

DOI:10.12158/j.2096-3203.2020.03.024

基于弹性网络模型的月度用电量预测方法

胡春风, 田世明, 苏航

(中国电力科学研究院有限公司, 北京 100192)

摘要:由于现有月度用电量预测所选影响因素较少,无法较为全面地反映与用电量强关联的因素,同时针对高维数据变量筛选和高精度预测等突出难题,文中提出了一种弹性网络用电量预测模型。为了考虑更为全面的影响因素,建立了用电量、气象、经济、交通4类,共340个变量的数据集。首先对8年96个点的高维变量数据进行弹性网络因子筛选,然后使用Granger因果关系分析找出了用电量数据与其它数据的关联关系,对一年范围内的全社会月度用电量使用弹性网络进行预测,预测结果的平均绝对百分误差为3.07%。为验证该模型的有效性,对比向量自回归(VAR)模型,反向传播(BP)模型和最小绝对值收缩和选择算子(Lasso)预测的效果,验证了文中所提方法预测精度较高。

关键词:弹性网络;最小绝对值收缩和选择算子;Granger因果关系;因子筛选;用电量预测

中图分类号:TM71

文献标志码:A

文章编号:2096-3203(2020)03-0166-07

0 引言

针对大区域进行月度用电量预测对电力市场运行和地区电量平衡等具有重要作用,是电力计划部门、用电、营销部门的重要工作,目的是为了合理安排电力系统的中期运行计划,降低运行成本,提高供电可靠性。

目前月度用电量预测方法^[1]主要有:回归分析法^[2]、动平均法、神经网络法、灰色系统法,以及组合预测^[3-4]等。向量自回归(vector autoregression, VAR)模型是处理多元时间序列的经典模型,文献[5]采用VAR模型进行了季度性预测,但只进行了简单预测并未得出预测效果结论。回归分析法以及动平均法拟合效果较好,但预测效果不佳。神经网络法模型复杂、参数较多,适用于数据量大的场合,数据量小容易在训练集上过拟合使预测精度大大降低。灰色系统法^[6-9]与通过大量样本进行分析的方法相比,不要求原始数据越多越好,但应用于长期预测时,对随机波动性较大的数列拟合效果可能不好。组合预测模型的拟合方差和预测相对误差总体上一般优于单一模型,但受随机因素的影响较大。文献[10]利用自回归移动平均模型(autoregressive integrated moving average model, ARIMA)和反向传播(back propagation, BP)神经网络模型进行用电量预测,但这一方法主要适用于数据量小、影响因素也较少的情况,对高维数据变量筛选以及高

精度预测方法的研究还有缺失。

使用最小绝对值收缩和选择算子(least absolute shrinkage and selection operator, Lasso)进行电力相关的预测技术已经相对成熟^[11-14]。文献[15]使用Lasso算法对电力负荷及相关天气因素大数据进行高维数据特征提取,简化预测模型,加快预测速度。弹性网络^[16]与Lasso相比,可以在持续收缩的同时自动进行特征选择,由于 L_2 -范数的作用,弹性网络表现出了比Lasso更好的稳定性。弹性网络模型可以挑选出成组相关的特征,也可以选择多于样本数的特征以使特征达到饱和^[17]。目前尚未有文献使用弹性网络进行用电量预测。

文中对用电量、气象、经济、交通4类,共340个变量,96个时间点(每年12个月,1个月1个点,共8a)的月度数据进行关联分析。相比于以往的月度预测,考虑的影响因子更为全面。首先对高维变量进行弹性网络因子筛选,并使用Granger因果关系分析找出了用电量数据与其它数据的关联关系,使数据维数大大降低。对一年范围内的全社会月度用电量进行预测,预测结果的平均绝对百分误差为3.07%。对比VAR模型、BP模型和Lasso,预测结果表明文中所选方法精度较高,验证了该方法的可行性及有效性。

