文使用 MATLAB 进行算法仿真,所有实验数据经过 MATLAB 进行处理。

实验分析流程主要分为两部分。第一部分是优选策略的性能分析,主要针对特征评价准则合理性和优选策略的效果进行分析;第二部分是用电行为分析方法的性能分析,主要针对本文方法的分析性能及对用户数据集的自适应性进行分析。

3.1 特征优选策略性能实验分析

针对优选策略每一次选择的迭代过程详细分析,其第一次特征选取过程分析见附录 A 表 A1 和图 2。根据附录 A 表 A1 可知,本文所提优选策略将会选取评价价值最高的日平均负荷(编号 5),而由图 2 可知使用日平均负荷进行分析时,分析所得准确率为 80.4%,会远高于使用其他特征进行分析。可见,各类特征的评价价值的大小与使用该类特征进行分析的准确率所表现趋势基本一致。

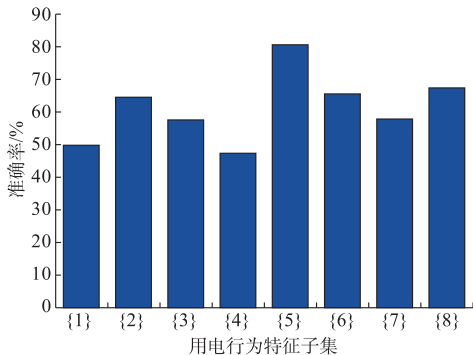


图 2 第一次特征选择时各特征子集聚类准确性对比图
Fig.2 Comparison diagram of clustering accuracy of different feature subset in first feature selection process

对迭代过程中的第二次特征优选过程进行进一步分析,其结果见附录 A 表 A2 和图 3,其中日平均负荷特征作为已选特征无需再次评价。与第一次选择过程类似,根据附录 A 表 A2,本文方法会选用由谷电系数(编号 2)和日平均负荷组成的特征子集,而由图 4 可知使用该特征子集分析,其准确率远高于其他特征子集。之后的选择过程的分析与之类似,不再进一步赘述。结果表明本文所提策略的评价准则能很好地反映对应特征子集的有效性。

依照本文所提优选策略分析用电行为特征,所得特征选取序列的编号依次为 5,2,3,1,8,6,7,4,若将其依次添加至优选特征子集则准确率的变化趋势如图 4 所示。

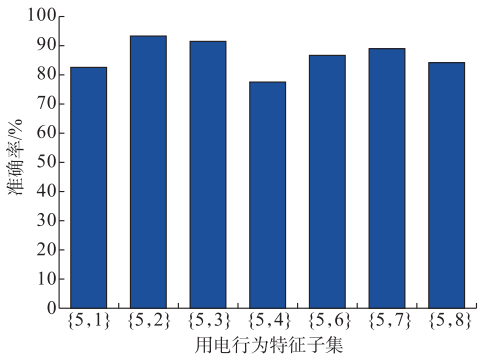


图 3 第二次特征时各特征子集聚类准确性对比图
Fig.3 Comparison diagram of clustering accuracy of different feature subset in second feature selection process

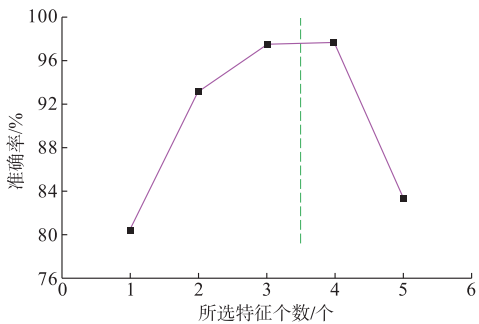


图 4 特征选择过程中准确率变化趋势
Fig.4 Accuracy change trend in feature selection process

图 4 中随着优选特征子集中所选特征的增加,准确率逐渐上升,但由于特征间关联密切,增加特征所提供有效信息减少,冗余信息增加,上升幅度逐渐减缓,在图 4 中特征个数为 4 时达到最高。当特征个数达到 5 个时准确率出现了下降,因为随着特征的增加,可能出现由冗余信息的干扰过大导致准确率下降的现象。根据本文所提优选策略,如图中绿色虚线所示,在选取 3 个特征后达到阈值,终止运算时所获得的优选特征子集编号为 5,2,3,即特征个数为 3 时就终止了选择,此时所得的优选特征集的聚类分析准确率为 97.47%。若进一步增加特征,则在增加大量计算量的同时并不能大幅提升分析性能,其结果表明本文所提优选策略能针对用户样本自适应地选出有效可靠的优选特征子集用于用电行为分析。

