

# 基于改进 $K$ -means 算法的电力短期 负荷预测方法研究

荀超<sup>1</sup>, 陈伯建<sup>2</sup>, 吴翔宇<sup>2</sup>, 项康利<sup>3</sup>, 林可尧<sup>1</sup>, 肖芬<sup>1</sup>, 易杨<sup>4</sup>

(1. 国网福建省电力有限公司, 福建 福州, 350000; 2. 国网福建省电力有限公司电力科学研究院, 福建 福州, 350000; 3. 国网福建省电力有限公司经济技术研究院, 福建 福州, 350000; 4. 福州大学电气工程与自动化学院, 福建 福州 350108)

**摘要:** 现有方法预测电力短期负荷时忽略了对其进行聚类优化处理, 导致预测耗时较长, 短期负荷预测精度偏低。为此, 提出一种基于改进  $K$ -means 算法的电力短期负荷预测方法。该方法利用改进后的  $K$ -means 算法聚类处理电力负荷大数据, 使用聚类后获得的训练样本构建循环神经网络 RNN 拓扑结构, 然后通过对 RNN 神经网络模型设置最优权值, 实现电力负荷的短期预测。实验结果表明, 所提方法具有高预测效率和高短期负荷预测精准度。

**关键词:**  $K$ -means 算法; 数据聚类; RNN 神经网络模型; 电力负荷大数据; 预测方法

DOI:10.19781/j.issn.1673-9140.2022.01.011 中图分类号: TM715 文章编号: 1673-9140(2022)01-0090-06

## Research on short-term power load forecasting method based on improved $K$ -means algorithm

XUN Chao<sup>1</sup>, CHEN Bojian<sup>2</sup>, WU Xiangyu<sup>2</sup>, XIANG Kangli<sup>3</sup>, LIN Keyao<sup>1</sup>, XIAO Fen<sup>1</sup>, YI Yang<sup>4</sup>

(1. State Grid Fujian Electric Power Co., Ltd., Fuzhou 350000, China; 2. Electric Power Science Research Institute, State Grid Fujian Electric Power Co., Ltd., Fuzhou 350000, China; 3. Economic and Technological Research Institute, State Grid Fujian Electric Power Co., Ltd., Fuzhou 350000, China; 4. College of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China)

**Abstract:** The existing methods ignore the clustering optimization process when predicting the short-term load of electricity, which leads to a long prediction time and a low accuracy of short-term load prediction. Therefore, a short-term load forecasting method based on improved  $K$ -means algorithm is proposed. This method uses the improved  $K$ -means algorithm to cluster the big data of power load, uses the training samples obtained after clustering to construct the RNN topology structure of the recurrent neural network. Then the optimal weights are set for the RNN neural network model to realize short-term forecast of the power load. The experimental results show that the proposed method has high forecasting efficiency and high short-term load forecasting accuracy.

**Key words:**  $K$ -means algorithm; data clustering; RNN neural network model; power load big data; forecasting method

随着电网智能化程度和信息化程度的不断提高, 由电网维修和运行产生的数据不断增加, 对大数据处理和存储提出了更高的要求, 增大了管理大数

据的难度, 导致在信息传输、系统可靠性、信息存储和集成等方面增加了智能电网面临的挑战<sup>[1-2]</sup>。电网建设、电力营销和电力设备等产生的数据是智能

电网中主要存在的电力数据,通过对这些电力数据进行分析,可以对电力短期负荷进行预测,从而保障智能电网的运行性能<sup>[3-4]</sup>。当前电力短期负荷预测方法存在预测效率和精度不高的问题,需要对预测方法进行深入的研究<sup>[5]</sup>。文献[6]提出基于集合经验模态分解(ensemble empirical mode decomposition, EEMD)和神经网络的电力短期负荷预测方法,通过 EEMD 样本熵分解原始的电力负荷大数据,融合各子序列的预测结果,实现电力负荷大数据的短期预测;文献[7]提出基于径向基函数神经网络(radical basis function, RBF)的电力短期负荷预测方法,该方法利用主成分分析法消除冗余信息,提取电力负荷大数据的特征向量,在 RBF 网络中结合历史负荷数据和提取的特征向量建立负荷大数据预测模型,实现电力负荷大数据的预测;文献[8]提出基于改进随机森林的电力短期负荷预测方法,采用遗传算法优化随机森林决策树,利用优化后的随机森林决策树实现电力负荷大数据的预测;文献[9]提出基于深度学习的多特征短期电力负荷预测,对比了多种算法下短期电力负荷预测结果的情况,选择最佳算法进行负荷预测。但以上方法均忽略了对其进行聚类处理,导致预测耗时较长、短期负荷预测精度需进一步提高。