1 弹性网络及Granger因果关系检验模型

1.1 弹性网络

弹性网络是岭回归和Lasso的动态混合。岭回归和Lasso的差别在于惩罚项不同,前者使用的是系数值的平方和(见式(1)),而Lasso使用的是系数

收稿日期:2019-11-03;修回日期:2019-12-15

基金项目:国家电网有限公司总部科技项目“支持电力大数据分析的核心算法改进及其实用化技术”(520940180016)

绝对值的和(见式(2))。

$$\operatorname{argmin} \left[\sum_{i=1}^N (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^k \beta_j^2 \right] \quad (1)$$

式中: N 为样本数量; y_i 为因变量真实值; \hat{y}_i 为因变量预测值; λ 为惩罚参数; k 为样本特征数; β 为回归系数; $\lambda \sum_{j=1}^k \beta_j^2$ 为 L_2 正则化项。

$$\operatorname{argmin} \left[\sum_{i=1}^N (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^k |\beta_j| \right] \quad (2)$$

式中: $\lambda \sum_{j=1}^k |\beta_j|$ 为 L_1 正则化项。

岭回归的优点是可以提高预测准确度,但因为它不能使任何一个特征系数为 0,只能使系数无限趋于 0,所以模型较难解释。引入 L_1 正则化项,使得 Lasso 不仅可以像岭回归那样收缩变量,还可以进行变量筛选,Lasso 可以把某些待估系数精确地收缩到 0,极大地提高了模型的解释性。但在高共线性或高度两两相关的情况下,Lasso 可能会将某个预测特征强制删除,这会损失模型的预测能力^[18]。所以,2 个模型有不同的适用场合,依赖于输入特征的某个子集的模型往往用 Lasso 表现更好,而很多不同变量的系数具有较大分散度的模型往往在岭回归下有更好的表现。

弹性网络为:

$$\min \left[\frac{1}{2N} \sum_{i=1}^N (y_i - \beta_0 - x_i^T \beta)^2 + \lambda P_\alpha(\beta) \right] \quad (3)$$

其中:

$$P_\alpha(\beta) = (1 - \alpha) \frac{1}{2} \|\beta\|_2^2 + \alpha \|\beta\|_1 \quad (4)$$

式中: λ 为复杂参数,控制着压缩的程度(0 表示没有判罚, ∞ 表示完全判罚); α ($0 \leq \alpha \leq 1$) 为弹性网络混合参数,控制着结果中有多大程度是岭回归和 Lasso 回归, α 等于 0 是完全的岭回归,等于 1 是完全的 Lasso 回归; x_i 为自变量; $\|\beta\|_2^2$ 是岭回归项; $\|\beta\|_1$ 是 Lasso 项。

弹性网络的强大之处在于,既能做到岭回归不能做的特征提取,也能实现 Lasso 不能做的特征分组^[19]。其中,特征提取是指通过某种变换(通常指正交变换),将原始数据从高维空间映射到低维空间,进而在低维空间中找到最能表征原始数据本质信息的少量综合特征,通常将其作为一种数据预处理方法来克服高维数据的“维数灾难”问题。由于输入数据特征的品质直接决定预测算法的性能,因此如何提取出有效的特征是至关重要的。特征分组是将特征按其相关性进行分组,将高维的特征分

为一组组低维的特征。

进行弹性网络建模时,需要解出复杂参数 λ 和混合参数 α ,可以按照如下规则试验这 2 个超参数。选择最优模型的准则是均方误差(mean squared error, MSE)最小。

α 从 0 到 1,每次增加 0.1; α 必须被绑定在 0 到 1 之间。

λ 从 0 到 0.8,每次增加 0.02;0.8 的 λ 值稍大于岭回归 λ 值(0.78,调参得到)以及 Lasso λ 值(0.5,调参得到)最大值的一个设定值。

1.2 Granger 因果关系检验

Granger 因果关系检验能够检验出两个变量的因果关系强弱。所以在使用弹性网络进行第一步因子筛选后,为了使所筛选变量更为合理,采用 Granger 因果关系检验进行第二步因子筛选。

