



## 基于回归分析的迎峰度夏用电量分析

陈雨泽<sup>1</sup>, 李磊<sup>2</sup>, 方学民<sup>2</sup>, 刘建<sup>1</sup>, 赵加奎<sup>3</sup>, 刘玉玺<sup>1</sup>, 左松林<sup>4</sup>

(1. 国网信息通信产业集团有限公司, 北京 100085; 2. 国家电网有限公司, 北京 100031;  
3. 国家电网有限公司大数据中心, 北京 100031; 4. 国网安徽省电力有限公司, 安徽 合肥 340001)

**摘要:** 利用回归分析算法, 提出了一种迎峰度夏期间最高气温与日用电量关系分析方法, 首先利用皮尔森相关系数对度夏期间的最高气温和日用电量相关程度进行定量分析, 然后利用回归分析算法建立最高气温和日用电量的拟合函数, 并利用最小二乘法进行求解, 得到最高气温和日用电量相关关系的定量分析结果。研究表明, 提出的度夏期间用电量分析方法能够定量地给出气温变化导致的用电量变化情况, 并能根据分析结果对未来几天的日用电量进行准确预测。

**关键词:** 用电量; 迎峰度夏; 皮尔森相关系数; 回归分析; 最小二乘法

**中图分类号:** TN915

**文献标识码:** A

**doi:** 10.11959/j.issn.1000-0801.2019274

## An analysis method for electricity consumption in peak load period of summer based on regression analysis

CHEN Yuze<sup>1</sup>, LI Lei<sup>2</sup>, FANG Xuemin<sup>2</sup>, LIU Jian<sup>1</sup>, ZHAO Jiakui<sup>3</sup>, LIU Yuxi<sup>1</sup>, ZUO Songlin<sup>4</sup>

1. State Grid Information & Telecommunication Group Co., Ltd., Beijing 100085, China

2. State Grid Corporation of China, Beijing 100031, China

3. State Grid Big Data Center, Beijing 100031, China

4. State Grid Anhui Electric Power Co., Ltd., Hefei 340001, China

**Abstract:** A method for analyzing the relationship between daily maximum temperature and electricity consumption during peak load period was proposed. Firstly, the relationship between maximum temperature and electricity consumption was quantified by Pearson correlation coefficient. Secondly, the functional relationship was established using regression and solved by least squares method. The quantified result of the relationship between maximum temperature and electricity consumption was then acquired. The empirical research shows the proposed method for electricity consumption analysis during peak load period of summer can quantify the variation of daily electricity consumption caused by temperature change and predict electricity consumption of the next few days according to the analysis result.

**Key words:** electricity consumption, peak load period of summer, Pearson correlation coefficient, regression analysis, least squares



## 1 引言

电力行业迎峰度夏期间,电力客户用电需求达到高峰,电力企业供电压力陡增。为确保电网设备安全、平稳运行,有效缓解用电高峰期的紧张局面,需要提前根据天气情况对迎峰度夏期间的电力负荷、日用电量等指标进行定量分析。此外,掌握迎峰度夏期间公司供电区域用电、售电情况,了解迎峰度夏期间公司相关业务的运行特点还能够对公司营销、营业的后续工作开展提供很好的借鉴。

近年来,随着电力行业数据规模的不断攀升,已有大量学者利用数据挖掘、机器学习等方法对迎峰度夏期间的用电、售电情况进行了分析。参考文献[1-4]通过相关数据对湖北省 2010 年和 2011 年迎峰度夏期间的用电负荷、日用电量、设备运行状况、设备故障情况以及电网运行存在的问题进行了量化分析,但仅通过简单的线性关系对温度上升时用电负荷及日用电量等指标的变化进行了估计。参考文献[5-6]利用数据挖掘和数据分析方法,通过分析历史用电数据与气象因素数据的关系对不同地区的日用电量和用电负荷进行了分析。参考文献[7-9]分别研究了武汉市、上海市以及江苏省的日用电情况与最高温度、最低温度和平均温度等气象因素的关系并建立了日用电量函数模型。上述方法均能在一定程度上对日用电量进行量化分析,但它们大多仅采用简单的线性模型来拟合日用电量与气象因素的函数关系,函数的拟合效果仍有提升的空间;且这些方法的区域特点比较明显,模型存在一定的局限性。