为解决上述现有方法中存在的问题,进一步提高电力短期负荷预测的效率和精度,本文提出一种基于改进 K-means 算法的电力短期负荷预测方法。基于改进 K-means 算法的数据分类,采用循环神经网络(recurrent neural networks, RNN)模型,在 RNN 循环神经网络模型中输入最优权值,实现大数据环境下的电力短期负荷预测,可提高预测的效率和短期负荷预测精度。

## 1 基于改进 K-means 算法的数据分析

K-means 算法核心思想:将  $n$  个数据对象划分为  $K$  个类,并且使每类中的所有数据对象到该类的聚类中心点的平方和最小,但其聚类时间比较长,为实现对数据的快速聚类,保留 K-means 算法的效率同时将 K-means 的应用范围扩大到离散数据,其

K-means 改进算法的计算过程如下。

1)从整个样本  $X$  中,令  $I=1$ ,在  $X$  中随机挑选  $K$  个数据对象作为初始聚类中心  $m_j(I)$ ,其中,  $j=1, 2, \dots, K$ 。

2)设  $d(i, j)$  为  $K$  个聚类中心  $m_j(I)$  与电力负荷大数据样本  $X$  中每个对象  $x_i$  之间的距离,即

$$d(i, j) = \sqrt{(x_{i1} - m_{j1})^2 + (x_{i2} - m_{j2})^2 + \dots + (x_{iK} - m_{jK})^2} \quad (1)$$

利用式(1)寻找  $d(i, j)$  的所有  $(i, j)$  取值对应的欧式距离中最小的欧式距离  $d$ ,在聚类中心  $m_j(I)$  相同的簇  $S_j$  中存储对象  $x_i$ <sup>[10]</sup>。

3)设  $m_j(I+1)$  为新聚类中心点,即

$$m_j(I+1) = \frac{1}{N_j} \sum_{x_i \in S_j} x_i \quad (2)$$

式中  $N_j$  为数据对象在第  $j$  个类中的数目。

4)设置判断准则,判断是否满足准则,如果满足则进行下一步,如果不满足则转到步骤 2 中。

5)输出电力负荷大数据的聚类结果。通常情况下利用判断准则对是否终止循环进行确定,即①聚类中心点相同,当第  $I$  次迭代和第  $I-1$  次迭代获得的划分结果相同时,认为划分合理,结束迭代;②离散度准则函数,当第  $I$  次迭代对应的聚类离散度与第  $I-1$  次迭代对应的聚类离散度之间的差值小于计算得到的收敛极限值时,认为划分合理,停止迭代<sup>[11]</sup>。聚类离散度函数为

$$J_c(I) = \sum_{j=1}^K \sum_{x_i \in S_j} [x_i - (x_i - m_j(I))]^2 \quad (3)$$

设置收敛极限值  $\xi$ ,当  $|J_c(I) - J_c(I-1)| < \xi$  时,停止迭代。

对其聚类方法进行改进,通过混沌理论优化自由搜索算法进行计算,具体过程如下。

①通过随机值策略获得初始化种群,即

$$x_{0ji} = X_{\min i} + (X_{\max i} - X_{\min i})r_{ji}(0, 1) \quad (4)$$

式中  $i$  为第  $i$  个变量; $j$  为第  $j$  只个体,  $j=1, 2, \dots, m$ ,其中  $m$  为个体数目; $n$  为搜索空间在 Elman 网络中对应的维数,即变量在目标函数中的总数; $X_{\min i}$ 、 $X_{\max i}$  分别为第  $i$  维变量在 Elman 网络中存在的最小值和最大值; $r_{ji}(0, 1)$  为随机数。