考虑 2 个时间序列 X 和 Y ,由 Granger 提出的关于 X 和 Y 的 Granger 因果检验用来评估 X 过去的观测值对于预测 Y 是否有用^[20]。

$$Y_t = a_0 + a_1 Y_{t-1} + \dots + a_p Y_{t-p} + b_0 + b_1 X_{t-1} + \dots + b_p X_{t-p} + \varepsilon_t \quad (5)$$

式中: a, b 为系数; p 为 X 过去的观测值个数; ε_t 为误差项。

如果系数 b_1, \dots, b_p 都显著不为 0,则称变量 X 是 Y 的“Granger 原因”。这个检验虽然在原理和计算上都很简单,但在计量经济学研究中很有用。

2 算法流程及输出评估

2.1 算法流程

通过人工或者传统方法选择变量有一定难度,随着计算机技术的快速发展和计算能力的提高,使用计算机进行自动筛选越来越受到重视。所以文中首先建立一个变量尽可能多的数据集,再用大数据的方法对变量进行筛选。

综合以上算法介绍,基于弹性网络模型的月度用电量预测方法流程如图 1 所示。

步骤一:对输入数据进行预处理。针对数据中存在缺失值的问题,采用多重插补(multiple interpolation, MI)法处理。MI 是一种基于重复模拟的处理缺失值的方法,在面对复杂的缺失值问题时,MI 是最常选用的方法,它将从一个包含缺失值数据集中生成一组完整的数据集,每个模拟数据集中,缺失数据将用蒙特卡罗法来填补。此时,标准的统计方法便可应用到每个模拟的数据集上,通过组合输出结果给出估计结果,以及引入缺失值的置信区间。缺失值处理后再进行归一化处理。

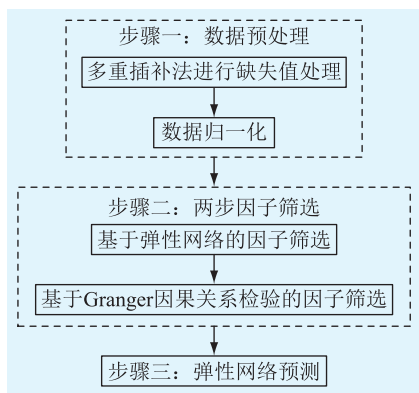


图1 算法流程

Fig.1 Flow chart of the proposed method

步骤二:对缺失值处理后的数据进行弹性网络和 Granger 因果关系检验两步因子筛选。

步骤三:使用弹性网络模型针对筛选出来的变量进行月度用电量预测。

2.2 输出评估

对于预测效果的评估采用相对误差绝对值和平均绝对百分误差。

假设用 $y_i, i \in [1, n]$ 表示真实值,用 $\hat{y}_i, i \in [1, n]$ 表示预测值,则相对误差的绝对值定义为:

$$e_i = \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad i \in [1, n] \quad (6)$$

平均绝对百分误差 (mean absolute percentage error, MAPE) 是相对误差绝对值之和的平均值,与单纯的相对误差相比,避开了正负相对误差不能相加的问题。通过求平均值,也反映了预测相对误差的平均水平,它是经常用到的数值预测评估指标,一般认为, MAPE 小于 10%, 预测效果已经不错了^[21]。其定义如下:

$$M_{MAPE} = \frac{1}{n} \sum_{i=1}^n |e_i| = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (7)$$

式中: n 为样本数。

由于后面需对误差进行累加,故在缺失值处理后,对所有指标按均值和标准差进行标准化处理。

3 算例分析

3.1 数据源及数据预处理

以下数据均为 2011—2018 年的月度数据。用电量数据来自上海统计局,包含上海市的用电量信息:全社会用电量、第一产业用电量、第二产业用电量、第三产业用电量、城乡居民生活用电量。

气象数据来自中国气象数据网上海宝山站,包含:极大风速、最低气压、最低气温、最高气压、最高

气温、降水量、平均气压、平均 2 min 风速、平均气温、平均水汽压、平均相对湿度、平均最低气温、平均最高气温、日降水量 0.1 mm 日数、日照时数、最大风速、最大日降水量。

