

计及经济因素的混合模型电量预测研究

徐久益¹, 李自明¹, 姚剑峰², 杨晓雷², 屠一艳², 李杰¹

(1. 国网浙江桐乡市供电有限公司, 浙江 桐乡 314500;

2. 国网嘉兴供电公司, 浙江 嘉兴 314000)

摘要:精确的电量预测是进行电网规划和建设的重要依据。为解决月度用电量预测方法思路单一、预测模型通用性不强、预测精度低等问题,通过对区域用电量与经济的历史数据研究分析,提出了一种计及经济因素影响的混合模型月度用电量预测方法。该方法将回归分析法和指数平滑法这两种预测方法相结合,建立电量预测模型,取得了较高的预测精度;最后,将本模型运用于浙江嘉兴地区月用电量预测,结果表明,该模型预测平均绝对百分误差为2.31%,从而验证了该模型的有效性与通用性。

关键词:月度电量预测;混合模型;经济因素;回归分析法;指数平滑法

中图分类号:TM74 文献标志码:A 文章编号:1003-6954(2021)01-0083-06

DOI:10.16527/j.issn.1003-6954.20210117

Research on Electricity Consumption Prediction under Mixed Model Considering Economic Factors

Xu Jiuyi¹, Li Ziming¹, Yao Jianfeng², Yang Xiaolei², Tu Yiyan², Li Jie¹

(1. State Grid Tongxiang Electric Power Supply Co., Ltd., Tongxiang 314500, Zhejiang, China;

2. State Grid Jiaxing Electric Power Supply Company, Jiaxing 314000, Zhejiang, China)

Abstract: Precise electricity consumption prediction provides an important basis for layout and construction of power grid. In order to solve the problems of single method, weak universality and low accuracy of prediction models, a monthly electricity consumption prediction method under mixed model on the basis of economic factors is proposed through analyzing the historical data of electricity consumption and economy. The regression analysis method and exponential smoothing method are combined to set up a mixed model of electricity consumption prediction, which contributes to the high prediction accuracy. At last, the proposed model is applied to the monthly electricity consumption prediction of Jiaxing area. The results show that its mean absolute percentage error (MAPE) is 2.31%, which verifies the effectively and universality of the model.

Key words: monthly electricity consumption prediction; mixed model; economic factor; regression analysis method; exponential smoothing method

0 引言

电力系统规划设计和运行管理依赖用电量数据。同时,用电量又是经济变化的“风向标”。因此,月度用电量的准确预测既是实现电网安全、经济运行的基础,也对把握宏观经济发展趋势起到至关重要的作用^[1-2]。

月度用电量预测是电力需求侧管理重要的市场

基金项目:国家电网有限公司QC项目(JDTX-2019F)

预测技术手段,精确的用电量预测对于平衡区域电力电量、精准的电网投资、电力现货市场交易、有效实施电力需求侧管理和节能降耗战略等^[2]都有着十分重要的意义。

传统的电量预测方法如回归分析法、时间序列预测法、灰色预测、神经网络等预测方法^[3-6],其预测精度已经不能满足当前电网发展需求。目前,电量预测技术和理论不断涌现,一些学者对月度用电量的预测方法进行研究,在传统预测方法上面不断改进,也取得新的成果。文献[7]在传统时间序列

预测模型基础上,提出了考虑经济因素时滞效应的月度电量预测方法。文献[8]从影响月度用电量的因素出发,利用计量经济学中的协整检验以及格兰杰因果检验来分析经济因素与用电量的关系,提出一种基于协整-格兰杰因果检验和季节分解的月度电量预测方法。近年来,中国的经济发展进入转型期,各地区月度用电量的变化也呈现差异性和多样性,因此,这些预测方法的运用也存在一定的局限性。基于此,通过对区域历史数据的研究分析,挖掘其发展的规律和特点,并结合两种算法优势建立电量预测模型,创新性地提出了一种计及经济因素的混合模型电量预测新方法,为地区月度用电量预测提供了一种新选择。