②设  $P_j$  为信息素,其计算式为

$$P_j = \frac{f_j}{\max f_j} \quad (5)$$

$$S_j = S_{\min} + (S_{\max} - S_{\min})r_j(0,1) \quad (6)$$

式(5)~(6)中  $S_{\min}$  为灵敏度对应的最小值,存在  $S_{\min} = P_{\min}$ ;  $S_{\max}$  为灵敏度对应的最大值,存在  $S_{\max} = P_{\max}$ 。利用计算得到的初始信息素,获得初始搜索结果<sup>[12]</sup>。

③设  $x_{Oj}$  为灵敏度,其计算式为

$$x_{Oj} = \begin{cases} x_k, & P_k \geq S_j \\ x_j, & P_k < S_j \end{cases} \quad (7)$$

在新一轮中,根据式(7)计算的结果,在 Elman 网络中选择搜索的起始点。

④经过初始化处理后,个体在 Elman 网络中随机产生  $T$  个坐标,得到最佳适应度坐标,即

$$\begin{cases} x_{tji} = x_{Oji} - \Delta x_{tji} r_{tji}(0,1) \\ \Delta x_{tji} = R_{ji}(X_{\max i} - X_{\min i}) r_{tji}(0,1) \end{cases} \quad (8)$$

式中  $t$  为当前搜索过程中的搜索小步;  $R_{ji}$  为搜索空间中第  $j$  个体在 Elman 网络中的空间邻域搜索范围,其取值范围为  $R_{ji} \in [R_{\min}, R_{\max}]$ 。

⑤在适应度优劣的基础上确定普通粒子和精英粒子<sup>[13]</sup>。基于改进  $K$ -means 算法的电力短期负荷预测方法通过 Tent 映射确定精英粒子,即

$$g(x) = \begin{cases} 2x, & 0 \leq x \leq \frac{1}{2} \\ 2(1-x), & \frac{1}{2} < x \leq 1 \end{cases} \quad (9)$$

通过式(9)获得  $K$  个点对应的混沌点列,原精英粒子在 Elman 网络中的位置对应适应度值最好的点。

⑥释放信息素,并存储最佳个体。

⑦设置终止条件,若符合设置的终止条件,则输出搜索结果。

## 2 电力短期负荷预测方法

采用 RNN 神经网络模型,分组对上述聚类后电力负荷大数据进行训练,实现电力负荷的短期预测。RNN 循环神经网络又指 Elman network 结构的循环神经网络,其中 Elman 网络数据由 3 部分构成,分别是输出层、输入层、隐层等<sup>[14]</sup>。RNN 循环神经网络结构如图 1 所示。

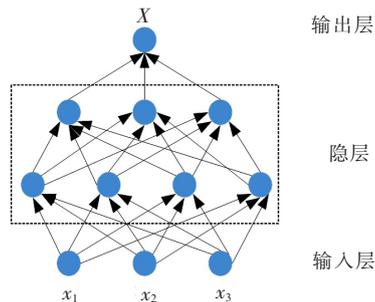


图 1 RNN 循环神经网络结构

Figure 1 Structure diagram of RNN recurrent neural network

设  $w_{1i,j}$ 、 $w_{2i,j}$ 、 $w_{3i,j}$  分别为节点从输出层到隐层、从隐层到输入层、从输入层到输出层对应的权值;  $M$  为输入 Elman 网络的节点数量,本文中输入 Elman 网络的节点数为 300;  $N$  为输出 Elman 网络的节点数量,输出 Elman 网络的节点数为 100;  $L$  为隐层中存在的节点数量,本文取该节点数量为 400。

