

基于深度数据感知的工业园区短期负荷预测

胡博¹,李桐²,王义贺¹,崔嘉³,杨俊友³,许军金³

(1. 国网辽宁省电力有限公司, 辽宁 沈阳 110015;

2. 国网辽宁省电力有限公司电力科学研究院, 辽宁 沈阳 110006;

3. 沈阳工业大学电气工程学院, 辽宁 沈阳 110870)

摘要:结合自动编码器的特征提取能力与长短期记忆(LSTM)良好的时序预测优势,提出一种基于LSTM自编码的短期负荷预测模型。首先,基于深度模型学习能力优势,构建自动编码器的工业园区负荷数据特征提取模型。其次,给出基于格布拉斯准则的异常值判断方法和拉格朗日样条插值的缺失值补全法。最后,利用Tensorflow架构调用Keras库搭建实验平台,对轻工业负荷、重工业负荷和光伏发电工业用户负荷分别展开预测实验,实验结果验证LSTM自编码模型预测短期工业负荷的有效性。

关键词:负荷预测;工业负荷;深度学习;长短期记忆;自动编码器

中图分类号:TM712 **文献标识码:**A **DOI:**10.19457/j.1001-2095.dqed22913

Short-term Load Forecast of Industrial Park Based on Deep Data Aware

HU Bo¹, LI Tong², WANG Yihe¹, CUI Jia³, YANG Junyou³, XU Junjin³

(1. State Grid Liaoning Electric Power Co., Ltd., Shenyang 110015, Liaoning, China; 2. Electric Power

Research Institute of State Grid Liaoning Electric Power Co., Ltd., Shenyang 110006,

Liaoning, China; 3. School of Electrical Engineering, Shenyang University of Technology, Shenyang 110870, Liaoning, China)

Abstract: Combined with the feature extraction ability of self-coding and the time series analysis ability of long short-term memory (LSTM), a short-term load forecasting model based on LSTM self-coding was proposed. First of all, based on the basic principle and flow of the depth model, the functional characteristics of the automatic encoder were given. Secondly, the outlier judgment method based on Goblas criterion and the missing value completion method based on Lagrangian spline interpolation were proposed. Finally, using Tensorflow architecture to call Keras library to build an experimental platform to carry out forecasting experiments on light industrial load, heavy industrial load and photovoltaic industrial user load respectively, which verifies the conclusion that LSTM self-coding model is more suitable for short-term industrial load forecasting.

Key words: load forecasting; industrial load; deep learning; long short-term memory (LSTM); automatic encoder

随着智能电网的发展,电网大数据时代已经到来。据我国国家能源局在2019年发布的电力行业统计数据显示,工业用电占总用电量的70%以上。因此,分析与掌握工业区域负荷预测,是电网经济安全运行的有力保障。

国内外学者对负荷预测做了大量研究。文献[1]利用自回归积分滑动平均模型实现负荷预测。文献[2]通过傅里叶展开模型分析负荷并预

测。文献[3]通过卡尔曼滤波法建立负荷预测模型。文献[4-7]提出多种解决负荷预测问题的机器学习模型。文献[8-12]提出多种深度学习模型。虽然上述文献负荷预测的精度随着深度学习算法的发展得到进一步的提升,尤其是在有大数据作为信息支撑时,深度学习可以实现信息的充分利用,但是,深度学习保证精度的前提是数据真实完整,当数据信息坏点较多时,利用这种

基金项目:辽宁省“兴辽英才计划”(XLYC1902090)

作者简介:胡博(1972—),男,博士,教授级高工,Email:hubo@ln.sgcc.com.cn

携带假信息的数据会对预测精度造成影响。而且,目前的深度学习模型没有涉及到针对工业负荷这种受多种因素干扰的预测研究,对数据的挖掘也不够彻底。

综上,目前预测研究多针对传统大规模电网负荷,鲜有专门针对工业园区负荷的预测研究。同时,工业园区负荷占比大,可调度潜力强,且运行具有一定的规律,但数据的采集和深度感知是亟需解决的难题。因此,需要一种能够对历史负荷数据和各种影响因子深度挖掘的模型,实现隐藏信息的有效利用,进而实现负荷预测精度的提高。