1 数据分析与挖掘

电量预测的核心是根据区域用电量的历史数据以及对用电量产生影响的经济、气候等因素的历史数据,建立数学模型来表述用电量发展变化规律。因此,在建立数学模型之前,首先需要对历史数据进行分析 and 处理,研究和挖掘出历史数据发展的内在规律性^[9]。

1.1 数据分析

所用数据来源于中国中西部某区域电网公司,如表 1 和表 2 所示。

1.2 数据预处理

通过分析半年度电量与半年度经济总量的关系

表 1 中西部某区域电网公司近年各月电量情况

单位:GWh

| 年份 | 1月 | 2月 | 3月 | 4月 | 5月 | 6月 | 7月 | 8月 | 9月 | 10月 | 11月 | 12月 | 合计 |
|-------|------|------|------|------|------|------|------|------|------|------|------|------|--------|
| 2014年 | 4456 | 4314 | 4698 | 4809 | 4890 | 4905 | 5357 | 5259 | 4937 | 4709 | 4761 | 4330 | 57 425 |
| 2015年 | 5460 | 4243 | 5023 | 5419 | 5110 | 5565 | 5920 | 6125 | 5678 | 5438 | 5477 | 5605 | 65 063 |
| 2016年 | 5782 | 5287 | 5786 | 5821 | 5805 | 5948 | 6003 | 6428 | 5749 | 5544 | 6122 | 6592 | 70 867 |
| 2017年 | 6140 | 6260 | 6497 | 6457 | 6128 | 6342 | 6899 | 6897 | 6470 | 6300 | 6211 | 6038 | 76 639 |
| 2018年 | 6312 | 6218 | 6522 | 6535 | 6333 | 6455 | 6901 | | | | | | |

表 2 中西部某区域相关经济数据

| 经济数据 | 2014年 | | 2015年 | | 2016年 | | 2017年 | |
|--------------|-------|--------|-------|--------|-------|--------|-------|--------|
| | 1—6月 | 1—12月 | 1—6月 | 1—12月 | 1—6月 | 1—12月 | 1—6月 | 1—12月 |
| 规模以上工业增加值/亿元 | 2830 | 6188 | 3002 | 6625 | 2792 | 6430 | 2965 | 6383 |
| 固定资产投资/亿元 | 4844 | 10 519 | 5782 | 12 576 | 6340 | 13 997 | 6795 | 15 147 |
| 社会消费品零售总额/亿元 | 2841 | 5827 | 3186 | 6589 | 3532 | 7332 | 3933 | 8114 |
| 出口总值/美元 | 1435 | 3731 | 1813 | 4375 | 1574 | 4438 | 1862 | 4837 |

可以发现,电量以及相关 4 类经济数据具有较强的时序特征,但 2015 年和 2016 年上半年的电量与经济总量的增长率存在明显的不一致性,其余年份对应的电量与经济总量增长率基本保持一致。

针对以上非一致性问题,将对原始数据进行预处理,主要是数据清洗,即对原始数据中的明显错误值、缺失值、异常值、可疑数据,选择合适的方法进行“清理”,使“脏”数据变为“干净”数据,也包括对重复记录进行删除。这里通过比较多种方法进行数据预处理,鉴于篇幅,仅介绍最终选择的箱形图检测方法,箱形图原理如图 1 所示,即对于一组数据,计算中位数、上下四分位数和上下限,若数值超过上下限则为异常值。下、中、上四分位数的位置计算方法为

$$f(Q_i) = i(n+1)/4 \quad (1)$$

式中, $n=12$,表示数据个数。经电量数据检测发现,存在异常的月份有 2014 年 7 月,2015 年 2 月、8 月,2016 年 2 月、8 月和 12 月,共 6 个月份。