经济数据来自上海统计局,包含上海市经济信息:工业生产情况、规模以上工业总产值(按行业分)、规模以上工业企业主要经济效益指标、规模以上工业主营业务收入(按行业分)、规模以上工业利润总额(按行业分)、规模以上工业主要产品产量、房地产开发经营、进出口商品总额、外商直接投资、国际旅游入境人数、星级饭店、居民消费价格指数、工业生产者出厂价格指数、工业生产者购进价格指数。

交通数据来自上海统计局,包含:货物运输总量(铁路、水运、公路、机场)、港口货物吞吐量(进港量、出港量)、国际标准集装箱吞吐量(进港量、出港量)、机场旅客吞吐量。

克强指数包括用电量、铁路货运量和贷款发放量,是用于评估中国 GDP 增长量的指标。文中在收集数据时作为被筛选的对象考虑了交通数据,是因为它间接影响用电量。

经识别,数据中存在缺失值,采用 2.1 节介绍的 MI 方法进行缺失值处理,得到用电量数据、气象数据、经济数据、交通数据共 340 列、96 行。取 2011—2017 共 7 a 84 个时间点的数据作为训练集,2018 年 12 个时间点的数据作为验证集。

3.2 基于弹性网络的因子筛选

考虑篇幅,此处仅用气象数据、交通数据、经济数据对全社会用电量数据进行弹性网络建模。对于全社会用电量使用弹性网络建模时,当 α 取 0.9、 λ 取 0.72 时, MSE 最小。图 2 为对全社会用电量弹性网络建模时的交叉验证 (cross validation, CV) 统计图^[18]。使用 K 折交叉验证法,结果中包含每次拟合的 λ 值和响应的 MSE,由于训练集中数据量的原因,只分 3 折。图 2 曲线表示 $\log\lambda$ 和 MSE 之间的关系,曲线上每个红点都对应灰色的误差棒^[22],上方坐标表示模型中特征的数量。2 条垂直的虚线分别表示取得 MSE 最小值的 $\log\lambda$ (左侧虚线)和距离最小值的一个标准误差的 $\log\lambda$ (右侧虚线)。此处选择左侧虚线来尽量保留变量的数目,由图 2 可以看出保留变量 29 个。具体保留变量内容见表 1, C_1 为弹性网络建模各变量的系数; C_2 为 Granger 因果关系检验的系数。由表 1 可以看出:

- (1) 气象量保留了 1 个:平均水汽压。
- (2) 经济量保留 24 个:其中第二产业经济量保

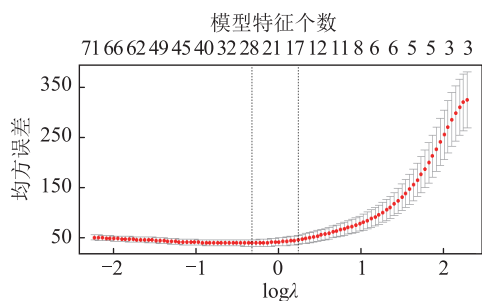


图2 对全社会用电量弹性网络建模时的 CV 统计

Fig.2 The CV statistics chart of the whole society electricity consumption elastic network modeling

留 13 个;第三产业经济量保留 9 个。第二产业是用电量最多的产业,所以它与用电量的关联性最强,第三产业次之。外向型经济量保留 2 个,这与上海作为外向型经济城市的实际情况是相符的。

(3) 交通量保留了 2 个:国际标准集装箱出港

量和机场旅客吞吐量,这也与上海作为外向型经济城市的实际情况是相符的。

变量保留情况与城市的发展情况有关,即经济情况不同的城市筛选出的变量是不同的,但方法是普适的。

3.3 基于 Granger 因果关系检验的因子筛选

以表 1 中保留的 29 个变量为解释变量,以全社会用电量为目标变量进行 Granger 因果关系检验。Granger 因果检验值越小,表明因果关系越显著。分析时着重考虑值小于 0.2 的因果关系。由表 1 可以看出,总是先于全社会用电量变化的变量有 6 个,分别为(按因果关系从强到弱排列):食品制造业总产值、其他工业原材料及半成品类工业生产者购进价格指数、发电量、文化娱乐居民消费价格指数、国有经济总产值、电力热力生产和供应业总产值。