工业园区负荷具有强耦合性、非线性,且具有一定的冲击性,以及历史数据缺乏的明显特点。自动编码器正好对负荷数据具有非线性特征提取能力,将工业园区历史负荷数据自动编码,能够进一步突出工业园区历史负荷数据的时序性。本文利用自动编码器的特征提取性能,结合LSTM良好的时序数据处理能力,提出LSTM自编码模型,实现工业园区短期负荷预测。

1 问题描述及框架体系

工业负荷在总用电负荷中占比很大,如果能

够实现短期工业负荷的准确预测,将对电网安全稳定运行提供重要保障。但与其他类型的负荷预测不同,工业负荷预测更为复杂,难度更大,主要有两方面原因:第一,如今的负荷预测已经成为根据多变量时序数据预测的问题,短期的负荷预测更是受多种灵活变量影响,尤其对于工业负荷而言,其受天气、季节、电价等外在因素的影响更为突出;第二,工业用户的负荷数据噪声很大,使普通模型难以应对。

本文的总体研究流程与框架如图1所示。为提升工业区域短期负荷预测的精度,将主要研究以下方面:

1)对历史数据进行数据预处理,主要包括缺失值的补全、异常值的判断与替换以及输入预测模型前的数据归一化处理;

2)考虑到工业负荷的多样性,尤其是不同类型工业负荷的特点与各种外在因素密切相关,同时不同因素产生的干扰又各不相同,对工业负荷分类,选取影响工业负荷的主要外在因素;

3)智能设备为负荷预测提供充足的数据支撑,研究如何基于数据挖掘技术有效利用历史数据,充分提取数据内部深层次的特征。

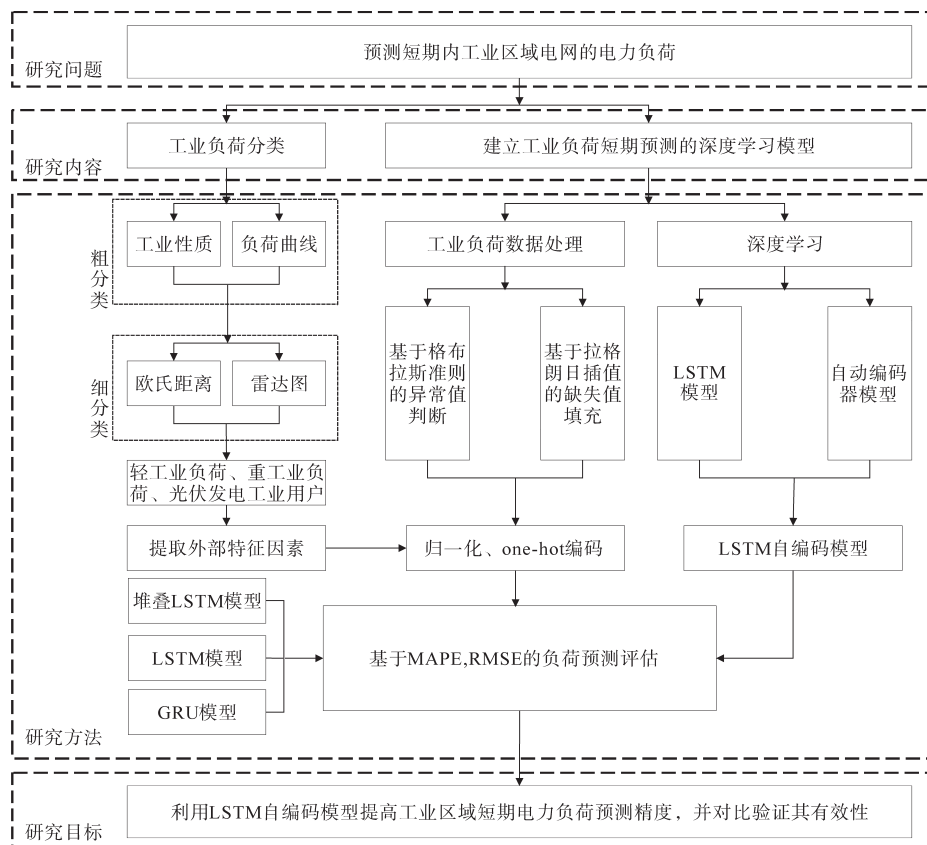


图1 总体研究框架

Fig.1 Overall research framework

2 LSTM自编码的工业园区负荷模型

2.1 LSTM工业园区负荷模型

LSTM是SimpleRNN的一种变体,在结构上改良,可以实现对长短期依赖信息的学习。因此LSTM被广泛应用于具有时序数据的预测问题,本文首次将LSTM方法引入到工业园区负荷预测。LSTM内部结构如图2所示。