设存在  $P$  组训练样本,即输出—输入对  $(x_i^{(k)} - y_i^{(k)})$ , 其中,  $k=1,2,\dots,P$ ;  $i=1,2,\dots,M$ ;  $j=1,2,\dots,N$ 。  $h_{i,\text{net}}^k$  为节点在隐层中的输入,  $h_{i,o}^k$  为节点在隐层中的输出,即

$$h_{i,\text{net}}^k =$$

$$\sum_{i=1}^M w_{1i,j} x_i^{(k)} + \dots + \sum_{v=1}^L w_{3v,i} c_{v,o}^{(k)} \quad (10)$$

$$h_{i,o}^k = f(h_{i,\text{net}}^k) \quad (11)$$

其中,  $c_{v,o}^{(k)}$  是对隐层一步的延时,即输出表达式为

$$c_{v,o}^{(k)} = h_{i,o}^{(k-1)}, v=t \quad (12)$$

设  $y_j^k$  为样本对应的输出向量,其计算式为

$$y_j^k = \sum_{i=1}^L w_{2i,j} h_{i,o}^{(k-1)} \quad (13)$$

根据输出向量构建训练指标函数,即

$$E = \frac{1}{2} \sum_{k=1}^P \sum_{j=1}^N (Y_j^{(k)} - y_j^{(k)})^2 \quad (14)$$

式中  $Y_j^{(k)}$  为第  $k$  个输出向量。

设  $w_{i,j}^{k+1}$  为连接点在 Elman 网络中对应的权值,其计算式为

$$w_{i,j}^{k+1} = w_{i,j}^k + \eta \delta_i^{(k)} o_j^{(k)} \quad (15)$$

式中  $\eta$  为训练步长,在区间  $(0,1)$  内取值,初始步长可以取 0.1,此后按 0.2,0.3,0.4, $\dots$ ,0.9 进行取值分别计算;  $o_j^{(k)}$  为各层节点  $j$  对应的输出;  $\delta_i^{(k)}$  训练误差,通过下述公式调整各层对应的权值:

$$\delta_i^{(k)} = \mathbf{Y}_j^{(k)} - \mathbf{y}_j^{(k)} \quad (16)$$

$$\omega_{2i,j}^{(k+1)} = \omega_{2i,j}^{(k)} + \eta \delta_i^{(k)} h_{t,o}^{(k)} \quad (17)$$

$$\begin{aligned} \omega_{1i,t}^{(k+1)} &= \\ \omega_{1i,t}^{(k)} + \eta \sum_{j=1}^N \delta_j^{(k)} \omega_{2i,j}^{(k)} f'(h_{t,\text{net}}^k) x_i^{(k)} \end{aligned} \quad (18)$$

$$\begin{aligned} \omega_{3v,t}^{(k+1)} &= \\ \omega_{3v,t}^{(k)} + \eta \sum_{j=1}^N \delta_j^{(k)} \omega_{2i,j}^{(k)} f'(h_{t,\text{net}}^k) h_{t,o}^{(k-1)} \end{aligned} \quad (19)$$

### 3 电力短期负荷预测过程

基于改进 K-means 算法的电力短期负荷预测方法的预测流程如图 2 所示。

电力短期负荷的预测过程如下。

1)利用 K-means 聚类算法对电力负荷大数据进行聚类处理,获得训练样本,构建 RNN 网络拓扑结构。

2)对种群进行初始化处理。

3)通过个体解码计算对应的初始权值,在训练样本数据的基础上对 RNN 神经网络进行训练,将期望和预测输出误差作为个体适应度值  $f_{ij}$ ,即

$$f_{ij} = \frac{K}{E} \quad (20)$$

式中  $E$  为期望输出误差; $K$  为常数。

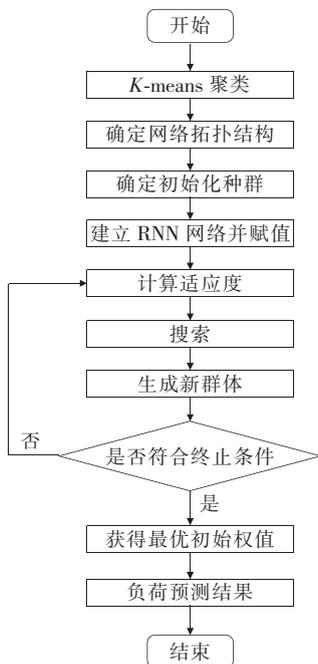


图 2 电力短期负荷预测流程

Figure 2 Flow chart of power load short-term forecasting

4)搜索、计算个体对应的灵敏度  $S_j$  和初始信息素  $P_j$ , 通过对比结果对连接权值进行优化。

5)设置终止条件,对输出的负荷预测值与电力负荷的实际值进行判断比较。若满足预先设定的最小误差,则停止迭代。判断是否满足条件,若满足则进行下一步,若不满足则返回步骤 3 中。RNN 神经网络中存在的初始权值通过混沌自由搜索算法进行优化,获得最佳权值<sup>[15]</sup>。