本文根据上述研究存在的问题,提出一种迎峰度夏期间日用电量的量化分析方法。首先利用电力公司迎峰度夏期间日用电量与天气数据通过皮尔森相关系数分析数据间的关系;然后利用回归分析算法构建相关性较强的指标间的函数关系;并基于可决系数衡量构建的函数关系与原始数据样本间的拟合情况,最终给出电力公司迎峰

度夏期间日用电量分析方法,并结合实际数据对未来几天的日用电量进行预测。实验结果表明,该方法对迎峰度夏期间日用电量预测的精度较高,可为电网设备迎峰度夏期间的安全运行提供数据支撑,帮助电力公司充分做好迎峰度夏准备工作。

## 2 数据分析算法

### 2.1 皮尔森相关系数

皮尔森相关系数又称相关系数<sup>[10-11]</sup>,它描述了两个定距变量间联系的紧密程度。样本的简单相关系数一般用  $r$  表示,如式(1)所示:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

其中, $n$  为样本量, $\bar{x}$ 、 $\bar{y}$  分别为两个变量的均值, $x_i$  和  $y_i$  为当前变量,其中, $i \in n$ 。 $r$  描述的是两个变量间线性相关强弱的程度。 $r$  的取值在-1 与 1 之间,若  $r > 0$ ,表明两个变量是正相关,即一个变量的值越大,另一个变量的值也会越大;若  $r < 0$ ,表明两个变量是负相关,即一个变量的值越大,另一个变量的值反而会越小。 $r$  的绝对值越大,表明相关性越强。

### 2.2 回归分析算法

回归分析是一种预测建模的技术,其中,被估计的目标变量是连续的。回归应用的例子包括利用其他经济学指标预测股市指数,基于高空气特征流预测一个地区的降水量,根据广告开销预测公司的总销售额。

假设  $D$  指包含  $n$  个预测的数据集  $D = \{(x_i, y_i) | i = 1, 2, \dots, n\}$ 。每个  $x_i$  对应于第  $i$  个预测的属性集(又称作说明变量),而  $y_i$  对应于目标变量(或因变量)。回归任务的属性可以是离散或者连续的。

回归分析的算法较多,主要包括简单线性回归算法、最小二乘法、SVM 回归算法、神经网络

回归算法、 $L_{1/2}$ 稀疏迭代回归算法等。

## 2.3 最小二乘法

最小二乘法是回归分析中的一种常用方法。已知第2.2节中所述的包含 $n$ 个预测的数据集 $D$ ，通过最小二乘法可以拟合出一条曲线使误差的平方和最小。对于一次拟合曲线计算如下。

(1) 计算式(2)和式(3)的值具体如下：

$$h_{11} = n, h_{12} = h_{21} = \sum_{i=1}^n x_i, h_{22} = \sum_{i=1}^n x_i^2 \quad (2)$$

$$l_1 = \sum_{i=1}^n y_i, l_2 = \sum_{i=1}^n x_i y_i \quad (3)$$

其中， $n$ 为数据条数。

(2) 根据式(2)和式(3)的值构建如下计算式：

$$H \cdot A = L \quad (4)$$

其中， $H = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}$ ， $A = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$ ， $a_0$ 为拟合曲线常数项系数， $a_1$ 为一次项系数，即为方程组：

$$\begin{cases} h_{11} \cdot a_0 + h_{12} \cdot a_1 = l_1 \\ h_{21} \cdot a_0 + h_{22} \cdot a_1 = l_2 \end{cases} \quad (5)$$

