

基于智能电网的高校学生用电模型分析与研究^{*}

郑贵林, 柳帅

(武汉大学 动力与机械学院自动化系, 武汉 430072)

摘要: 基于校园智能电网对学生用电数据进行了挖掘分析, 用基于箱形图的回归分析法分析了学生的用电习惯和特征, 构建了学生用电模型, 得到不同温度下学生用电的时间序列特征。对学生的用电模型进行研究分析, 为校园电网的电力调度和管理提供了更准确的数据支持, 对电力需求侧管理有着重要意义。

关键词: 智能电网; 用电模型; 箱形图; 回归分析; 数据挖掘分析

DOI: 10.19753/j.issn1001-1390.2019.017.006

中图分类号: TM711

文献标识码: B

文章编号: 1001-1390(2019)17-0031-05

Analysis and research on electric consumption model of college students based on smart grid

Zheng Guilin, Liu Shuai

(School of Power and Mechanical Engineering, Wuhan University, Wuhan 430072, China)

Abstract: This paper analyzed the data of electricity consumption of students based on smart grid in college, and analyzed students habits and characteristics of power consumption by using regression analysis based on boxplot, and built the power consumption model of students and obtained the power consumption characteristic curve in time series at different temperature. The analysis of power consumption model of students has great importance on power demand-side management and provides data support for power demand side management.

Keywords: smart grid, power consumption model, boxplot, regression analysis, data mining analysis

0 引言

随着近年来智能电网的发展, 迎来了电力大数据时代, 电力数据的挖掘分析成为一门热门话题。近年已有一些学者对电力用户的用电数据进行了分析研究: 文献[1]利用并行 k-means 聚类算法得到了不同电力用户的用电模型, 文献[2]中研究了居民用电负荷特性, 分析了居民经济条件对峰谷电价的影响。文献[3]中采用优化的 k-means 聚类算法分析了工业用户的用电行为。

而广大学生也是不可忽略的电消费者, 对学生的用电行为进行研究, 能为电网在进行错峰用电管理, 用电用户分类研究, 峰谷电价研究等方面提供数据支持。

将来学生用电也纳入智能电网的需求侧管理中, 实现智能用电, 低碳环保。

2014 年 9 月份某大学信息学部 and 医学部共 5 栋学生公寓(其中 2 栋为男生公寓, 3 栋为女生公寓)进行了智能化改造, 将基于互联网的能源管控系统融入了学生公寓中, 公寓中所有房间的用电数据以及用电相关数据能实时上传到服务器并存储在数据库。这些数据有着海量、高频、分散等特点, 并且数据之间有着关联性, 隐藏了学生用电的行为习惯。

基于校园智能电网, 对学生用电数据进行挖掘分析, 用基于箱形图的回归分析法构建了学生用电模型, 得到了学生的用电习惯和特征, 为智能电网中用电识别以及电力调度提供了数据支持。

1 能源管控系统介绍

整个系统结构如图 1 所示, 根据所实现的功能不同, 主要划分为数据管理层、网络通信层、现场设备层、控制对象层四部分。

^{*} 基金项目: 国家自然科学基金资助项目(61272114); 国家海洋局“海洋可再生能源专项”资助项目(GHME2013JS01); 智慧校园能源管控系统(NGII20160617); 基于 IPv6 的农业大数据野外信息采集系统(NGII20160312)

数据管理层即数据库和服务器,功能有两部分,一是发送控制命令或者采集指令。二是负责数据存储整合,将现场返回的原始数据处理转化成电数据和电相关数据,分类存储到数据库。

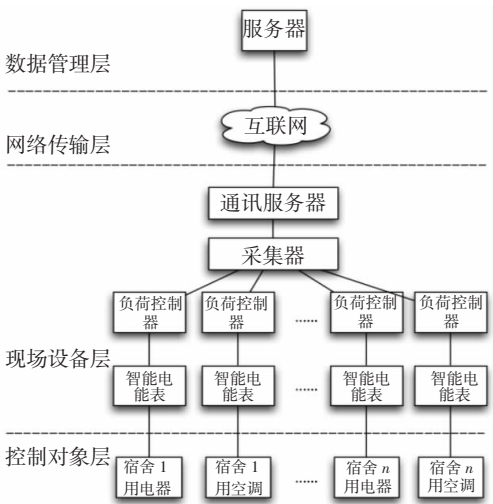


图 1 系统结构图

Fig. 1 System structure diagram

网络通信层即互联网,其主要功能是让数据通过互联网在服务器与现场设备之间交互传输。

现场设备层主要由智能负荷控制器和智能电表两部分构成,智能电表负责现场用电数据的采集,而智能负荷控制器则负责对现场用电的控制。

控制对象层主要就是学生公寓中的各种用电设备,包括学生的日常用电器以及空调。

各层负责不同功能,每一层为其下一层提供服务。控制对象层的用电器的用电情况被现场设备层的设备识别采集以后,经过网络通信层上传到数据管理层;数据管理层发出的控制指令也通过网络通讯层到达现场设备层。并由现场设备层执行命令^[4-5]。