表 1 弹性网络和 Granger 因果检验变量选择
Table 1 Elastic network and Granger causal test variable selection

变量名	变量类型	C_1	C_2
平均水汽压	气象量	0.776 782 132	0.635 455 406
国际标准箱集装箱出港量	交通量	0.000 869 613	0.522 933 048
机场旅客吞吐量	交通量	0.041 182 055	0.755 961 271
发电量	电力量	0.699 423 671	0.109 435 949
国有经济总产值	经济量	0.002 719 318	0.154 772 233
食品制造业总产值	二产经济量	0.011 984 078	0.023 621 26
烟草制品业总产值	二产经济量	0.000 201 915	0.247 704 373
家具制造业总产值	二产经济量	0.230 489 017	0.405 346 582
石油加工、炼焦和核燃料加工业总产值	二产经济量	-0.024 836 341	0.553 111 68
有色金属冶炼和压延加工业总产值	二产经济量	-0.051 173 351	0.438 951 13
电力、热力生产和供应业总产值	二产经济量	0.051 562 422	0.174 784 574
水的生产和供应业总产值	二产经济量	1.006 527 108	0.603 894 032
精品钢材制造业利润总额	二产经济量	-0.002 028 488	0.585 976 069
家具制造业利润总额	二产经济量	-0.122 769 909	0.884 963 104
水的生产和供应业利润总额	二产经济量	0.138 888 002	0.967 736 228
服装产量	二产经济量	0.003 485 408	0.233 716 978
乙烯产量	二产经济量	0.281 616 228	0.854 660 995
其他工业原材料及半成品类工业生产者购进价格指数	二产经济量	1.201 890 004	0.025 952 216
住宅销售面积	三产经济量	0.001 659 305	0.353 025 000
五星级饭店平均房价	三产经济量	-0.042 051 471	0.847 591 877
五星级饭店平均房价增长率	三产经济量	0.253 874 239	0.514 469 04
医疗保健居民消费价格指数	三产经济量	0.034 337 991	0.789 531 98
菜类居民消费价格指数	三产经济量	0.051 211 212	0.506 214 755
水产品居民消费价格指数	三产经济量	-0.367 245 503	0.411 728 607
干鲜瓜果类居民消费价格指数	三产经济量	-0.124 112 062	0.943 555 994
衣着居民消费价格指数	三产经济量	-0.262 572 608	0.643 337 031
文化娱乐居民消费价格指数	三产经济量	0.101 316 403	0.146 485 612
一般贸易进口总额	外向型经济量	0.008 714 142	0.903 056 695
高新技术产品进口总额	外向型经济量	0.066 202 624	0.917 091 419

以下对筛选出的变量进行合理性分析:

(1) 食品制造业总产值、其他工业原材料及半成品类工业生产者购进价格指数、电力热力生产和供应业总产值属于第二产业经济量。上海市第二产业发达,与用电量关联性是合理的。

(2) 发电量和用电量是一对具有供需关系的变量。

(3) 文化娱乐居民消费价格指数属第三产业经济量,上海服务业发展活跃,与用电量关联性是合理的。

(4) 克强指数表明,经济和电量密不可分,上海作为中国的一线城市,经济发达,这种关系会尤为明显,所以筛选出国有经济总产值是合理的。

通过弹性网络建模和 Granger 因果关系检验后,保留了先于全社会用电量变化的变量,数据维数由 340 降为 6,实现高维数据变量筛选的功能,大幅降低了计算的复杂度。

3.4 弹性网络预测

根据弹性网络和 Granger 因果关系检验两步因子筛选出 6 个变量,使用弹性网络对 2018 年 12 个月全社会用电量进行预测,预测结果见表 2 和图 3。