由此，即可解得 $a_0$ 和 $a_1$ 。

对于二次拟合曲线计算如下。

(1) 计算式(6)~式(9)的值：

$$h_{11} = n, h_{12} = h_{21} = \sum_{i=1}^n x_i \quad (6)$$

$$h_{13} = h_{22} = h_{31} = \sum_{i=1}^n x_i^2 \quad (7)$$

$$h_{23} = h_{32} = \sum_{i=1}^n x_i^3, h_{33} = \sum_{i=1}^n x_i^4 \quad (8)$$

$$l_1 = \sum_{i=1}^n y_i, l_2 = \sum_{i=1}^n x_i y_i, l_3 = \sum_{i=1}^n x_i^2 y_i \quad (9)$$

其中， $n$ 为数据条数。

(2) 根据式(6)~式(9)的值构建方程，如式(10)：

$$H \cdot A = L \quad (10)$$

其中， $H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$ ， $A = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}$ ， $a_0$ 为拟合

曲线常数项系数， $a_1$ 为一次项系数， $a_2$ 为二次项系数，即为方程组：

$$\begin{cases} h_{11} \cdot a_0 + h_{12} \cdot a_1 + h_{13} \cdot a_2 = l_1 \\ h_{21} \cdot a_0 + h_{22} \cdot a_1 + h_{23} \cdot a_2 = l_2 \\ h_{31} \cdot a_0 + h_{32} \cdot a_1 + h_{33} \cdot a_2 = l_3 \end{cases} \quad (11)$$

由此，即可解得 $a_0$ 、 $a_1$ 和 $a_2$ 。

## 2.4 可决系数

可决系数是衡量样本回归线与样本观测点拟合程度的统计量，用 $R^2$ 表示。 $R^2$ 的计算公式为：

$$R^2 = \frac{\left[ \frac{\sum_{i=1}^n (y'_i - \bar{y})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (y'_i - \bar{y})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \right]^2 \quad (12)$$

其中， $n$ 为样本量， $y$ 为样本真实值， $y'$ 为拟合值， $\bar{y}$ 为样本所有观测点的平均值。样本观测点与拟合的回归线靠得越近，回归拟合程度越好，由样本回归做出解释的离差平方和与总离差平方和越相近， $R^2$ 值越大；反之，回归拟合程度越差，由样本回归做出解释的离差平方和与总离差平方和相差越大， $R^2$ 值越小。

## 3 迎峰度夏用电量分析方法

首先分析电力公司迎峰度夏期间日用电量与日最高气温间的相关性。利用日用电量数据与日最高气温数据进行基于皮尔森相关系数的相关性分析，用量化的方法检查数据间的相关关系。其计算方法如式(13)所示：

$$r_{qt} = \frac{\sum_{i=1}^n (q_i - \bar{q})(t_i - \bar{t})}{\sqrt{\sum_{i=1}^n (q_i - \bar{q})^2 \sum_{i=1}^n (t_i - \bar{t})^2}} \quad (13)$$

其中， $r_{qt}$ 为日用电量与日最高气温的皮尔森相关系数， $n$ 为样本量即迎峰度夏天数， $\bar{q}$ 、 $\bar{t}$ 分别为日用电量和日最高气温在 $n$ 天内的均值， $q_i$ 和 $t_i$ 为第 $i$ 天的日用电量和日最高气温样本，其中， $i \in n$ 。

完成相关性分析后，利用回归分析算法对电力



公司迎峰度夏期间的日用电量  $\{q_i | i=1, 2, \dots, n\}$  与日最高气温  $\{t_i | i=1, 2, \dots, n\}$  之间建立函数关系。以日最高气温为输入，日用电量为输出，采用最小二乘法进行函数参数拟合，首先计算二次拟合曲线：

$$h_{11} = n, h_{12} = h_{21} = \sum_{i=1}^n t_i \quad (14)$$

$$h_{13} = h_{22} = h_{31} = \sum_{i=1}^n t_i^2 \quad (15)$$

$$h_{23} = h_{32} = \sum_{i=1}^n t_i^3, h_{33} = \sum_{i=1}^n t_i^4 \quad (16)$$

$$l_1 = \sum_{i=1}^n q_i, l_2 = \sum_{i=1}^n t_i q_i, l_3 = \sum_{i=1}^n t_i^2 q_i \quad (17)$$