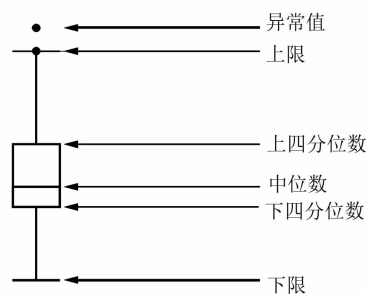


图 1 箱形图检测法的基本原理

检测出异常值后,需要通过分箱法对异常值进

行平滑处理,具体实现方法:若存在纵向两个数据则取纵向两个数据的均值 a ,再取横向两个数据的均值 b ,再取 a 、 b 的均值作为异常点的替代值。最终将上述存在异常的 6 个月用电量数据依次替换为 5082、5021、5821、5518、6194、5822 GWh。

2 建立月度售电量预测新方法

2.1 模型的建立

建立合适的电量预测模型,是电量预测核心,针对本项目原始数据分析,2014—2018 年电量曲线波动不大、趋势平稳,且原始样本数据量小。而现代预测方法如神经网络、时间序列^[10]等需要大量数据进行训练才能得到比较准确的模型,对于本项目数据量较小的情况下,预测效果不好。

研究采用传统的回归分析法和指数平滑法^[11]相结合来建立电量值的预测模型,优点是这两种预测方法比较成熟,模型参数设置较少,几乎不需要人为调参,且在数据量较小的情况下仍能得到比较准确的结果。本项目所提模型和算法的整体流程如图 2 所示。

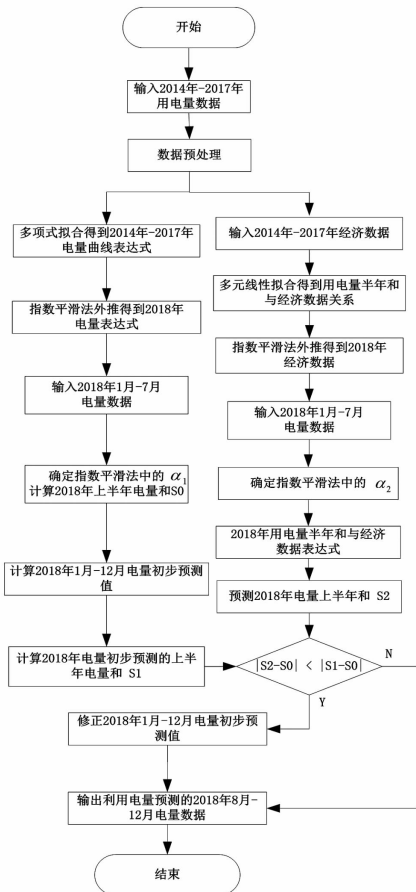


图 2 月用电量预测模型流程

1) 区域用电量曲线具有一定的周期性、时序性,一般以年为单位,每年电量曲线变化趋势相同。选取同样次数的多项式分别拟合 2014—2017 年的用电量曲线,多项式的系数采用最小二乘法^[12]原则确定,即误差平方和最小,得到用电量与时间的关系式。多项式次数的选择采用自动寻优,选择效果最好的值。自动寻优找出最优拟合次数 9 次。

2) 指数平滑法是一种数据预测方法,利用预测数据的变化在短期内保持历史数据的惯性特征,当前的数据与历史数据有关,且离现在时刻越近,对当前数据影响越大。本模型采用线性指数平滑模型:

$$F_{t+m} = a_t + b_t m \quad (2)$$

式中:

$$a_t = 2S_t^{(1)} - S_t^{(2)}; b_t = \frac{\alpha}{1 - \alpha}(S_t^{(1)} - S_t^{(2)}) \quad (3)$$

将步骤 1 中得到的 2014—2017 年的多项式函数的系数运用指数平滑法外推,得到 2018 年用电量与月份的关系:

$$f(x) = 0.000\ 081\ 38\ x^9 - 0.004\ 741\ 82\ x^8 + 0.116\ 814\ 56\ x^7 - 1.584\ 389\ 12\ x^6 + 12.920\ 345\ 13\ x^5 - 65.031\ 307\ 13\ x^4 + 199.920\ 938\ 01\ x^3 - 358.480\ 374\ 09\ x^2 + 337.644\ 679\ 21\ x - 62.318\ 446\ 43 \quad (4)$$