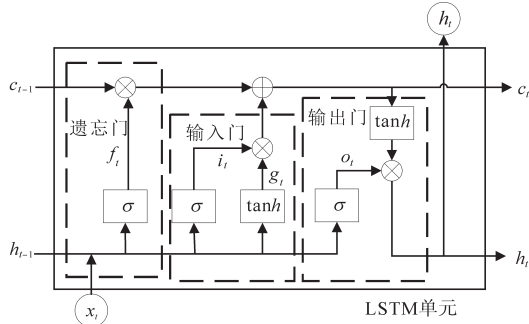


图2 LSTM结构图

Fig.2 LSTM structure diagram

2.2 自动编码器

自动编码器由输入层、隐藏层和输出层构成,其结构如图3所示。输入数据经过隐藏层后压缩,由于数据维度的降低,数据间的内部关联信息被不断提纯,再通过输出层将压缩后的数据还原为输入。整个过程的核心就是隐藏层的特征提取,基本思想就是数据表达的有效转换。

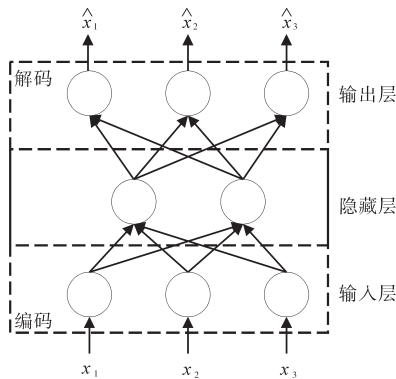


图3 自动编码器结构图

Fig.3 Automatic encoder structure diagram

2.3 LSTM自编码模型结构

本文提出的基于LSTM自编码的短期负荷预测模型结构设计如下:

1)第1层:输入LSTM层,将包含历史负荷和外部因素数据的 n 维输入矩阵 $X = [x_1, x_2, \dots, x_n]$ 重构为三维矩阵[样本,步长,特征],送入第1个LSTM层,在第 t 时刻其输出为

$$H_t = LSTM(X_t) \quad (1)$$

2)第2层:RepeatVector层,RepeatVector(k)是将第1个LSTM层的输出作为本层的输入,实现 k 次重复输入量,输出3D张量形式[None, k ,Features]。

3)第3层~第5层皆为中间LSTM层,值得一提的是,这3层LSTM要设置return_sequences参数为True,使LSTM能返回序列,以便作为下一个LSTM的输入。

4)第6层为TimeDistributed层,它将Dense层独立且均匀地应用在每一个时间步上,其输入为上层LSTM返回的序列。

基于LSTM自编码的短期负荷预测模型其结构如图4所示。

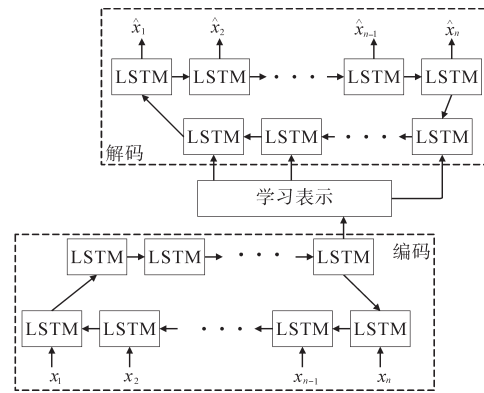


图4 LSTM自编码模块

Fig.4 LSTM self-coding module

2.4 模型预测数据处理

工业负荷的数据集噪声较大,为了最大程度保留原始数据,以下三步处理模式:

1)异常值判断:终端测量设备在收集记录信息时可能存在失准,数据集中连续变量会突现异常值,对电力负荷历史数据基于格布拉斯准则法判定异常值。

2)缺失值填充:由于设备故障或是检修,会导致时序性数据在某一时刻或一段时间内缺失记录值,对于各变量的数据序列中存在的缺失值,采用Lagrange三次样条插值补全。

3)归一化处理:不同指标的量纲不同,各类特征指标变化范围差异较大,因此要对收集的数据归一化处理。

对于时间序列数据,采用的标准化公式为

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (2)$$

2.5 预测流程图

工业负荷预测难度很大,主要体现在两方面:一方面,工业区域负荷种类多样,受多种外部

因素影响,且每种不同类型负荷受不同外部因素影响的大小各不相同,尤其对于短期的负荷预测而言,外部因素的干扰权重变得更大,传统的预测方法和一般的深度学习算法难以有效准确地预测;另一方面,工业负荷数据噪声很大,对数据的有效判定清洗,可以避免模型的错误学习,提高负荷预测精度。

传统 LSTM 作为常用的预测模型,对于一般的负荷预测或是对电网整体负荷预测效果尚可,对于工业区域这种极易受外部因素影响的负荷类型,难以取得理想的预测效果,这是因为 LSTM 难以学习多种变量之间的内部联系,即使通过人工提取变量特征也会丢失一些隐藏信息。

针对上述问题,本文提出一种基于 LSTM 自编码的深度学习模型,结合自动编码器的特征提取能力与 LSTM 良好的时序预测优势,实现对多种变量数据的深度挖掘,提取其与历史负荷之间的信息,充分利用数据集可以实现工业区域电网负荷的精确短期预测。基于 LSTM 自编码的短期负荷预测流程如图 5 所示。

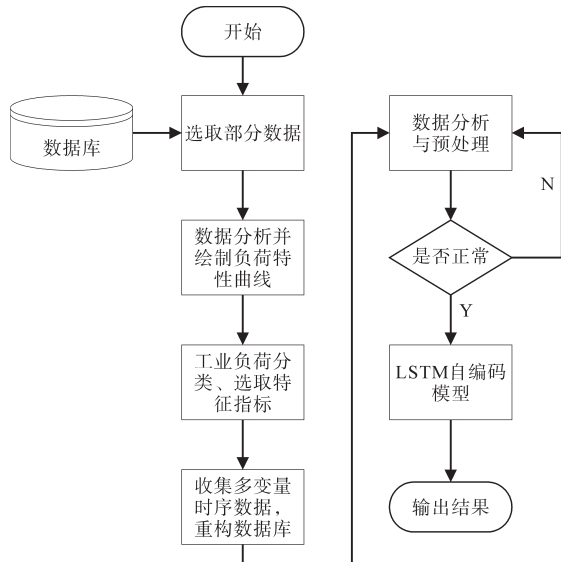


图5 基于 LSTM 自编码的短期负荷预测流程图

Fig.5 Short-term load forecasting flow chart based on LSTM self-coding

3 算例分析

3.1 实验准备

3.1.1 数据分析与预处理

基于调研的辽宁省某市内 12 家工厂的历史负荷数据,数据时间长度为从 2019 年 7 月 31 日至 2019 年 11 月 7 日,时间跨度为 100 d。选取纺织公司作为轻工业负荷代表,选取水泥公司作为重

工业负荷代表,选取商贸公司作为光伏发电工业用户代表,每个数据集中 90% 作为训练集,10% 作为测试集。

同时,收集该地区 2019 年 7 月 31 日至 11 月 7 日共计 100 d 的温度、湿度记录,将每天 24 点数据通过插值法构成 96 点数据,即每 15 min 一组,按文中提出的基于格布拉斯准则异常值判断,基于 Lagrange 完备三次样条插值缺失值填充。此外,工作日类型、峰谷分时电价对于工业用电影响很大,采用 one-hot 编码处理,若待预测日为工作日,待预测时刻为谷时电价,则补充其输入特征为[工作日,非工作日,峰时,平时,谷时],其对应编码为 [1,0,0,0,1]。特征选取并收集完毕后,重构数据库,每时刻特征维度为 8 维,为后续训练模型做好准备。