根据上述值构建如下方程组：

$$\begin{cases} h_{11} \cdot a_0 + h_{12} \cdot a_1 + h_{13} \cdot a_2 = l_1 \\ h_{21} \cdot a_0 + h_{22} \cdot a_1 + h_{23} \cdot a_2 = l_2 \\ h_{31} \cdot a_0 + h_{32} \cdot a_1 + h_{33} \cdot a_2 = l_3 \end{cases} \quad (18)$$

其中， $a_0$  为拟合曲线常数项系数， $a_1$  为一次项系数， $a_2$  为二次项系数。通过求解方程组，即可解得二次拟合函数参数  $a_0$ 、 $a_1$ 、 $a_2$ ，若解得的二次项系数  $a_2$  为负，则计算一次拟合曲线：

$$h_{11} = n, h_{12} = h_{21} = \sum_{i=1}^n t_i, h_{22} = \sum_{i=1}^n t_i^2 \quad (19)$$

$$l_1 = \sum_{i=1}^n q_i, l_2 = \sum_{i=1}^n t_i q_i \quad (20)$$

根据上述值构建如下方程组：

$$\begin{cases} h_{11} \cdot a_0 + h_{12} \cdot a_1 = l_1 \\ h_{21} \cdot a_0 + h_{22} \cdot a_1 = l_2 \end{cases} \quad (21)$$

其中， $a_0$  为拟合曲线常数项系数， $a_1$  为一次项系数。通过求解方程组，解得一次拟合函数参数  $a_0$ 、 $a_1$ ，即得到电力公司迎峰度夏期间日用电量随日最高气温变化的模型。

解得拟合函数参数后，根据可决系数评判拟合函数的拟合程度，可决系数在 0 到 1 之间，值越大，表示函数对数据点的拟合程度越高；值越小，表示函数对数据点的拟合程度越低，可决系数在 0.5 以上表示拟合的效果较好。可决系数的计算方法

如式 (22) 所示：

$$R^2 = \frac{\left[ \sum_{i=1}^n (q'_i - \bar{q}') (q_i - \bar{q}) \right]^2}{\sqrt{\sum_{i=1}^n (q'_i - \bar{q}')^2} \sqrt{\sum_{i=1}^n (q_i - \bar{q})^2}} \quad (22)$$

其中， $R^2$  表示拟合函数可决系数， $n$  为样本量即迎峰度夏天数， $\bar{q}$  和  $\bar{q}'$  分别为日用电量真实值以及拟合值在  $n$  天内的均值， $q_i$  和  $q'_i$  分别为第  $i$  天的日用电量真实值以及拟合值。

#### 4 实证研究

本文利用 2015 年迎峰度夏期间国家电网公司经营区域内 23 个电力公司的日用电量数据与其经营区域内日最高气温数据进行分析。某电力公司迎峰度夏期间日用电量及其日最高气温如图 1 所示。通过图 1 可以看出，迎峰度夏期间日用电量与日最高气温之间存在较强的正相关，在日最高气温高于一定值后，日最高气温越高，日用电量越大。首先对日用电量与日最高气温进行相关性分析，利用式 (13) 计算皮尔森相关系数，该电力公司迎峰度夏期间日用电量与日最高气温的皮尔森相关系数为 0.81，相关性较强。对各电力公司进行相关性分析均可得到类似的较强正相关关系，由此表明，迎峰度夏期间电力公司日用电量受日最高气温变化影响较明显。