然后便可计算出 2018 年 1—12 月电量预测值,这是一个初始预测值,后续需要做修正。

3) 采用多元线性回归法^[13],将用电量与其影响因素(即 4 类经济因素)进行拟合。用多元线性表达式来拟合半年电量和与这些经济数据的关系,表达式为

$$h(x_1, x_2, x_3, x_4) = ax_1 + bx_2 + cx_3 + dx_4 \quad (5)$$

式中: $h(x_1, x_2, x_3, x_4)$ 为半年电量和; x_1, x_2, x_3, x_4 分别为规模以上工业增加值、固定资产投资、社会消费品零售总额和出口总值。2018 年上半年的经济数据与用电总量关系拟合表达式为

$$f(x) = 0.043\ 239\ 8\ x_1 + 0.014\ 314\ 24\ x_2 + 0.072\ 597\ 81\ x_3 - 0.326\ 030\ 59\ x_4 - 46.704\ 427\ 381 \quad (6)$$

4) 作为一个多输入系统而言,考虑越多输入量对系统的输出预测越准确,但是为了简化电量预测模型,需要找到影响程度较大的因素,这就需要进行相关性分析。

相关性分析是指对两个或多个具备相关性的

变量元素进行分析,从而衡量两个变量因素的相关密切程度。相关性的元素之间需要存在一定的联系或者概率才可以进行相关性分析^[14]。分析不同元素的相关性需要计算它们之间的相关系数,其定义为

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x}_i)(y_i - \bar{y}_i)}{\sqrt{\sum_{i=1}^n (X_i - \bar{X}_i)^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y}_i)^2}} \quad (7)$$

式中: r 为样本相关系数; n 为样本总数; x 、 y 为样本观测值,在本项目中 x 为电量, y 为经济性因素。

需要计算各经济指标因素与电网电量的相关系数,找出对电网电量的影响大且影响较为稳定的因素,相关性结果如表 3 所示。

表 3 4 类经济因素与电量的相关性

| 经济因素 | 规模以上 工业增加值 | 固定资产 投资 | 社会消费品 零售总额 | 出口 总值 |
|----------------|---------------|------------|---------------|----------|
| 与电网电量的 相关系数 | 0.955 | 0.998 | 0.999 | 0.989 |

从表 3 可以看出,4 类经济数据与电网电量的相关度均较高,相关系数均大于 0.95,其中社会消费品零售总额相关度最高,因此,将 4 类经济数据作为电量预测建模研究的关键因素。

5)用半年电量和修正初始预测值,如果经济数据得到的 2018 年上半年电量和与步骤 2 中得到的初始预测数据的 2018 年上半年电量和相比,更接近实际数据,那么经济数据得到的 2018 年下半年电量和也会更接近,故用经济数据得到的电量和对步骤 2 得到的初始预测数据做修正。如果经济数据得到的 2018 年上半年电量和更不接近于实际数据,则不对初始电量预测数据进行修正。

修正的具体方法为整体倍数法:

$$\beta = (S_1 + S_2) / 2 \quad (8)$$

式中: S_1 为经济数据得到的 2018 年下半年电量和; S_2 为步骤 2 中得到的 2018 年下半年电量和,对于 2018 年 8—12 月的初始预测数据都乘以 β ,得到本模型最终的预测结果。

3 算例分析

3.1 模型的验证

为了验证所提方法的实际运用效果,对给定的中国中西部某区域电网公司 2018 年 1—7 月的月用

电量进行了预测,与目前使用最广泛的 BPNN 预测、灰色预测进行比对,并计算平均绝对百分误差 (mean absolute percentage error, MAPE)。三者预测结果见表 4。

由表 4 可以看出,所提方法预测平均绝对百分误差最小,为 0.83%,由此可见,所提方法预测精度得到了有效提高,且具有更高的稳定性。

所建模型对未知的 2018 年 8—12 月的电量进行预测,并用经济数据对预测结果进行修正,结果如表 5 所示。

3.2 模型的应用

3.2.1 力助电力部门辅助决策

在已知浙江嘉兴地区月用电量数据与 4 类经济数据的前提下,将所建模型运用于与中西部区域电量波动、气候条件以及经济结构完全不同的嘉兴地区,对预测精度及通用性进行验证。表 6 为浙江嘉兴地区 2018 年 8—12 月份月用电量预测值。