3.1.2 实验设置

实验硬件设备显卡为 GTX1650,基于 python 语言编程,通过 Tensorflow 框架作背景,利用 Keras 库搭建深度学习模型,实现神经网络层的灵活叠加。

3.1.3 评价指标

为了全面准确地评估模型预测性能,参照国家电网对负荷预测的评价指标,选择以下两种指标:

1) 平均绝对百分比误差 (mean absolute percentage error, MAPE):

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left[\frac{x(t) - y(t)}{x(t)} \right] \times 100\% \quad (3)$$

式中: n 为预测点个数; $x(t)$, $y(t)$ 分别为待预测日第 t 时刻的实际测量值与模型预测值。

2) 均方根误差 (root mean square error, RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n [x(t) - y(t)]^2} \quad (4)$$

3.2 实验结果分析

采用 LSTM、门控循环单元 (gated recurrent unit, GRU)、堆叠的 LSTM 以及提出的基于 LSTM 自编码网络 4 种模型实验,比较这 4 种模型的预测结果,给出实验结论。

搭建好网络模型后,训练过程中采用控制变量法调优,逐一调整模型参数、神经元个数、LSTM 神经网络层数,同时平衡性能与训练时间,最终确定 batch_size 为 512,训练周期设置 100,特征数为 8,步长为 192,即构建 [192, 96] 的滑动窗口,根据前 2 d 数据预测后一天负荷,同时设置神

经元个数为50,并在每层LSTM后设置dropout为0.2,防止过拟合。

3.2.1 轻工业负荷

对于轻工业负荷数据,表1为预测日的测试结果,记录模型对应MAPE与RMSE。

表1 轻工业负荷预测指标

Tab.1 Light industry load forecasting indicators

指标	LSTM自编码	堆叠LSTM	LSTM	GRU
MAPE/%	1.96	1.97	2.03	2.37
RMSE	25.54	25.97	26.58	30.60

从表1可以看出,提出的基于LSTM自编码的短期负荷预测模型相对于堆叠LSTM模型、LSTM模型和GRU模型预测精度有所提高,但是幅度并不大,并未突出提出模型的优势,原因可能在于轻工业负荷受外界因素干扰不大,受工作日类型影响稍明显,而对于峰谷分时电价、温度、湿度等外部因素关联较小。

轻工业用户用电负荷相对较平均,波动次数多但起伏不大,单日预测负荷曲线如图6所示。

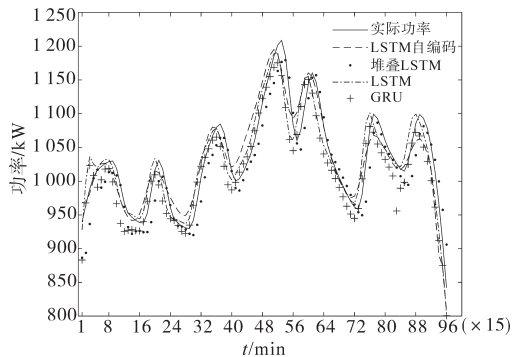


图6 轻工业负荷单日预测曲线

Fig.6 Daily forecast curves of light industrial load

由图6可以看出,提出的LSTM自编码与堆叠自编码、LSTM和GRU模型相比并不突出,在曲线上的表现性能相近,这是由于轻工业负荷受外部因素影响一般,数据内部信息通过一般神经网络挖掘即可,可以说问题的简易性限制文中提出的模型的表现。

3.2.2 重工业负荷

对于重工业负荷数据,表2记录不同模型对应的MAPE与RMSE。

表2 重工业负荷预测指标

Tab.2 Heavy industry load forecasting indicators

指标	LSTM自编码	堆叠LSTM	LSTM	GRU
MAPE/%	7.25	8.13	15.46	13.08
RMSE	194.82	267.34	434.53	493.21

由表2可知,在重工业负荷短期预测中,提出的基于LSTM自编码预测模型无论是在MAPE指标,还是在RMSE指标上,都要明显优于堆叠LSTM,LSTM和GRU模型。文中模型平均预测精度高达92.75%,比其余三者分别高0.88%,8.21%,5.83%。RMSE比其余三种模型降低72.52,239.71,298.39。