然后，利用各电力公司 2015 年 6—8 月日用电量与日最高气温数据通过回归分析算法建立函数关系。先对日最高气温相同的样本对应的日用电量数据取平均值作为创建函数关系的新样本，将不同日期的日用电量数据均衡化，使取平均值后的日用电量与短期经济变化无关；以日最高气温为输入，日用电量为输出，利用最小二乘法进行函数参数拟合，按照式 (18) 和式 (21) 构建的拟合函数参数的方程组对参数进行计算。函数拟合结果表明，在 23 个电力公司中，迎峰度夏期间日用电量与日最高气温之间为二次函数关系的公司有 19 个，

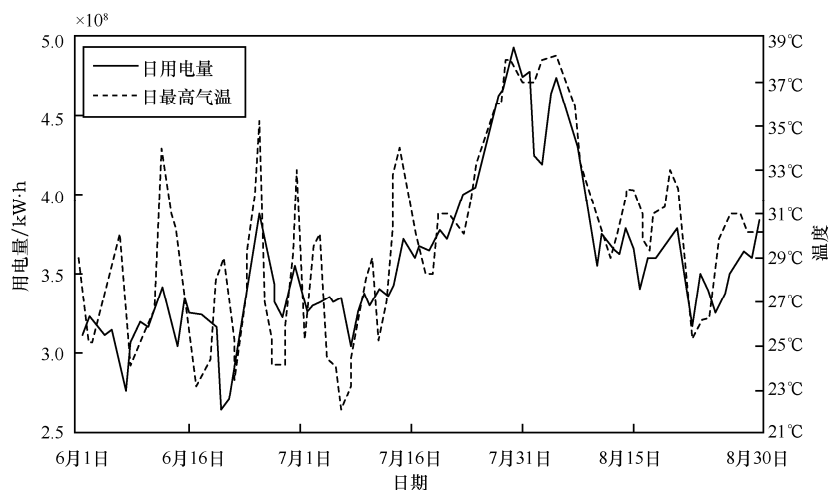


图1 某电力公司迎峰度夏期间日用电量与日最高气温

为一次函数关系的公司有4个。

再利用式(22)计算可决系数以评判拟合函数的拟合程度。某电力公司迎峰度夏期间日最高气温与日用电量的函数关系如图2所示。通过图2可以看出,该电力公司的函数拟合可决系数为0.97,拟合程度较高,且该公司迎峰度夏期间日用电量与日最高气温存在明显的二次函数关系,二次函数对称轴为日最高气温  $t=24.45^{\circ}\text{C}$ ,以二次函数的对称轴为阈值,当日最高气温超过这个阈值后,日用电量随日最高气温的增加而增加。对各电力公司日用电量与日最高气温的拟合函数计算可决系数可知,利用本文提出的拟合方法,可以较好地量化电力公司迎峰度夏期间日用电量随日最高气温变化的关系。

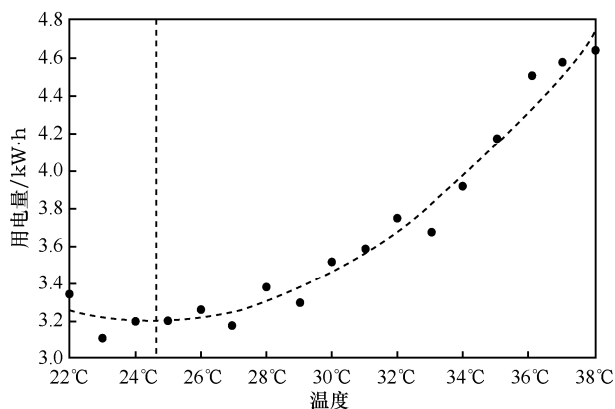


图2 某电力公司迎峰度夏期间日最高气温与日用电量的函数关系

利用迎峰度夏期间日最高气温与日用电量的函数关系,可以对未来几天的日用电量进行预测。利用本文提出的分析方法得到的某电力公司2015年6月1日—8月25日的日最高气温和日用电量的函数关系如图3所示,利用该函数计算从8月25日的最高气温变化为8月26—28日的最高气温时的用电量变化量,将该用电量变化量与8月25日的日用电量相加即可得到8月26—28日的日用电量预测值。例如,8月26日的日用电量预测值可由式(23)计算:

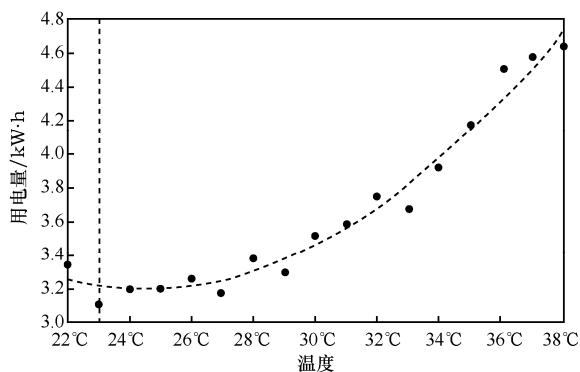


图3 某电力公司2015年6月1日—8月25日的日最高气温与日用电量的函数关系

$$q_{26} = q_{25} + (q_{t_{26}} - q_{t_{25}}) \quad (23)$$

其中,  $q_{t_{25}}$  为8月25日的日用电量实际值,  $q_{t_{26}}$  为将8月26日的最高气温预报值代入日最高气温和日用电量的函数关系时计算得到的日用电量值,



$q_{r25}$  为将 8 月 25 日的日最高气温实际值代入日最高气温和日用电量的函数关系时计算得到的日用电量值。根据 8 月 26—28 日的日用电量实际值计算得到预测误差分别为 0.22%、1.75% 和 3.42%，预测精度较高。如果进一步考虑节假日、空气湿度等其他因素，可以得到更好的预测结果。

## 5 结束语

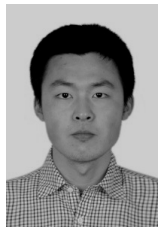
本文提出了一种电力公司迎峰度夏期间日用电量的分析方法，该方法首先通过电力公司迎峰度夏期间日用电量数据与日最高气温数据进行了相关性研究，并利用皮尔森相关系数对相关程度进行了量化；再利用回归分析算法对日用电量数据与日最高气温数据之间建立了函数关系并通过最小二乘法求解得到了日用电量随最高温度变化的定量分析结果。通过实证研究发现，迎峰度夏期间电力公司的日用电量与其经营区域内日最高气温之间存在较强的正相关，经过函数拟合，其函数关系均呈二次函数或一次函数关系，且存在一个阈值使日最高气温高于该阈值时日用电量随日最高气温的增长而增加。实验结果表明，利用该方法建立的函数关系在电力公司迎峰度夏期间气温发生变化时能够很好地对日用电量的变化情况进行量化，还能够对未来几天的日用电量进行预测，为电网设备迎峰度夏期间的安全运行提供数据支撑，对电力公司充分做好迎峰度夏准备工作、提高电力公司营销、营业管理水平具备一定的指导意义。此外，在电力公司迎峰度冬期间，也可以采用类似方法对温度变化时用电量的变化情况进行量化。