经计算,所建模型预测的 MAPE 为 2.31%,证明其具有较高的预测精度且通用性强,可适用于不同地区的电量预测。

通过月用电量预测,进一步勾勒出嘉兴地区电量预测曲线,如图 3 所示,可更加精准地对地区用电量特性进行分析,为电网公司或电力部门提供辅助决策^[15],包括:为发展策划部门提供规划、决策、投资的依据;为财务部门提供电力现货交易市场的清晰判断;为营销部门提供增供扩销形势的判断。

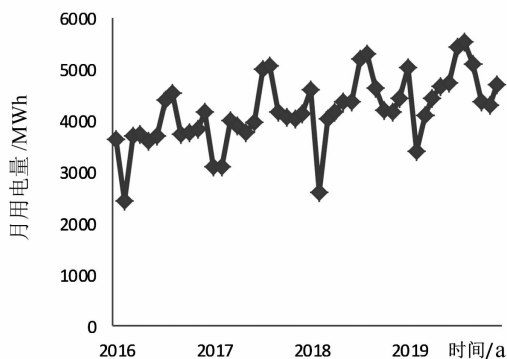


图 3 嘉兴地区月用电量预测曲线

3.2.2 提升主动服务能力

所建模型运用于某水泥厂大用户月用电量预测,可以了解用户电力消耗发展趋势,预测曲线如图 4 所示。

通过大用户电量预测曲线分析,供电公司可以为企业客户提升主动服务能力^[16],包括提前感知客

表4 不同预测方法预测结果对比

| 时间 | 原始数据/ GWh | BPNN | | 灰色预测 | | 所提方法 | |
|----------|--------------|--------------|------------|--------------|------------|--------------|------------|
| | | 模型数据/ GWh | 相对误差/ % | 模型数据/ GWh | 相对误差/ % | 模型数据/ GWh | 相对误差/ % |
| 1月 | 6312 | 6313 | 0.02 | 6331 | 0.30 | 6318 | 0.10 |
| 2月 | 6218 | 4317 | 30.57 | 6251 | 0.53 | 6375 | 2.52 |
| 3月 | 6522 | 5803 | 11.02 | 6345 | 2.71 | 6521 | 0.02 |
| 4月 | 6535 | 5791 | 11.38 | 6289 | 3.76 | 6474 | 0.93 |
| 5月 | 6333 | 6313 | 0.32 | 6253 | 1.26 | 6225 | 1.71 |
| 6月 | 6455 | 6313 | 2.20 | 6436 | 0.29 | 6424 | 0.48 |
| 7月 | 6901 | 6915 | 0.20 | 6797 | 1.51 | 6896 | 0.07 |
| 平均绝对误差/% | | 7.95 | | 1.48 | | 0.83 | |

表5 2018年8—12月的电量预测结果修正

单位:GWh

| 时间 | 修正前预测值 | 修正后预测值 |
|-----|--------|--------|
| 8月 | 6917 | 6933 |
| 9月 | 6465 | 6480 |
| 10月 | 6324 | 6339 |
| 11月 | 6417 | 6432 |
| 12月 | 6291 | 6306 |

表6 嘉兴2018年8—12月月用电量预测值

单位:GWh

| 时间 | 实际值 | 预测值 |
|-----|------|------|
| 8月 | 5309 | 5269 |
| 9月 | 4637 | 4273 |
| 10月 | 4225 | 4185 |
| 11月 | 4183 | 4181 |
| 12月 | 4431 | 4339 |