学生公寓中用电线路分成生活照明用电和空调用电两路,所以系统也从学生的生活用电和空调用电两部分实时监控和管理学生的用电情况,并在过去四年中系统运行稳定正常,实时采集了每一间宿舍的用电信息并存储到云端,包括电压、电流、功率、功率因数、电度数、账户资费、充值记录等情况。公寓管理员不仅能通过人机交互页面实时监控每间房间的用电数据和状态,也能发送指令通过智能负荷控制器直接控制某房间的总闸。

2 数据预处理和数据分析

2.1 数据预处理

2014 年 9 月~2017 年 12 月,合计 42 个月,系统采

集并存储了各个宿舍的实时用电数据,这些数据就是用来进行分析的对象。但是这些原始数据比较“脏”,若是直接对这些数据进行分析,会影响分析结果的正确性和精确性,所以需要对这些数据进行预处理。预处理主要包括以下内容:

数据集成,将多个数据库中多个表的关联数据提取集成到一个数据库中,方便数据处理。

数据采样转换。将数据转化为容易分析的形式。比如原始数据中的电表读数是每隔数分钟采集一次,如表 1 所示。但是这样数据量较大,而且用电量分析不需要精确到分钟。所以我们对其进行降采样,将电量数据的频率转化为小时,如表 2 所示。这样降低时间复杂度,方便后面步骤的处理分析。

表 1 某房间原始电能表数据

Tab. 1 Original electric data for a room

时间	电度数
...	...
2016/5/31 9:00:42.	660.31
...	...
2016/5/31 9:57:24	660.44
2016/5/31 10:07:12	660.45
...	...
2016/5/31 10:52:37	660.46
2016/5/31 11:00:49	660.47
...	...
2016/5/31 11:50:00	660.48
2016/5/31 12:04:46	660.48
...	...

表 2 降采样后电能表数据

Tab. 2 Electric data after down-sampling

时间	电度数
...	...
2016/5/31 9:00:00.	660.31
2016/5/31 10:00:00	660.45
2016/5/31 11:00:00	660.47
2016/5/31 12:00:00	660.48
...	...

数据清洗,在数据分析中,需要对明显不合理的数据进行清洗。例如由于某段时间网络或者设备出现了问题,服务器无法从现场采集用电数据,就会出现空白数据。若是不做处理,在计算用电量时,需要对电度数做差分计算,这个时候可能会在空白

数据时间点之后出现非常大的用电量,这是不合理的。为了降低空白数据对分析结果的影响,文中使用了线性插值的方式来填补这些空白^[6],如表3所示。对不合理的数据进行清洗以后,能确保数据分析结论的科学、有效、精确。

表3 数据插值

Tab. 3 Data interpolation

时间	插值前电度数	插值后电度数
2014/9/18 2:00:00	43.6	43.6
2014/9/18 3:00:00	None	43.727
2014/9/18 4:00:00	None	43.854
2014/9/18 5:00:00	43.98	43.98

数据验证,检查数据各个字段的最大值、最小值、平均值、空缺数量、记录个数、中位数等,并以此来评估最后的数据中是否满足分析的需求。如果依然出现不合理数据,需要进行数据清洗。某房间的数据验证如表4所示,其满足分析的需求。

表4 某房间用电数据验证

Tab. 4 Electric data verification for a room

	电流/A	电压/V	功率/W	用电量 kW/h
记录个数	29 052	29 052	29 052	29 052
空缺数量	1 254	1 254	1 254	1 254
最大值	8.97	246.8	1476.4	1.38
最小值	0	0	0	0
平均值	0.59	101.20	23.90	0.08

所有房间进行数据验证以后,发现有的房间空缺数量过多,就算使用插值弥补空缺,也会带来较大误差,所以最终从748个房间中选出了690个房间的用电数据进行下一步分析。

2.2 用电数据分析

2.2.1 统计学生用电与月份关系

将690间宿舍的照明用电和空调用电分别按照月份划分,求出每个月的平均每间宿舍的用电量,统计结果如图2所示。照明用电平均每个月每间宿舍的照明用电量为28.92°,空调用电为49.80°。其中2月、7月、8月是寒暑假时间,学生大部分会回家,所以这段时间的用电量明显偏少。