3.2 用户行为聚类性能实验分析

将文献[14]的分析方法作为对比方法 1,与本文方法相比主要区别为其所用特征子集为日负荷率、谷电系数、平段用电百分比、峰时耗电率,将使用全部特征的分析方法作为对比方法 2,结果如表 1 所示。

从对比方法 1 与本文方法的对比可发现,两者准确率存在较大差异,而对比方法 1 在文献[14]中分析效率较高,此处分析效率较低是因为对比方法 1 所用特征集并不适用于分析本文所用的数据集,故面对不同的用户数据集选取合适的特征集是很有必要的,而本文方法所优选出的特征子集能很好地适应用户数据集的分析。同时由对比方法 2 和本文方法对比,可以发现本文方法确实能有效地从特征库中精简有效的特征集,实现减少了特征集的冗余。

表 1 实验方法的性能对比

Table 1 Performance comparison of experiment methods

方法	准确率/%	聚类迭代次数/次	总聚类时间/s	主要计算时间/s
对比方法 1	73.33	23	0.909	0.031 7
对比方法 2	97.20	10	0.405	0.032 4
本文方法	97.47	8	0.315	0.031 6

根据表 1 所示,以计算时间来反映计算量的大小,其中主要计算时间为计算各目标离聚类中心的距离的过程所花时间。由此可以发现本文方法的迭代次数、总聚类时间都小于对比方法。而由于用户数据较少以及程序运算的影响,计算时间的对比不太明显,但仍看出其略少于其他方法。该结果表明特征集的有效性一定程度上影响了聚类的迭代次数,而其精简性则影响每次迭代过程中的计算时间,故通过特征集的优选能有效降低分析过程中的计算量。

依据本文所提方法,采用聚类分析所得的用户用电行为分析结果如图 5 所示,图中用电量进行了归一化处理,用户被分为 6 类,各类用户的典型用电规律曲线如图中红线所示。

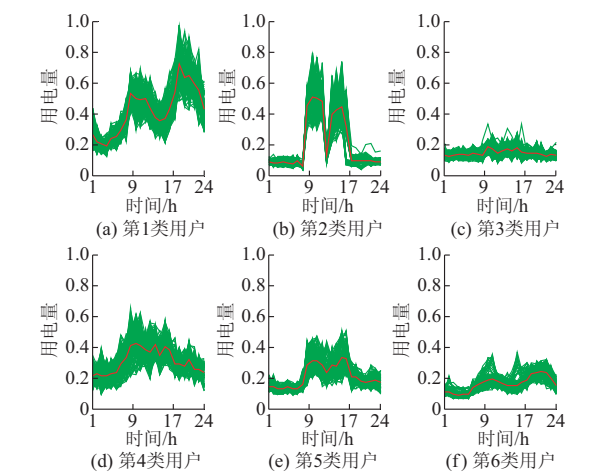


图 5 聚类结果及用户用电规律曲线
Fig.5 Clustering result and electricity consumption rule curves of users

据此对各类用户用电行为进行分析,结果见附录 A 表 A3,其中第 1 类用户和第 2 类用户具有较大的负荷调整空间,可以根据用户特点建模,通过对各类用户分析建模有助于电网对用户负荷的调度策略提供支撑,其次对同类用户的用电行为的影响机理进行分析,可以用于自身用电方式的改进。以上部分不属于本文探讨范围,在此不再赘述。

4 结语

本文基于特征信息量对用户用电行为特征有效性进行了定量分析,并提出了基于互信息与相关系数的用户用电行为特征优选策略,以此完成对用户用电行为的分析实验。上述工作提高了用户用电行为分析的有效性与可靠性,能有效地挖掘电力大数据的潜在价值信息,为相关政策及服务的制定提供有力的信息支撑。但本文方法仅提出了一种有效的挑选特征的方法,如何更有效地挑选有效特征需要进一步研究,同时本文仅是研究对现有特征的挑选,除了对现有特征进行挑选进行研究以外,还可以从深度学习等机器学习的角度进一步研究自主寻找新特征的算法。

附录见本刊网络版 (<http://www.aeps-info.com/aeps/ch/index.aspx>)。

参考文献

[1] 胡江溢,祝恩国,杜新纲,等.用电信息采集系统应用现状及发展趋势[J].电力系统自动化,2014,38(2):131-135.DOI:10.7500/AEPS20130617005.
HU Jiangyi, ZHU Enguo, DU Xingang, et al. Application status and development trend of power consumption information collection system[J]. Automation of Electric Power Systems, 2014, 38(2): 131-135. DOI: 10.7500/AEPS20130617005.