重工业负荷在一天内变化次数很少,但变化幅度大,重工业类用户的用电行为受到峰谷分时电价制约,使重工业负荷在短期内会发生突变。图7为4种模型对重工业负荷单日预测的曲线,可以明显看出,基于LSTM自编码模型紧密贴合实际负荷曲线的变化,说明该模型的精度较好。这也证明LSTM自编码特征提取能力和时序预测能力的有效结合,起到实质性的作用。

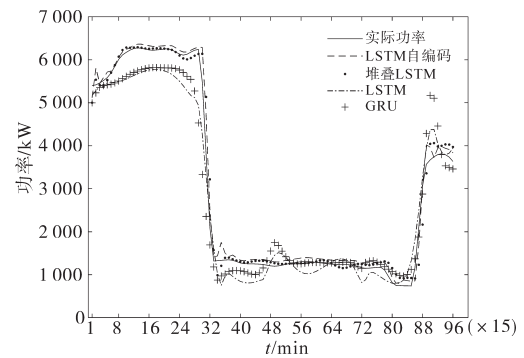


图7 重工业负荷单日预测曲线

Fig.7 Daily forecast curves of heavy industry load

3.2.3 光伏发电工业用户

光伏发电用户的单日预测指标如表3所示。

表3 光伏发电工业用户负荷预测指标

Tab.3 Load forecasting indicators of photovoltaic power generation industry users

指标	LSTM自编码	堆叠LSTM	LSTM	GRU
MAPE/%	11.21	19.91	52.79	54.98
RMSE	12.95	23.76	44.94	62.08

从4种模型对比可以发现,LSTM自编码模型精度达到88.79%,比堆叠LSTM模型提高8.7%,比LSTM模型提高41.58%,比GRU模型提高43.77%。RMSE较另外三种模型分别降低10.81,31.99,49.13。

图8给出4种模型对光伏发电的工业用户单日预测的结果,LSTM和GRU都出现较大的偏差,堆叠的LSTM模型表现尚可,LSTM自编码最为接近实际负荷曲线。

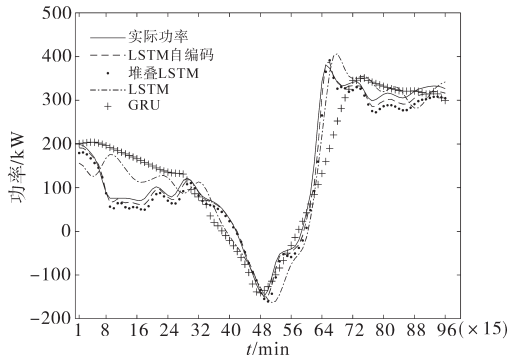


图8 光伏发电工业用户负荷单日预测曲线

Fig.8 Single-day load forecast curve for photovoltaic power industry users

4 结论

本文提出基于LSTM自编码的深度学习模型,用以实现对工业区域负荷的准确预测,经过算例分析得出以下结论:

1)基于自动编码器强大的数据挖掘能力,有效地提取隐藏层数据特征,减少工作量,明显提高传统深度学习对数据的预测精度,克服深度学习模型对数据挖掘的不彻底性。

2)LSTM自编码有效抑制各种外部因素对历史负荷数据的影响问题,深度挖掘历史负荷数据,实现隐藏信息的最大利用,工业负荷的预测精度高达92.75%,比GRU模型提高0.41%~43.77%。

3)LSTM与自动编码器的有机结合,通过相关程序达到更加智能的操作,具有更高的稳定性,使得预测结果更加精确,工业负荷预测中均方根误差比传统LSTM模型降低1.04~239.71。

4)LSTM自编码模型方法对工业负荷数据预测优势明显,比堆叠LSTM模型在工业负荷预测中平均绝对百分比误差降低0.01%~8.7%。

参考文献

[1] 艾欣,周志宇,魏妍萍,等. 基于自回归积分滑动平均模型的可转移负荷竞价策略[J]. 电力系统自动化,2017,41(20):26-31,104.
 Ai Xin, Zhou Zhiyu, Wei Yanping, et al. Bidding strategy for time-shiftable loads based on autoregressive integrated moving average model[J]. Power System Automation, 2017, 41(20): 26-31, 104.