6)负荷预测。在 RNN 神经网络中输入最优权值作为优化参数组合,获取最优初始权值,实现电力短期负荷的预测。

### 4 实验与结果

为了验证基于改进 K-means 算法的电力短期负荷预测方法的整体有效性,在 Matlab 平台上进行测试。输入 Elman 网络的节点数量为 300,输出 Elman 网络的节点数量为 100,隐层中存在的节点数量为 400。负荷预测样本数量为 1 000 个。对基于 EEMD-样本熵和 Elman 神经网络的文献[6]方法、基于大数据简约及 PCA 改进 RBF 网络的文献[7]方法与本文方法进行对比测试,对比不同方法预测电力负荷大数据消耗的时间,测试结果如图 3 所示。

由图 3 中的数据可知,在多次迭代中本文方法预测电力负荷大数据消耗的时间均少于文献[6]和文献[7]方法消耗的时间。因本文方法采用改进后的 K-means 算法对电力负荷大数据进行了聚类处理,获得了大量训练数据,减少了预测消耗的时间,提高了预测效率。

为进一步验证方法的整体有效性,采用上述方法进行电力负荷大数据预测测试,分别对正常工作日负荷(周一到周五)、休息日负荷(周六、周日)、节

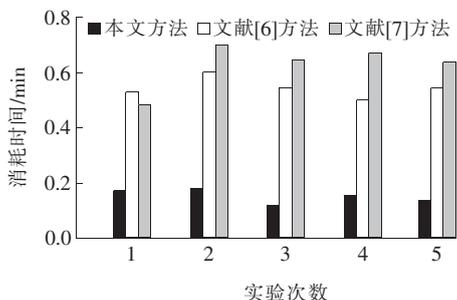


图 3 不同方法的时间消耗

Figure 3 Time consumption of different methods

假日负荷这些不同类型日负荷的预测结果,并与另外2种方法比较,进一步验证本文方法在提高短期负荷预测的精度方面的优势。对比结果如图4~6所示。

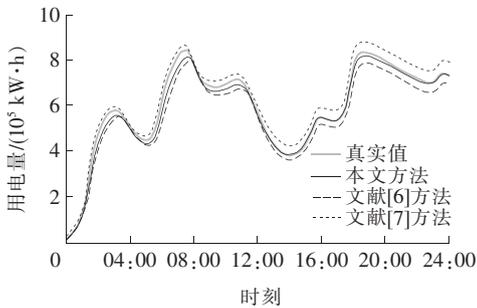


图4 工作日负荷预测结果

Figure 4 Working day load forecast results

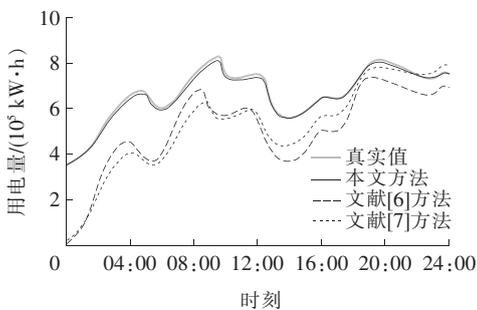


图5 休息日(周六、周日)负荷预测结果

Figure 5 Rest days (Saturday and Sunday) load forecast results

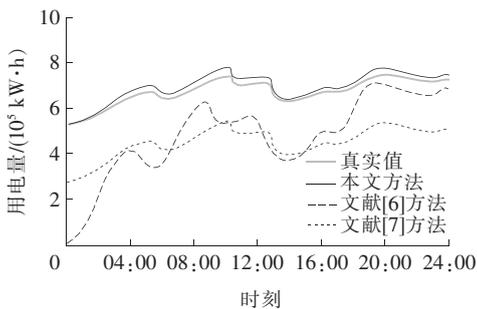


图6 节假日负荷预测结果

Figure 6 Holiday load forecast results

分析图4~6中的负荷预测对比曲线可知,采用本文方法获得的预测曲线与真实值曲线相差不大,精度始终接近96%。文献[6]、[7]的精度最高为92%,此方法获得的预测曲线与真实值曲线之间的差距较大。因为本文方法利用K-means聚类后获得的训练样本对RNN神经网络进行训练,通过混沌自由搜索算法获得最佳权值,实现电力负荷大数据的精准预测,验证了本文方法的预测精度。