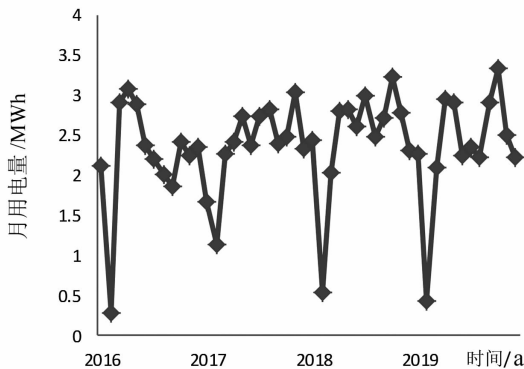


图4 大用户月用电量预测曲线

户的容量变化需求、有的放矢地为客户提供增容等业务支持、辅助企业确定节能方法、为用户提供有竞争力的差异化服务等用途。

4 结 语

将回归分析法和指数平滑法这两种预测方法相结合,建立电量预测模型,模型中所有参数都采用算法自动寻优,减少了人工干预带来的误差。通过对中西部区域实际历史数据精细和深入的挖掘,先用历史电量数据初步预测未来几个月电量,再用与电量相关性很高的经济因素修正电量初步预测值,新方法实际预测精度得到了有效提高。

算例分析的结果表明,预测 MAPE 为 0.83%,而模型在客观条件完全不同的嘉兴地区进行预测,MAPE 也仅为 2.31%,证明模型具有通用性。所提预测方法为地区电量预测工作提供了一种可供选择的新思路,同时,也为供电公司加强电力需求预测和供电负荷管理等工作提供了有益的方法支撑和必要的决策手段。

参考文献

- [1] 陈国平,梁志峰,董昱. 基于能源转型的中国特色电力市场建设的分析与思考[J]. 中国电机工程学报, 2020,40(2): 369-378.
- [2] 黄宏和,吴臻,琚军,等. 基于数据挖掘技术的配电网精准投资策略研究[J]. 浙江电力,2019,38(3):92-97.
- [3] 徐俊,徐文辉,曾鑫,等. 基于大数据挖掘电量预测方法的创新及应用[J]. 电力大数据,2018,21(10):57-61.
- [4] 程超. 基于时间序列法和回归分析法的改进月售电量预测方法研究[D]. 重庆:重庆大学,2016.
- [5] 钟全辉,张以全,肖少华,等. 基于灰色预测理论的区域电量概率预测方法及其应用[J]. 浙江电力,2018,37(1):19-22.

- [6] 李瑞杰. 基于 BP 神经网络的地区售电量预测新方法研究[D]. 西安:西安科技大学,2018.
- [7] 郭鸿业,陈启鑫,夏清,等. 考虑经济因素时滞效应的月度负荷预测方法[J]. 电网技术,2016,40(2):514-520.
- [8] 刘俊,赵宏炎,刘嘉诚,等. 基于协整-格兰杰因果检验和季节分解的中期负荷预测[J]. 电力系统自动化,2019,43(1):73-80.
- [9] 赵亮,刘友波,余莉娜,等. 基于深度信念网络的光伏电站短期发电量预测[J]. 电力系统保护与控制,2019,47(18):11-19.
- [10] 张凤珠. 基于历史数据的光伏发电量预测方式的组合应用[J]. 通信电源技术,2020,37(5):190-191.
- [11] 邵臻. 基于半参数统计模型的中长期电量负荷预测应用研究[D]. 合肥:合肥工业大学,2012.
- [12] 张颖,王琳,王丽华,等. 基于最小二乘法和聚类的用电数据异常分析算法[J]. 河北电力技术,2019(5):4-6.
- [13] 史松奇. 试析现代统计分析方法及其应用[J]. 科技经济导刊,2019,27(24):214.
- [14] Yuc N, Mirowski, Ho T K. A Sparse Coding Approach to Household Electricity Demand Forecasting in Smart

Grid[J]. IEEE Transactions on Smart Grid, 2017, 8(2):738-748.

- [15] 李翔,欧阳森,冯天瑞,等. 一种基于用电行业分类的中长期电量预测方法[J]. 现代电力,2015,32(6):86-91.
- [16] 葛风雷. 基于线性回归的电网夏季降温负荷与气温相关性分析[J]. 农村电气化,2017(4):9-11.