表 2 2018 年上海市全社会用电量预测结果

Table 2 Shanghai's total social electricity consumption forecast results in 2018

时间	实际值/ (亿 kW·h)	预测值/ (亿 kW·h)	相对误差 绝对值/%
2018-01	151.40	148.84	1.69
2018-02	112.87	110.71	1.91
2018-03	120.89	118.83	1.70
2018-04	106.84	107.24	0.37
2018-05	119.03	117.47	1.31
2018-06	126.79	123.15	2.87
2018-07	163.32	154.11	5.64
2018-08	169.55	157.76	6.95
2018-09	136.53	131.41	3.75
2018-10	110.10	106.69	3.10
2018-11	109.52	105.75	3.44
2018-12	139.82	134.03	4.14

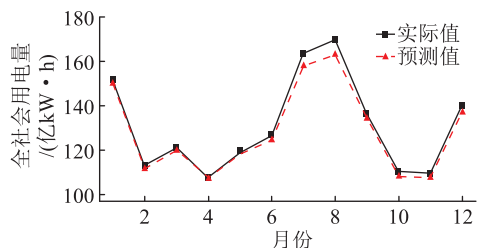


图 3 2018 年上海市全社会用电量预测结果

Fig.3 Shanghai's total social electricity consumption forecast results in 2018

由表 2 可以看出相对误差绝对值百分比在 7% 以内,从图 3 可以看出预测值曲线和实际值曲线基本吻合,取得了较好的预测效果。

3.5 预测结果对比

为了验证方法的可行性和有效性,将弹性网络模型、处理多变量时间序列的经典模型 VAR 模型、BP 神经网络以及 Lasso 的预测结果进行对比,预测曲线见图 4,预测误差对比如表 3 所示。

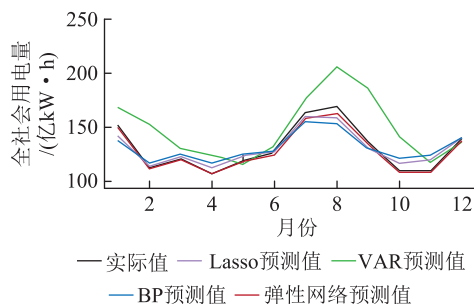


图 4 不同预测方法曲线比较

Fig.4 Curve comparison of different prediction methods

表 3 不同方法预测结果 MAPE 对比

Table 3 Comparison of different methods for predicting MAPE

方法	$M_{MAPE}/\%$	方法	$M_{MAPE}/\%$
VAR	15.10	Lasso	5.71
BP	10.08	弹性网络	3.07

不同模型的结果对比可以看出,文中所提弹性网络预测模型效果较佳,与实际值曲线吻合度高。

4 结语

文中通过分析气象、经济、交通多种因素对用电量的影响,提出了一种基于弹性网络模型的月度用电量预测方法。通过弹性网络和 Granger 因果检验两步因子筛选实现高维数据变量筛选功能,降低计算复杂度。对比 VAR 模型、BP 神经网络以及 Lasso,弹性网络模型可以挖掘出各种潜在关联因素和用电量之间的关联关系,以强关联因素数据作为输入,预测精度更高,从而为月度用电量预测提供一种新思路。

参考文献:

[1] 康重庆,夏清,刘梅. 电力系统负荷预测第二版[M]. 北京:中国电力出版社,2017:142-159.
KANG Chongqing, XIA Qing, LIU Mei. Second edition of power system load forecasting [M]. Beijing: China Electric Power Press, 2017: 142-159.