为进一步验证方法的聚类处理后的效果,在

1000个数据样本中,选取100个样本数据,数据样本通过五角星形状、三角形以及圆形表示。分别采用上述几种方法进行测试,对比不同方法预测电力短期负荷大数据聚类结果,如图7所示。

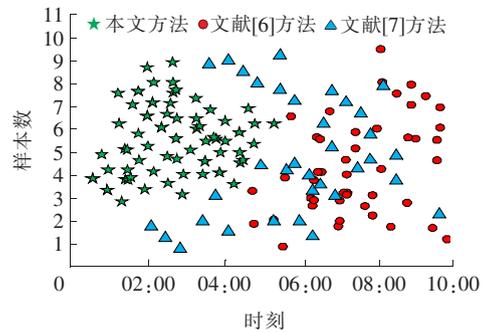


图7 不同方法的聚类结果

Figure 7 Clustering results of different methods

分析图7中的数据可知,采用本文方法获得的数据紧密聚集,聚类性较好,采用文献[6]、[7]方法获得的数据比较分散,聚类性较差。因为本文采用混沌理论优化自由搜索算法进行聚类优化,通过改进后的K-means算法对电力负荷大数据进行了聚类分析。

## 5 结语

电力负荷大数据的复杂性随着电力系统的发展不断增加,目前电力负荷大数据的预测方法已经难以满足电力系统的需求,使电力负荷大数据预测成为目前亟需解决的问题。当前电力短期负荷预测方法存在预测效率低和预测精度低的问题,现提出一种基于改进K-means算法的电力短期负荷预测方法,利用改进后的K-means算法对电力负荷大数据进行分类处理,通过RNN神经网络高效、精准的实现电力负荷大数据的预测,可以满足现有电力系统无法解决负荷大数据复杂性的需求,提高短期负荷预测的精度,为电力系统的发展奠定了基础。

## 参考文献:

- [1] 程津,黎燕,夏向阳,等.基于双组合预测的经济—电力负荷预测模型[J].电力科学与技术学报,2018,33(3): 18-22.  
CHENG Jin, LI Yan, XIA Xiangyang, et al. Economic-electricity conduction prediction model based on dual combination prediction[J]. Journal of Electric Power