[2] 中国电机工程学会信息化专委会.中国电力大数据发展白皮书[R].2013.

[3] 宋亚奇,周国亮,朱永利.智能电网大数据处理技术现状与挑战[J].电网技术,2013,37(4):927-935.
SONG Yaqi, ZHOU Guoliang, ZHU Yongli. Present status and challenges of big data processing in smart grid[J]. Power System Technology, 2013, 37(4): 927-935.

[4] DORR D. Data analytics and applications newsletter (DMD and TMD demonstrations)[R]. California, USA: Electric Power Research Institute, 2014.

[5] C3 energy releases new version of C3 enterprise customer analytics software and announces C3 energy analytics platform[EB/OL]. [2013-03-06]. <http://www.marketwired.com/press-release/c3-energy-releases-new-version-c3-enterprise-customer-analytics-software-announces-c3-1764822.htm>.

[6] BELHOMME R, TRANCHITA C, VU A, et al. Overview and goals of the clusters of smart grid demonstration projects in France[C]// IEEE Power and Energy Society General Meeting,

- July 24-29, 2011, San Diego, CA, USA: 8p.
- [7] QU Zhijian, CHEN Ge. Big data compression processing and verification based on Hive for smart substation[J]. Journal of Modern Power Systems and Clean Energy, 2015, 3(3): 440-446.
- [8] 李扬,王蓓蓓,李方兴.灵活互动的智能用电展望与思考[J].电力系统自动化, 2015, 39(17): 2-9. DOI: 10.7500/AEPS20150730004.
- LI Yang, WANG Beibei, LI Fangxing. Outlook and thinking of flexible and interactive utilization of intelligent power [J]. Automation of Electric Power Systems, 2015, 39(17): 2-9. DOI: 10.7500/AEPS20150730004.
- [9] 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.
- LIU Keyan, SHENG Wanxing, ZHANG Dongxia, et al. Big data application requirements and scenario analysis in smart distribution network [J]. Proceedings of the CSEE, 2015, 35(2): 287-293.
- [10] 史常凯,张波,盛万兴,等.灵活互动智能用电的技术架构探讨[J].电网技术,2013,37(10):2868-2874.
- SHI Changkai, ZHANG Bo, SHENG Wanxing, et al. A discussion on technical architecture for flexible intelligent interactive power utilization[J]. Power System Technology, 2013, 37(10): 2868-2874.
- [11] 朱文俊,王毅,罗敏.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,2016,40(12):21-27.DOI:10.7500/AEPS20160316007.
- ZHU Wenjun, WANG Yi, LUO Min. Distributed clustering algorithm for awareness of electricity consumption characteristics of massive consumer[J]. Automation of Electric Power Systems, 2016, 40(12): 21-27. DOI: 10.7500/AEPS20160316007.
- [12] 冯晓蒲,张铁峰.基于实际负荷曲线的电力用户分类技术研究[J].电力科学与工程,2010,26(9):18-22.
- FENG Xiaopu, ZHANG Tiefeng. Research on electricity users classification technology based on actual load curve[J]. Electric Power Science and Engineering, 2010, 26(9): 18-22.
- [13] 陈伟,乐丽琴,崔凯,等.典型用户负荷特性及用电特点分析[J].能源技术经济,2011,23(9):44-49.
- CHEN Wei, YUE Liqin, CUI Kai, et al. Analysis on power load and consumption characteristics of representative consumers[J]. Energy Technology and Economics, 2011, 23(9): 44-49.
- [14] 刘继东,韩学山,韩伟吉,等.分时电价下用户响应行为的模型与算法[J].电网技术,2013,37(10):2973-2978.
- LIU Jidong, HAN Xueshan, HAN Weiji, et al. Model and algorithm of customers' responsive behavior under time-of-use price [J]. Power System Technology, 2013, 37(10): 2973-2978.
- [15] 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.
- [16] 赵莉,侯兴哲,胡君,等.基于改进 K-means 算法的海量智能用电数据分析[J].电网技术,2014,38(10):2715-2720.
- ZHAO Li, HOU Xingzhe, HU Jun, et al. Improved K-means algorithm based analysis on massive data of intelligent power utilization[J]. Power System Technology, 2014, 38(10): 2715-2720.
- [17] 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 Yongxiu, WANG Bing, XIONG Wei, 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.
- [18] 庄池杰,张斌,胡军,等.基于无监督学习的电力用户异常用电模式检测[J].电机工程学报,2016,36(2):379-387.
- ZHUANG Chijie, ZHANG Bin, HU Jun, et al. Anomaly detection for power consumption patterns based on unsupervised learning[J]. Proceedings of the CSEE, 2016, 36(2): 379-387.
- [19] 郭晓利,于阳.基于云计算的家庭智能用电策略[J].电力系统自动化,2015,39(17):114-119.DOI:10.7500/AEPS20150310017.
- GUO Xiaoli, YU Yang. A residential smart power utilization strategy based on cloud computing[J]. Automation of Electric Power Systems, 2015, 39(17): 114-119. DOI: 10.7500/AEPS20150310017.
- [20] 王守相,孙智卿,刘喆.面向智能用电的家庭能量协同调度策略[J].电力系统自动化,2015,39(17):109-113.DOI:10.7500/AEPS20140625016.
- WANG Shouxiang, SUN Zhiqing, LIU Zhe. Co-scheduling strategy of home energy for smart power utilization [J]. Automation of Electric Power Systems, 2015, 39(17): 109-113. DOI: 10.7500/AEPS20140625016.
- [21] 庄绪强.基于云计算技术的用户用电智能分析技术研究[J].自动化与仪器仪表,2016(2):187-189.
- [22] 杨浩,张磊,何潜,等.基于自适应模糊 C 均值算法的电力负荷分类研究[J].电力系统保护与控制,2010,38(16):111-115.
- YANG Hao, ZHANG Lei, HE Qian, et al. Study of power load classification based on adaptive fuzzy C means[J]. Power System Protection and Control, 2010, 38(16): 111-115.
- [23] 薛承荣,顾洁,赵建平,等.基于用户用电特性及供电成本分摊的销售侧电价机制研究[J].华东电力,2014,42(1):168-173.
- XUE Chengrong, GU Jie, ZHAO Jianping, et al. Electricity retail tariff mechanism based on customers' electrical characteristics and cost apportionment[J]. East China Electric Power, 2014, 42(1): 168-173.
- [24] 杨永标,颜庆国,王冬,等.居民用户智能用电建模及优化仿真分析[J].电力系统自动化,2016,36(3):46-51.DOI:10.7500/AEPS20150924010.
- YANG Yongbiao, YAN Qingguo, WANG Dong, et al. Intelligent electricity consumption modeling and optimal simulations for residential users[J]. Automation of Electric Power Systems, 2016, 36(3): 46-51. DOI: 10.7500/AEPS20150924010.