[2] 王惠中,周佳,刘轲. 电力系统短期负荷预测方法研究综述[J]. 电气自动化,2015,37(1):1-3,39.
 Wang Huizhong, Zhou Jia, Liu Ke. Summary of research on the short-term load forecasting method of the electric power system [J]. Electrical Automation, 2015, 37(1): 1-3, 39.

[3] 石文清,吴开宇,王东旭,等. 基于时间序列分析和卡尔曼滤波

波算法的电力系统短期负荷预测[J]. 自动化技术与应用, 2018, 37(9):9-12,23.

Shi Wenqing, Wu Kaiyu, Wang Dongxu, et al. Eclectic power system short-term load forecasting model based on time series analysis and Kalman filter algorithm[J]. Automation Technology and Applications, 2018, 37(9): 9-12, 23.

[4] 李飞. 基于ESO的电压平衡器模型预测控制方法[J]. 电气传动,2020,50(5):61-66,74.
 Li Fei. The predictive control method of voltage balancer model based on ESO[J]. Electric Drive, 2020, 50(5): 61-66, 74.

[5] 吴玮,程国栋,王贵峰,等. LCL并网逆变器多变量模型预测控制策略[J]. 电气传动,2020,50(11):7-12.
 Wu Wei, Cheng Guodong, Wang Guifeng, et al. Multi-variable model predictive control strategy of LCL grid-connected inverter [J]. Electric Drive, 2020, 50(11): 7-12.

[6] 陶雪峰,陈媛媛,叶磊,等. 模块化多电平变换器快速模型预测控制器设计[J]. 电气传动,2020,50(10):72-78.
 Tao Xuefeng, Chen Yuanyuan, Ye Lei, et al. Design of fast model predictive controller for modular multilevel converter[J]. Electric Drive, 2020, 50(10): 72-78.

[7] 张兴旺,杨康佳. 基于有功电流预测算法的三相VSR软启动策略研究[J]. 电气传动,2020,50(8):74-78,92.
 Zhang Xingwang, Yang Kangjia. Research on three-phase VSR soft start strategy based on active current prediction algorithm [J]. Electric Drive, 2020, 50(8): 74-78, 92.

[8] Wen Lulu, Zhou Kaile, Yang Shanlin, et al. Optimal load dispatch of community microgrid with deep learning based solar power and load forecasting[J]. Energy, 2019, 171(2):1053-1065.

[9] 丁立,和阳,吉敬华,等. 基于新型SMO的无位置传感器PMSM模型预测控制[J]. 电气传动,2020,50(7):25-30.
 Ding Li, He Yang, Ji Jinghua, et al. Sensorless model predictive control based on new sliding mode observer for PMSM[J]. Electric Drive, 2020, 50(7): 25-30.

[10] 陈洪波,王璨,徐斌,等. 基于LSTM网络的短期负荷预测[J]. 电工技术,2019(9):27-29.
 Chen Hongbo, Wang Can, Xu Bin, et al. Short-term load forecasting based on LSTM network[J]. Electrical Technology, 2019(9): 27-29.

[11] 陈振宇,刘金波,李晨,等. 基于LSTM与XGBoost组合模型的超短期电力负荷预测[J]. 电网技术,2020,44(2):614-620.
 Chen Zhenyu, Liu Jinbo, Li Chen, et al. Ultra short-term power load forecasting based on combined LSTM-XGBoost model[J]. Power Grid Technology, 2020, 44(2): 614-620.

[12] 陆继翔,张琪培,杨志宏,等. 基于CNN-LSTM混合神经网络模型的短期负荷预测方法[J]. 电力系统自动化,2019,43(8):131-137.
 Lu Jixiang, Zhang Qipei, Yang Zhihong, et al. Short term load forecasting method based on CNN-LSTM hybrid neural network model[J]. Power System Automation, 2019, 43(8): 131-137.

收稿日期:2020-09-17
 修改稿日期:2020-10-27