[2] 王雁凌,吴梦凯. 经济新常态下基于偏最小二乘回归的中长期负荷预测模型[J]. 电力自动化设备, 2018, 38(3): 133-139.
WANG Yanling, WU Mengkai. Medium and long term load fore-

- casting model based on partial least-square regression under new normal economy [J]. *Electric Power Automation Equipment*, 2018, 38(3):133-139.
- [3] 赵腾, 王林童, 张焰, 等. 采用互信息与随机森林算法的用户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.
- [4] DUAN J D, QIU X Y, MA W T, et al. Electricity consumption forecasting scheme via improved LSSVM with maximum correntropy criterion[J]. *Entropy*, 2018, 20(2):112.
- [5] 生雅明. 基于时间序列分析的电力需求预测研究[D]. 大连:大连理工大学, 2015.
SHENG Yaming. Research on electric power demand forecasting based on time series analysis[D]. Dalian: Dalian University of Technology, 2015.
- [6] 王雁凌, 吴梦凯, 周子清, 等. 基于改进灰色关联度的电力负荷影响因素量化分析模型[J]. *电网技术*, 2017, 41(6):1772-1778.
WANG Yanling, WU Mengkai, ZHOU Ziqing, et al. Quantitative analysis model of power load influencing factors based on improved grey relational degree [J]. *Power System Technology*, 2017, 41(6):1772-1778.
- [7] WU L F, GAO X H, XIAO Y L, et al. Using a novel multi-variable grey model to forecast the electricity consumption of Shandong province in China [J]. *Energy*, 2018, 157:327-335.
- [8] WANG Z X, LI Q, PEI L L. A seasonal GM(1, 1) model for forecasting the electricity consumption of the primary economic sectors [J]. *Energy*, 2018, 154:522-534.
- [9] XU N, DANG Y G, GONG Y D. Novel grey prediction model with nonlinear optimized time response method for forecasting of electricity consumption in China [J]. *Energy*, 2017, 118:473-480.
- [10] CRISTINA N, IULIA S, IOANA F, et al. Energy consumption forecasting using ARIMA and neural network models [C]// IEEE International Symposium on Electrical & Electronics Engineering, 2017.
- [11] TIBSHIRANI R J. Regression shrinkage and selection via the Lasso [J]. *Journal of the Royal Statistical Society: Series B (Methodological)*, 1996, 73(1):273-282.
- [12] ZIEL F, LIU B D. Lasso estimation for GEFCom2014 probabilistic electric load forecasting [J]. *International Journal of Forecasting*, 2016, 32(3):1-10.
- [13] 王金甲, 彭汝佳. 在线自适应 Lasso 罚向量自回归模型的风电功率预测[J]. *燕山大学学报*, 2018, 42(6):532-538, 551.
WANG Jinjia, PENG Rujia. An online adaptive Lasso penalty vector autoregressive model for wind power prediction [J]. *Journal of Yanshan University*, 2018, 42(6):532-538, 551.
- [14] 张航飞. 基于 Lasso-PCA 及改进自适应遗传神经网络的电力负荷预测研究[D]. 秦皇岛:燕山大学, 2018.
ZHANG Hangfei. Research on power load forecasting based on Lasso-PCA and improved adaptive genetic neural network [D]. Qinhuangdao: Yanshan University, 2018.
- [15] 张淑清, 任爽, 陈荣飞, 等. 基于大数据简约及 PCA 改进 RBF 网络的短期电力负荷预测 [J]. *计量学报*, 2018, 39(3):392-396.
ZHANG Shuqing, REN Shuang, CHEN Rongfei, et al. Short-term power load forecast based on big data reduction and PCA-improved RBF network [J]. *Acta Metrologica Sinica*, 2018, 39(3):392-396.
- [16] ZOU H, HASTIE T. Regularization and variable selection via the elastic net [J]. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2006, 67(2):301-320.
- [17] 张颖. 基于改进扩展弹性网络的多类别特征选择方法研究 [D]. 合肥:安徽大学, 2017.
ZHANG Ying. Research on multi-class feature selection with modified extended elastic net [D]. Hefei: Anhui University, 2017.
- [18] CORY L. 精通机器学习基于 R [M]. 北京:人民邮电出版社, 2018:60.
CORY L. Mastering machine learning with R [M]. Beijing: Posts & Telecom Press, 2018:60.
- [19] JARED P L. R 语言实用数据分析和可视化技术 [M]. 北京:机械工业出版社, 2015:222.
JARED P L. Advanced analytics and graphics [M]. Beijing: China Machine Press, 2015:222.
- [20] 吴喜之, 刘苗. 应用时间序列分析 R 软件陪同 [M]. 北京:机械工业出版社, 2018:193.
WU Xizhi, LIU Miao. Applied time series analysis with R [M]. Beijing: China Machine Press, 2018:193.
- [21] 游皓麟. R 语言预测实战 [M]. 北京:电子工业出版社, 2016:269.
YOU Haolin. R in predict action [M]. Beijing: Publishing House of Electronics Industry, 2016:269.
- [22] 吴林涛, 张军朋. Origin8.0 在物理实验数据处理中的应用 [J]. *大学物理实验*, 2012, 25(1):61-63.
WU Lintao, ZHANG Junpeng. Application of Origin8.0 in data processing in physics experiment [J]. *Physical Experiment of College*, 2012, 25(1):61-63.