作者简介:

徐久益(1991),男,工程师,主要从事电网调度运行与控制工作;

李自明(1989),男,工程师,主要从事电网调度运行与控制工作;

姚剑峰(1977),男,高级工程师,主要从事电网调度运行工作;

杨晓雷(1981),男,高级工程师,主要从事电网调度运行及新能源管理工作;

屠一艳(1983),女,工程师,主要从事电网调度负荷预测及新能源管理工作;

李杰(1989),男,工程师,主要从事电力系统继电保护工作。

(收稿日期:2020-06-24)

(上接第 46 页)

- [36] 张立佳,徐国宁,赵向阳,等. 基于神经网络的老化锂电池 SOC 估算方法的研究[J]. 电源学报,2020,18(1):54-60.
- [37] 刘征宇,杨俊斌,张庆,等. 基于 QPSO-BP 神经网络的锂电池 SOC 预测[J]. 电子测量与仪器学报,2013,27(3):224-228.
- [38] 刘征宇,杨俊斌,张庆,等. 基于 QPSO-BP 神经网络的锂电池 SOC 预测[J]. 电子测量与仪器学报,2013,27(3):224-228.
- [39] 尹安东,张万兴,赵韩,等. 基于神经网络的磷酸铁锂电池 SOC 预测研究[J]. 电子测量与仪器学报,2011,25(5):433-437.
- [40] 雷肖,陈清泉,刘开培,等. 电动车电池 SOC 估计的径向基函数神经网络方法[J]. 电工技术学报,2008(5):81-87.
- [41] 齐智,吴锋,陈实,等. 利用人工神经网络预测电池 SOC 的研究[J]. 电源技术,2005,29(5):325-328.
- [42] 陈小军,韩晓娟,梁亮. 基于小波复合模型的锂电池阻抗特性研究[J]. 电源技术,2012,36(1):67-70.
- [43] 刘征宇,黎盼春,朱诚诚,等. 基于组合模型的锂电池参数辨识和电池荷电状态在线联合估计[J]. 中国机械工程,2020,31(10):1162-1168.
- [44] 李荣波,陈勇. 基于 AFFDRLS-AHIF 对动力锂离子电池单体 SOC 估计[J]. 北京交通大学学报,2020,44(2):129-135.
- [45] 吴忠强,尚梦瑶,申丹丹,等. 基于神经网络和 MS-

AUKF 算法的蓄电池荷电状态估计[J]. 中国电机工程学报,2019,39(21):6336-6344.

- [46] 莫易敏,叶鹏,骆聪,等. 基于改进的双卡尔曼滤波的锂电池 SOC 估计[J]. 电源技术,2020,44(5):732-735.

[47] 何志刚,李尧天,盘朝奉,等. 基于 AIUKF 的锂离子电池 SOC 估算[J]. 电源技术,2020,44(4):518-521.

[48] 李超然,肖飞,樊亚翔,等. 基于门控循环单元神经网络和 Huber-M 估计鲁棒卡尔曼滤波融合方法的锂离子电池荷电状态估算方法[J]. 电工技术学报,2020,35(9):2051-2062.

[49] 魏孟,李嘉波,叶敏,等. 基于高斯混合回归的锂离子电池 SOC 估计[J]. 储能科学与技术,2020,9(3):958-963.

[50] 贺林,胡敏康,石琴,等. 一种分阶段锂离子电池荷电状态估计算法[J]. 电力电子技术,2020,54(2):8-11.

[51] 刘树林,崔纳新,李岩,等. 基于分数阶理论的车用锂离子电池建模及荷电状态估计[J]. 电工技术学报,2017,32(4):189-195.

[52] 孙冬,陈息坤. 基于离散滑模观测器的锂电池荷电状态估计[J]. 中国电机工程学报,2015,35(1):185-191.

[53] 蒋波. 带有模型参数提取的电池管理系统监控平台的设计[D]. 天津:天津大学,2011.

作者简介:

王晓辉(1990),男,硕士研究生,主要研究方向为变电运维技术。

(收稿日期:2020-08-21)