从变化趋势上看,照明用电较平稳,除了寒暑假的月份,受季节性影响较小;空调用电的变化幅度很大,且用电集中在温度最高和最低的月份。所以接下来分

析用电量与温度之间的关系。

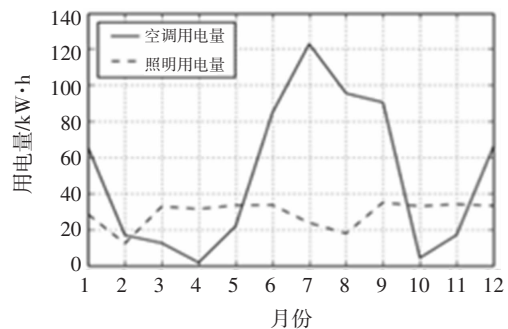


图2 用电与月份关系

Fig. 2 Relationship between month and power consumption

2.2.2 宿舍用电与气温的关系

为了探寻学生用电量与温度的关系,首先要获取到城市的历史气温信息,获取的历史温度信息来源于:<http://www.tianqihoubao.com/>。统计期间,某市的日平均气温最低为-3℃,最高为30.5℃。此外,为了得到学生用电与温度之间的关系,使用了基于箱形图的回归分析来计算学生的用电模型。

箱形图能很好的反映一组数据分散情况和其中异常值,每个箱中一般有6个数据节点,分别是上边缘、上四分位数、中位数、下四分位数、下边缘以及异常值。将690间宿舍不同温度下的空调用电量和照明用电量进行统计,得到箱形图如图3、图4所示。图中的圆点代表的是异常值,每个温度下有690个数据点,异常值个数为5~30个,小于5%。异常值的产生可能是假期期间有学生留校产生了用电量等原因,所以为了保证回归分析的准确,对异常值进行剔除处理。剔除以后对每个温度下的数据计算平均值,作为这个温度下的平均用电量,结果如图5所示。

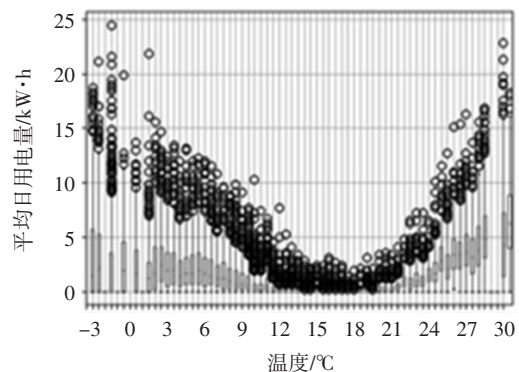


图3 空调用电量箱形图

Fig. 3 Boxplot of air conditioning power consumption

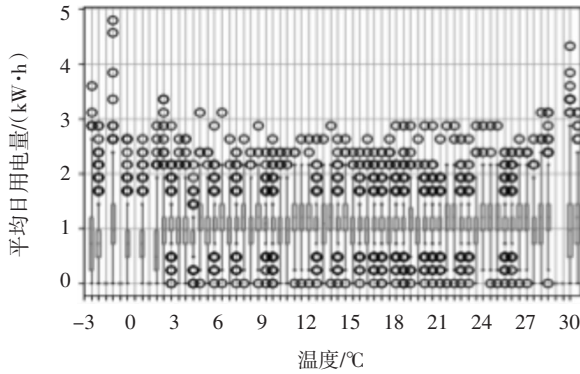


图4 照明用电量箱形图

Fig. 4 Boxplot of lighting power consumption

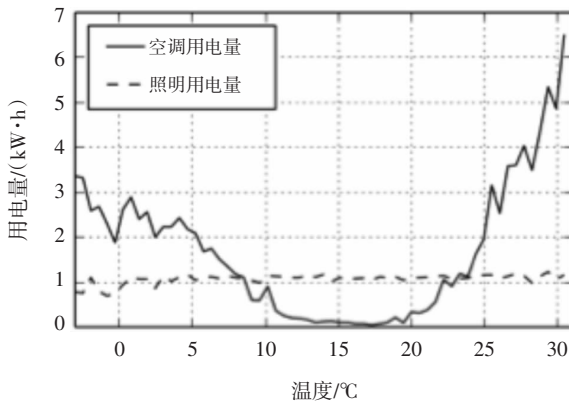


图5 日用电量与当日平均温度

Fig. 5 Average daily power consumption at different temperature

2.2.3 回归分析

回归分析是用来研究表达两个或者两个以上变量之间相关关系数学方式的一种统计分析方法,最终建立一个表达变量之间相关关系的数学表达式^[7-15]。文中寻求学生用电量与温度之间的关系,即利用回归分析法得到函数 $y = Q(t)$,使得式(1)取得最小值,使函数 $Q(t)$ 能最佳拟合用电量与温度的关系。