## 参考文献:

- [1] 刘勇, 李淼, 李大虎, 等. 2010 年湖北电网迎峰度夏运行分析[J]. 湖北电力, 2010, 34(S2):1-6.  
LIU Y, LI M, LI D H, et al. The analysis of Hubei Power Grid operation in peak load period of summer 2010[J]. Hubei Electric Power, 2010, 34(S2):1-6.
- [2] 季斌, 黄松泉. 2011 年迎峰度夏湖北城市配网运行分析[J]. 湖北电力, 2011, 35(S1):34-41.  
JI B, HUANG S Q. The analysis of Hubei urban power grid operation in peak load period of summer 2011[J]. Hubei Electric Power, 2011, 35(S1):34-41.
- [3] 程建翼, 陈明珠. 2011 年迎峰度夏湖北电网可靠性分析[J]. 湖北电力, 2011, 35(S1):74-76.  
CHENG J Y, CHEN M Z. The analysis of the reliability of Hubei Power Grid in peak load period of summer 2011[J]. Hubei Electric Power, 2011, 35(S1):74-76.
- [4] 张俊, 陈卫勇, 喻子易, 等. 2011 年迎峰度夏农网运行情况分析[J]. 湖北电力, 2011, 35(S1):42-50.  
ZHANG J, CHEN W Y, YU Z Y, et al. The analysis of rural power grid operation in peak load period of summer 2011[J]. Hubei Electric Power, 2011, 35(S1):42-50.
- [5] 胡元, 徐沐阳, 程丽君. 考虑气象因子的温州市日用电量指数预报[J]. 中国科技信息, 2014, (23):55-56.  
HU Y, XU M Y, CHENG L J. Forecasting daily power consumption of Wenzhou considering meteorological factors[J]. China Science and Technology Information, 2014, (23):55-56.
- [6] 郭广芬, 陈正洪, 汪金福, 等. 华中区域夏季日用电量气象预报模型研究[J]. 华中师范大学学报(自然科学版), 2009, 43(2):327-331.  
GUO G F, CHEN Z H, WANG J F, et al. Research on meteorological forecast model of daily power consumption during summer in central China[J]. Journal of Huazhong Normal University (Natural Science), 2009, 43(2):327-331.
- [7] 洪国平, 李银娥, 孙新德, 等. 武汉市电网用电量、电力负荷与气温的关系及预测模型研究[J]. 华中电力, 2006, 19(2):4-7.  
HONG G P, LI Y E, SUN X D, et al. Relation and forecast model research between the electricity consumption, the load and the air temperature in Wuhan[J]. Central China Electric Power, 2006, 19(2):4-7.
- [8] 王桂新, 沈续雷. 气温变化对上海市日电力消费影响关系之考察[J]. 华北电力大学学报(社会科学版), 2015, 93(1):35-41.  
WANG G X, SHEN X L. Impact of temperature change on daily electricity consumption in shanghai[J]. Journal of North China Electric Power University (Social Sciences), 2015, 93(1):35-41.
- [9] 赵彤, 孙大雁, 葛诚, 等. 江苏电网夏季气温与用电量敏感性关系初探[J]. 电力需求侧管理, 2004, 6(6):20-21.  
ZHAO T, SUN D Y, GE C, et al. Research of sensitive relation between summertime temperature and electric power consuming in Jiangsu province[J]. Power Demand Side Management, 2004, 6(6):20-21.
- [10] SHEN X. Symmetry recurrence in protein sequence and structure with pearson's correlation coefficients[C]//Proceedings of the 2012 IEEE International Conference on Biomedical and Health Informatics, January 4-7, 2012, Hong Kong, China. Piscataway: IEEE Press, 2012: 832-835.

- [11] LEO E, LOET L. The relation between pearson's correlation coefficient  $r$  and salton's cosine measure[J]. Journal of the American Society for Information Science and Technology, 2009, 60(5):1027-1036.

#### [作者简介]



陈雨泽（1988-），男，国网信息通信产业集团有限公司工程师，主要研究方向为电力营销大数据分析。

李磊（1989-），男，国家电网有限公司工程师，主要研究方向为电力营销大数据分析。

方学民（1972-），男，国家电网有限公司高级工程师，主要研究方向为电力信息化及电力大数据分析。

刘建（1989-），男，国网信息通信产业集团有限公司工程师，主要研究方向为电力营销大数据分析。

赵加奎（1979-），男，博士，国家电网有限公司大数据中心教授级高级工程师，主要研究方向为电力信息化及电力大数据分析。

刘玉玺（1984-），男，博士，国网信息通信产业集团有限公司高级工程师，主要研究方向为电力营销大数据分析。

左松林（1981-），男，国网安徽省电力有限公司高级工程师，主要研究方向为电力信息化及电力大数据分析。