- Science and Technology, 2018, 33(3): 18-22.
- [2] 刘南艳, 贺敏, 赵建文. 基于大数据平台的电力负荷预测[J]. 现代电子技术, 2018, 41(20): 161-164.  
LIU Nanyan, HE Min, ZHAO Jianwen. Electric load forecasting based on big data platform[J]. Modern Electronic Technology, 2018, 41(20): 161-164.
- [3] 刘琪琛, 雷景生, 郝珈玮, 等. 基于 Spark 平台和并行随机森林回归算法的短期电力负荷预测[J]. 电力建设, 2017, 38(10): 84-92.  
LIU Qichen, LEI Jingsheng, HAO Jiawei, et al. Short-term power load forecasting based on Spark platform and parallel random forest regression algorithm[J]. Electric Power Construction, 2017, 38(10): 84-92.
- [4] 徐晴, 周超, 赵双双, 等. 基于机器学习的短期电力负荷预测方法研究[J]. 电测与仪表, 2019, 56(23): 70-75.  
XU Qing, ZHOU Chao, ZHAO Shuangshuang, et al. Research on short-term power load forecasting methods based on machine learning[J]. Electrical Measurement & Instrumentation, 2019, 56(23): 70-75.
- [5] 祝学昌. 基于 IFOA-GRNN 的短期电力负荷预测方法研究[J]. 电力系统保护与控制, 2020, 48(9): 121-127.  
ZHU Xuechang. Research on short-term power load forecasting method based on IFOA-GRNN[J]. Power System Protection and Control, 2020, 48(9): 121-127.
- [6] 陈艳平, 毛弋, 陈萍, 等. 基于 EEMD-样本熵和 Elman 神经网络的短期电力负荷预测[J]. 电力系统及其自动化学报, 2016, 28(3): 59-64.  
CHEN Yanping, MAO Yi, CHEN Ping, et al. Short-term power load forecasting based on EEMD-sample entropy and Elman neural network[J]. Journal of Electric Power System and Automation, 2016, 28(3): 59-64.
- [7] 全新宇, 张宇泽, 张长生, 等. 基于广义回归神经网络的有源配电网网供负荷预测方法[J]. 供用电, 2020, 37(12): 40-45.  
TONG Xinyu, ZHANG Yuze, ZHANG Changsheng, et al. Load forecasting method of active distribution network based on generalized regression neural network[J]. Distribution & Utilization, 2020, 37(12): 40-45.
- [8] 李刚, 邹波. 改进随机森林的电力负荷预测方法[J]. 机械设计与制造, 2019, 57(10): 103-105+109.  
LI Gang, ZOU Bo. Improved random forest power load forecasting method[J]. Machine Design & Manufacturing, 2019, 57(10): 103-105+109.
- [9] 龙雯, 王同喜. 基于深度学习的多特征短期负荷预测[J]. 电脑知识与技术, 2021, 17(16): 186-187+194.  
LONG Wen, WANG Tongxi. Multi-feature short-term load forecasting based on deep learning[J]. Computer Knowledge and Technology, 2021, 17(16): 186-187+194.
- [10] 陈杰尧, 黄炜斌, 马光文, 等. 基于相似性识别的短期负荷动态预测方法[J]. 电网与清洁能源, 2020, 36(4): 1-7+13.  
CHEN Jieyao, HUANG Weibin, MA Guangwen, et al. A short-term load dynamic prediction method based on similarity recognition[J]. Power System and Clean Energy, 2020, 36(4): 1-7+13.
- [11] 莫文火, 陈碧云. 基于邻域关系矩阵的电力大数据增量式属性约简研究[J]. 高压电器, 2020, 56(8): 192-197+204.  
MO Wenhua, CHEN Biyun. Incremental Attribute reduction of electric power big data based on neighborhood relation matrix[J]. High Voltage Apparatus, 2020, 56(8): 192-197+204.
- [12] 蔡冬阳, 赵申, 周玮, 等. 光伏影响下考虑气象负荷分解和 LSSVM 的负荷预测[J]. 计算技术与自动化, 2020, 39(4): 81-85.  
CAI Dongyang, ZHAO Shen, ZHOU Wei, et al. Load forecasting considering meteorological load decomposition and LSSVM under the influence of photovoltaic[J]. Computer Technology and Automation, 2020, 39(4): 81-85.
- [13] 魏晓川, 王新刚. 基于气象大数据的城市电力负荷预测[J]. 电测与仪表, 2021, 58(2): 90-95.  
WEI Xiaochuan, WANG Xingang. Prediction of urban electricity load based on meteorological big data[J]. Electrical Measurement & Instrumentation, 2021, 58(2): 90-95.
- [14] 李正浩, 李孟凡. 基于深度学习的智能型负荷预测方法的研究[J]. 智慧电力, 2020, 48(10): 78-85+112.  
LI Zhenghao, LI Mengfan. Smart load forecasting method based on deep learning[J]. Smart Power, 2020, 48(10): 78-85+112.
- [15] 蒋铁铮, 尹晓博, 马瑞, 等. 基于  $k$ -means 聚类 and 模糊神经网络的母线负荷态势感知[J]. 电力科学与技术学报, 2020, 35(3): 46-54.  
JIANG Tiezheng, YIN Xiaobo, MA Rui, et al. Bus load situation awareness based on the  $k$ -means clustering and fuzzy neural networks[J]. Journal of Electric Power Science and Technology, 2020, 35(3): 46-54.