$$\sum_{i=1}^n (\hat{y}_i - Q(t_i))^2 \quad (1)$$

拟合以后使用均方误差根 σ 来作为回归函数的评价指标。

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - Q(t_i))^2}{n}} \quad (2)$$

观察照明用电量的数据和空调用电量数据,可以发现照明用电量与温度的关系可以适合使用一次线性回归方程,空调用电量与温度的关系适合使用

多次线性回归方程。回归以后,最终得到每日照明用电量与温度关系见式(3),空调用电量与温度关系见式(4):

$$Q(t) = 0.0062t + 0.9718 \quad (3)$$

$$Q(t) = 7.155 t^3 \times 10^{-4} - 1.389 t^2 \times 10^{-2} - 0.1284t + 2.683 \quad (4)$$

曲线拟合以后,得到的拟合用均方根误差来进行精度评价,计算得到照明用电量的回归函数均方根误差 $\sigma = 0.093$,空调用电量回归函数均方根误差 $\sigma = 0.2955$ 。均方根误差都很小,说明拟合效果十分良好。将拟合后的曲线画出来如图6所示。

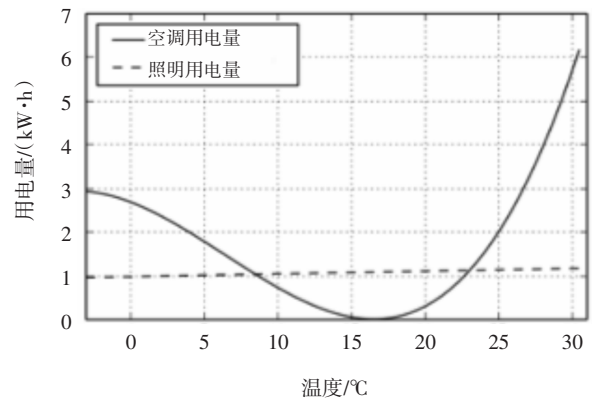


图6 拟合后的曲线

Fig. 6 Curves after fitting

通过以上分析,可以发现学生的生活用电量与温度没有太大的关系,每天基本是1 kW·h左右。而空调的用电量与温度有明显的关系,只有当温度偏高或者偏低的时候,空调才会被使用,这点也从模型曲线上很好的反应出来。此外,夏天的空调用电量比冬天的增长趋势更高。这是因为学校位于武汉,夏天的温度更为炎热,对空调的需求更大,所以夏天的空调耗电量比冬天多。

2.2.4 时间序列特征提取

根据前面分析结果,可以得知学生的生活照明用电每天几乎都是固定的,将690间房间的生活用电量根据一天24 h划分,可以得到学生每日的生活照明用电模型曲线如图7所示。同样,可以知道学生的空调用电量与温度有着非线性关系,而且高温低温下用电量增长趋势不同,所以将温度划分为三个区间 $(-\infty, 10)$, $[10, 22]$, $(22, +\infty)$, 对不同温度区间下的空调用电量根据24 h划分,最终得到学生每日的空调用电模型曲线如图8所示。