作者简介:



胡春风

胡春风(1994),女,硕士在读,研究方向为智能配用电大数据应用研究(E-mail: m15600465900@163.com);

田世明(1965),男,硕士,教授级高级工程师,从事能源互联网、大数据等相关工作;

苏航(1995),男,硕士在读,研究方向为智能配用电大数据研究。

Monthly electricity consumption forecasting method based on elastic network model

HU Chunfeng, TIAN Shiming, SU Hang

(China Electric Power Research Institute, Beijing 100192)

Abstract: Since the existing monthly electricity consumption forecast has fewer influencing factors, and it is unable to comprehensively reflect the factors associated with strong electricity consumption. An elastic network electricity consumption forecasting model for high-dimensional data variable screening and high-precision prediction is proposed. The volume prediction model analyzes the monthly data of 340 variables and 96 time points for electricity consumption, economy, transportation, and meteorology. By using elastic network to screening for high-dimensional variables, and Granger causality analysis to find out the dependence of electricity consumption data and other data, the monthly electricity consumption of the whole society in a year is predicted. And the mean absolute percentage error of the prediction results is 3.07%. Compared with the VAR model, BP model and Lasso, the feasibility and effectiveness of the method are verified.

Keywords: elastic network; least absolute shrinkage and selection operator (Lasso); Granger causality; factor screening; electricity consumption forecasting

(编辑 陈静)

(上接第 157 页)

Space charge behavior of retired high-voltage XLPE cables

ZHAO Yifeng¹, LIU Gang¹, XIE Yue¹, HUANG Jiasheng², ZHU Ningxi¹

(1. School of Electric Power, South China University of Technology, Guangzhou 510640, China;

2. Guangzhou Power Supply Bureau of Guangdong Power Grid Co., Ltd., Guangzhou 510310, China)

Abstract: Two retired high-voltage cross-linked cables with service years of 16 and 32 are chosen to conduct a 180-day prequalification test to research the diversity of space charge behavior and evaluate the reliability of reusing these cables for practical operation. Space charge behavior in cross-linked polyethylene (XLPE) insulation of the samples before and after aging test is tested by the pulse electro acoustic (PEA) method, and space charge characteristics of the samples is investigated. Fourier transform infrared spectroscopy (FTIR) and X-ray diffraction (XRD) experiments are used to observe changes on microstructure and aggregation structure of each sample. Experiments reveal the potential relationship between the structural changes and space charge behavior, including the characteristic of migration, accumulation and dissipation of space charge. The results show that the impurities causing severe accumulation of space charge inside the samples of the cable with 16 service years. After the accelerated aging test, reduction of impurities and annealing effect decreases the accumulation of space charge and speeds up dissipation process. On the contrary, the accumulation and dissipation of space charge inside the cable with 32 service years is moderate. After the accelerated aging test, the oxidative degradation and destruction of crystalline structure result in aggravation of space charge accumulation and more moderate dissipation process of space charge.

Keywords: cable; cross-linked polyethylene (XLPE); space charge; Fourier transform infrared spectroscopy (FTIR); X-ray diffraction (XRD); trap

(编辑 吴楠)