(下转第 83 页 continued on page 83)

[25] 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]. 中国电机工程学报, 2016, 36(3): 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, 36(3): 604-614.

陆俊(1976—), 男, 通信作者, 博士, 副教授, 主要研究方向: 信息处理理论、电力通信和通信网络。E-mail: lujun@ncepu.edu.cn

朱炎平(1992—), 男, 硕士研究生, 主要研究方向: 电力通信及信息处理。E-mail: zyp2016_gz@163.com

彭文昊(1993—), 男, 硕士研究生, 主要研究方向: 电力信息处理。E-mail: 1054518964@qq.com

(编辑 万志超)

Feature Selection Strategy for Electricity Consumption Behavior Analysis in Smart Grid

LU Jun, ZHU Yanping, PENG Wenhao, SUN Yi

(School of Electrical and Electronic Engineering, North China Electric Power University, Beijing 102206, China)

Abstract: In face of the problem with mass electricity data processing, a strategy of electricity consumption behavior feature selection is put forward to reduce computing complexity and improve validity. Firstly, the validity and the relevance of electricity consumption feature are analyzed by its information entropy and related coefficient. On this basis, an evaluation criterion for electricity consumption feature is designed. Then a feature selection strategy is built to reduce the high dimensions of the feature vectors by reducing the redundancy of feature space. And the condensed feature set is devised using this strategy to select features. According to the feature selection strategy, a method of electricity consumption behavior analysis is put forward to make electricity consumption behavior analysis optimized. Finally, the simulation results of this method achieved by electricity consumption data show that it can effectively improve the accuracy of clustering and reduce computation time.

This work is supported by State Grid Corporation of China.

Key words: electricity consumption behavior analysis; feature selection; mutual information; clustering analysis; smart power utilization