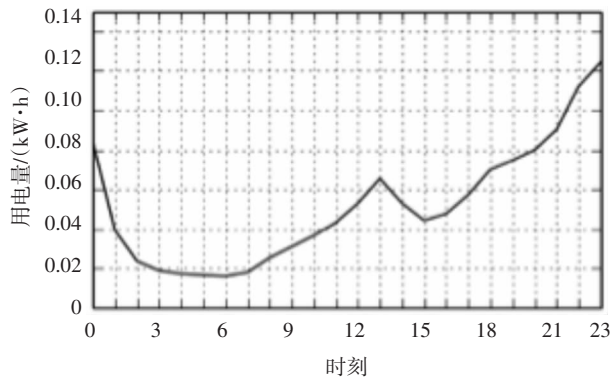


图7 照明用电的时间序列特征曲线

Fig. 7 Lighting power consumption characteristic curve in time series

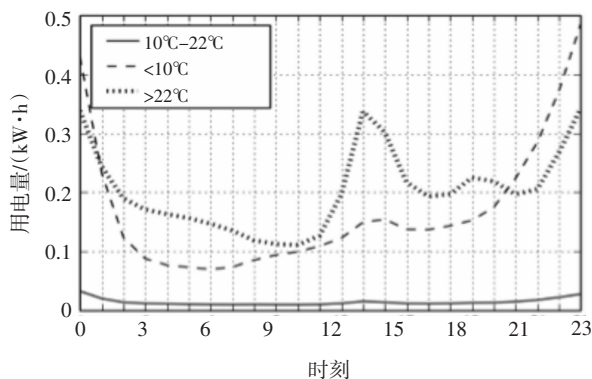


图8 不同温度下空调用电的时间序列特征曲线

Fig. 8 Air conditioning power consumption characteristic curve in time series at different temperature

通过图8,我们可以发现:(1)学生用电高峰期集中在晚上和中午,这与学生的作息時間一致;(2)空调用电量在不同温度下有不同特征,表现为低温时夜间用电量更多,高温时日间用电量更多。

通过对比文献[1]中研究到的居民用电的五种居民用电模型曲线图,可以发现学生的用电规律也较为明显,将来在电网中进行用户分类,根据这些特征很容易将学生用户区分出来。

2.2.5 分析总结

大学生学生公寓的生活用电电器主要是照明和风扇为主,还有部分电脑^[8]。所以学生的生活照明用电量有很强的周期性且比较稳定,与温度无关。所以对学生每日生活照明用电量进行能耗预测,建立能耗值的时间序列 Y_t ,应有 $Y_t = TC_t + I_t$,其中 TC_t 是时间项, I_t 为不规则因子。

而学生的空调用电受温度的影响较大,结合回归分析得到的空调用电量与温度的模型,对学生每日的

空调用电量进行能耗预测,空调的能耗时间序列 Y_t 应有: $Y_t = TC_t + S_t + I_t$

式中 TC_t 是时间项; I_t 为不规则因子; S_t 是温度项。

所以学生的用电是很有规律并且有周期性,十分易于识别与预测。未来随着智能电网的普及和发展,通过以上分析得出的学生用电特征,能实现学生用电的识别,并能将学生用电纳入需求侧管理,从而围绕峰谷电价,错峰用电等展开研究,最终实现节约能耗的目的。

3 结束语

基于学生公寓能源管控系统中的智能电网,从生活用电和空调用电两个方面,对2014年9月~2017年12月,合计42个月的用电数据进行了挖掘分析。通过使用基于箱形图的回归分析法构建了反映学生用电与温度之间关系的数学模型。基于此,绘出了学生的用电曲线,得到了不同温度下学生的时间序列用电特征,分析学生的用电规律和习惯。

根据这些情况,校园用电网以及当地电网能够对电网进行合理的调度,并对当地的实时电价以及峰谷电价进行研究,将学生用电纳入需求侧管理中,从而实现智能电网更加智能,节能环保。

参考文献

- [1] 张素香,刘建明,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, et al. Cloud computing-based analysis on residential electricity consumption behavior[J]. Power System Technology, 2013, 37(6): 1542-1546.
- [2] 林启开,王珂,等.峰谷电价下居民用电聚合响应特性分析[J].电力工程技术,2017,36(3):88-93.
Lin Qikai, Wang Ke, et al. Analysis on the polymer response Characteristics of Residents under the Peak and valley Electricity Price[J]. Jiangsu Electrical engineering, 2017, 36(3): 88-93.
- [3] 徐磊,杨秀,张美霞.基于数据挖掘的工业用户用电行为分析[J].电测与仪表,2017,54(16):68-74.
Xu Lei, Yang Xiu, Zhang Meixia. Industrial users of electricity behavior analysis based on data mining[J]. Electrical Measurement & Instrumentation, 2017, 54(16): 68-74.
- [4] 郑贵林,张丽,迟进武.互联网+建筑能源管控系统设计与实现[J].自动化与仪表,2015,30(12):1-5,19.
- [5] 郑贵林,张丽.基于智能负荷控制器的建筑电气优化布线研究[J].武汉大学学报(工学版),2015,48(3):417-423.
- [6] 丁晓,孙虹,郑海雁,等.基于配用电大数据的短期负荷预测[J].电力工程技术,2018,37(3):21-27.
- [7] 赵广社,张希仁.数据挖掘中的统计方法概述[J].计算机测量与控制,2003,(12):914-917.

(下转第127页)